A characterisation of comics and sequential art for supporting the evolution of digital formats for comics : a crowdsourcing approach

Mihnea Tufis

▶ **To cite this version:**

Mihnea Tufis. A characterisation of comics and sequential art for supporting the evolution of digital formats for comics : a crowdsourcing approach. Image Processing [eess.IV]. Université Pierre et Marie Curie - Paris VI, 2017. English. NNT : 2017PA066590 . tel-01922460

HAL Id: tel-01922460
https://theses.hal.science/tel-01922460

Submitted on 14 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

École Doctorale Informatique, Télécommunications et Électronique

Laboratoire d'Informatique de Paris 6

A characterisation of comics and sequential art for supporting the evolution of digital formats for comics - A crowdsourcing approach

Présentée par

Mihnea TUFIȘ

Thèse de Doctorat en Informatique

Présentée et soutenue publiquement le *20 avril 2017*

Devant un jury composé de:

Jean-Gabriel GANASCIA	Directrice d'Études	Directeur de Thèse
Sihem AMER-YAHIA	Directrice de Recherche	Rapporteur
Karell BERTET	Maître de Conférences	Rapporteur
Marie-Jeanne LESOT	Maître de Conférences	Examinateur
Jochen LAUBROCK	*Head of Junior Research Group*	Examinateur
Guillaume RASCHIA	Maître de Conférences	Examinateur

Except where otherwise noted, this work is licensed under:
https://creativecommons.org/licenses/by-nc-nd/3.0/

Table of Contents

I. Introduction and Context **5**

1. Introduction **7**
1.1. Context 8
1.2. Problem statement 14
1.3. Proposed approach 15
1.4. Roadmap 17

2. Comics and Sequential Art **19**
2.1. A brief history of comics and sequential art 21
2.1.1. Sequential art in the history of humanity 21
2.1.2. A brief history of modern comics 26
2.2. Reading and collecting comics 43
2.2.1. Elements of the comics medium 43
2.2.2. Fans, collectors and readers 61
2.2.3. Reading comics in the age of digital 64

II. Digital publishing of comics **71**

3. Digital publishing of comics: State of the art **73**
3.1. The publishing environment – an overview 73
3.2. Standards for digital publications 77
3.2.1. Standards for presentation and distribution 78
3.2.2. Adding semantics to digital publications 80
3.2.3. Bibliographic standards for comics 85
3.3. Putting everything together 101

4. New standards for comics' digital publishing **103**
4.1. EPUB 103
4.2. Development of the new *EPUB* standards for comics 106
4.2.1. Development context – Advanced and Hybrid Layouts 106
4.2.2. AHL new standard specification #1: EPUB Multiple Rendition Publications 107
4.2.3. AHL new standard specification #2: EPUB Region Based Navigation 112

4.2.4. New standard specification for image-based publications: EPUB Sequences . . . 115
4.2.5. AHL new standard specification #4: EPUB Previews . . . 121
4.3. New standards for publication annotations . . . 122
4.4. A look into EPUB 3.1 . . . 125
4.4.1. Limits of the standard and future directions . . . 126
4.5. Enhancements to *CBML* – the semantic format for comics . . . 127

III. Crowdsourcing comics annotations 133

5. Crowdsourcing – literature review 135
5.1. A review of crowdsourcing initiatives . . . 135
5.2. General . . . 139
5.3. Applications . . . 152
5.4. Evaluation . . . 163

6. Comics++: a new system for crowdsourcing comics annotations 181
6.1. Data requirements for the description of comics . . . 182
6.2. *Comics++*: Crowdsourcing system overview . . . 184
6.3. *Comics++*: Choosing the "crowd" . . . 187
6.4. *Comics++*: System development and implementation . . . 189
6.4.1. Technical choices . . . 191
6.4.2. Implementing the crowdsourcing platform . . . 192

7. Data Analysis of Comics++ 209
7.1. Experiment design . . . 210
7.2. Experimental results and analysis of Comics++ . . . 211
7.2.1. Analysis of special panels . . . 212
7.2.2. Analysis of page structure . . . 214
7.2.3. Analysis of transcriptions . . . 220
7.2.4. Analysis of element types . . . 223
7.2.5. Crowd analysis . . . 228

8. Conclusions 233
8.1. Future work . . . 236

A. Appendix 239
A.1. FRBR . . . 239
A.2. EPUB Sequences – Charter Draft . . . 242
A.3. Comics++: Platform deployment . . . 242
A.3.1. Deploying the platform . . . 242
A.3.2. Comics++: The mobile prototype . . . 244

A.4. Comics Code Authority 245

Bibliography **251**

List of figures **261**

List of tables **273**

Abstract

In the span of just more than a century, graphic narrative and sequential art in all their cultural flavours (e.g., American comics, Japanese *manga*, Franco-Belgian *bande dessinée*), have made the journey from the periphery of culture to establishing themselves as an artistic medium at the frontier between literature and graphic art. During this journey they changed our lives, ignited the imagination of billions and inspired artists and writers to push the limits of creativity.

From newspaper funnies to critically acclaimed, award-winning graphic novels, from childish masked superheroes and pulp horror stories to the exploration of the self and the role of art, comics evolved in terms of form and theme, developed a language of their own and established themselves as a medium for complex, varied and challenging storytelling. Their recognition as a literary form in the late 1980s ignited the interest of academia to study and better understand comics; the latest developments in research fields such as digital humanities or machine learning are providing scholars with a suitable toolbox for asking increasingly complicated questions about all forms of sequential art. At the same time throughout the world, comics are confirming or claiming their place in the cultural patrimony through their presence in specialised museums, libraries, collections and research centres.

From a publishing perspective comics have made their mark on the book industry and are now occupying relevant positions in all important book markets. They seem to have sailed clear off the troubles that have recently hit the publishing industry worldwide and dealt well in aftermath of the economical downturns of 2008. In this, their fast reaction to the smartphone and tablet revolution of 2010 and their subsequent transition to digital media have been recognised as important supporting factors.

Whether creating enhanced digital content, supplying annotated corpus for academic research or indexing it for libraries and collections, the future of comics and sequential art seems tightly linked to the ability of fully describing and encoding them in an appropriate electronic format. While such formats exist for different applications related to comics in the areas described (publishing, research, cultural heritage) each of them comes with a high degree of domain and end-user specificity, which makes it difficult to swiftly apply them across all these domains.

Even resolving the problem of a multi-purpose format for comics, we are left with a more complicated issue: retrieving the relevant information from comics corpora and

persisting it in the format of our choice. Digitally available comic books come either in the form of electronic publications (most of them, web-based) or as flat pictures resulted from scanning and digitally reassembling print editions. While there have been recent significant improvements of machine learning approaches for mining comics content, such algorithms are still under-performing as a result of the artistic complexity, variety and specificity of the comics medium.

To adapt to these issues, we introduce a crowdsourcing-based solution for annotating comics, designed to capture their compositional, structural and content variety (i.e., characters, events, places, onomatopoeia, visual expressions, visual cues, panel types etc.). The data we collect is subsequently persisted using *Comic Book Markup Language (CBML)*, a *Text Encoding Initiative (TEI)*-based format adapted to the description of comics and sequential art.

At the same time, we address the shortcomings of digital publishing standards for image-based publications by drafting an extension to the *EPUB* format in which we aim at better formalising digital reading and more effectively capturing all recent evolutions in the design and consumption habits of digital image-based publications. Also, with the help of our different partners we analyse *CBML*'s compatibility with digital publishing formats, its application to academic research and industrial projects, and its use in the Galleries, Libraries, Archives, Museums (GLAM) sector.

Part I.

Introduction and Context

Chapter 1.

Introduction

"Comics escape a paper-box and electronic questions pop out."
— Sarah Boxer, New York Times (2005)

Comics as we know them today, are the result of a turbulent history, one of great transformations, during which they navigated between marginalisation and the perils of disappearance, but ultimately found their way to critical success and a place in contemporary culture. From their beginnings in newspaper pages at the turn of the 20th century, comics started their transformation into a medium exploring its potential for remarkable storytelling, often pushing the boundaries of taste and challenging the "establishment" ("a virulent art"), eventually turning into what we gracefully call today "the ninth art".

In the United States, comics had to reinvent themselves, following the intervention of the government and the self-imposed regulatory measures in the early years of the Cold War, which severely stalled their evolution and earned them the reputation as a childish and simplistic form of entertainment. They recovered by engaging the youthful, post-war generation in what is now known as the Silver Age of superheroes. The late 1960s and the turbulent early 1970s saw the rise of the quirky "underground" movement, which took on complex themes such as the social transformations in American society, the human condition and the role of art. During the 1980s comics creators pursued their explorations in terms of themes and form, culminating with what is known today as *the 1986 "tri-fecta"*: Art Spiegelman's Pulitzer Award winning *Maus*, Frank Miller's revival of Batman in *The Dark Knight Returns* and Alan Moore's innovative *Watchmen* earned comics the century sought legitimacy as a literary form, extended the boundaries of a now clear medium and inspired today's generation of artists.

The Franco-Belgian *bandes dessinées (BD)* mirrors the path of American comics, in which they found a great deal of inspiration starting with the 1920s. World War II determined local authors to dedicate their time and skill to imitating and continuing famous American stories (*"Superman", "Flash Gordon"*), whose distribution had been disrupted by the Nazi occupation. Thus deepening their understanding of the comics art and comics creation

process, these authors would soon move on to create their own original characters. After the war, much like in the US, *bandes dessinées* turned into a medium for children and youth and quickly became the subject of combined Catholic and Communist government regulations. But just like across the Atlantic the reactionary cultural movements from the 1960s marked the appearance of adult comics – the *BD adulte* – an evolution into a medium that was now taking on more complex, often political themes.

In Japan, the oldest forms of graphic narrative date back to the 12th century, whereas other scholars pin the origins of *manga* sometime in the 18th century, at the end of which the first use of the terms is also observed. Modern Japanese *manga* originates in the post-World War II period, with strong influences attributed to comic books brought by the American occupation troops. At the same time, the new Japanese constitution of 1947 prohibited all forms of censorship and thus encouraged artistic creativity resulting in an accelerated development of the medium, in opposition to the difficulties that comics and BDs were having in the US and the Franco-Belgian space, respectively. As a consequence, in the following decades Japanese *manga* went on to develop their own graphical identity (e.g., *gegika* – aesthetic realism or Osamu Tezuka's staple "cinematographic technique") and to quickly explore a wide array of genres (*shōnen manga* destined to male readers and *shōjo manga* destined to female readership) and subgenres (romance, superhero, science-fiction, supernatural, sports, war, history and the tremendously successful "ladies" comics). A particular genre, stemming from the liberty of expression and the fast evolution of the medium, is *seijin manga*, in which authors explore sexual and erotic themes, in ways which at times have challenged the moral standards of Japanese society and have eventually led to the creation of self-regulation mechanisms in the beginning of the 1990s.

1.1. Context

The publishing market for print and digital comics

The recognition of their literary value earned comics their entrance in book-shops, the exposure to a bigger audience and the diversification of their readership. Comics thus moved away from their niche market and started competing in the general book market. Throughout the 1990s and 2000s, comics saw an unprecedented development, especially in the United States and Europe (especially France, Belgium, United Kingdom, Spain, Italy), with yearly record breaking numbers in both publishing volume and sales. Japan is a particular case: *manga* has historically taken a big share of the national publishing market, doing well even after the continuous contraction of the domestic book market which has been ongoing since 1996. In addition, *manga* sales received an important boost in the late 1990s and early 2000s, by reaching new readership, particularly the young generations in US and Europe. But in 2008, just like in other economic sectors, the situation turned dire for the publishing industry worldwide. For comics in France and the United States, this stopped the yearly ascending trends for newly published titles, but remarkably it didn't

lead to any shrinking of their own market. Many speculate that a possible explanation behind this is the tablet and smartphone "revolution" of 2010, which brought with it a new paradigm for producing and consuming comics and sequential art. Granted, web comics have long existed (successful web comics, for that matter!), but what tablets and later smartphones brought along, is a screen format closer to that of print pages, as well as the digital technologies able to enhance the comics reading experience.

In Japan, the early adoption of digital technologies didn't have the same effect on publishing as it did in the US or Western Europe. "Scanlations" – unlicensed scanned *manga* translated by fans – were the first form of digital content adapted to mobile hand held devices, a while before any other licensed editorial enterprise in this direction. Despite the obvious licensing infringement of this kind of content, its key qualities – cheap, fast available, quality translations – played an important role in the growth of the US and European *manga* market in the decade between 1996 - 2006. And beyond all related controversy, "scanlations" show an undeniable truth about Japanese publishing industry: "*[m]anga* is hurting the way that all print media is hurting but in some ways it's worse, because *manga* is ill-equipped to adapt to New Media."[Thompson, 2012, Vakili, 2012].

Zooming out, the global publishing context describes "a world where, quite literally, the avenues through which we discover new books are shrinking – bookstores closing, library funding down, disproportionate market share to one digital books retailer" [Vakili, 2012], a world in which print and digital comics are trying to reinvent themselves as an artistic outlet and carve their way through as a successful competitor on the market for literature.

From the onset, digital technologies and expectations for high quality content mutually influenced the evolution of one another, leading to technological innovation and improved user experience. To this end, industry stakeholders (authors, readers, publishers, editors, technology producers) quickly realised the benefits of focusing their efforts towards the creation of standardised reading solutions. For many years digital reading occurred almost exclusively on computers and involved the consumption of a word processor file (e.g., Microsoft's *DOC(X)*), which in many cases allowed for acceptable text and media rendering. While Adobe's *PDF* standard raised the quality in terms of formatting and document rendering, it was the advent of mobile hand-held devices that triggered the e-book revolution and the appearance of a number of dedicated formats (e.g., Amazon's *AZW(3)*, Apple's *iBook*). At the same time, the steady growth of the World Wide Web nurtured the development of open source technologies, file formats and standards. Web browsers, the underlying *HTML* standard, *XML*-based formats (such as *OPF*) quickly became the most widespread solutions for publishing, packaging and reading web publications.

All these solutions however, are each addressing different subsets of aspects related to digital publishing. For instance, the case would be that an author would deliver the content to an editor in *DOC* format; subsequently, this could be converted into an *XML-based* document for internal use by the publishing house, before being handed over for packaging in view of web (as in-browser *HTML*) or device-specific (as open source – *HTML* or *XML-*

based, or proprietary formats – *Kindle* or *iBook*) publication. In order to address resulting interoperability issues, industry stakeholders combined their efforts to develop *EPUB*, "the distribution and interchange format standard for digital publications and documents based on Web Standards"[1], and also the open format supported by the largest number of e-readers (albeit not by Amazon's Kindle devices).

But like other formats, *EPUB* turned out to be poorly equipped for the demands of comic books publishing, underlining the technological gap with respect to the expectations of both comics readers and authors. To this end EPUB 3.0.1 and its planned 3.1 release address issues such as identifying comics-specific elements (e.g., panels, balloons, captions, onomatopoeia, etc.) and basic rendering of regions of interest (ROI). As an active part of several of the EPUB 3.1 working groups, we believe these initiatives directed the standard towards the right direction; we do, nevertheless, feel that there is so much more to the comic book experience that e-book formats such as *EPUB* have yet to catch. Due to their structure (sequences of pages, sequences of panels separated by "gutters") comics primarily engage their audience visually through text and image, but also establish a deeper link by freeing the readers' imagination with the help of the "gutter" – the space between two consecutive panels. This subtle, yet powerful transitional space might just hold the key for the next step in the digital evolution of comics.

Comics in cultural heritage

The 1970s underground and alternative culture movements, reflective of the social changes worldwide, influenced the evolution of comics into a medium of free speech and triggered their substantial transformations in terms of form, genre and subject matter that eventually led to the revival of the graphic novel, their critical acclaim in the late 1980s and their establishment as a literary form starting with the early 1990s. At the same time underground "comix" were the creative outlet of a new generation of authors and the object of interest of a new generation of readers many of whom would form the proactive fan-base whose engagement resulted in the creation of fan-magazines (fanzines), comics conventions and festivals starting with the mid 1970s. On the back of the quick success of such conventions (particularly festivals, which had less of a commercial approach), comics art themed museums, libraries and research centres soon appeared, with all forms of graphic and sequential art thus finding their value and appreciation as items of cultural heritage. The most important institutions working towards the preservation and valorisation of comics art are functioning in places with a long tradition of "the ninth art", namely Belgium, France, Japan, the United Kingdom and the United States.

"[T]he world's largest and most comprehensive research institution dedicated to original and printed comic books, daily strips, Sunday strips, editorial cartoons, graphic novels, magazine cartoons, sports cartoons" is the The Billy Ireland Cartoon Library & Museum, functioning since 1977 at the Ohio State University. The size of its collection is overwhelming: 450,000 original cartoons, 36,000 books, 51,000 serial titles and approximately 2.5 million

[1] http://www.idpf.org/epub

clippings. The other major American institution dedicated to the preservation of comic books is the Comics Art Collection of the Michigan State University (MSU), which is host to more than 200,000 items – "the largest library comic book collection in the world". In addition, one of the most important actions at the MSU Comics Art Collection is the creation of their extensive online catalogue. The British Cartoon Archive (BCA) at the University of Kent is one of the first organisations of this kind. Established in 1973 following a thematic exhibition at the National Portrait Gallery, it hosts 130,000 original drawings by approximately 350 artists and a library of comic books and magazines. Starting with 1990, the BCA curates a searchable catalogue of over 140,000 cartoon images from their own collections. As of 2006, also in the UK, The Cartoon Museum in London is dedicated to the preservation, exhibition and research of all British forms of cartoon art. It includes close to 10,000 books, comics and original cartoons and prints and also issues its own catalogue.

In France, the most important institution of the kind is the *Cité Internationale de la Bande Dessinnée et de l'Image (CIBD)* in Angoulême. Founded in 1983, following the increasing success of the Angoulême International Comics Festival, in place since 1974, the CIBD is an impressive centre hosting a comics museum, a patrimonial and a public library, and a documentation centre. Its varied activities are focused on the conservation, exhibition and dissemination of comics art, including the large scale digitisation of their corpus. And starting with 1989, the Belgian Comic Strip Centre (*Centre belge de la bande dessinée / Belgisch Stripcentrum*) in Brussels features a library and a research centre dedicated to promoting the cultural value of comic strips.

In Japan, the Kyoto International Manga Museum *Kyōto Kokusai Manga Myūjiamu* established in 2006 and operated by the Kyoto Seika University hosts over 300,000 items (including valuable Meiji period magazines) available within the Museum's gallery, research centre and the extraordinary "Wall of Manga" – a public library of 50,000 *manga* magazines.

To synthesise, Table 1.1 below gives an overview of other important GLAM institutions around the world with a focus on comics art.

Besides these, a number of similar institutions around the world are dedicated to the life and work of famous comics and graphic art creators, such as Chester Gould (creator of the *Dick Tracy* comics), Hergé – the famous Belgian creator of *Tintin*, famous Flemish-Belgian creator Marc Sleen, Taiwanese author Liu Hsing-chin and Japanese *manga* artists Hasegawa Machiko (creator of *Sazae-san*) and Shotaro Ishinomori. The global spread of these institutions, some of them dating as early as the 1970s, as well as the range of activities they propose in terms of conservation, collection, dissemination and research of comics art, is a testimony of the role that comics are gradually assuming in cultural heritage. To achieve their mission, some of these largest institutions are currently taking steps in the direction of digital conservation through the development of digital and online catalogues. For the moment, such catalogues provide with bibliographic metadata (authorship information,

Museum	Location	Since	Size
Caricature & Cartoon Museum	Basel, Switzerland	1979	3,000 original works
Caricature Museum (Museo de la Caricatura)	Mexico City, Mexico	1987	NA
Museum of Caricature (Muzeum Karykatury)	Warsaw, Poland	1978	20,000 pieces by Polish and foreign authors
Cartoon Art Museum	San Francisco, USA	1984	7,000 pieces
Gibiteca Antonio Gobbo	Belo Horizonte, Brazil	1992	18,000 copies
The Netherlands Strip Museum (Het Nederlands Stripmuseum)	Groningen, Netherlands	2004	approx. 30,000
Comic book, illustration and animation museum (Wow Spazio Fumetto)	Milano, Italy	2011	over 9,000
Finnish Comics Center (Suomen Sarjakuvakeskus)	Helsinki, Finland	2008	NA
Swedish Comic Book Library (Serieteket)	Stockholm, Sweden	2004	approx. 30,000

Table 1.1.: A list of other worldwide GLAM institutions dedicated to comics.

publication date, editorial data) which, while difficult to obtain, satisfy only a few of the diverse necessities of users engaging with such collections.

To the extent of their possibilities, some online libraries for comics, such as the Digital Comics Museum[2], ComicBook+[3], Grand Comics Database[4], Comic Book Database[5] or those of established publishers such as Marvel, DC or Dark Horse are seeking to expand their databases with content-based information, such as character names. While tremendously useful for users to query content based on desired characters, additional such scenarios could be easily envisioned based on place, events or object indexing in online catalogues of most GLAM institutions.

[2] http://digitalcomicmuseum.org/
[3] http://comicbookplus.com/
[4] http://www.comics.org/
[5] http://www.comicbookdb.com/

Comics in academic research

What about an academic approach to comics? Before the 1980s this question would have raised a great deal of suspicious eyebrows. We know that throughout their history comics have fought a double battle: one against their reputation as a youth and children entertainment medium, the other against the censorship of the Code [6]. Not only that comics have survived this battle, but as a result they established themselves as an artistic mean and paved the way for their evolution – today the term "graphic novel" suggested by Will Eisner in "A Contract with God" [Eisner, 1978] is defining the sub-genre at the inter-twining between comic and novels. And it is the same Will Eisner the first to dedicate his time and talent to writing instructional books about comics. Works like "Comics and Sequential Art" [Eisner, 1985], "Graphic Storytelling and Visual Narrative" [Eisner, 2008b] and "Expressive Anatomy for Comics and Narrative" [Eisner, 2008a] have become essential guides for aspiring professionals, scholars or simply comics aficionados. Subsequently the matter started to receive more and more academic awareness [7], while theoretic works like Scott McCloud's "Understanding Comics" [McCloud, 1993], through its simple, yet methodical approach, uncover increasingly more research paths for humanities researchers. Obviously, the fields of digital humanities and data science offer powerful scientific tools to pose more inquisitive questions about literature, graphical art and, why not?, comics! But while the first two have started to crystallise their own toolboxes, comics are again tributary to their challenging nature. The mixture of words and images, the thing that sets them apart from other arts is, in the case of the scholar endeavour, what makes them hard to be tackled.

Currently, numerous universities worldwide are offering a range of academic programs on the topic of comics and sequential art, from practical courses meant to educate the aesthetic and technical skills of future authors, to scholar approaches based on art, history or social studies. And with the acceptance of comics as a medium at the boundary between graphic art and literature, scholars now have the necessary tools to study them from both these perspectives.

Recently, comics scholars are exploring more scientific, data-centred approaches in their efforts to further understand comics, thus integrating comics studies within the research field of digital humanities (DH). Exciting research questions concerning text and graphic art stylistics pertaining to different authors, periods, genres or cultures, racial and gender studies, to name just a few, could soon be much easier to investigate once reasonably annotated corpora would become available. To this end, teams from the image processing community have been working for the past ten years on developing content extraction algorithms adapted to the particularities of comic books. [Tolle and Arai, 2013] and particularly [Rigaud et al., 2013] achieved good results on frame and text segmentation in different

[6] Comics Code Authority (CCA)

[7] The National Association of Comics Art Educators keeps an up to date list of United States based programmes.

types of comics. The team from the *L3I* Laboratory[8] at La Rochelle University is currently working on both image-processing and deep-learning approaches for character extraction. Obviously, all these endeavours would greatly benefit from the existence of an already annotated corpus of comics which could be used as a gold standard in evaluating automatic extraction techniques. Equally interesting work by the "Hybrid Narrativity" group at the Potsdam University (Germany) is combining techniques from the digital humanities and cognitive science to further the empirical understanding of graphic narration. Some of their most exciting results come from eye-tracking experiments recording the visual engagement of the reader with text and image in graphic novels.

A digital format for semantic encoding of comics

Whether in publishing, academic research or supporting Galleries, Libraries, Archives, Museums (GLAM), the future of comics seems to be tightly linked to digital media. All these new and evolving demands from comics content highlight the necessity for the development of a flexible, multi-purpose digital format for encoding comics. From this standpoint, the most widespread solutions to date are separately addressing these necessities: *EPUB, CBR/CBZ* focus on digital publishing needs, *Text Encoding Initiative (TEI)* stems from the text-centric digital humanities research community, while the numerous bibliographic and library standards (*ISBN, ISTC, ISSN, FRBR*) deal with different, sometimes overlapping aspects of book as an object, while failing to take into account the bibliographic particularities of comics.

This is where *Comic Book Markup Language (CBML)* [Walsh, 2012] really stands out as one of the best open source solutions dedicated to comics, able to address requirements beyond bibliographic metadata or content presentation. Based on *TEI*, *CBML* provides with a rich schema, perfectly adapted to the particularities of the comics medium. Moreover, its *XML* reliance means that the format is fully equipped to be integrated as a complementary interchange and distribution solution for digital publishing (see *EPUB*) or computer processing tasks ranging from simple parsing to text mining and information retrieval.

1.2. Problem statement

Taking into account the large comics corpora, the diversity of their subject matter and more importantly, their high degree of specificity, we acknowledge the challenge of the data acquisition effort for encoding comics. Of course, machine learning allows us to achieve good results on tasks such as image labelling, object recognition, and significant improvements in character recognition following very recent advances in deep learning. At the same time techniques based on optical character recognition (OCR) help us with text identification and transcription. But the application of these techniques on comics corpora turns out to be a bit more complicated. Comics page segmentation cannot always rely on crisp

[8] http://l3i.univ-larochelle.fr/

separation between structural elements, such as panels or balloons. As we will see, well crafted comics base their art on the sense of a shared experience between author and readers, while structure and composition come together to serve the act of reading. Thus, elements such as panel borders give a sense of space (or lack thereof), canonical balloon shapes express different types of communications (speech, thought, narration), simple lines have the ability of conveying the sense of speed, motion or smell, while onomatopoeia form a sound vocabulary of their own. These elements are all naturally easy to interpret even by the most inexperienced readers, yet pose tremendous difficulties for state of the art computer algorithms.

The goal of our work is to realise a comprehensive characterisation of comics and sequential art publications in order to support the evolution of corresponding digital formats. The term "comprehensive" refers both to the bibliographic characterisation (e.g., author, publisher, date etc.) as well as encoding the content of such a publication. The motivation for this work is three-fold:

- enhancing the rendering of digital comics, by adapting digital publication formats (i.e. *EPUB*) to the problematic of image-based publications;
- providing with a corpus of annotated comics to support academic research on the topic (e.g., training sets for machine learning and deep learning content extraction algorithms, digital humanities approaches to comics);
- indexing comics content to support activities of private collectors and professionals from the GLAM sector.

Reading comics is clearly an engaging experience, establishing a personal relationship between the reader, the book and its author. Moreover, reading a comic book involves all senses in a decoding exercise that furthers this relationship. Thus, we aim at involving the human factor to support our global objective, that of encoding comics. Traditionally, this was done with expert annotators, incurring large costs for training and financially motivating them [Kilgariff, 1998], as well as a high turnaround time on the task. Obviously, not feasible for large scale annotations such as those that we wish to acquire. So, could we actually use non-experts to annotate our large corpus? And if yes, would their results be reliable?

1.3. Proposed approach

Enter crowdsourcing! The term was coined by Jeff Howe and Mark Robinson, editors at "Wired Magazine", during discussions they had in 2005 about a seemingly new workforce paradigm businesses were using [Saffire, 2009]. It appeared as if companies were "outsourcing" tasks to "the crowd". Later, in 2008, Darren Brabham became the first to use the term in a scholar research article by defining it as an "online, distributed problem-solving

and production model" [Brabham, 2008]. Since then, crowdsourcing established itself as a problem solving methodology, with a strong data collection component. It was used to different degrees of success in knowledge discovery and management, distributed human intelligence tasking, broadcast search or peer-vetted creative production. Its main features – large, differently skilled crowds, cheap financial incentives, small realisation time – recommend it for tasks such as large corpora annotation [Nowak and Rüger, 2010]. Additionally, framing it within game or competition scenarios usually alleviate the financial incentives and move them into different reward areas: being the best at something one likes, learning a new skill, discovering a new hobby or being part of a community turn out to be influential factors in attracting large crowds to such endeavours.

To achieve our data collection goal, we develop an online crowdsourcing engine for annotating comics. The tasks are designed to mirror the page reading experience for the participating "crowd" of non-expert annotators, who are subsequently invited to engage with the platform in either of two ways:

- marking structural (panel layout, splash pages, meta-panels) or content (characters, places, events, onomatopoeia) elements;
- transcribing content elements previously marked by them or other participants.

However, crowdsourcing applied to data collection efforts comes at a cost. Usually this is reflected in the effort put in to detect and subsequently deal with bad annotations. [Brodley and Friedl, 1996] suggest that reasons for annotation noise can be as varied as: subjectivity, data dependent error and inappropriate feature information used for labelling. Initially, solutions relied largely on the creation and use of gold-standard corpora, but as [Kilgariff, 1998] rightly observes without high agreement between annotators gold standards quickly turn into fool's gold. Moreover, in settings in which high volume of annotations are demanded on a large number of features, building a gold standard seems costly and highly unlikely to be achieved. It is why inter-annotator agreement (IAA) measures are becoming more and more widespread in evaluating the quality of non-expert annotations. These statistics based measures inspired from computational linguistics range from simple percentage agreement to measures accounting for chance agreement and are some of the most reliable tools for assessing annotation quality.

[Nowak and Rüger, 2010] notes that one of their shortcomings is their lack of sensitivity to image annotation tasks. Thus, our secondary research objective is to test our custom algorithm for evaluating the quality of annotators. Our method, adapts inter-annotator agreement measures proposed by [Artstein and Poesio, 2008], and applies them only to a subset of quality annotators. In our model, the quality of an annotator is a function of i) task difficulty, ii) annotator expertise given the task and iii) annotator motivation for the task. We record the interactions between our "crowd" of annotators and the crowdsourcing platform and use them as proxies for the three aforementioned factors.

To summarise, the present work proposes a comprehensive characterisation of comics, by defining an approach to encode the content of such publications. To this end, our contributions are two-fold and consist in:

1. The development of new open standards for rendering and semantically encoding comics, more specifically:
   - development of new EPUB-specific open standards for rendering image-based publications;
   - development of a new EPUB-specific open standard for specifying transitions and sequences in image-based publications;
   - development of extensions for the CBML open standard, to encode additional comics-specific content (places, events, objects, splash pages);
2. The development of *Comics++*, a crowdsourcing platform for annotating comics, with the purpose of generating semantically reach comics encodings;
3. An experimental study on *Comics++* with the goal of understanding participants' engagement with the platform.

1.4. Roadmap

The remainder of this thesis is structured as follows: Chapter 2 is dedicated to comics and other forms of sequential art and includes a brief history of the medium, a motivation for considering comics as objects of cultural heritage, as well as an introduction to sequential art theory, laying the ground for the subsequent technical choices related to annotating comics corpora. Part II is fully dedicated to digital formats for comics publishing, with Chapter 3 providing an overview of the different publishing and bibliographic standards available and their application to digital publishing of comics, and Chapter 4 dedicated to our contributions to the development of new open standards for digital publication of comics. Part III is fully dedicated to *Comics++*, our crowdsourcing platform for annotating comics corpora. To this end, in Chapter 5 we realise an extended literature review covering the state of the art of crowdsourcing; Chapter 6 presents the technical design, development and deployment of *Comics++*, whereas Chapter 7 discusses an experimental study conducted on *Comics++*, with the purpose of understanding participants' engagement with the platform. Finally, we conclude our work in Chapter 8, not before sketching several future research directions in both comics studies and crowdsourcing.

Chapter 2.

Comics and Sequential Art

"Comics . . . are sometimes four-legged and sometimes two-legged and sometimes fly and sometimes don't . . . to employ a metaphor as mixed as the medium itself, defining comics entails cutting a Gordian-knotted enigma wrapped in a mystery . . . "

— R.C. Harvey (2001)

We set off with what we wish to be a compact, yet comprehensive review of the history of graphic sequential art and its evolution in what is today one of the most complex, fascinating and yet still under-researched and sometimes misunderstood art forms. For those readers who are asking themselves "why comics?", we hope this chapter offers at least part of the answer. Treat it as an extension of the motivation to our work, use it as a launching ramp to get curious about comics and sequential art and gradually discover the way in which they personally engage you and stimulate your sensibility. For those who are passionate about comics and those for which comics are the object of their professional interest, this chapter should be merely regarded as a brief pass through comics history (and pre-history, for that matter), and while we tried to base our discussion on a selection of great comics, we undoubtedly omitted others. Understanding the history of comics throughout the first part of this chapter and the complexity of the medium in the second part, we wish to establish the premise that comics have long moved away from being a marginal form of entertainment, they have stepped into our culture and are therefore very much deserving of the scholar attention that other forms of art have long benefited.

So what exactly do we mean when we talk about comics? Most people around the world got their imagination caught at least once into the captivating mixture of images and text, whether they called it *bande dessinée, manga, manhwa, fumetti* or *comics*. In the following, together with our reader we will explore the history of sequential art, from the cave paintings in Lescaux to the successful publishing industry that has become today. We will then dive into the medium itself, explore its tools and how they come to service the story by engaging the reader's senses. Finally, we will talk about the lure of collecting comics – what makes them so desirable by private collectors of all ages and how did it

evolve from a personal habit (sometimes border-line obsession) into specialised collections hosted by museums, libraries or dedicated centres.

It is thus important to clarify what comics are and adhere to a formal definition. In *Understanding comics* [McCloud, 1993], an insightful comic book **about** comics, Scott McCloud takes it to himself to develop a definition of comics, covering all their possible forms: strips, comic books, graphic novels. He builds his definition starting from *"sequential art"*, the all-encompassing term used by [Eisner, 1985] and develops it incrementally by bringing under scrutiny each intermediate step (Figure 2.1).

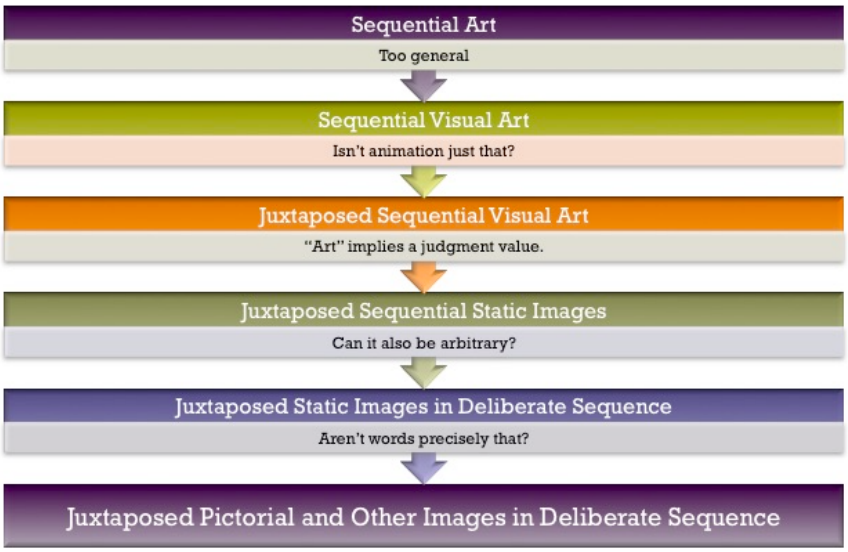


Figure 2.1.: Building a definition for comics. [McCloud, 1993] starts from the general definition given in [Eisner, 1985] and gradually tries to refine it.

The author finally settles for the following definition:

> "Fixed pictorial and other kind of images willingly juxtaposed in sequences, in order to transmit information and/or to trigger an aesthetic reaction by the reader."

The etymology of terms related to graphic narrative in different cultures – comics, *manga* (jp. whimsical images), *bandes dessinées* (fr. drawn strips), *fumetti* (it. little puffs of smoke) – is reflective of their initial perception and bares reference to either their subject matter or some of their structural properties. McCloud's definition, on the other hand, makes no such reference and for him, comics are not anchored to either their subject, form, production tools (i.e., pencil, ink, technical pens, brushes) or support (i.e., paper, print). Twenty five years later, McCloud's views on setting comics free from production tools and paper support are now fully achievable in today's age of digital.

2.1. A brief history of comics and sequential art

2.1.1. Sequential art in the history of humanity

Born from the desire to assert their individuality, share their experiences or express their emotions, humans produce art from the early stages of their existence. From a child's random scribble in the dirt when sad or bored to the anguish in van Gogh's late paintings, from a peasant's song during a hard day's work to the sublime voice of Maria Callas on the stage of *La Scala*, from the vigorous moves of a Māori *haka* to the grace of a ballerina, all art forms and expressions seem to have a common catalyst – the human need for communication. In oral folklore and literature, this need takes the form of storytelling and the first such forms involved graphic productions, rather than written ones. McCloud dedicates an entire chapter of his "Understanding Comics" [McCloud, 1993] to examine the evolution from image to word, formalising this into "The Big Triangle"[1] (Figure 2.2).

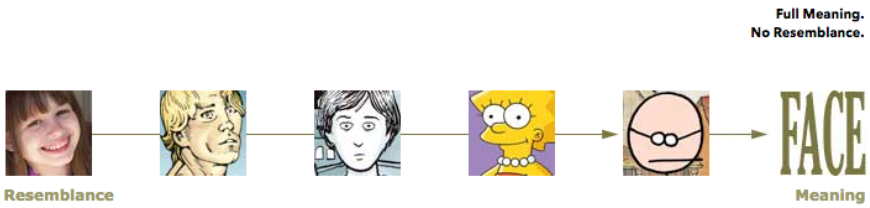


Figure 2.2.: The lower part of "The Big Triangle" illustrates the evolution of graphic representation from photographic resemblance (left) to iconic abstraction (4th tile from left to right), all the way to full meaning (right), i.e., words.
Source: Scott McCloud, *Understanding Comics* [McCloud, 1993, p. 52-53].

He considers that images have evolved through abstraction in two different directions: a semantic one, whose ultimate abstraction is the written word and a pictorial one whose level of abstraction is still being explored in modern and contemporary art. Working on his hypothesis, the role of comics – a "mixture" of words and images – is to bring these two representations back together, after millennia of self development. Taking a look back in history, we might be surprised by the many instances of sequential art that have been produced since the dawn of humanity. The cavemen were the first creators of graphical depictions, particularly of daily-life objects and events (e.g., hunting). Surely the Lascaux cave paintings are exponential for the "genre", but such examples dating back up to 40,000 years ago are spread around the world. This is indicative of a point that we will come back to later in this chapter – **graphic storytelling appeals to a sense of commonality, both in terms of the experience that is being depicted and how this is being achieved**.

[1] http://scottmccloud.com/4-inventions/

Figure 2.3.: Various stages in the production of grains. The Mortuary Chapel of Menna, superintendent of the estates of the king and of Amen, at the middle of the 18th dynasty, Thebes.
Credit: Guity Novin's A History of Graphic Design – The Online TextBook.

We take a giant leap in time to the Egyptian graphic art[2], which brings us a step closer to today's common interpretation of sequential art (Figure 2.3). Notice the separation of the painting into three horizontal tiers, which are read bottom-to-top, the lower tier left-to-right, the middle one right-to-left and finally the top tier left-to-right again.

These depictions begin to use two elements that will define the consumption pattern of modern sequential art – structure and reading order – which would further refine, with the establishment of panel separation and the controlled reading order mostly due to the physical properties of the supporting medium (e.g., papyrus or wooden scrolls, stone walls, textile cloth etc.). For example, in the beginning of the 2nd century A.D., the 30 meters high *Trajan's Column* (Figure 2.4) was erected in Rome to depict the epic scenes in the two dramatic wars that resulted in the Roman conquest of Dacia[3]. It is one of the early examples of scene fragmentation in graphic narrative, with the spiral structure of the column clearly imposing the reading order of the sequence.

1,000 years later, the *Bayeux Tapestry* (Figure 2.5) describes in great pictorial detail the Norman conquest of Britain. This too features a clear reading order (left-to-right) synchronised with the intended chronological order of the scenes. The scene panels don't feature clear borders, but a division by subject matter does stand out. Perhaps the most

[2]Hieroglyphs are excluded by this example. Although they are pictorial representations, they translate to actual words, syllables and sounds.
[3]Ancient kingdom occupying the territory of present-day Romania and Moldova, as well as parts of Bulgaria, Hungary, Serbia and Ukraine.

Figure 2.4.: Trajan's Column – Battle scene between the Roman and Dacian armies
Attributed to Apollodorus of Damascus – Conrad Cichorius: "Die Reliefs der Traianssäule", Erster Tafelband: "Die Reliefs des Ersten Dakischen Krieges", Tafeln 1-57, Verlag von Georg Reimer, Berlin 1896.
Credit: Wikipedia, Public Domain.

Figure 2.5.: The messengers with Guy, with portrayal of medieval agriculture in the border.
Credit: Wikipedia, CC BY-SA 3.0.

interesting element, moving us one step closer to modern comics, is the inclusion of textual captions which complement the graphical content.

Across the Atlantic, at about the same time, a number of codices featuring chronicles and historical accounts of pre-Colombian civilisations were also making use of rich juxtaposed graphical elements.

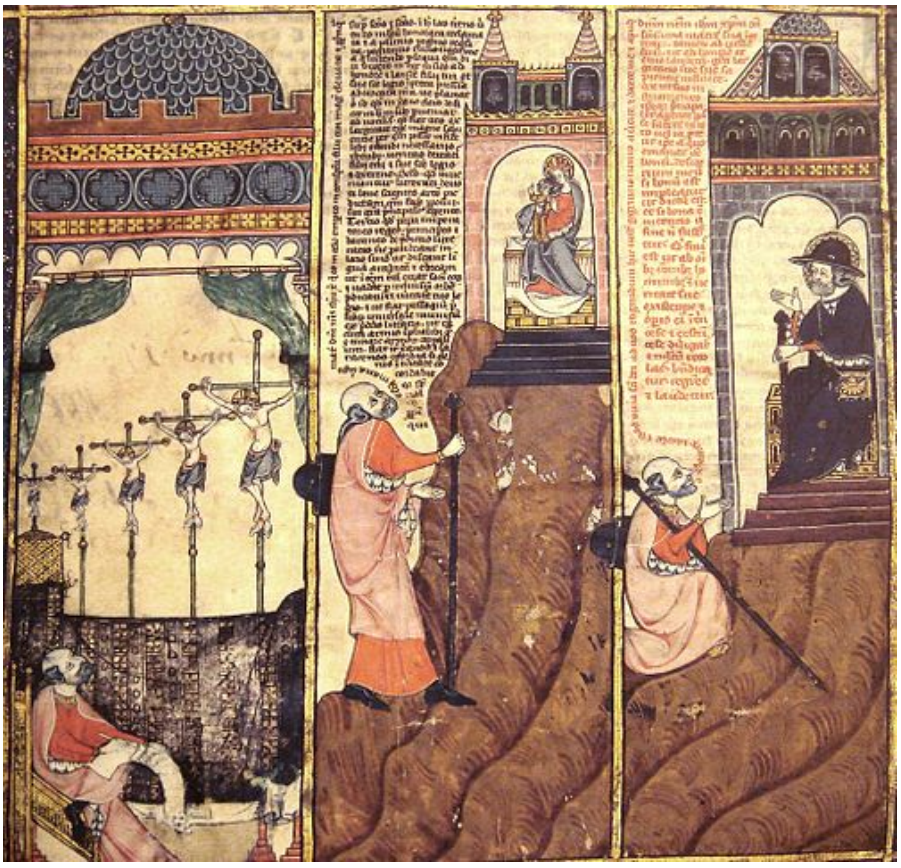


Figure 2.6.: "Life of Raymond Lull" – 14th century manuscript featuring elements such as panels and speech areas similar to speech balloons.
By Thomas le Myésier – Karlsruhe, Badische Landesbibliothek, Cod. St. Peter perg. 92, fol. 1v. Manuscript, reproduced in "An illustrated history of the Knights Templar", James Wasserman
Credit: Wikipedia, Public Domain.

Also in Europe, pre-print early Christian art including hand written religious books (Figure 2.6) and manuscripts featured an abundance of graphic art supporting the narrative or dogmatic text. This was organised in structures whose shapes, separation and implied reading sequence bare striking resemblances to simple comic book pages.

The most pervasive use of graphic narrative in religious art are perhaps the mural paintings in Christian churches. Through its simplicity, this art form has been historically used by the clergy as an education tool to help pass the religious message to the predominantly illiterate masses. One such example is shown in Figure 2.7 in which we can observe that

Figure 2.7.: Mural painting at Voroneţ Monastery (Romania) – 15th century, featuring sequences of panels. Anonymous painter.
Credit: Wikipedia, Public Domain.

the story (the martyrdom of a local saint) is depicted by layering the composing paintings into registries identical to panel tiers. Even more interesting, these registries form a grid that gives birth to its own type of "gutters".

From ancient Egypt to the Americas and from the cavemen to the great Renaissance artists, on rock, paper, leaves or walls, graphical storytelling seems to have always found a way to surge to the surface of human culture. And it is its simplicity and appeal to the "sense of commonality" that seem to convey its historic and geographic universality.

2.1.2. A brief history of modern comics

We take another leap in time and continue our history of comics during the 19th century, when Swiss cartoonist Rodolphe Töpffer creates *Les Amours de Mr. Vieux Bois* (Geneva, 1837), published in the United States as *"The Adventures of Mr. Obadiah Oldbuck"* (New York, 1842) and printed in comic book format. The use of interdependent text and pictures for the first time in modern history qualify R. Töpffer as the parent (some say precursor) of comics. Goethe himself acknowledged not only the talent of the Swiss cartoonist, but also the limitless potential of the newly employed medium:

> "If for the future he (Töpffer) would choose a less frivolous subject and restrict himself a little, he would produce things beyond all conception."

The beginning of the 20th century marked an important step in the development of comic books, with the establishment and growth of comic strips. In the United States, these were distributed by syndicates to newspapers and published in dedicated "funnies" sections or supplements. Two comic strips masters distinguish themselves during this period: Winsor McCay through his wonderful layout experiments supporting fantastic narratives in strips such as the *Little Nemo in Slumberland* and George Herriman, who inspired generations through sensitive, poetic creations and the meaningful use of colour; his best known strip, *Krazy Kat*, is considered one of the best comic strips of all time.

The early 1930s in America mark the transition from comic strips to a new form of publication – comic books. In 1929, *The Funnies*, a tabloid format comic magazine is the first to publish **original comic strip material**, albeit for a duration of 18 months. Only two years later, in 1933, Harry Wildenberg and Maxwell Gaines (Eastern Color Printing Company) convince Procter & Gamble to commission and distribute *Funnies on Parade*, a 36 page paper-cover magazine, featuring **reprinted comic strips**. Based on their great success, they decided to try a new distribution outlet - in 1934, they created *Famous Funnies* and distributed them at news-stands at the price of 10 cents per magazine. The latter proved to be a success and soon comic strip syndicates joined the two. These three publications proved that there was a market for such content. Despite featuring largely reprinted content, they provided a vehicle for the transition from comic strips to comic books.

On a market over saturated with reprints, a former military, writer and entrepreneur, Major Malcom Wheeler-Nicholson, decided to try something new. In February 1935, he brought to market *New Fun: The Big Comic Magazine*, a publication that was factually living up to it's title: it featured new, original material, written by the Major himself and printed in tabloid-size[4], a format larger than what would later become the standard for comic books. In March 1937, the Major launched his third title, a magazine focusing on single-subject stories – Detective Comics #1 (Figure 2.8) or... DC!

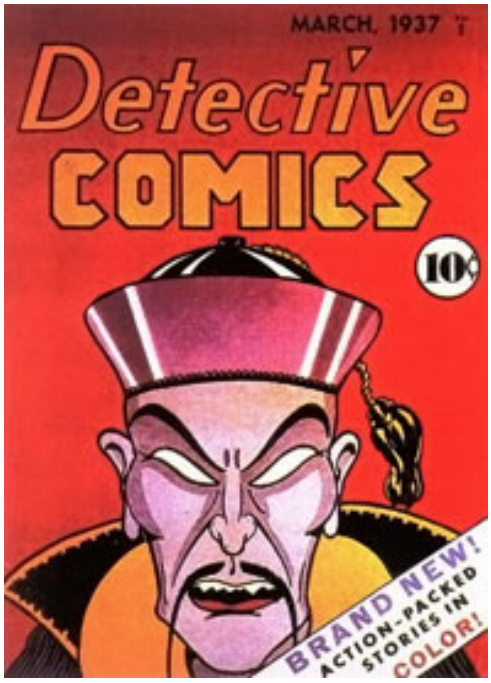


Figure 2.8.: Cover of Detective Comics #1 (Mar. 1937). Art by Vin Sullivan. Retrieved from http://thebatman.bravepages.com/comics/dc001.jpg, Fair use.

Unfortunately, soon after the Major would be the protagonist of a story of deceit that is now part of comics history: because of continuous economic problems, some due to the difficulties of promoting untested content on a young, still cautious market, the Major was forced to take in Harry Donenfeld (pulp magazines publisher and the owner of a printing company he owned money to) as a partner in his own publishing company. Comics historian Gerard Jones recounts [Jones, 2005, p. 125]:

> "In early 1938, Harry Donenfeld send him [the Major] and his wife on a cruise to Cuba to 'work up new ideas'. When they came home, the major found the lock to his office door changed. In his absence, Harry had sued him for nonpayment and pushed Detective Comics, Inc. into bankruptcy court. There a judge named Abe Mennen, one of Harry's old Tammany buddies, had been appointed interim president of the firm and arranged a quick sale of its assets to Independent News. Harry gave the major a percentage of More Fun Comics as a shut-up token and wished him well."

Today, Major Malcom Wheeler-Nicholson is acknowledged for his double fold contribution to the history of comics: the first to publish all-original material comic books and, ironically, being the creator of what would eventually become the financial media giant DC Comics.

In June 1938, *Action Comics #1* (Figure 2.9) hosts the first appearance of a character that would completely reshape modern culture – Superman. Co-created by Jerry Siegel and Joe Shuster, it is the result of a longer creative process (started in 1933), which saw

[4]Tabloid refers to the compact newspaper format, typically about 11 x 17 in.

him evolve from a seed story[5] into a comic strip character before finally featuring in his own story. Somewhat similar to the character most of us know today, he did present some important differences, which some believe partly explain his time tested success: his superpowers were less developed (he would jump over buildings and lift heavy objects), but compensated with a sarcastic sense of humour, especially directed at the villains, who would find themselves in awe at his formidable physical feats.

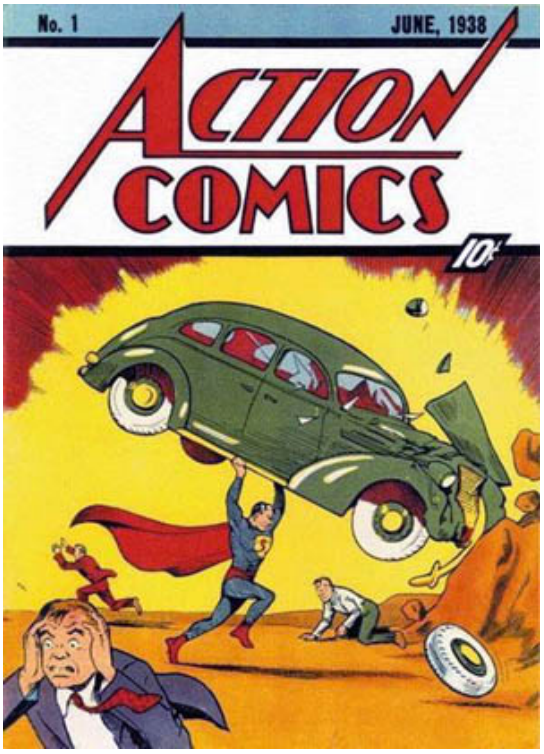


Figure 2.9.: Cover of *Action Comics #1* (Jun. 1938, DC Comics). Art by Joe Shuster (art), and Jack Adler (colour). Source: "Action Comics #1" at The Grand Comics Database. Retrieved October 31, 2006. (similar file if not the one originally uploaded.) Fair use.

Superman would prove to be a pioneer character, the spark that would ignite the superhero genre in comic books:

- Batman, *Detective Comics #27* (May 1939),
- Wonder Man, *Wonder Comics #1* (May 1939),
- The Sub-Mariner, *Marvel Comics #1* (October 1939),
- The Human Torch, *Marvel Comics #1* (October 1939),

[5] "The Rise of Superman", in *Science Fiction Magazine* (1933) features a scientist creating a secret formula allowing humans to develop superpowers. However, those who had these powers did not use them to do good.

- The Flash, *Flash Comics #1* (January 1940),
- Captain Marvel, *Whiz Comics #2* (February 1940),
- Green Lantern, *All American Comics #16* (July 1940),
- Captain America, *Captain America Comics #1* (March 1941),
- Wonder Woman, *All Star Comics #8* (December 1941).

Later, this fascinating character would be the only one to successfully cross mediums and feature in radio series, television series and cinema.

Figure 2.10.: The first appearance of Captain America in *Captain America Comics #1* (Mar. 1941). Cover art by Joe Simon (inks and pencils) and Jack Kirby (pencils).
Retrieved: Grand Comics Database, Fair use.

Between 1941-1944, during the American involvement in World War II, despite paper shortages, the comic book industry doubled its sales figures up to 20 million copies per month and the superheroes genre in particular saw an unprecedented rise in readership. Comics in general and especially superheroes were tremendously popular amongst the army service personnel, but the two major publishers reacted differently to these preferences, in ways that will shape their publishing future for decades to come. DC aligned its flagship superheroes – Superman (the most popular reading altogether!), Batman and Wonder Woman – with America's war effort. However, none of their characters were directly participating on the front and the message they were sending was that the troops were perfectly able to handle the combat on their own. At the same time, Timely (the precursor of Marvel Comics) chose for their superheroes to have a more "hands-on" involvement in the war. The stories of Captain America (debuted in *Captain America #1* (March 1941) – Figure 2.10) and Timely's other superheroes (Sub-Mariner and the Human Torch) would take place on the front line, and feature them fighting America's enemies in occasionally brutal scenes and pages abounding with racial stereotypes. However, irrespective of the publishing strategy one thing is certain: World War II turned superhero comics into American pop-culture institutions and as a result gave them the boost they needed to truly take off as a business.

We remember Max Gaines, one of the publishers who in the early 1930s printed *Funnies on Parade*, the magazine that triggered the transition from strips to actual comic

books. Immediately after the war, his new company Education Comics (EC) was printing educational comic books that were distributed in schools and churches. In 1947, after his sudden death and his son's William takeover of the business, the company changed its name (Entertaining Comics) and publishing direction. We mention EC not only because they were one of the important publishers during the Golden Age of comics, but because we find their change of name, from "education" to "entertainment", illustrative for the trend that comics had taken at the time. EC were also the first to give credit to their contributing authors, encouraging them to sign their stories and to establish idiosyncratic styles at a time when writers and artists were usually employed under work for hire contracts and the copyright for their characters and stories was passing to the publishing entity [6].

Figure 2.11.: Cover of *Atomic War! #1* (Nov. 1952), one of the many comics which reflected Cold War paranoia in 1950s America.
Credit: Wikisource, Public Domain.

In the 1950s most comics magazines would feature one main story together with several secondary stories, often representatives of very different genres. Most of the time, readers would seek and buy a magazine only for its main story, reading the rest of it with diminished interest. To attract a larger audience, magazine editors would put all their efforts into the design of their covers, often featuring an appealing illustration and the promise of "weird", "incredible", "horror" content, despite the different types of stories bundled within. Freshly out of WW2 and on the brink of entering the Cold War, American society was still dealing with anxiety and fear, while the prospect of nuclear warfare and the "red mennace" propaganda were slowly seeding mass paranoia. Like other media, comics too were capitalising on these fears and their content and especially covers (illustrations, title keywords) were reflective of the social context (Figure 2.11).

Quality stories were indeed out there (e.g., some of EC's contained clever plot lines and good development of the story), but most of them "were vulgar and poorly drawn, with gratuitous violence and crude depictions of women and minorities" [Dooley, 2013a].

A look at the readership numbers for those years, which tell us that approximately 80% of children and 90% of teenagers read comics, help us better understand the context of the

[6] Jerry Siegel and Joe Shuster lost all their rights on Superman after receiving a $130 check. DC claimed that they were paid under a work-for-hire agreement. Several legal disputes between them and their heirs on one side and DC on the other side, were finally resolved in favour of the publishing company. As it stands today, DC owns all rights for Superman.

upcoming events that would impact the development of comics for decades to come. In 1954, Dr. Frederic Wartham published what is today seen as the most vehement anti-comic book research. In *Seduction of the Innocent*, the American psychiatrist advocates that comics fuel a rise in juvenile delinquency and are paramount in the corruption of America's youth. Wartham's research came to support the activities of the freshly established Senate Subcommittee on Juvenile Delinquency under presidential aspirant Senator Estes Kefauver. The Subcommittee's hearings [Quarles, 2012] saw interventions of both Dr. Wartham and EC owner, William Gaines[7], with one exchange in particular perfectly summarising the entire debate:

Figure 2.12.: *Crime SuspenStories #22* (May 1954, EC Comics), art by Johnny Craig. The (in)famous issue used by Sen. Kefauver during the Senate hearing on comics. Retrieved: Grand Comics Database, CC BY 3.0.

> **Kefauver**: Here is your May 22 issue [of *Crime SuspenStories #22* – Figure 2.12]. This seems to be a man with a bloody axe holding a woman's head up which has been severed from her body. Do you think that is in good taste?
>
> **Gaines**: Yes, sir; I do, for the cover of a horror comic. A cover in bad taste, for example, might be defined as holding the head a little higher so that the neck could be seen dripping blood from it and moving the body over a little further so that the neck of the body could be seen to be bloody.
>
> **Kefauver**: You have blood coming out of her mouth.
>
> **Gaines**: A little.

[Dooley, 2013b] notes that "most recently, research by University of Illinois professor Carol Tilley uncovered definitive proof that Wertham falsified crucial facts, distorted quotes, omitted relevant circumstances, and otherwise, quote: played fast and loose with the data he gathered on comics". The industry's reaction to government and public pressure was the establishment later in 1954 of the Comics Code Authority (CCA) under the supervision of judge Charles F. Murphy[8]. The Code regulated every aspect of comics content (see Appendix A.4) and insured the self-censorship of most publications. Even though most editors adhered to it, a few (EC amongst them) chose not to and let their own readers be the judge for the quality of their content. The aftermath of the Code's adoption include 18

[7]The son and heir of previous EC owner, Max Gaines.

[8]Following the appearance of the CCA, judge Charles F. Murphy became known as the Comics "Czar"

publishers leaving the industry during the 1950s, which led to a steep decline of monthly comic book titles, from 696 in 1952 to 300, just after the Code. The resulting self-censorship, based on the provisions of the Code "child-proofed" comics and insured that for the following decades the medium would address to children or teenagers at best.

As mentioned, not everyone complied, and publishers such as Dell and Gold Key stayed out of the CCA, believing that their established brands alone are a good enough guarantee for the quality of their content. On a larger scale though, **three main reactions** stood out in the years following the Code: the Silver Age of superheroes, the underground movement and the evolution of the distribution chain towards a direct market.

The Silver Age of Superheroes

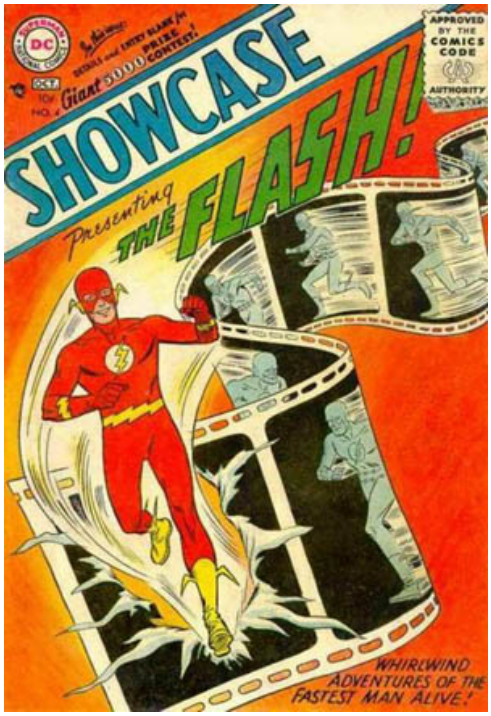


Figure 2.13.: Cover of *Showcase #4* (Sep.-Oct. 1956, DC Comics), featuring the first appearance of The Flash. Art by C. Infantino (pencils) and J. Kubert (inks)
Source: "Showcase #4" at The Grand Comics Database. Retrieved Dec. 25, 2004, Fair use.

The adoption of the CCA had negative effects over all types of comics, including the superhero genre which had attained a good level of popularity in the years before and during the war. Two other important factors were contributing to this decline: first, the rise of a new entertainment media – television and second, the ageing of the mass of readers that supported the 1940s comics boom, who by the late 1950s were outgrowing the genre. Marvel in particular had a difficult time in promoting their characters (Captain America, Human Torch, Sub-Mariner), whose story lines had been tremendously popular during the war years due to their direct "participation" to the conflict, but were failing to adapt to the post-war realities[9].

In 1956, after a seven years hiatus, DC editor Julius Schwartz relaunched the Flash, in *Showcase #4* (Figure 2.13), the first in a series of notable superhero **reboots** such as those of the Green Lantern (*Showcase #22*, 1959), Hawkman (*The Brave and the Bold #34*, 1961) and The Atom (*Showcase #34*, 1961). The character reboot technique and its successful application to the Flash gave new life to the American superhero genre

[9]DC's publishing policy with respect to Superman and Batman, who you might recall, were supporting America's war effort without actually getting involved in the combat, paid well in the post-war years, when their characters could easily be adapted to other type stories. Superman in particular played once again on his original light hearted humour.

and marked the beginning of the period known today as the Silver Age of superheroes.

The greatest boost to the reviving superhero genre was likely given by DC's introduction of The Justice League of America, in *The Brave and the Bold #28* (March 1960, Figure 2.14a), marking the reappearance of superheroes as a team (Wonder Woman, Flash, Green Lantern, Aquaman, Martian Manhuter and occasionally even Batman and Superman). To some extent, the Justice League too was a reboot of DC's Golden Age Justice Society of America. Comics scholars speculate that the success of such concepts is partly linked to the social context of late 1950s in America and in particular the increasing belief that organisations of any kind (government, military, companies) are much more effective than the individual alone. As a matter of fact, the Justice League became so popular that six months following its launch it got its own comic book title.

(a) **(b)**

Figure 2.14.: Teams of superheroes were instrumental in DC's and Marvel's efforts to relaunch the genre. Their success greately contributed to the Silver Age of superheroes.
a *The Brave and the Bold #28* (Mar. 1960, DC Comics), the first appearance of the Justice League of America. Art by Mike Sekowsky and Murphy Anderson. Retrieved: Grand Comics Database, Fair use.
b Cover of *Fantastic Four #1* (Nov. 8, 1961). By J. Kirby (art), S. Lee (writer). Source: Marvel, Fair use.

At the same time, Marvel was slowly trying to recover from their post-war downfall and in November 1961 released *Fantastic Four #1* (Figure 2.14b), their response to DC's superhero team concept. The Fantastic Four were the result of a fruitful cooperation between author Stan Lee and illustrator Jack Kirby, the team that would be at the origin of many other great Marvel superheroes:

- Hulk, in *The Incredible Hulk #1* (May 1962),
- Thor, in *Journey into Mistery #83* (August 1962),
- Iron Man, in *Tales of Suspense #39* (March 1963),
- X-Men, in *X-Men #1* (September 1963),
- The Avengers, in *The Avengers #1* (September 1963).

The creative duo and their working methodology would prove so successful, that Marvel would adopt it as a staple technique – "the Marvel method" and encourage its use by all their authors and artists. One such example is Spider-Man, the result of the collaboration between Stan Lee and Steve Ditko, which first appeared in *Amazing Fantasy #15* (August 1962). But there was more to Marvel's success than just the creative method. All Marvel superheroes are a product of science (or science hazard for that matter): the Fantastic Four had been exposed to cosmic radiation, the Hulk is the result of an experiment gone wrong, X-Men are mutants, Iron Man is a bionic man, Spider-Man had been bit by a radioactive spider and the examples can go on... Their appeal stems from a unique combination between their hazardous origins, reflective of the curiosity and fear of American society towards science, as well as its fascination with science-fiction, and the social allegories that these groups of superheroes embody:

- The Fantastic Four is a **family** of misfits,
- Spider-Man is an allegory of **adolescence**,
- The X-Men represent a **school** of misfits and exiles,
- The Avengers are the allegory for **corporations**.

Indeed, Marvel's superheroes are intriguing not because of their superpowers, but because of the common "problems" they have, which make them humans at least in the same measure as they are super-humans. Dealing with all sorts of problems (human and super-human), Marvel sets the new post-war standard of bundling stories of different genres into the same issue. If the Silver Age was ignited by the revival of superheroes by DC, it is Marvel which breathes new life into comics, activates a new reading audience and will eventually drive comics to the heights of American entertainment industry.

Underground Comix

Shortly after his testimony and the fiasco during the Senate Hearings on Juvenile Delinquency of 1954, William Gaines, the owner of EC, gradually lost interest in their superhero comics publications and focused his efforts in promoting *Mad*, EC's satirical magazine whose staple approach was to question everything and leave all endings possible.

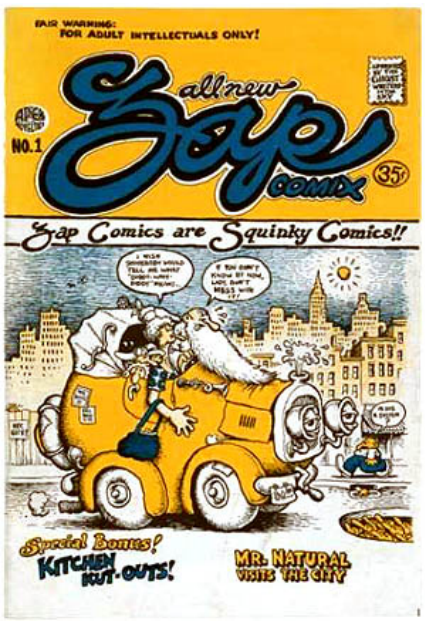
Not playing by the CCA rules, the magazine[10] was at the origins of what would become the second important reaction to the Code – the underground movement.

The precursor of underground "comix" was *Zap* magazine. Debuting in 1968, its first issue kicks off with its famous cover imagery - "Mr. Natural Visits the City". *Zap* is challenging the authority from the first lines of its cover, mocking CCA's seal of approval with its own inscription: *"Fair Warning: For Adult Intellectuals Only"*. In one of the stories of *Zap #1* (Figure 2.15a), creative director and later underground-symbol Robert Crumb introduces "Whiteman". This story is a warning about the shackles in which the 1960s individual (and art itself!) have been constrained by societal standards and expectations. It is also a parody of stereotypes: Crumb tackles racist issues by "having fun" with and making fun of racist stereotypes. Superheroes don't escape his sharp pencil either, as the stereotype of their alternate identities is satirised through Whiteman's inability to conceal his lust (here, his true identity). Underground "comix" and their anthologies (*Feds'N'Heads, Bijou Funnies, Young Lust* – see Figure 2.15 – are representative for the counter-culture movement of the 1960s.

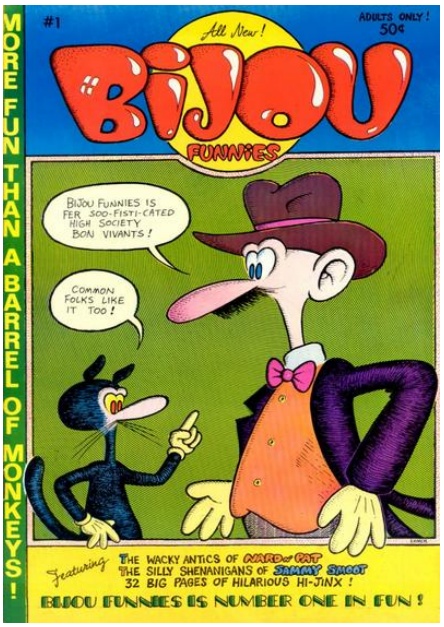
By the late 1960s the virulent line of underground "comix" finally got their editors into trouble and the movement started to fade away after a controversial incest story, published by the same Robert Crumb in *Zap #4* (1969). But the seeds were planted and their legacy was carried forward by publications such as *Wimmen's Comix* (1972-1992), the most important all-female "comix" magazine, which under the direction of Trina Robbins – herself one of the key figures of the underground movement, went beyond feminist subjects and took on issues such as sexuality, politics and the relationship between art and social change. Underground is also responsible for cultivating the autobiographic and confessional genres in comic books. One of the greatest achievements in this sense is Justin Greene's disturbing, yet soothing, *Binky Brown Meets the Holy Virgin Mary* (1968), a sensitive confessional story about growing up while dealing with Obsessive Compulsory Disorder (OCD) and the neurosis he developed as a child receiving a strict Catholic education. In 1975, Art Spiegelman and Bill Griffiths launched *Arcade* – "a life raft for everybody [from the dying underground movement] could climb onto", a magazine hosting "the best of old and new comics". Despite its short life, this anthology featured some of the greatest creative minds of the 1960s and symbolised an important evolutionary step towards the alternative comic books of the 1980s.

By the 1970s, the adolescent readers from the 1960s (primarily affected by the introduction of the Code) were mature people who had grown up with the challenging creative endeavours of the underground movement, thus developing different reading needs from those of the Silver Age generation. This is the context in which Will Eisner, made his return to comics authorship. Here is his own account of his participation at the 1971 *Comic Art Convention* in New York City, where he met members of the underground movement:

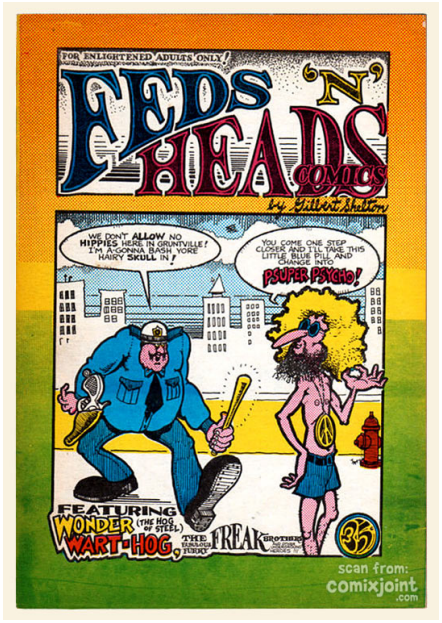
[10] *Mad* is still in circulation today, making it one of the longest lasting magazines.



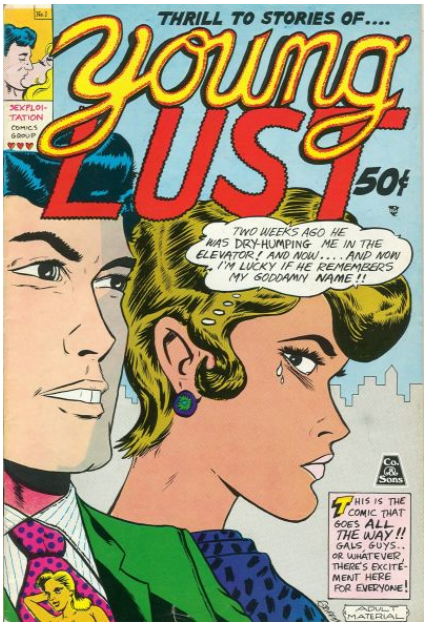
(a)



(b)



(c)



(d)

Figure 2.15.: Covers of some of the most important underground "comix". Although short lived, most of them played an important role in the development of the movement and subsequent evolution of comics.
a Mr. Natural visits the City – The famous cover of *Zap Comix #1* (Feb. 1968, Apex Novelties). The flagship magazine of the "comix" underground movement with its own "seal" reading *Fair warning: For Adult Intellectuals Only*, an irony to CCA's seal of approval. Art by R. Crumb. Source: "Zap Comix #1" at The Grand Comics Database. Retrieved Feb. 3, 2009., Fair use.
b Cover of *Bijou Funnies #1* (summer 1968, Bijou Publishing Empire), one of the earliest underground "comix". Artwork by J. Lynch. Source: (WP:NFCC#4), Fair use.
c Cover of the only issue of superhero parody "comix" *Feds'n'heads* (Feb. 1968). Just like *Zap*, it features a seal-resembling notice of its own *For Englightened Adults Only!*. Art by G. Shelton. Credit: `http://comixjoint.com/fedsnheads-1st.html`, Retrieved Nov. 25, 2016, Fair use.
d Cover of satirical "comix" *Young Lust #1* (Oct. 1970, Company & Sons). The title itself is a parody of 1950s romance comics *Young Love*. Art by Bill Griffith and Jay Kinney. Source: (WP:NFCC#4), Fair use.

> "I came back into the field because of [convention organiser] Phil Seuling. I remember [him] calling me in New London, [Connecticut], where I was sitting there as chairman of the board of Croft Publishing Co. My secretary said, "There's a Mr. Seuling on the phone and he's talking about a comics convention. What is that?" ... I came down and was stunned at the existence of the whole world. [...] That was a world that I had left, and I found it very exciting, very stimulating. I went down to the convention, which was being held in one of the hotels in New York, and there was a group of guys with long hair and scraggly beards, who had been turning out what spun as literature, really popular "gutter" literature if you will, but pure literature. And they were taking on illegal [sic] subject matter that no comics had ever dealt with before. [...] I came away from that recognising that a revolution had occurred then, a turning point in the history of this medium."

Will Eisner had already made his mark during the Golden Age of comics with his masterful working of the page and his unique narrative style in stories such as *The Spirit*. Returning to comics in the early 1970s he had decided to make use of his craftsmanship by experimenting with longer, more sophisticated storytelling, and in October 1978 published *A Contract with God*, largely accepted as the first modern graphic novel. While this assertion is debatable (we have seen that similar forms of storytelling had been promoted in the 1920s by Rodolphe Töpffer and Lynd Ward), Will Eisner is undoubtedly responsible for refining the technique, choosing more complex subject matter as well as the form which led to the critical acclaim that would help establish comics as a literary art form. Another fine example in this sense is *The Dreamer* [Eisner, 1986], "a graphic novella set during the dawn of comic books", in which Eisner borrows from the autobiographical exercise of the underground movement, which admittedly had prompted his comeback as a writer and combines it with his special gift for historical accounts. From a graphical perspective, *The Dreamer* is a wonderful exhibition of Eisner's craftsmanship, in the way he works the page as a compositional unit, playing with its elements and putting them in the service of the story. His borderless panels loosen up the page grid without complicating the composition of the page. Instead, they reveal a sense of flow which help the story achieve what all great comics do – engage the reader with the story (see Figure 2.24).

The 1986 Trifecta

1986 is a special year in the history of comics. It is the pinnacle of almost a century of evolution, the year when comics truly come of age through the famous "trifecta": Art Spiegelman's *Maus*, Frank Miller's *Batman – The Dark Knight Returns* and Alan Moore's *Watchmen*.

Maus. "A quiet triumph, moving and simple – impossible to describe accurately, and impossible to achieve in any other medium but comics"; this fragment from *Washington Post* encompasses the magnitude and complexity of Art Spiegelman's Pulitzer Award winning graphic novel. Art Spiegelman acknowledges his friend's Justin Greene graphic novel *Binky*

Brown Meets the Holy Virgin Mary) as the source of inspiration for the confessional exercise that *Maus* represents. The story itself is built on two levels: the confessional story by Art's father, Vladek, a holocaust survivor and their complicated father-son relationship. The epigraph of the novel is a quote by Hitler: "The Jews are undoubtedly a race, but they are not human" which Spiegelman uses as a premise and tries to deconstruct – "*reductio ad absurdum*" – by the end of the novel. Thus, he creates (one of) the most powerful (and controversial) allegories in the comics medium by drawing his characters as "humans who look like animals" (the Jewish are mice, the Germans are cats, the Polish are pigs and the Americans are dogs). The story is so deep and personal, so tragic and brutal that the author needs this allegory to explain the unexplainable (Figure 2.16).

Figure 2.16.: "Humans who look like animals" – the main allegory driving graphic storytelling in *Maus*. In these panels the author draws attention on the absurdity of categorising people and creating racial-based stereotypes. In the end, he acknowledges that "the metaphor self-destructs".
Source: *Maus*, Art Spiegelman [Spiegelman, 2011, p. 210], Fair use.

Connecting the two levels of the story, we witness Spiegelman's effort to document and write *Maus*. The novel is written in front of us and becomes a quest by the author to connect with his past and to investigate his role as an artist. Spiegelman is afraid that "reality is too complex for comics" and that details risk of being distorted or left out. He admits he is confused when choosing an animal to draw his Jewish-converted French wife (is she a frog – the metaphor for the French – or a mouse, since she converted to Judaism?) or for those Germans who were imprisoned themselves (are they cats, just like the persecuting majority or... mice, like all those persecuted?). In the end, Spiegelman acknowledges the self-destruction of the *Maus* allegory and with it, he points out to the absurdity of classifying people along racial stereotypes.

In the years following its publication, *Maus* has received critical acclaim and has with it an impressive number of awards, including a special Pulitzer Prize in 1992. This was the first time a production of the comic medium was recognised by the literary community. Shortly after its complete publication, *Maus* became the first comic book to be featured in the pages of the *New Yorker*. Comics scholar William Kuskin reflects on the importance of this gesture and the way in which it underlines this turning point in comics history:

> "If comics derive authority from literary culture, Spiegelman changed that, installing a new comic sensibility into the heart of literary culture itself – the New Yorker."

Frank Miller's **The Dark Knight Returns**. During the 1970s, DC and Marvel's superhero productions saw a relative decline, mostly due to the demographic change of their readership which, as we saw, had by then reached adulthood and were subsequently seeking different type of comics content. In this context, Frank Miller's merit is that of revitalising the superhero genre and consolidating comics' position in mainstream culture through his involvement in a number of subsequent cinema projects (e.g., *Sin City, 300*). Throughout

Figure 2.17.: The cover of Book Two of *The Dark Knight Returns* series – "Dark Knight Triumphant". Towards the end of Book One, Bruce Wayne finally embraces his identity as Batman, with repercussions in both the story and its artistic delivery. From this point on, Batman is committed to his mission of defending Gotham at all cost, even if this transforms him into a near-villain. The brutality of his actions is enhanced by the rugged "big-foot" drawing style adopted by the artist.
Source: *The Dark Knight Returns*, Frank Miller [Miller, 1986, p. 78], Fair use.

the four parts series, Frank Miller recovers Batman from his past incarnations: the parody style of the 1960s (a rich playboy running around in his tights) or the naturalist drawing style of the 1970s. He gradually transitions to the "big-foot" drawing style (Figure 2.17), whose rugged expressiveness is fully accomplished in the final two parts, in which the reader

is witness to a cruel, brutal, and bitter Batman. The author delivers us a hero that has no problem in breaking his own code of honour and make use of fire guns to deliver *his* justice.

The Dark Knight Returns is not the typical Batman story. In it, Frank Miller explores the graphic possibilities of the medium and ties together several sensitive themes, some of them strikingly similar to those of *Maus*: the loss of parents and the discovery of the (ageing) self.

The Watchmen. The final piece of the 1986 comics "trifecta" is Allan Moore's *Watchmen*, a comic book about comics and at the same time an allegory of their history, confirming the coming of age of the superhero genre. For instance:

- Hollis Mason, one of the characters, admits he became a superhero after reading the Superman comics in 1938, the year in which Superman appears in real life;
- *Watchmen* follows the story of a group of heroes forming the Minutemen (a reference to the Golden Age of comics) and their progression to becoming the Crimebusters (a reference to the Silver Age of superheroes);
- also, the action takes place around the fictional Keene Act – a bill prohibiting the activities of all superheroes or masked vigilantes that were not affiliated with the U.S. government. This act echoes the real life Comics Code Authority and its impact on comics;
- in a wonderful exercise of pushing the boundaries of the form, Allan Moore incorporates a comic book within the *Watchmen* universe. In the series, *Tales of the Black Freighter* is authored by a certain Joe Orlando, a masterfully crafted reverence to Joe Orlando, the real life illustrator and comics editor for EC and later DC Comics.

Watchmen are more than references to comics history; they draw their essence and their inspiration from it. Initially, Allan Moore wanted to recycle a series of 1946 superhero characters that DC had acquired from Charlton Comics in 1986. His trial story had a good reception, but DC editors saw more potential for these characters in other stories, leading to Moore eventually developing a new team of superheroes. When unmasked, their alternate persona are not "real" people (in the sense that Superman is Clark Kent or Batman is Bruce Wayne), but reminiscences of other superheroes (those of Charlton Comics).

With Will Eisner's revival of the graphic novel and the 1986 "trifecta" comics reach their maturity as a medium and receive the much sought critical recognition for the skillful use of their format in approaching complex subject matter. On the long term, this period's legacy would be the development of a new generation of artists who will continue to push the boundaries of the medium in the 1990s and 2000s.

One of these is Joe Sacco, the author of a series of journalistic investigations, taking the form of comic books. In his 1996 *Palestine*, Sacco reports the Israeli-Palestinian conflict from a perspective less known to the wide public – that of the Palestinians. Sacco tells

real stories about ordinary people, without the claim of uncovering the truth. At a first glance, the sheer reality of these stories and their intensity could seem incompatible with the comics medium. How could one use cartoons to depict centuries old problems? Just like Art Spiegelman in *Maus* solves this apparent inadequacy by using the allegory of animals, so does Sacco by employing parody. His irony is present in the way he depicts real people (disdainful Israeli soldiers, screaming Palestinians, even the curious affection Palestinians have for serving tea), including himself (he almost vilifies his desire to capture stories no matter what) and is strengthened by his fascination with their faces. If animals are Spiegelman's narrative vehicle, Joe Sacco tells a story of faces. He "hunts" them everywhere and doesn't shy away to depict all their expressiveness and the range of human experiences they are able to convey. His graphic style doesn't desist from cultivating the same irony, by employing a gargantuan style (much like Frank Miller in the last part of *The Dark Knight Returns*) and exaggerating features. Nevertheless, his deep, dense line style and narrative richness immediately remind the reader of the depth and density of the stories that he's telling. In the end, *Palestine* leads nowhere; it has no resolution much like the complicated history between Israeli and Palestinians. It is nevertheless, a compelling account about national security and the dangerous parallel worlds it creates, for such divided worlds indeed lead nowhere.

Figure 2.18.: The cover of *Ultimates #13*, art by Brian Hitch. Notice the modern, dynamic and vividly coloured drawings depicting the characters in cinematic postures.
Credit: Wikipedia, Fair use.

The 1986 "trifecta" also proved the potential of the superhero genre in developing great stories around complex characters. In 2002, British creative duo Mark Millar and Bryan Hitch authored *The Ultimates*, a modern take on *The Avengers* and an editorial effort by Marvel to solve the narrative incoherence of their superhero universe. But beyond that, *The Ultimates* is a story about the transformations in the American society and a critique of the overwhelming media entertainment. Millar's heroes are much more scarred than Moore's *Watchmen* as they find themselves becoming the icons of a world that finds it increasingly more difficult to distinguish between celebrity and news. *The Ultimates* are promoting hedonism (be gorgeous, be beautiful, be famous and especially be angry!) and in their actions show carelessness to all consequences (be rude, be arrogant!). Its graphical realisation is equally impressive. Millar's worship of the cinema reflects both in his "movie star" characters, as well as in the structure, composition and content of his panels, which mirror the wide screen. *The Ultimates* features large, action-packed panels and an *agon*

of suspended animation (much like in the film production of *Matrix*) in a burst of aggressive colours (see Figure 2.18).

As comics scholar William Kuskin notices, *The Ultimates* offers both a critique and a message. It is a view over American culture and a key to being more understanding to other cultures.

We conclude our overview of comics and their history with one of the most sensitive and truly brilliant exploits of the past decade – Allison Bechdel's graphic memoir, *Fun Home*. Subtitled *A Family Tragicomic*, the story is deeply influenced by the feminist comics from the 1970s and explores issues of paternity and artistic creativity, just like 1986 *Maus*. Because of this, at a first glance, it is hard to believe that this graphic novel has only appeared as late as 2006. *Fun Home*[11] is a game of hide and seek, a layered plot alternating

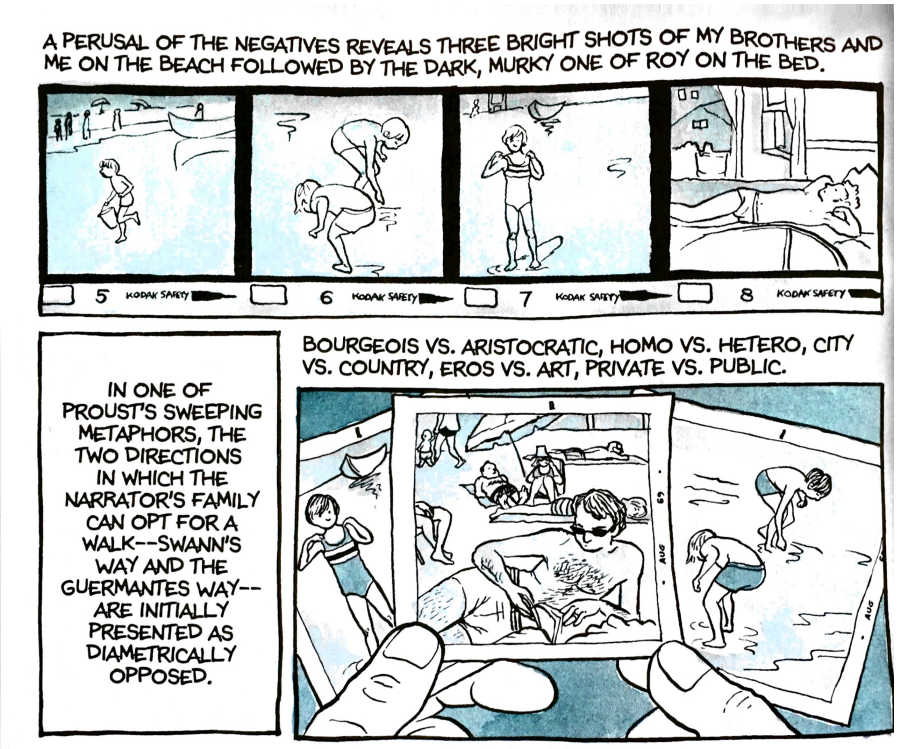


Figure 2.19.: An illustrative page of Allison Bechdel's graphic style and original use of photographic art in *Fun Home*. Source: *Fun Home*, Allison Bechdel [Bechdel, 2006, p. 102], Fair use.

between past and present, between Bechdel's self-discovery and the discovery of her father's secret life. The story is triggered by the suicide of Allison's father and goes on to explore sensitive issues such as sexual orientation, gender roles and at the same time, it explores the role that literature and art play in shaping one's self. *Fun Home*'s narrative vehicle is a simulacrum of photos that superposes the gradual uncovering of a dysfunctional family. If

[11] A pun on "funeral home" – her father's business.

photographic play is key to the narrative (Allison takes photos or discovers photos taken by her father – Figure 2.19), then photographic art is defining Bechdel's graphic style. Indeed, some of her panels are actual reproductions of personal photos and they appear as revelatory pieces in her quest to uncover her father's real identity, posthumously.

Bechdel's graphic style seems influenced by the "underground" and alternative movements from the 1970s. Sean Wilsey notices her drawing style combines "the detail and technical proficiency of R. Crumb with a seriousness, emotional complexity and innovation completely its own" [Wilsey, 2006].

We conclude here our short journey through the history of comics. We tried, to the best of our abilities and taking into account the limited space, to give our reader a good understanding about the origins of sequential art, from its most ancient forms to its evolution into comics. We placed the development of the medium in its historical context, thus understanding how comics and sequential art have always had the intrinsic quality of accurately describing events and (in modern times) changes in our society. Comics have grown together with their readers and in the past four decades not only that they came of age, but they matured into a complex medium, capable of more than entertainment, they broke the barriers of their own medium and those of different forms of authority – governmental and lately those of literature and academia. Comics need no more to convince us of their seriousness, it is time to treat them as such!

2.2. Reading and collecting comics

2.2.1. Elements of the comics medium

We resume our journey through the world of comics and sequential art and in this section we discuss the theoretical grounds for creating comics. We will discover the complexities behind the design of a panel, the way in which multiple panels compose a page structure and the role of the gutter between two panels. We will discuss the graphical means for expressing time, movement and space, as well as the drawing cues allowing artists to convey sensations that go beyond the visual. While the previous section provided with a well documented, historical view on why comics should be studied, the present section lays the theoretical foundation for our future work in encoding and studying the comics medium.

To understand why comics are a medium, we turn to [Grossberg et al., 2005] and their framework which defines media across three dimensions: **the technology** used to experience media, **the institutional context** that produces and disseminates the media and **a cultural form** (i.e., formats, structures, ways of telling stories), in one word the "product" of a media organisation. Looking at comics through this three-way lens, all aspects certifying them as a media are definitely there: the print or digital technology, used for their production and distribution by the publishing houses or self-publishing authors. The

final aspect, as [Gordon, 2016] remarks, is a bit more complicated due to the specificity in which comics are perceived and assimilated in different cultures; this can be summarised by the ensemble of ideas that they cultivate in forms such as comic strips, comic books, *manga* magazines, graphic novels etc. [Grossberg et al., 2005] concludes:

> "What holds the three aspects of media – technology, institutions, and cultural forms – together, and what provides the unity of the concept of media, is the idea of mediation."

The comics medium is realised through a universal vocabulary, which we find across all cultural forms. At a macro-level, this vocabulary is formed by the juxtaposition of words and images, while the intended order of reading (the sequence) is achieved by dividing the story into smaller narrative units – **the panels**. This division is done by the artist and ideally depends only on the artistic intent; practically, it is limited by the spatial properties of the physical support (traditionally – the page, more recently – different types of screens). The composition of panels gives birth the **grid**, which superposes the page, and a by-product – **the ellipsis**.

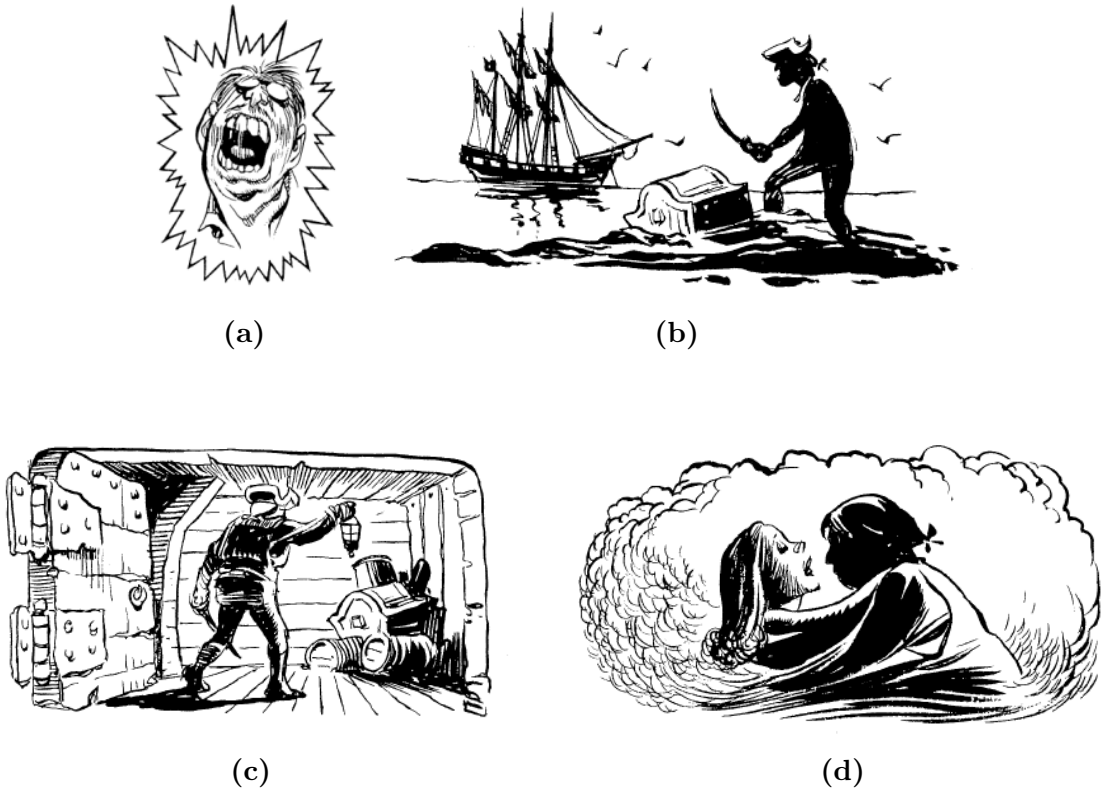
Panels

Panels are the core structural elements for developing the graphic narrative in the comics medium and a representative instance of the pact between the artist and the reader – that comics are based on common human experiences, which the author is able to render and the reader is able to recognise.

Panels are created by breaking down narrative events and encapsulating them to compose a narrative flow. This process, Will Eisner insists, should be the result of artistic intent and not that of a technological process. He also stresses several properties of the panels [Eisner, 1985, Chapter 4], which need to be taken into account when dividing and deploying them to form the grid and ultimately the book itself:

- **Panels are a vehicle for conveying time**. In comics, the reader experiences the present (in the box currently being "read"), but at the same time can glance at both the past and the future.
- **Panels are a means of controlling the reading act.** The point above shows a strength of the medium; however, if not properly mastered this strength can undermine the aforementioned pact between the artist and the reader. Eisner recognises that the greatest challenges for an artist is to deal with the tendency of the reader's eye to wander on the page. The solution is an addendum to "the pact" in the form of another cooperation between the artist and the reader which is based on the cognitive disciplines and the reading convention (e.g., left-to-right or right-to-left, top-to-bottom).
- **The border reinforces the message.** For instance, straight borders are associated with events in the present tense, a wavy edge can imply an action in the past, while

irregular borders are used to express or amplify strong emotions. Scalloped edges are typically used for expressing thoughts, dreams or things of the imaginary, whereas non-frames, although sensitive to integrate in the page composition, help convey a feeling of unlimited space and at the same time give fluidity to the narration (see Figure 2.20).



(a) **(b)**

(c) **(d)**

Figure 2.20.: The type of the panel border supports the action within.
a The jagged, irregular border implies an emotionally intense action.
b Border-less panels suggest an open, vast or limitless space.
c The border can sometime be a structure (window, door, portal) and confines the characters – the opposite effect of border-less panels.
d Wavy borders are used to depict thoughts, memories, imaginary or future actions.
Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 48-49], Fair use.

- **The emotional function of the panel.** Both the physical properties and the content within the panel can be used to trigger an emotional response from the reader – Figure 2.21.

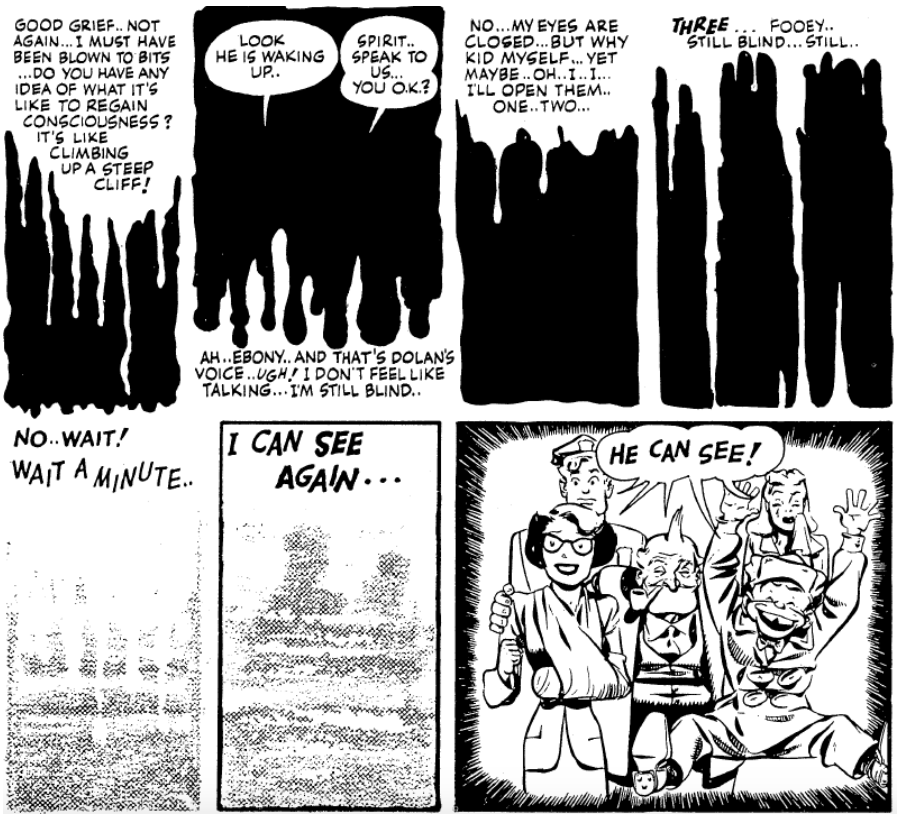


Figure 2.21.: In this sequence of panels from *The Spirit* story "Fluid X", we experience an intense emotional moment through the eyes of the character.
Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 61], Fair use.

The Ellipsis

The content of a panel is a frozen instance of a continuous action and represents the image that the artist deems suitable to engage the reader's sense of shared experience. The rest of the action, that which is not depicted in the panel completely depends on the reader, in what is certainly the most subtle, most curious and perhaps most powerful features of the comics medium. **The ellipsis** is the blank space between two panels, in which the imagination of the reader overrides the graphical or literary skills of the artist or the author. Scott McCloud rightfully calls this space "the blood in the gutter" (Figure 2.22).

Figure 2.22.: One of the best illustrations of the concept of "ellipsis" and the engagement it demands from the reader, for it is the reader who "drops the axe", and decides who screams and why.
Source: Scott McCloud, *Understanding Comics* [McCloud, 1993, p. 66], Fair use.

The ellipsis is both structurally and functionally dependant on panels. This means that the ellipsis is not drawn, but instead is the graphical and functional by-product of drawing two consecutive panels one next to the other: structurally it couldn't exist without their borders, whereas functionally the reader would have difficulties imagining anything without the content of the two consecutive panels. McCloud goes further and identifies different types of links between panels (see Figure 2.23).

1. Moment-to-moment (Figure 2.23a);
2. Action-to-action: follows a series of actions of the same subject (Figure 2.23b);
3. Subject-to-subject: focuses on the same idea or scene. The transition needs the reader's participation in order to be meaningful (Figure 2.23c);
4. Scene-to-scene: transports the reader over time and space. May require deductive the reader to engage in deductive reasoning (Figure 2.23d);

5. Standpoint-to-standpoint: generally bypasses time and focuses on different aspects of the same idea. By allowing the eye to wander, it requires more reading "discipline" from the reader and more craftsmanship from the artist (Figure 2.23e);
6. Continuity solution: no apparent relationship between the panels (Figure 2.23f).

(a) Type 1: Moment-to-moment

(b) Type 2: Action-to-action

(c) Type 3: Subject-to-subject

(d) Type 4: Scene-to-scene

(e) Type 5: Aspect-to-aspect

(f) Type 6: Continuity solution

Figure 2.23.: The six types of transitions between adjacent panels.
Source: Scott McCloud, *Understanding Comics* [McCloud, 1993, p. 70-72], Fair use.

The Page and the Grid

Page composition follows the fragmenting of the action and its framing into panels. The page comes together by taking into account the narrative flow, the reading conventions and obviously the message it wants to communicate (the action within, the mood it sets, the emotion it conveys). The physical properties of the panels (size, shape, border) dictate the rhythm of the narrative and establish what some are calling the "poetry of the page". This is achieved by gradually building up narrative tension: first from panel to panel, the tension is partially resolved at the end of a tier. Then, the "residual" tension building up by the end of the page is resolved to a certain extent by the reader's gesture of **turning the page**.

The panels thus come together to compose a grid-like structure that superposes the page. Will Eisner was the first to award special attention to the grid, by using it beyond its basic structural role and experimenting with its potential to support and blend in with the narrative itself.

A special case is that of the **meta panel**, that provides structural support by encapsulating a number of regular panels on the same page and offering them both graphical and narrative background. This allows for setting different dimensions of the stage and developing multi-levelled actions. Figure 2.24 shows a meta panel from Will Eisner's graphic novella *The Dreamer*, depicting the scene in which Billy Eyron (the author's persona in the book) and his associate Jimmy Samson select artists for their new comics workshop. We are in the early of comics, most authors work for hire, and Eyron and Samson decide to start the first in-staff studio, where artists receive salaries. Besides being a historic "document", the page has deep artistic value: it puts in perspective Samson's pragmatic approach to producing more comics against Eyron's fear that their enterprise is becoming a comics "sweatshop". This example helps us get across two main points: first, the intricacy and creative power of meta-panels and second, the role of the grid as both a structural and narrative element. Indeed, the featured page is "a poem of the grid", about the ambivalence of its structure, as both a necessary container (in one panel, Eisner is drawing the panels himself) and a limitation to the form (the presence of the wall and windows, which almost imprison the artists working in a space resembling to a comics "sweatshop"). At the same time, this page is an allegory of the ambivalence of the Golden Age of comics, when copyright issues were rarely solved in favour of the creators, who were essentially the underpaid "dreamers" of an emerging art form.

This theme – the limitations of the comics form – is further echoed by the authors of the 1986 "trifecta" who likely became aware of it as a result of their dedication to experimenting and pushing the boundaries of the comics medium.

Figure 2.24.: A wonderfully complex page from the graphic novella *The Dreamer*, an account of Will Eisner from the early years of American comic books. The page seeks to find the balance between its freedom as a meta-panel with its seemingly borderless panels and the dissimulated references (the window, the wall, even the panels drew by Eyron) to the structural and artistic limits of the panel.
Source: Will Eisner, *The Dreamer* [Eisner, 1986, p. 23], Fair use.

One of Art Spiegelman's innovations in *Maus* is to have his characters interact with a comic story – *Prisoner on the Hell Planet* [12] – that is drawn within the actual graphic novel. If *Maus* is a story of complicated paternity, *Prisoner on the Hell Planet* is contracted account of the role Artie's mother had on his shaping as an artist in his quest for truth. The story's last tier (Figure 2.25) refer to the consequences of her actions for the author's persona, which himself trapped in a prison. At the same time, just like Eisner in *The Dreamer*, this is an allegory of the limits of the form, as the grid identifies itself with the prison.

Figure 2.25.: The last tier from *Prisoner on the Hell Planet* – the comic story enclosed in *Maus*, Art Spiegelman. A wonderfully built allegory of both the complicated origins of *Maus* – paternity and maternity fuel Spiegelman's artistic endeavour – as well as a reminder of the limits of panels and the comics medium.
Source: *Maus*, Art Spiegelman [Spiegelman, 2011, p. 105], Fair use.

In the same way, in the first part of Frank Miller's *Dark Knight Returns*, a troubled, ageing Bruce Wayne struggles with his urges to fight crime. The final scenes echo the author's reclaim of the superhero from its farcical, decaying past, as Wayne gives in and decides to re-become the Batman. At the same time, a monstrous bat breaks through a grid-looking window (Figure 2.26), an anticipation of the new rule-breaking Batman and at the same time an allegory of the new age that comics were stepping in.

[12] In this story, he gives up on the animal allegory that he uses throughout *Maus*, unmasks his characters and allows them to assume their human faces.

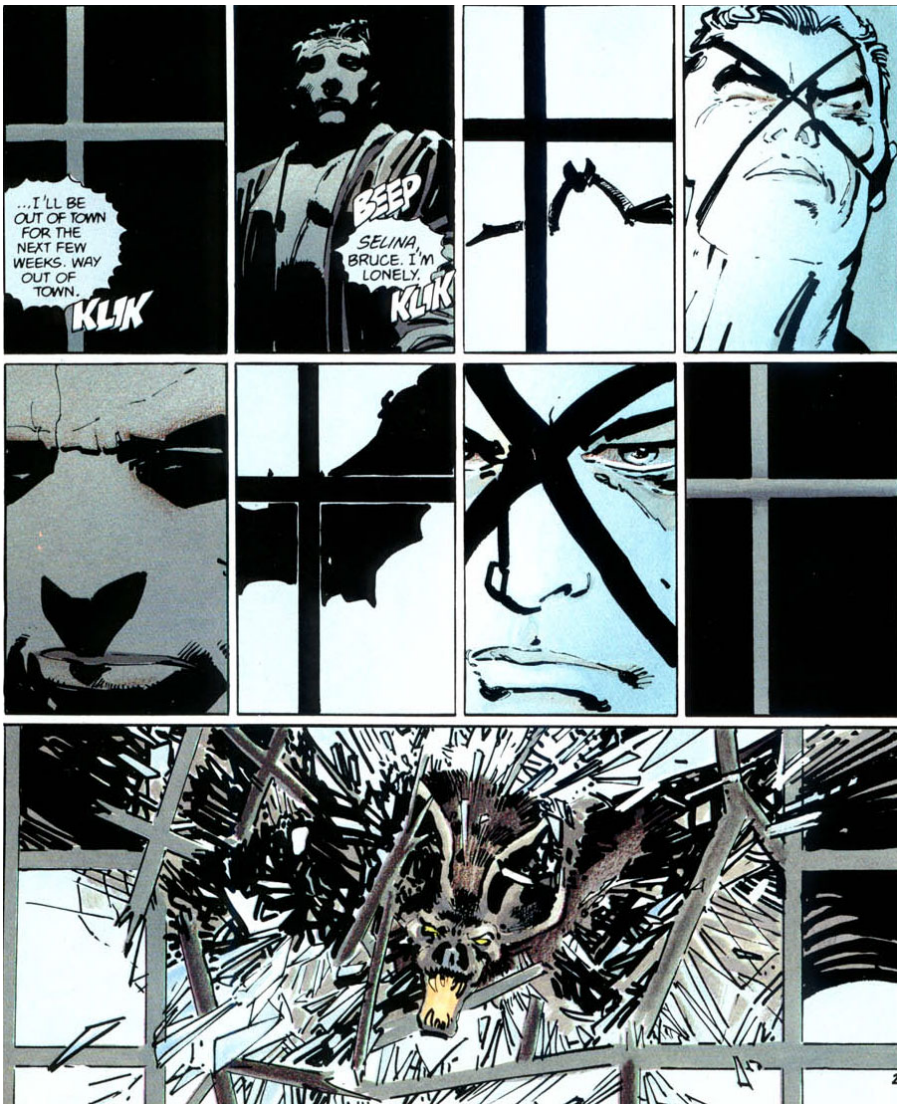


Figure 2.26.: The catalyst page in Frank Miller's Book One of *The Dark Knight Returns*. The bat breaking through the window symbolises the end of Bruce Wayne's struggle to accept his role as the Batman, protector of the law. At the same time, the bat is breaking the grid, a metaphor for Miller's reinvention of the Batman.
Source: F. Miller, *The Dark Knight Returns* [Miller, 1986, p. 26], Fair use.

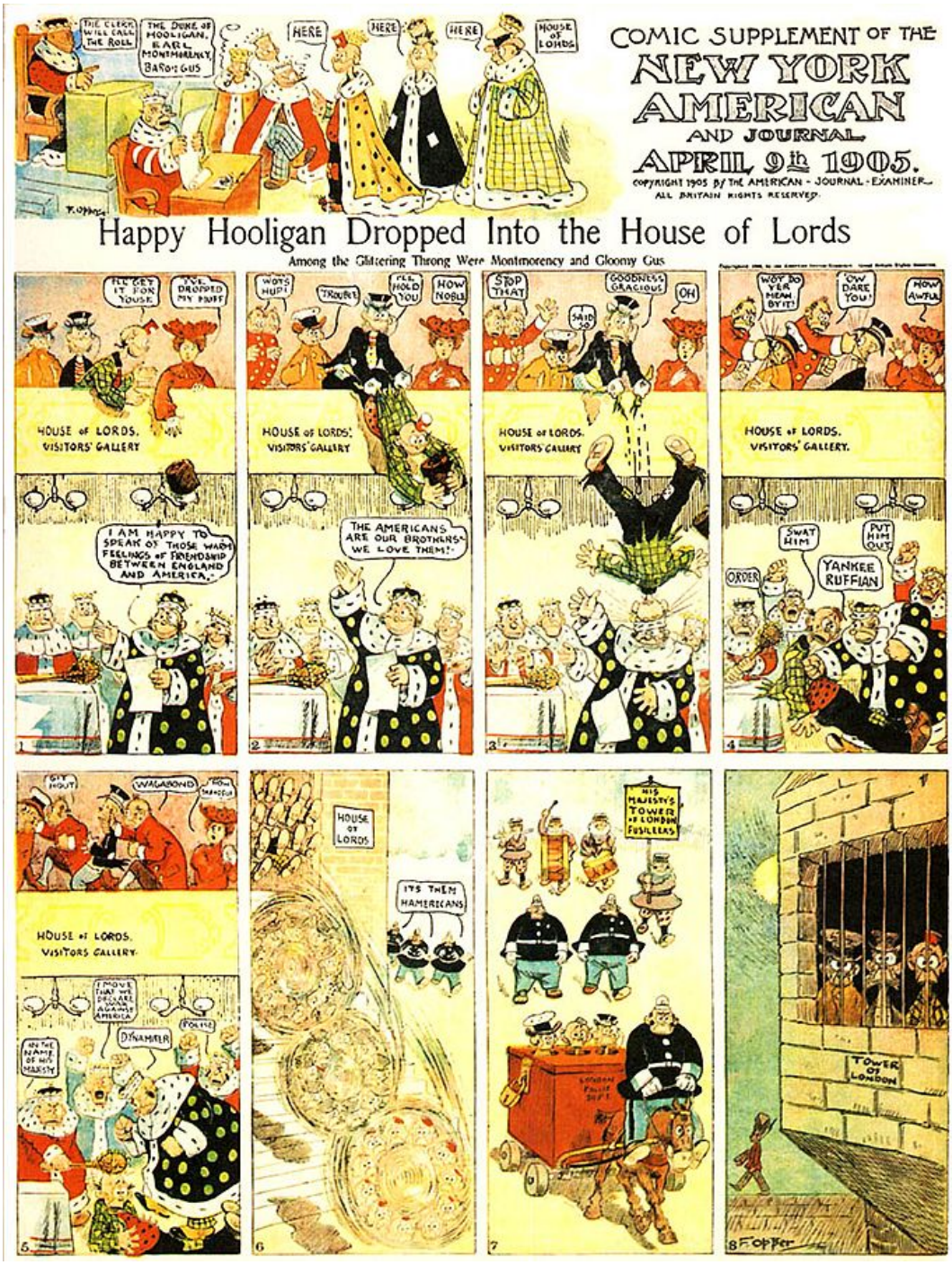



Figure 2.27.: An early use of elements of the main comics medium (panels, balloons, motion lines) in F. Opper's *Happy Hooligan* (1905). Copyright 1905 American Journal-Examiner – marked at top right portion of page. Source: Frederick Burr Opper, *The Happy Hooligan* (Apr. 5, 1905) Retrieved: http://chnm.gmu.edu/aq/comics/spectlex1.html#spect21, Public domain

Balloons

The formal properties of comics (panels, balloons and motion lines) became a convention in American newspaper comic strips around 1901. Figure 2.27 gives an example of one of the earliest modern uses of all these elements of the medium.

Writing in comics doesn't have the explanatory, nor the descriptive role, as in the case of figure captions for example. In comics, the writing complements the graphic by providing a "soundtrack" for the image, in the same way in which sound has added a new dimension to cinema.

Writing in comics occurs in two forms: **balloons** and stylised **onomatopoeia**. Balloons are, together with panels, the most widespread and versatile elements of the medium. Their versatility resides in their capacity to render not only speech, but also abstract notions such as emotions, sensations and very often, thoughts. The styling also plays an important role in the way in which the message is intended to be received by the reader. For instance, when the imagery is too spectacular (and thus distracting) or when the author simply wants to stress the importance of certain words within a balloon, these will usually be highlighted (in bold or italic letters). Figure 2.28 presents two interesting instances of styling the letters in Spiegelman's *Maus*.

Figure 2.28.: Two interesting instances of styling the letters in *Maus*. Notice word telegraphing and the special styling of the letter "S" resembling that used by the Nazi SS.
Source: *Maus*, Art Spiegelman [Spiegelman, 2011, p. 176], Fair use.

First, notice how ideas are underlined by highlighting certain words (**COMIC STRIPS, COMPLEX** are bolded), a technique which Eisner calls "telegraphing". On a more subtle note, throughout all the novel, the styling of the letter "S", especially instances of double-"S" (e.g., "I GUE**SS** I BIT OFF MORE THAN I CAN CHEW." in the second panel) resembles to that of the Nazi *SS* service, a constant reminder of how history drives this story, even in panels which evoke a present action. Just like the border of a panel conveys the emotional

Figure 2.29.: A few of the most common types of balloons: speech, thought, caption, scream.

dimension to the scene it contains, so do balloon lines give different meaning to the type of communication within (Figure 2.29). For instance, a round-shaped balloon is used for dialogue, a scalloped balloon typically expresses thoughts or unspoken communication, whereas irregular balloons are reserved for loud communications or broadcasting. A special case is that of captions, which express the intervention of the narrator and may or may not be enclosed within (usually) rectangular areas.

Sometimes, the personality or particular traits of a character are captured by the special styling of a balloon (i.e., a special font, a certain background colour, a particular border etc.), as is the case in Figure 2.30.

Characters, places, events

Amongst elements of the comics vocabulary, **characters** are especially endowed with the capability of reflecting on shared experiences. This is achieved through the drawing of the characters at a high degree of abstraction, with the purpose of allowing the reader to substitute him- or herself to the character, thus getting more immersed in the reading experience. This abstraction process has been theorised by Scott McCloud as he explains the transition of images from realistic to iconic representation. Figure 2.31 describes precisely this process of taking a realistic photographic image and abstracting it (by stripping away

details, conceptualising forms, exaggerating features, etc.), but without losing its basic meaning.

Figure 2.30.: In this panel from *The Avengers #32* (2000), the android's speech balloon has a different styling: the straighter shape, the metallic colour and the particular type of border are all suggestive of its mechanical nature.
Source: *The Avengers #32*, K. Busiek & G. Pérez (2000), Fair use.

In comics, just like in any other graphic art, characters are defined by two dimensions: the face (and its visual expression) and the body (and the combination between posture and gesture). [Eisner, 1985, Chapter 5] recommends artists to work with a limited "dictionary" of human gestures (Figure 2.32) that are easily recognisable by the reader.

Concerning body posture, Eisner stresses that the challenge for the artist is to choose from a sequence of many related movements, the exact posture that renders the narrative eloquent enough, such that the reader can infer the previous movements leading to it (see Figure 2.33).

When properly employed, these elements (posture, gesture, visual expression) put together are able to transmit to the reader much more than simple text would. Thus, when juxtaposed to imagery, "words add depth of meaning and nuance" [Eisner, 1985, chapter 5].

To some extent, **places** are subject to the same process of simplification, without losing their ability to be recognised by the reader to the extent intended by the artist. For instance, the image in Figure 2.34 is able to convince the reader (the mosque, the street landscape, the people's clothes) that the action takes place in no other place than Istanbul, Turkey.

Obviously, sometimes this is done with the help of text, either through the intervention of a character (in a speech or thought balloon) or that of the narrator (through caption text). Other times, the author doesn't want to specifically locate the action in a named

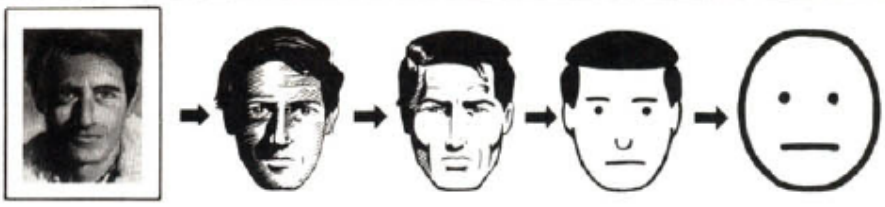


Figure 2.31.: The abstraction process of a drawing, from photographic realism to a simple, commonly understood representation of a human face.
Source: Scott McCloud, *Understanding Comics*, [McCloud, 1993, p. 29], Fair use.

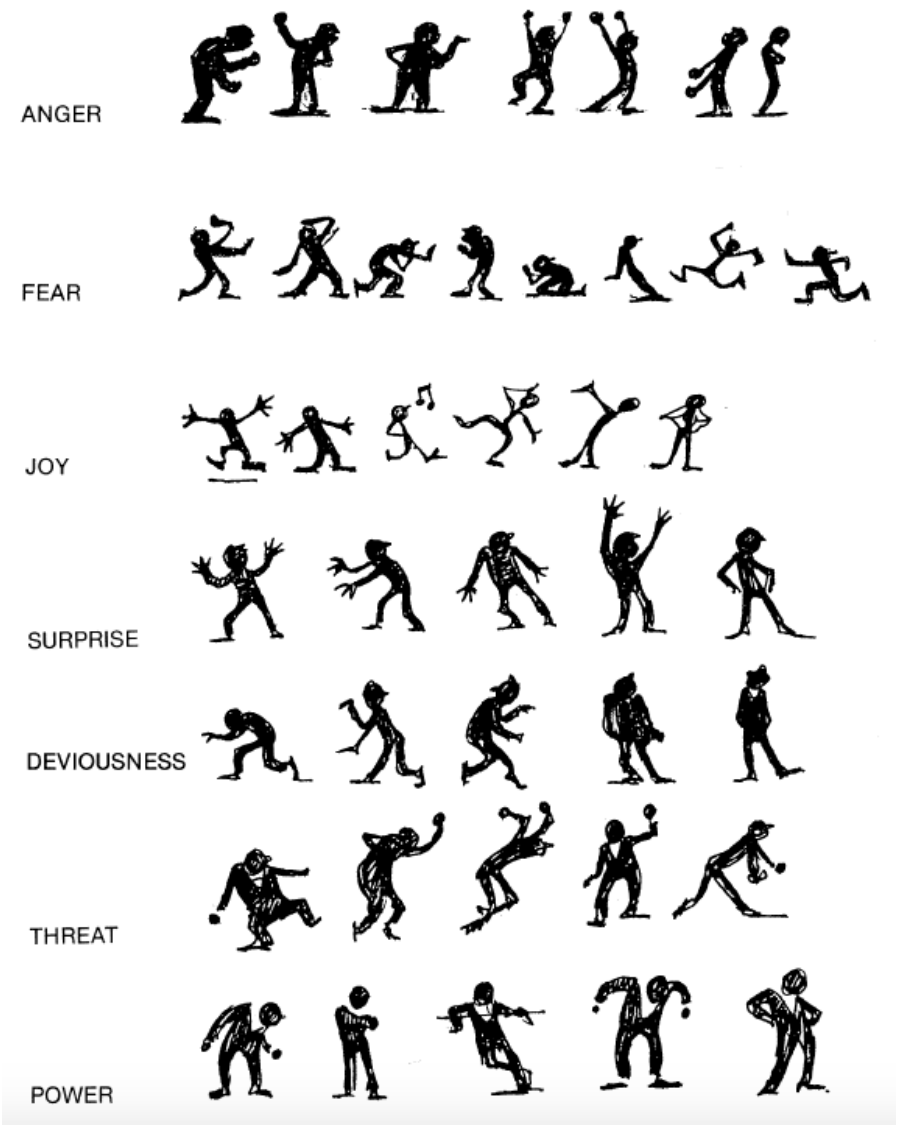


Figure 2.32.: A micro-dictionary of common gestures and postures.
Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 105], Fair use.

Figure 2.33.: Depending on the message the artist wants to convey, a single representative posture is selected out of a sequence of continuous movements. Had the focus been on the finality of the action, the artist would have chosen the last posture.
Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 109]

Figure 2.34.: A panel from Will Eisner's story "P'gell from Paris". It describes two of the characters overlooking a street in Istanbul. The city is referenced both by name (in the dialogue), but also through a series of cultural cues (the streets, the landscape etc.)
Source: "P'gell from Paris" in *Spirit #21*, Retrieved: Digital Comics Museum, Fair Use

place, therefore depicting it only through some easily recognisable, yet general elements (e.g., drawing a large amount of sand, a palm tree and a camel will generally identify a desert landscape, whether in Africa or Arabia).

A very interesting interplay, which has developed quite differently in the three main geographies of sequential art, occurs between the drawing of characters and that of places and landscapes. [McCloud, 1993] notes that in European comics, characters have traditionally been drawn as iconic (i.e., abstract) as possible and set against realistic landscapes, usually allowing the reader to immerse in the comics experience by both identifying with the character and getting lost in the medium. In American comics this effect has been used to a lesser extent, whereas in Japan it has long been a national trait of *manga*. In the 1980s however, Japanese creators took the idea of photo-realism a step further and used objectification either to convey physical properties of objects (e.g., weight, texture) or to stress the role of villain characters and prevent readers from identifying with them.

Special elements

Onomatopoeia are often stylised textual interpretations, of sounds. Wood chopping, water boiling, meat being cooked, firing a weapon, taking a step, breathing, rain falling... virtually every action or event, natural or artificial can be expressed through an onomatopoeia (Figure 2.35).

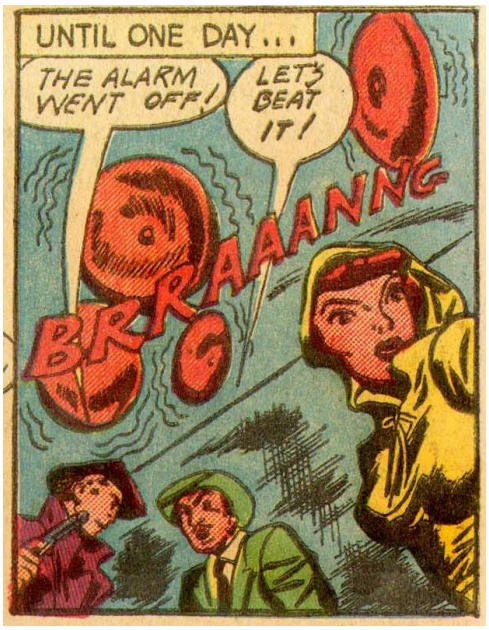


Figure 2.35.: Onomatopoeia describing the sound of an alarm ringing.
Source: "The Payoff – Prison!!", in *Startling Terror Tales #12* (1952). Retrieved: Digital Comics Museum, Fair use.

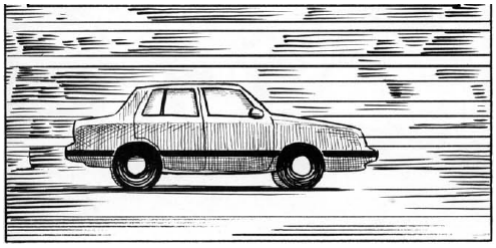
In Japan in particular, the extent to which sound symbolic and mimetic words are used and the variety of their application (i.e., animate, inanimate objects, states and conditions of the external world or of psychological nature) has long caught the interest of linguistics scholars, while editorial efforts were dedicated to assembling them in dedicated "dictionaries".

Motion lines are often used in comics to depict complex motion inside the confines of a single panel and are often used to visually mark the perceived trajectory of a movement.

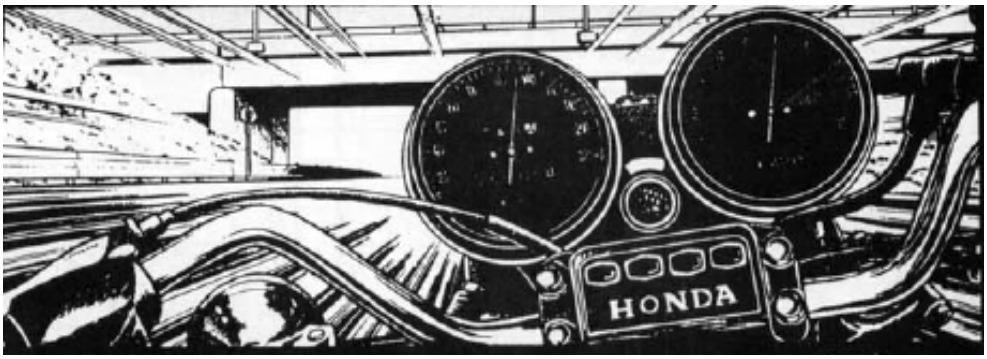
While this technique universally appeals to the reader's sense of motion and velocity, it is differently deployed in the main comics cultures. While in Europe it used to be rarely utilised, in American comics the perception is that of a fixed camera recording a movement (Figure 2.36a), hence the simplistic motion lines. In Japan, the technique lends more from photographic technology and the movement is recorded

from the perspective of a moving camera (Figure 2.36b), while the background loses focus. Later, *manga* development pushed the limits of this technique, giving birth to even more spectacular perspectives, in which the reader becomes part of the movement by being placed "in the driver's seat" (Figure 2.36c).

(a)



(b)



(c)

Figure 2.36.: Different uses of motion lines.
a Fixed camera perspective – Simple motion lines. Source: "The Grumly Murder" in *The Spirit #21*, Will Eisner (Jun. 1950), Retrieved: Digital Comics Museum, Fair use.
b Moving camera perspective. Source: *Understanding comics*, Scott McCloud [McCloud, 1993, p. 113], Fair use.
c "Driver's seat" perspective. Source: *Understanding comics*, Scott McCloud [McCloud, 1993, p. 114], Fair use.

The exploration of these elements of the medium and the theoretical understanding of comics is at the base of our research endeavour. Our work is the result of the richness of the comics medium in terms of both form and content, a richness which sometimes raises the complexity of the empirical analysis of comics and sequential art.

2.2.2. Fans, collectors and readers

French scholar Jean-Paul Gabillet distinguishes different levels of engagement with comics:

> "[C]omic book readers may be fans and collectors, but not all readers engage in this level of activity. Likewise, most fans may be collectors, but not all collectors are necessarily fans. And not all fans engage in the same types and array of activities, such as publishing fanzines and attending conventions."

Understanding the different ways in which the public engages with comics is a step forward towards understanding the role of comics as part of culture and the fascination they exert over people, whether they are readers or not. Conclusions about these topics are still highly speculative and empirical analyses of audience engagement with comics are desperately needed.

First of all, we need to understand what is a fan. This notion is not easy to define with respect to any media. Fandom is a volatile level of engagement and depends on the historical, societal and economical contexts, which we tried to cover in the earlier sections of this chapter. [Woo, 2012] defines the notion of fan with respect to the communities which were formed around comic-bookstores, conventions, fanzines and discussion forums and notes that in time, the aesthetic ideas expressed by these communities also came to have an impact on the production of comics. Fans are defined by two important traits: the object of their fandom (and its related material costs) and a different type of cost – the immaterial and affective investment they usually pour in these objects. Understanding fans doesn't stop at understanding the medium of their interest, and implies the exploration and analysis of their interactions (material, immaterial and affective) with this medium.

Media studies scholars tend to agree on mainly the same varieties of comics-oriented practises: collecting and reading. [Beaty, 2012, p. 155-156] considers "readers as undeveloped fans". Fandom on the other hand involves some sort of "strong involvement" and "one of the most rudimentary levels of participation is by engaging in collecting". [Woo, 2012] goes a step forward and describes a typology of collectors made of *completists*, *hobbyists* and *speculators*. In 2.1.2 we have seen that one of the factors that had hampered the development of comics for most of their existence was the unreliability of the old distribution channels, which mainly saw them sold in gas stations, department stores and newspaper stands. The nature of the market, which allowed retailers to return unsold copies and get refunds from the publishers, meant that the latter were running a high-risk business and that most publishing houses would generally print and distribute a limited number of copies. The readers were also affected since they had little to no choice in becoming collectors, while pursuing their weekly or bi-monthly hunt for their favourite magazine. At the same time, two different other practises developed in connection to the scarce [13] comics market: trading with other readers and the tertiary market of used comics.

[13] We consider the market scarce with respect to the seemingly higher demand some issues might have had at the time.

The 1970s saw the growth and establishment of a new distribution paradigm – **the direct market** of comic-bookstores. This model sought to take out all intermediaries from the distribution chain and have the publishers deal directly with the retailers – the comic-bookstores. But the most important trait of the direct market was the lowered risk that publishers were now taking. In this model, retailers were supposed to analyse and predict how many copies of each title they would acquire, since returns were not accepted any more. This lead to a very interesting shift: comic-bookstores workers and owners got to know their clients better, they deepened their knowledge about comics and eventually developed great skills of understanding the tastes and the needs of their customers; in a way, they became comics curators. A more convenient model for readers certainly, who in the direct market have greater choice over what they buy, its effects seem have had little impact on collectors practises. So what is driving them, as opposed to readers?

Qualitative interviews by [Woo, 2012] shows that *completists* are motivated by the nostalgic or sentimental value of comics, as well as the excitement of searching and pride of completing a series. Even more interesting, it seems that the serial nature of comics drives collectors habits beyond the bibliographic criterion, with some trying to acquire all issues featuring a certain character or all publications of a particular creator. For *hobbyists* on the other hand, collecting is at the core of their engagement with comics. Unlike *completists*, who at least at an early stage are motivated by their desire to read, *hobbyists* "value possession and preservation in the present, while deferring financial gain" [Tankel and Murphy, 1998, p. 59]. This sets them apart from a third category of collectors, the *speculators* who consider comics as an investment and who are not driven by the sentimental capital of comics, but rather by their potential value as collectibles. Interestingly, publishers sought to adapt to these patterns and based on their understanding of the motivation behind collectors' behaviour they have developed new publishing strategies. Reprinting comics material began to develop particularly due to the shift in the cultural perception of the medium, following the success of the graphic novel. These strategies resorted to different marketing tricks, such as multi-variant covers (hard cover, paperback, special print covers) and repackaging (alongside CDs with "extra" content or licensed merchandise), all attempting to convince collectors to acquire multiple such copies, in the hope of a later return on their investment. And then, there are the extreme forms of collection, such as those promoted by companies like Comics Guarantee. They propose collectors to send them their comics who are then evaluated and graded against an internal standard. The comics are subsequently sealed in a plastic case (*"slabbed"*, as collectors refer to this process) with the purpose of preserving them in exactly the same condition. This is an extreme example of how collecting practises are able to transform the original purpose of the collected object.

In [Woo, 2012] collectors and readers are placed at opposite ends of the engagement spectrum because "using" comics means reading them, but for collectors reading means "damaging" the object. Publishers have intervened in this regard as well and addressed the anguish of collectors by reprinting materials in the form of standalone or collected anthologies. Productions such as Marvel's *Essentials* or DC's *Showcase Presents* which

feature content that would normally interest collectors, were rolled out not for their historical, but were rather aimed at the reading public. Graphic novels either didn't spark so much interest amongst collectors; on the other hand, the complexity of their subject matter and the critical appraisal most of them enjoyed saw them rapidly populate the shelves of mass market bookstores. With this new distribution channel, a new and increased readership could access comics starting with the early 1990s, changing once more comics reading practises. As it turns out today, readers' engagement with comics is driven less by their attachment to the physical object and more by their search of a good story, a certain character [Woo, 2012] or, in the case of graphic novels a particular theme.

Conventions

Another expression of fandom and fans engagement with comics are the comics conventions (*comic-cons*). Initially small, these events reuniting fans, creators and experts alike, have always highlighted the sense of community that is formed around comic books. Today, they have evolved into a sizable industry, with proceedings necessitating large convention centres, stretching over several days and generating large revenues from selling tickets and merchandise. Forbes Magazine has dubbed the *San Diego International Comic-Con* as the "largest convention of its kind in the world"[14] and today almost every major city in America is hosting an yearly comic-con. In Europe, the *Festival International de la Bande Dessinée (International Comics Festival)* in Angoulême is the largest of its kind, whereas Japan's *Comiket* draws over half of million participants annually[15], the largest attendance of any event of the kind.

In the 1990s comic-cons developed beyond comics[16] and started promoting multiple facets of pop-culture: films, TV series, youth literature, video-games, toys, costume play. Conventions thus became increasingly more profit-oriented, with the largest of them currently being owned and produced by large corporations. Subsequently, this caused a noticeable divide between categories of participants: once mixing, fans are now clearly set apart from guest participants (invited speakers, authors, actors etc.); in addition to that, these conventions proved to be a fruitful meeting place for publishing professionals (authors, editors, publishing houses). By the mid-1990s, an effort to counter-act the excessive commercialization of conventions and reclaim them as comics-centred events, saw the appearance of a different brand of such events. These were based on the long lasting model of European festivals, like the *Lucca Comics & Games* (Italy) and the *International Comics Festival of Angoulême* (France), focusing almost entirely on the promotion of "alternative" or "independent" comics.

In spite of their multiple connections to other media, these events always circle back to their comics origins. This is perhaps one of the reasons for which comics are still referenced in their name (instead of *pop-cult-cons*). More recently, the *Denver Comic Con* has even

[14] [Ewalt, 2011]
[15] [McCarthy, 2006, p. 14]
[16] People attribute this shift to the success of Tim Burton's *Batman* movie (1989).

added a scholar section to its programme, inviting researchers to submit articles and take part in a regular academic conference attached to the original convention.

The seeds of comics fan culture can be found in this sense of refuge and belonging that comics have always cultivated amongst their readers. This has gradually evolved into different forms of communities around the comic book object: first, the "letter to the editor" section from the 1960s magazines; then, before the appearance of the direct market, the first comic conventions and fanzines (magazines edited and published through the efforts of the fans). Interestingly enough, the direct market didn't suppress these forms of communion; instead, the appearance of the comic-bookstores added itself as yet another space for fans and readers to share their common interest.

It is this sense of community, sharing and participation, that comics have in common with crowdsourcing and on which we rely when building our comics annotation platform. Moreover, [Woo, 2012] remarks that in pursuing the object of their interest, fans often develop expertise and with it, a form of criticism. We wish to harness this knowledge from our participants (fans, as well as readers) and would like to see whether the motivation factors that they usually exhibit in pursuing the object of their interest (a sense of affection, pleasure and even commitment or loyalty) will drive them into supporting our research project. For if this is the case, we will probably be able to conclude that crowdsourcing comics is yet another way (beyond fandom or readership) of engaging comics.

2.2.3. Reading comics in the age of digital

In 1990 Tim Berners-Lee laid the building blocks of the Internet (the *World Wide Web*, the *HTTP* protocol and *HTML* language) triggering the information age in which we are living today. The years that followed saw the rapid development of his technology and its impact in all spheres of human activity. Only a few years later Will Eisner, in the final chapter of the revisited version of [Eisner, 1985] discusses his take on the future development of comics given the new context. In it, he looks to the past evolution of comics from newspaper strips into comic books and discusses their (already!) imminent passage towards the digital. Interestingly, Eisner postulates there are four features of sequential art that should never change, regardless of the medium: the narrative (the story must adhere to a common reading convention), the composition (panels must be used for narrative purposes), the skillful creation of characters and the quality of the elements portrayed; at the same time, he hints that the major change that could affect the medium once gone digital, should likely be the addition of multimedia technologies (sound, motion, 3D). He also brings into discussion another topic that we elaborate on in Chapter 4, namely the rendition constraints posed by screens, as opposed to printed pages and discusses the impact this may have on the artistic process.

Ten years after, [McCloud, 2000] dedicates half of his work on **"Reinventing Comics"** to discussing the possibilities opened by the presence of comics in the online medium.

McCloud is convinced that the future of comics is in their marriage with the web and digital technologies and he goes forward to advocate that their future financial success is tightly coupled with the then emerging micro-payments.

Today, the concept of **digital comics** covers both digitalised versions of originally print comics, as well as digitally native content. Native content itself can refer to webcomics, designed for the size of a computer screen (experienced in a browser) and the lately developing mobile comics, usually experienced on small-screen mobile devices, by means of dedicated applications. Tablets in particular seem to have a great advantage due to their screen size being approximately equal to that of print comic books.

Several factors have contributed to the development of a market for digital comics, which involves a sizable part of the graphic narrative readership worldwide. First of all, as opposed to print retailers, who run on a limited stock, buying a digital comic eliminates the risk of readers not being able to purchase their desired comic. Moreover, in the case of American comics, digital editions are released simultaneously with the print ones and sometimes may include additional content. In what concerns pricing, there are usually little differences between print and digital comics and when this is the case, these are usually in the favour of the latter. This generally applies for digitised versions and is justified by the lack of paper and ink from the process. However, publishers cannot afford to allow for a too large of difference because of two main reasons: first, despite not being obvious at a first glance, digital comics too incur certain production and maintenance costs (scanning, file maintenance, tracking and uploading, multi-format support etc.); second, prices that are too small for the digital versions would undermine the print market, which is still the main revenue generator for comics publishers.

As with all revolutions, digital comics have also had to face their critiques. First of all, comics fans seem disappointed of all the complications caused by digital rights management (DRM), which are yet to be clear on issues such as the rights to transfer a purchase on different devices or typical fan behaviours such as trading issues with fellow fans or reselling them at discounted price (as it is the case of back issues in comic stores). Besides, there are also those "addicted" to physically experiencing the book as an object, through touch, smell and especially gestures such as turning the page. At the same time the creative opportunities opened by the constantly improving visual technologies weren't only good news. While some creators embrace the digital revolution and are pushing the boundaries of the form, others (fans, readers and authors alike) are worried about digital comics getting dangerously close to animation or film. We explore this topic in depth in 4.2.4, where we make an inventory of the most frequent (and some unusual) digital comics and give our own view with respect to the use of technology in the development of digital graphic narratives.

Finally, it is yet unclear how this new form of publication will impact the distribution mechanisms for comics. Once again, opinions are divided and they vary according to the role each stakeholder assumes within the industry. Large publishers seem to have embraced the model and adapted to the necessities of both creators and readers. Marvel and DC are both

present on ComiXology, which after the acquisition by Amazon in 2014 has consolidated its position as the largest digital comics store. Moreover, its "Submit" service offers creators the possibility to upload their work which, upon validation according to an internal quality standard, is published and sold[17] on the platform itself. In this respect, DriveThruComics has been "the first online retailer to specialise in downloadable comics." In place since 2004 they too offer, besides titles from big publishers, a means for creators to upload their work which can be purchased[18] upon validation by "minimum quality standards" (cover, preview, description). Other platforms didn't specialise as a market place. Instead, services such as Gumroad offer an e-commerce pipeline able to handle payments on behalf of creators who wish to sell their work through the platform[19]. Travis Nichols (content editor) describes Gumroad as an adaptation of the direct market model to present and most likely future necessities of comics distribution[20].

On the other side of the barricade, we have the comic-bookstore owners, who perceive the digital market as a direct threat. That is despite efforts from publishers to put a minimum cap on the prices of digital editions to avoid undermining the still dominating print market. Actialuna's Sequencity addressed this issue with their innovative bookstore recommendation system, in which the expertise of the bookstore professional is valued, thus opening the door to the digital market for comics for this professional category. Back in 2006 Dan Vado (president of respected independent publisher Slave Labor Graphics) seems to have had the right intuition about the role that digital distribution for comics (especially for self-publishing creators) and its relationship to print: "Digital [distribution] will still be there to support print, but it will replace the serialized comic book as a revenue stream and distribution method for many comic book creators"[21]. And while the need for comic-bookstores to adapt and reinvent themselves is obvious, it is perhaps worth to reflect on the real threat over their business – and this might just not be digital comics themselves, but the involvement of Web retailers in selling print comics.

Webcomics

When trying to understand digital comics, it is important to acknowledge the role webcomics have played in their development. In fact, from their first appearance in 1985 (*Witches and Stitches*) – marking the beginning of digital comics, webcomics have embraced and made use of most of the technological progress realized throughout the 1990s and 2000s in computer graphics, user experience and animation.

There are several factors that helped webcomics to attain the large success they still enjoy, even after the technological shift caused by the tablets and smartphone revolution

[17] ComiXology gets 50% of each purchase.
[18] DriveThruComics typically retains 30% from each purchase; this goes up to 40%, if the same title is published on other platforms, as well.
[19] Gumroad charges only 5% of each purchase, but instead leaves it to the authors to build and reach their audience.
[20] [Joe, 2014]
[21] [Soponis, 2006]

which started in 2010. First of all, there are more than 15,000 webcomics, and with them comes a great variety in terms of genre, topics and subject matter to suit whatever taste one might have. Webcomics embrace the web and its intrinsic freedom of speech, undoubtedly one of the reasons behind their large readership; they're raw and they can afford to tackle their subject head-on, as no censorship, no editor, no board of directors tampers with the message that is passed towards the readers. To this regard they remind of the freedom of speech assumed by underground "comix" in the 1970s. Mike Cariglitto, editor of successful webcomics *Penny Arcade*[22] or *Ctrl+Alt+Del*[23]: "No publisher or editor is telling [Web creators] what they can or cannot do. They speak in the audience's language, and nothing's holding them back [...]". Lastly, their close rapport with their fanbase also seems to contribute to their success. Most successful web cartoonists benefit from directly engaging with their fanbase, which in turn helps build a community around the publication. [Campbell, 2006], author of *The History of Webcomics* observes how today "we can e-mail our favourite cartoonists and most of them that aren't too big will respond. Almost every Web comic in existence has a forum."[24]

Despite this, most webcomics have had difficulties in turning online fandom into financial success, usually relying on donations and online advertisement as a source of income. More recently, there are reports of several dozen artists whose primary source of revenue are their webcomics. Amongst these, *Penny Arcade* – the most successful webcomic up until now, whose focus on video games managed to attract industry advertisement which, alongside licensed merchandise are currently generating enough revenue to support their creators and a professional team working with them. As of 2004, *Penny Arcade Expo*, the webcomic's own exhibition series is organized in cities in the United States and Australia, managing to attract sizeable crowds and generate enough revenue for the franchise to be able to redirect part of its profits to charity organisations. Other financially successful webcomics are science-oriented *xkcd*[25] or *PhD Comics*[26], an insightful and many times accurate depiction of the lives of graduate students. An interesting phenomenon is that of medium crossing, with successful webcomics getting print syndication, while other cartoonists choosing to serialize and publish their former print comics exclusively on the web. More so, in recent years, creators seem to have realized the viability of the "digital-first, print-after" approach, allowing for a cheap online gauging of their work's potential, before committing to the expenses incurred by publishing in print. These are, nevertheless, rare cases and most webcomic artists never cross mediums and the Web remains the only outlet for their work. The solution for them are online publishing platforms, such as Webcomic Nation[27] or Keenspot, whose main advantages are addressing the technological or promotional challenges related to Web publishing.

[22] https://www.penny-arcade.com/
[23] http://www.cad-comic.com/
[24] [MacDonald, 2005]
[25] http://xkcd.com/
[26] http://www.phdcomics.com/
[27] Ceased to exist in 2013.

Webcomics have established themselves as an important alternative to print comics. We have seen how they benefit from the freedom of speech and commercial advantages of the medium in which they are deployed. Their relationship to print comics is complex, in the sense that they both compete and support the print. "Web-comics are a new medium everywhere and it's impossible to judge whether their development has been adequate or not," says graphic novelist Samit Basu, author of *The Simoqin Prophecies and The Manticore's Secret*. "These things happen organically and they take time"[28].

Scanlations

We conclude this chapter with a brief look over an interesting, yet often overlooked phenomenon related (mainly) to *manga* and *manhwa*, that of **scanlations**. The term, a portmanteau of the words "scan" and "translation" refers to the "fan-made scanning, translation, and editing of comics from a language into another language." [29] Until recently, this phenomenon has been largely neglected or treated as "the dirty laundry" of the *manga* publishing industry, but with the advent of digital it managed to draw the attention of both media and academic institutions in an effort to understand its origins and the impact it has in today's publishing world.

How do they work? Scanlations can be considered one of the early mixtures between digital technology and graphic narrative. The idea for scanlations dates as early as the 1970s and their development throughout the 1980s and 1990s is tightly linked to that of the Internet. In an article for Japanese Times, [Macias, 2006] reports that often "scanlators" are motivated by their appreciation for *manga* and consider their effort a way to make the genre more accessible and more popular with a broader audience. Some "scanlators" remember having started out of frustration caused by the long delays it takes to publish a licensed translation issues, holding them months behind the release stream in Japan. Others say they do it to promote *manga* they love; this can be problematic in the case of niche publications such as Suehiro Mahuro's erotic or grotesque stories, which would obviously need to be censored and adapted prior to their publication in Europe or the United States. A truly remarkable aspect is the way in which most scanlation initiatives are organised. Despite being amateurs and fans, the form teams where a precise role is allocated to each member. A typical scanlation workflow involves the following roles:

- obtaining the pages from the print version of a *manga*;
- translator: translates the text into the desired language;
- proof reader: checks the accuracy of the translation;
- cleaner: removes the original text and makes graphical adjustments (scan noise, brightness, contrast);

[28] [Verma, 2015]
[29] https://en.wikipedia.org/wiki/Scanlation

- typesetter: fits the translated text into the 'clean' text areas and adds necessary text effects;
- quality controller: edits the final version.

As for the ethical aspect, "scanlators" attitude is diverse: some take pride in the work they are doing and there are genuine rivalries between scanlation groups. Others admit their activity is piracy (without any doubts), but do claim that because of them Japanese publishers start feeling the pressure and have become more proactive about releasing licensed translations. In addition, a tacit agreement seems to be in place between publishers and scanlators, who take down all related scanlations once a licensed translation is released [Macias, 2006]. This agreement seems however to have changed terms in recent years, when scanlation distribution shifted from small-size online channels and blogs towards large aggregator websites which, due to their sizable traffic, started generating important revenue from advertisement.

It is easy to condemn scanlations for their obvious copyright infringement, but what they really show is a problem at the core of the Japanese publishing industry, especially in the difficult post-2008 market: while they are clearly able to sell massively to their national audience, Japanese publishers are still indulging in a decades long lack of adaptation (one could call it ignorance!) to the demands of the American and European markets – fast, cheap and... digitally available *manga*. "Most Japanese publishers have no coherent digital strategy, and the extra step of licensing them in America makes them even slower to react to change."[30] Which brings us to our second issue – the pricing model practised by American publishers make licensed translations almost prohibitive for the average fan. More precisely, this is the unhappy combination between a bad pricing model and an American readership which is simply too young to even afford credit cards for paying the steep price of their *manga*[31]. Which brings us to the third problem of *manga*: its overseas readership. The readers themselves need to understand they cannot read their magazines for free. The second they access a scanlation via an aggregator website, the effect propagates over an entire ecosystem (author, editor, publisher, retailer), who is dependant on the readers to generate revenue for supporting future publications. And while paid advertisement may be an important alternative source of revenue, this revenue has to be generated towards the industry itself and not towards scanlation aggregators. Former editor Jake Forbes, remarks: "if free manga, powered by ads is the answer, fine; but let this solution be licensed" [Forbes, 2010].

In June 2010, the first real joint initiative of *manga* publishers saw the creation of the Manga Multi-National Anti-Piracy Coalition. This effort brought together all thirty-six members of the Japanese Digital Comics Association (including the three major houses –

[30] [Thompson, 2012]

[31] As of 2016, the yearly prices for some of the most sought out collections are: Naruto: $400; Fruits Basket: $250; Vampire Knight: $90; Negima: $300. Even more disturbing is that these prices are three to four times higher than those paid in Japan for the same titles [Forbes, 2010].

Kodansha, Shogakukan and Shueisha), the main US *manga* publishers (Vertical Inc, Viz Media, Tokyopop) and Yen Press, the *manga* and graphic novel imprint of the Hachette Book Group. Their first action was to issue a "cease and desist" document targeting the main scanlation aggregators. Perhaps the largest impact was the resulting shut down of the main aggregator OneManga (whose website was ranked in the Google Traffic Top 1000) in a matter of weeks, as well as the shut down of MangaHelpers which later that the year relaunched as a platform for self-publishing. While such actions seem to have had some initial effect, there is still a large number of scanlation aggregators out there and different forms of piracy are still largely affecting the *manga* and *anime* industries. There are important lessons that *manga* publishers seem to have difficulties in learning and scanlations point out to some them: first, the existence of sizable overseas markets for *manga* (namely those in the US and Europe), the possibility of releasing free licensed content on advertisement-supported platforms, as well as the availability and know-how of certain groups of fans who help create quality translations in a timely manner. But slowly, some publishers (Dark Horse, Digital Manga) are embracing the crowdsourcing potential of scanlations and are harnessing the expertise and motivation of "scanlators" to produce licensed material.

We conclude this chapter and the first part of this thesis. Our goal was to introduce our reader in the interesting, yet complicated world of comics. We sought to unveil the cultural background of comics, underlining their quality as a sequential art and its potential for developing complex graphic narratives. The subsequent analysis of comics as a medium presented its structural elements and its composition rules. We have then discovered how authors can use them to build stories and convey the intended message. Put in front of the author's originality, the "science of the panel" becomes less exact and a variety of structures are born only as the result of the creator's talent. It is this kind of elements that are difficult to extract and standardise for whatever purpose – digital publishing, indexing, analysis and research.

Part II.

Digital publishing of comics

Chapter 3.

Digital publishing of comics: State of the art

"A standard is a great thing; everyone should have one."
— Tech industry witticism

In this chapter we will present the reader with a conceptual toolbox allowing for the complete specification of a digital publication. We will identify and present the different aspects necessary for such a specification and will show how the currently existing standards and formats are equipped for only partial descriptions of publishing products (books, magazines, comics). While certain aspects of our discussion apply to print publications, this chapter is mainly concerned with digital publications.

The digital nature of a publication is defined by the digital nature of the packaging, distribution and consumption activities related to such a product. To this end digital publications are best described by three main components: the content, the presentation and the metadata. Digital publications cover a wide range of formats, including e-books (native or scanned), digital magazines or native apps. While they can be experienced both online and offline, their success in recent years is tightly linked to the ubiquity of hand-held mobile reading devices (tablets, e-book readers, smartphones). As for comics, their digital nature comes from their production and/or release media. We distinguish between originally printed comics, which are later scanned and released digitally, and comics that are digitally created (i.e. drawn on a computer or a tablet) and are either best experienced in digital media (by means of a reading device) or are eventually released in print.

3.1. The publishing environment – an overview

Before we proceed, it is important to understand the current context of the publishing industry, the role comics are playing within and the impact of digital publishing therein.

While an overview of the state of the publishing industry worldwide is beyond the scope of our work and would be suitable for a research project of its own, we are able nevertheless, based on statistics that we could obtain[1], to present you with a fairly good picture of how the publishing industry presents itself in the three regions where comics are historically an important component of the literary scene – France, the United States and Japan.

France

The data released by the *Syndicat National de l'Édition (National Publishing Union)* (SNE) [SNE, 2016] concerning the performance of the French publishing industry in 2015 indicate the first positive changes in the past five years in terms of both net revenue (2.667 billion , +0.6% compared to 2014 and volume (436 million units, +3.5%). Digital publishing (all supports and all editorial categories) accounts for a 6.5% of the total purchases in 2015, a 1.5% increase from 2014, a moderate change compared to the first years of the e-book "revolution".

As for the *bande dessinée*, the 5,255 titles (3,924 editorial novelties) accounted in the 2016 annual report of the *Association des Critiques et Journalistes de la Bande Dessinée* (ACBD) [Ratier, 2015] represent a 2.9% decrease compared to the previous year and only the second time such a contraction occurred since 2000. The 2015 net revenue for *bande dessinées* represents a steady 6.9% of the total French publishing market. The same report shows a steady 1% increase in the digital BD market, a constant over the past years, concluding that this form is still a marginal market and has yet to establish itself in France.

United States of America

The case of US publishing industry proved to be one of the most difficult to research, due to industry secrecy or prohibitive prices for related annual reports. The Bowker Reports[2] estimate that the number of published books went from 600,000 in 2007 to 1 million in 2014, with a staggering 700,000 of these titles being self-published (almost four times more than in 2010). And while the volume statistics paint a pretty good image, the sales data released by the Association of American Publishers (AAP) show quite the contrary. Indeed, in addition to the newly published books, the market also has to deal with previously unsold books – 13 million in the US, in 2014 – meaning that since their peak in 2007, sales have dropped or, in the best case, have stayed the same. Within, comics tell an interesting story, with 2015 being the most successful year since 1993, according to a joint estimate by

[1] This was a challenging task on its own since, as opposed to the film or music industry, the publishing industry makes it difficult to track such data as: periodical net revenues, periodical sale volumes, per-genre statistics etc. Even when available, this information is released in a very decentralised manner, usually in the form of extremely expensive reports.

[2] [Piersanti, 2016]

Comichron[3] and ICv2[4], which reported that sales revenues in North America (Canada and United States) increased by 10% compared to 2014, with graphic novels accounting for the most sizable progress of all book channels.

At the same time, digital publishing functioned at "full speed" between 2008 and 2012 and seemed well equipped to face the challenges of the economic crisis. But after a steady 2013, the digital market received a sizable blow, dropping by more than 10% in 2014 and thus strengthening beliefs that US publishing industry is in a perpetual struggle to reinvent itself. Or, as one American publisher recently summarised this:

> "The thin margins in the industry, high complexities of the business, intense competition, churning of new technologies, and rapid growth of other media lead to constant turmoil in bookselling and publishing (such as the disappearance over the past decade of over 500 independent bookstores and the Borders bookstore chain)."

The same trend was followed by digital comics in North America, which as opposed to the record breaking sales for print comics, reported a 10% decrease compared to 2014, after six years of continuous growth (see Figure 3.1). Publishers have tried to explain this by the near saturation levels of ownership of mobile hand-held devices, whose long-time users have already checked their comics wish-lists.

Japan

A very special case is that of Japanese publishing, a booming industry for more than thirty years, between the 1960s and its peak in 1996. Since then, the annual sales have reduced by 36% in 2013, with the print magazines sector taking the largest blow. Figure 3.2 gives a visual understanding of this evolution over time.

The consensus is that this has been influenced by two important events: the two economic recessions following the crisis in 1997 and 2008 that affected the Japanese economy and the diversification and pervasiveness of new information and communication technologies (video game consoles, the growth of the Internet, the advent of mobile hand-held devices), which as early as 1995 drastically changed the way in which Japanese spent their leisure time[5].

And there are voices saying that Japanese publishing is probably paying the price for its failure to adapt to the fast-paced technological changes.

[3] Comichron is the worlds largest public repository of comic-book sales figures, featuring data from the 1930s up until today about comic book and graphic novel circulation, cover prices, and market shares on its website, http://www.comichron.com.

[4] ICv2 is the #1 industry source on the business of geek culture, including comics and graphic novels, hobby games, and showbiz on its Website, www.ICv2.com, and in its magazine, *Internal Correspondence*.

[5] [nippon.com, 2015]

North American Digital Comics Sales			
Year	Digital Sales (in millions)	$ Change (in millions)	% Change
2009	$1	N/A	N/A
2010	$8	$7	700%
2011	$25	$17	213%
2012	$70	$45	180%
2013	$90	$20	29%
2014	$100	$10	11%
2015	$90	($10)	-10%
			© 2016 ICv2

Figure 3.1.: Digital comics sales in North America (2009-2015)
Source: ICv2

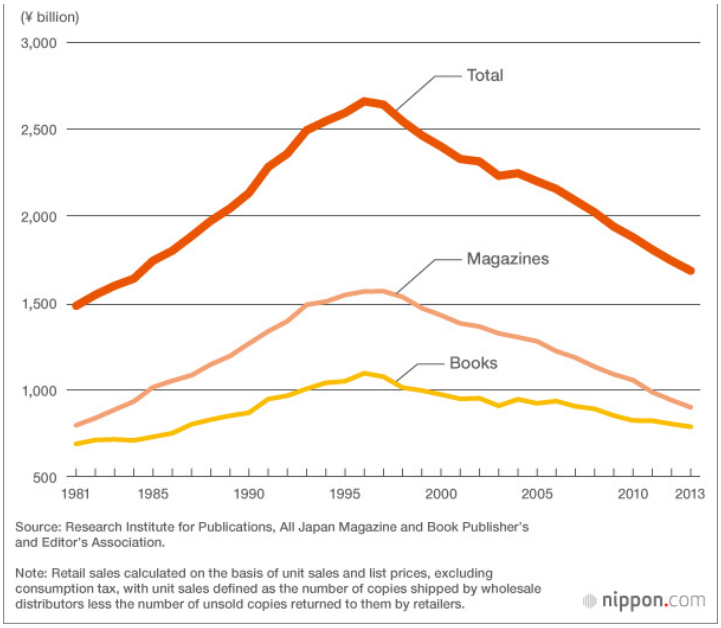


Figure 3.2.: Book and magazine sales in Japan, by value (1981-2013)
Credit: nippon.com

Its problem resides in the financial workings of this deeply entrenched system that for the past forty years delivers its product to market through a network having a number of wholesalers acting as the middleman between publishers and retailers:

- publishers ship new titles to wholesalers, who forward them to retailers;
- retailers pay only the quantity that they sell and can return the rest;
- this means, wholesalers are running on credit, while publishers try to minimise their loss, by putting more titles on the market.

As a consequence, just like the American market, the Japanese publishing industry hosts the paradox of dropping revenues – more severe in its case –, as opposed to constantly increasing volumes – quantity at the expense of quality. The 2014-2015 publishing guide by the Japan Book Publishers Association reads [JBPA, 2014]:

> "Comics (manga) and weekly magazines were the main staples of the publishing trade in Japan. However, the total sales of magazines has been decreasing, dismissing the status values as the most trusted information media by the emergence of new IT devices like smartphones and tablets."

Concerning *manga* publishing, the Oricon Annual Market Report[6] for 2014 shows that revenues from sales increased for a second year in a row, 4% higher than in 2013, while the volume increased by 1.3% compared to the same year. Comparing these numbers to the rest of the publishing industry, the market share for *manga* in 2014 represented 27% of the total industry revenues and 42% of its total volume. As for digital publishing, the data we managed to collect from [JBPA, 2014] indicate that the small growth rate from the past years continues – in 2013, the digital market represented 5.5% of the print market. Within, the largest segment is that of *manga*, accounting for 72% of all digital publications.

3.2. Standards for digital publications

The diversity of the publishing ecosystem and the different publishing needs of content creators have sparked the creation of a plethora of different standards, formats and tools for digital publications. One particular perspective, the one that not only interests us the most, but has also stood out throughout the recent years is the dichotomy text-based vs. image-based publications.

This translates into a considerable amount of standards and formats ranging from the very simple Comic Book Archive formats to the versatile Portable Document Format (*PDF*), from proprietary formats such as Amazon's *AZW* or Apple's iBook to IDPF's open standard *EPUB*. In the following, we will analyse the capacity of each of these formats

[6] [AnimeNewsNetwork, 2015]

to provide technical solutions for the creation, packaging and distribution of image-based publications, with an emphasis on comics and graphic novels.

3.2.1. Standards for presentation and distribution

Comic Book Archive Files. In the early years of digital publishing, Comic Book Archives represented the main format for packaging and distributing digital comics. The process involves scanning paper versions of comics, and packaging the resulting images, generally *PNG* or *JPG*, in basic archive formats (*7z, ACE, RAR, TAR, ZIP*) corresponding to different file name extensions (*.cb7, .cba, .cbr, .cbt, .cbz*). The rendition order for the pages is specified by naming each corresponding file with its page number, using pre-pending zeroes to guarantee a correct rendition (i.e. if the archive has 100 pages, page 1 corresponds to file `001.jpg`, page 15 to `015.jpg` and page 100 to `100.jpg`). Additional information, such as publication metadata, can be included either in accompanying *XML* files or the *ZIP* archive's comments section.

The format is compatible with all large platforms (Windows, iOS, Android, Linux) with specific viewer applications being developed for desktop, browser and e-book readers. The main limitation of the format is the content itself – plain flat images, which impose a fixed layout for their display, subsequently generating problems of image quality (bad scaling, bad aspect ratio) when presented on small screen devices.

Portable Document File (PDF). Developed by Adobe Systems in 1993, *PDF* has become ISO 32000 in 2008. The format has gained popularity due to its availability on different computer platforms and its particularly good results in terms of viewing and printing documents, thus becoming the most widespread format over the world wide web and in document management systems. Its support for vector and raster images made it a useful choice for image-based publications and several digital comics publishers are using it as either an internal or even publishing format.

Its most important limitation, one of critical importance from the perspective of image-based publications, is its initial development as a fixed-layout format. This has changed in 2012 with the release of PDF/UA, which allows for a better reflowable document rendering on small screen devices. Currently, *PDF*'s main problem is the limited support for displaying documents at more granular levels, such as regions of interest (i.e. for comics, this would map to panels, balloons, characters etc.).

Proprietary formats. While a considerable number of proprietary formats, have existed throughout the past two decades, with different degrees of success, two of them have swept the market as early as their release and are currently playing important roles in shaping the future of e-book formats.

Apple's *iBook* is an *EPUB*-based format, the main difference between the two coming from Apple's addition of proprietary *XML* namespaces and *CSS* tags, meaning *iBook*

is readable only by their own iBooks app. Their publishing eco-system also includes dedicated publishing tools, such as iBooks Author, allowing authors to create not only basic publications, but also enhance them by embedding widgets, which at the time of its market release *EPUB* was not supporting.

Amazon's latest Kindle release in 2011, came together with an update to their e-book format *AZW3*. Just like *iBook*, Amazon's format is largely relying on a subset of *HTML5* and *CSS3* for content and style specification, with proprietary added enhancements. Older Kindle readers still use *AZW*, Amazon's initial format, based on *Mobipocket* (an open e-book format built around *XHTML* and *Javascript*).

EPUB. *EPUB* is the open-standard for e-books developed and maintained by the IDPF community. Centered around web technologies and based on *HTML5*, *CSS3*, *SVG* and *Javascript*, the format has gathered popularity and has become the highlight of open-source publishing efforts since its launch in 2007. The *IDPF* community comprises a variety of actors from different areas – IT (Apple, Google, Sony), publishing (Barnes & Noble, Hachette, Kodansha, O'Reilly Media, Safari Books, Shueihsa) and other standardisation or publishing organisations (DAISY Consortium, Digital Comic Association, Electronic Book Publishers Association of Japan, W3C).

EPUB [7] is a collection of standards, normalising the content creation, packaging and distribution in the form of digital publications. The main advantages of this format are the interoperability between the standard and different compliant reading devices, its focus on solving the problems of rendering reflowable documents and packaging multiple renditions of the same publication.

The three main composing standards of *EPUB* are:

Publications [EPUB-Publications, 2014] defines the semantics and conformance requirements for each rendition of an EPUB publication;

Content Documents [EPUB-ContentDocs, 2014] defines profiles of *XHTML, CSS* and *SVG* for describing the contents of an EPUB publication;

Open Container Format [EPUB-OCF, 2014] defines a file format and a processing model for packaging the resources of an EPUB publication into a single EPUB container.

In addition to that, *EPUB* also defines specifications for synchronising text and audio (Media Overlays [EPUB-MediaOverlays, 2014]) and for referencing arbitrary content within an EPUB publication (Canonical Fragment Identifiers (CFI) [EPUB-CFI, 2016]).

The format's structure and capabilities, combined with its open-source nature have helped it garner popularity in the digital publishing world. And with image-based publications being an increasingly important part of the publishing market, authors and content creators have

[7] At the time this document was written, version 3.0.1 was the latest release, while version 3.1 was undergoing public member review.

become interested in the possibility of using *EPUB* to better serve their purposes. Therefore, the last five years have seen an unprecedented attention from the *EPUB* community to develop the format in this specific direction, which materialised with the specifications developed by the EPUB Advanced and Hybrid Layouts (AHL) working group: Previews [EPUB-Previews, 2015], Multiple Rendition Publications [Conboy et al., 2015b] and Region Based Navigation [Conboy et al., 2015a].

We will come back on all these topics and more in the chapter Chapter 4, reporting on our work within the AHL working group and other instances of the IDPF.

3.2.2. Adding semantics to digital publications

The three main concerns related to digital publications – packaging, rendering and distribution – are well addressed by the standards and formats discussed in the previous section. However, the steady growth of the digital publishing market and the changes in the way in which digital publications are consumed bring forward additional requirements concerning the e-book product: indexing, collecting, searching and annotation, all of which can only be supported through the addition of a rich semantic layer.

One of the most important formats currently in use is the *XML*-based open standard for describing content semantics, developed by the Text Encoding Initiative (TEI) [8] – a text-centric community of practice in the academic field of digital humanities. The *TEI* schema contains about 500 tags grounded in different academic fields connected to digital humanities and, while difficult for individual projects to make use of most of them at one time, the current practice is for different communities, entities or projects to choose their own subset schemas, using a *TEI* customisation mechanism.

Also, although richly equipped for annotating concepts in a variety of domains, *TEI* understands that there could be domain-specific aspects it has not foreseen, in which case it provides with an easy mechanism for extending the core schema with customisable elements. This is also the case of comic books, whose particularities require only a subset of the entire *TEI* schema, as well as an addition of new domain-specific elements.

Besides being the complicated medium we explored in Chapter 2, comics exhibit great ingenuity and sometimes playfulness in their use of graphical narration. Or, as comics scholar [Walsh, 2012] notices:

> "Comic books are often very formally self-conscious documents and express a fascination with their own bibliographic identitiescreators (comic book writers and artists) become characters in the narrative, editorial notes refer to episodes from prior issues, comic books and paratextual elements, such as advertisements, are parodied within the comic book narrative, publication milestones (such as

[8] http://www.tei-c.org/index.xml

(a) **(b)**

Figure 3.3.: Two examples of famous comics parodies:
a "Superduperman", in *Mad #4* (May 1953). By W. Wood, H. Kurtzman. A satirical story of Superman and Captain America. In these panels, the author ridicules the famous scene where Clark Kent (here, Clark Bent) changes into Super(duper)man inside a phone booth. Credit: Wally Wood, Fair use.
b The first appearance of Howard the Duck, a parody of "funny animals" (e.g., Donald Duck), later featuring in several other stories that satirise famous comics and important moments in comics history.
Source: "The Enchanter's Apprentice", in *Adventure into Fear #19* (Dec. 1973). By S. Gerber(w), V. Mayerik (p), S. Trapani (i), S. Goldberg (col.), A. Simek (let.), R. Thomas (ed). New York: Marvel Comics Group., Fair use.

> the first, fiftieth, or one hundredth issue are highlighted and celebrated. These literary, rhetorical, and commercial moves point to a self-awareness of the comic book as document and bibliographic object."

In Chapter 1 we discussed the increasingly important role of comics in different publishing markets, the research interest they are generating in fields such as digital humanities, literary studies or cognitive science, and how all these communities would greatly benefit from the development of semantic encodings for comics.

One such simple solution is *CoMet*[9], a very simple, open-source, *XML*-based encoding format for bibliographic and publishing metadata (e.g., title, publisher, editor, creators, series, volume etc.), whose biggest drawback is the lack of any capability for content markup. A slightly more complex solution is *E-Comics* [Arai and Tolle, 2010], an *XML*-based format primarily developed for adapting print content to mobile devices and thus allowing for the encoding of three types of information:

[9] http://www.denvog.com/comet/comet-specification/

- bibliographic metadata (e.g., title, chapter, creator, language, reading order etc.);
- panel information (e.g., panel number, order, URL, position);
- balloon information (e.g., balloon text, text position).

Since its focus is primarily (mobile-device) rendition, this format is somewhat limited as it lacks elements allowing for the description of a larger array of comics specific content, particularly characters, events, objects or onomatopoeia. This problem is partially solved by [Dunst et al., 2016], who developed *Graphic Narrative Markup Language (GNML)*, a *TEI*-based format allowing for a more detailed markup of panels, balloons, characters and onomatopoeia, including complex information, such as borders and shapes of elements.

Comic Book Markup Language (CBML)

However, the most complete and versatile comics encoding alternative up to date seems to be the *Comic Book Markup Language (CBML)* – developed by digital humanities scholar Andy Walsh (Indiana University – [Walsh, 2012]), adapting the richness of *TEI* to the annotation necessities of comics and sequential art. *CBML* is best adapted for representing, modelling and analysing 20th and 21st century comic books, daily comic strips, graphic novels, web comics, digital comics and **may** be adapted for virtually any other form of sequential art.

CBML relies on three native elements, which are enough for the basic decomposition of a comic book into panels and the separation between narrative and graphic components.

`<cbml:panel>` is used to describe the panel, as a structural component of a comic book page. Below, we give some of the most important attributes of this element:

- `@n`: specifies the inclusion of the current panel in a sequence of panels. This could be useful for rendering the content on small screen devices, on which panel to panel navigation is of paramount importance for obvious ergonomic reasons;
- `@characters`: allows for the enumeration of all the characters appearing in the current panel. The attribute is in fact a list of character identifiers, which should be initially described in a (potentially lengthy) `<teiHeader>`.
- `@ana`: this *TEI*-inherited attribute allows for the specification of a variety of analytic information pertaining to the panel that is being encoded. Used in conjunction with the `<interpGrp>` element, it can provide a useful tool for DH researchers studying particular aspects of a comic book page (e.g., the type of transitions from one panel to another, the shape of the panels in a page, the object in the central focal point of a panel, the drawing perspective of the depicted scene, etc.).

On top of this, one can easily group panels together, by using the `<div type=‘‘panelGrp’’>` construction, a technique which could be employed to describe local grouping of panels, such as double-page spreads.

While `<cbml:panel>` allows for the encoding of structural and graphic components of a page, the narrative is encoded using the other two main *CBML* elements: `<cbml:caption>` and `<cbml:balloon>`, the difference between them coming from their role in supporting either narration or dialogue respectively. Some of their most useful attributes are:

- `@rendition`: allows for the description of graphical features of speech balloons or captions (e.g., type of margin, colour, font type etc.)
- `@type`: this very useful attribute provides with a way to differentiate between different types of balloons (e.g. speech, thought, broadcast, etc.) as they have been described in Section 2.2.1. For a more granular distinction, the standard additionally provides with a `subtype` attribute (e.g. for broadcast communications: telecast, radiocast, telepathic, etc.).
- `@who`: identifies the character whose dialogue is being encoded.

In addition to these, the language is fully equipped for encoding and describing more sophisticated instances of the comics medium: comics with loose structures (allowing the reader to assume multiple reading paths), border-less panels (whose meaning and use was discussed back in Chapter 2.2.1), graphic elements crossing the panel boundaries or co-existing in more than one panel etc.

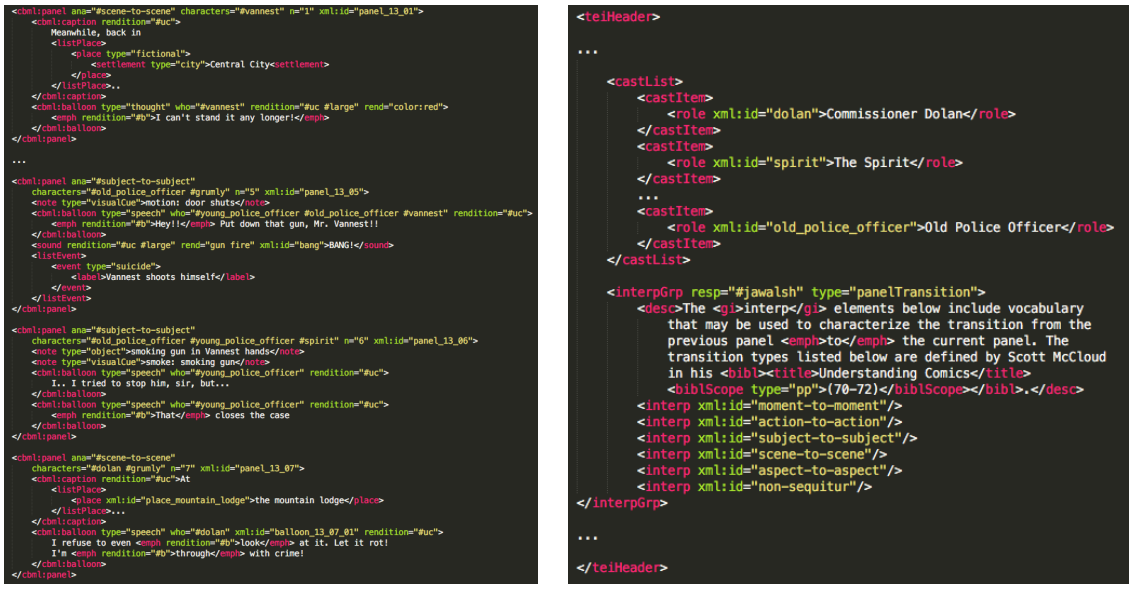
We conclude this section with a *CBML* encoding (Figure 3.5a) of a subset of panels (1, 5-7) for the comic book page in Figure 3.4, highlighting the use of *CBML*-specific elements and *TEI* stock elements, as well as the format's current limitations.

The *CBML*-specific elements are prefixed with the `cbml` namespace and encode the main structural elements of a comic book page: panels, balloons and captions. A page break is specified by using the *TEI* element `<pb>`. The example also presents the encoding of elements pertaining to the graphic nature of the medium:

- characters (the `@characters` attribute of the `<cmbl:panel>` element);
- places (using the *TEI*-specific `<settlement>`, `<place>` and `<placeList>` elements (see panel 1 for named fictive places and panel 7 for unnamed places);
- events (using the *TEI*-specific `<event>` and `<eventList>` elements (see panel 5);
- onomatopoeia and their meaning (using the *TEI*-specific `sound` and its `@rend` attribute (see panel 5).

The example also features the use of the `@ana` attribute of the `<cbml:panel>` element, which usually points to an interpretation (`<interp>`) defined inside an interpretation group (`<interpGrp>`), typically in the document's header. Figure 3.5b illustrates an encoding mechanism for the panel-to-panel transition typology, as it is defined in [McCloud, 1993, p. 70-72] and illustrated in Figure 2.23. This flexible mechanism can be generalised for

Figure 3.4.: Sample page from the story "The Grumly Murder" in *The Spirit #21*, Will Eisner (Jun. 1950), Retrieved: Digital Comics Museum, Fair use.



```
<cbml:panel ana="#scene-to-scene" characters="#vannest" n="1" xml:id="panel_13_01">
    <cbml:caption rendition="#uc">
        Meanwhile, back in
        <listPlace>
            <place type="fictional">
                <settlement type="city">Central City</settlement>
            </place>
        </listPlace>..
    </cbml:caption>
    <cbml:balloon type="thought" who="#vannest" rendition="#uc #large" rend="color:red">
        <emph rendition="#b">I can't stand it any longer!</emph>
    </cbml:balloon>
</cbml:panel>

...

<cbml:panel ana="#subject-to-subject"
    characters="#old_police_officer #grumly" n="5" xml:id="panel_13_05">
    <note type="visualCue">motion: door shuts</note>
    <cbml:balloon type="speech" who="#young_police_officer #old_police_officer #vannest" rendition="#uc">
        <emph rendition="#b">Hey!!</emph> Put down that gun, Mr. Vannest!!
    </cbml:balloon>
    <sound rendition="#uc #large" rend="gun fire" xml:id="bang">BANG!</sound>
    <listEvent>
        <event type="suicide">
            <label>Vannest shoots himself</label>
        </event>
    </listEvent>
</cbml:panel>

<cbml:panel ana="#subject-to-subject"
    characters="#old_police_officer #young_police_officer #spirit" n="6" xml:id="panel_13_06">
    <note type="object">smoking gun in Vannest hands</note>
    <note type="visualCue">smoke: smoking gun</note>
    <cbml:balloon type="speech" who="#young_police_officer" rendition="#uc">
        I.. I tried to stop him, sir, but...
    </cbml:balloon>
    <cbml:balloon type="speech" who="#young_police_officer" rendition="#uc">
        <emph rendition="#b">That</emph> closes the case
    </cbml:balloon>
</cbml:panel>

<cbml:panel ana="#scene-to-scene"
    characters="#dolan #grumly" n="7" xml:id="panel_13_07">
    <cbml:caption rendition="#uc">At
        <listPlace>
            <place xml:id="place_mountain_lodge">the mountain lodge</place>
        </listPlace>...
    </cbml:caption>
    <cbml:balloon type="speech" who="#dolan" xml:id="balloon_13_07_01" rendition="#uc">
        I refuse to even <emph rendition="#b">look</emph> at it. Let it rot!
        I'm <emph rendition="#b">through</emph> with crime!
    </cbml:balloon>
</cbml:panel>
```

(a)

```
<teiHeader>

...

    <castList>
        <castItem>
            <role xml:id="dolan">Commissioner Dolan</role>
        </castItem>
        <castItem>
            <role xml:id="spirit">The Spirit</role>
        </castItem>
        ...
        <castItem>
            <role xml:id="old_police_officer">Old Police Officer</role>
        </castItem>
    </castList>

    <interpGrp resp="#jawalsh" type="panelTransition">
        <desc>The <gi>interp</gi> elements below include vocabulary
            that may be used to characterize the transition from the
            previous panel <emph>to</emph> the current panel. The
            transition types listed below are defined by Scott McCloud
            in his <bibl><title>Understanding Comics</title>
            <biblScope type="pp">(70-72)</biblScope></bibl>.</desc>
        <interp xml:id="moment-to-moment"/>
        <interp xml:id="action-to-action"/>
        <interp xml:id="subject-to-subject"/>
        <interp xml:id="scene-to-scene"/>
        <interp xml:id="aspect-to-aspect"/>
        <interp xml:id="non-sequitur"/>
</interpGrp>

...

</teiHeader>
```

(b)

Figure 3.5.: a The *CBML* encoding of the comic book page in Figure 3.4.
b Customisation of the `<teiHeader>` element, containing a definition of both the list of characters (`<castList>`) and the description of panel-to-panel transitions typology (`interpGrp`) used by the `@ana` attribute of the `<cbml:panel>` element.

the definition of any kind of typology and opens up various avenues for analysis and interpretation of the graphic content of comics.

At the moment *CBML* is primarily focusing on structure, textual content and metadata, but falls short in what concerns the capability of encoding the more specific graphic elements of the comics medium. For instance, there is no solution for annotating objects (a topic on which *TEI* too has been stuck in debates since 2012), visual cues for smell or sound (i.e. movement lines), double-spread pages, whereas the description of loose structures such as splash pages and meta panels, even though achievable through the format's current schema, could be significantly improved through the addition of more specific attributes. We will return to this topic in Section 4.5, where we discuss several of our solutions to the format's limitations as well as directions for its future development.

3.2.3. Bibliographic standards for comics

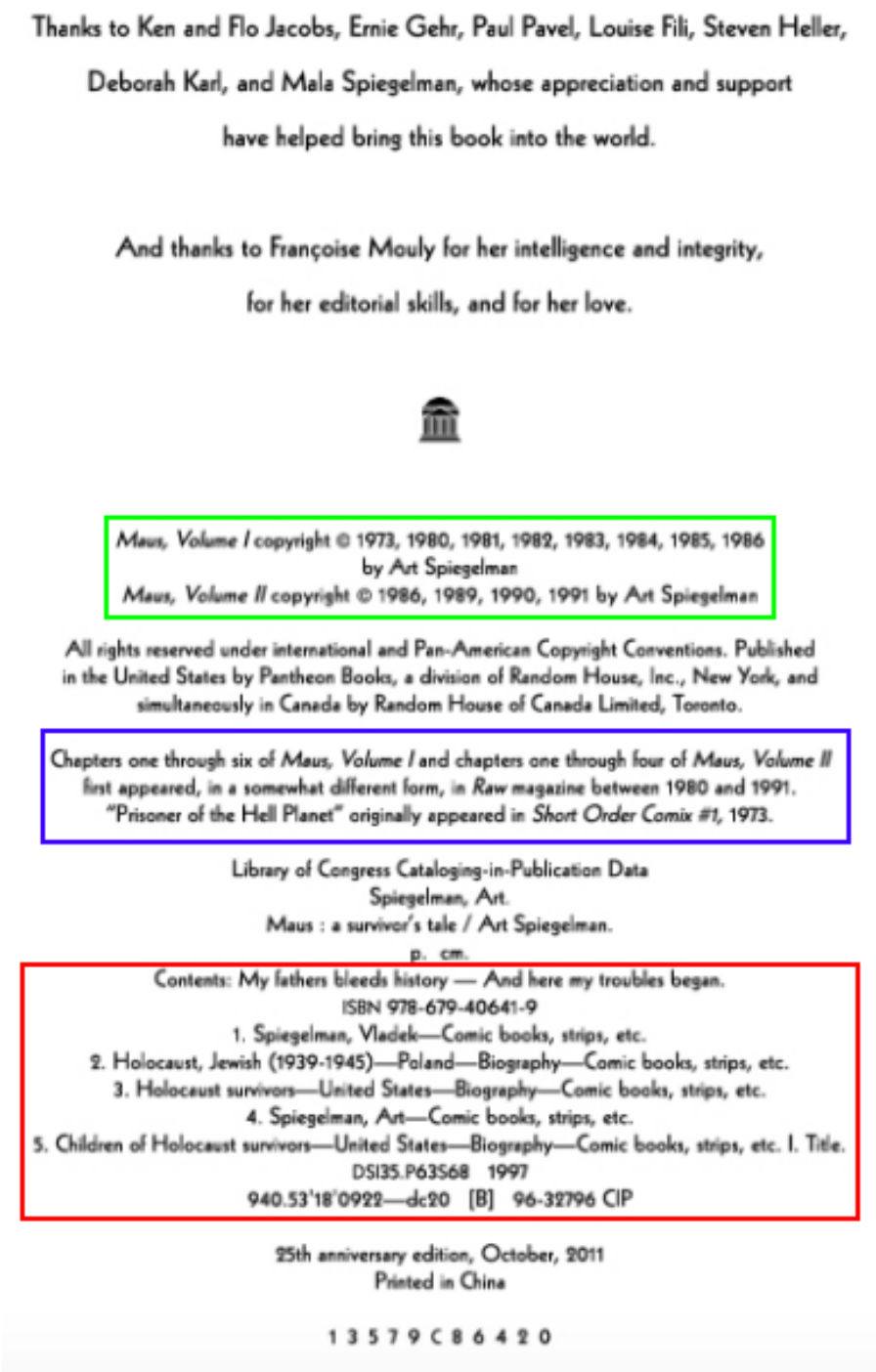
The previous sections of this chapter introduced some of the existing solutions to the rendition and encoding problems posed by comics publications. In the following, we investigate a third relevant aspect, that of bibliographic challenges specific to comics.

Such challenges originate, first and foremost, from the different forms of publishing and distribution of comic stories and comic books:

1. the space of an entire **issue** is dedicated to a **unique story** (e.g., most graphic novels are published as a one story book);
2. the same **story** extends over the entirety of a **series of issues** (e.g., *The Walking Dead* series);
3. an **issue** hosts **several stories**, which in turn extend over **several issues** themselves. For example, *Action Comics #1* (June 1938) contained eleven different features, such as: "Superman" (by Siegel and Shuster), "Chuck Dawson" (by H. Fleming), "Zatara Master Magician" (by Fred Guardineer), "Tex Thompson" (by Bernard Baily) etc. Some of these stories ("Zatara", "Tex Thompson") continued in subsequent issues of *Action Comics*, some were discontinued, while "Superman" eventually developed its own series and spin-offs.

This mix of variable, occasionally inconsistent publishing forms make the cataloguing, indexing and referencing of **bibliographic entities** related to comics stories and books particularly challenging: individual intellectual content (i.e. each story in a comic book), the issue itself (the collection of different stories forming an issue) and the related stories across different issues (serialized stories). This becomes even more intricate in the case of comics published using a mixture of these paradigms; for example, the stories across different issues can, at some point, be assembled in a volume of their own, as in the case of the Pulitzer award winning, graphic novel *Maus* [Spiegelman, 2011], whose chapters have first appeared as a series of stories in the underground "comix" magazine *Raw*. While struggling to find a publisher, the first six chapters were eventually released in the volume *"Maus: A Survivor's Tale* additionally subtitled "My Father Bleeds History" (very important!). Five years later, the last five chapters were assembled in a second volume, *And Here My Troubles Began*, before the two were put together and published in different editions (soft- and hard-cover, simple or with multimedia additions, for example interviews with the characters or the author). All this complicated "story of paternity" has made it very difficult for critics to classify the book, for booksellers to place it on their bookstore shelves and ultimately for librarians to catalogue it. An eloquent example of this is the provided by the mere examination of the front page of the 2011 anniversary hard-cover edition Figure 3.6, which presents not one, but several "unique" classifications according to the US Library of Congress Cataloguing.

A second delicate issue is related to the particularities of serialising comics productions. It is very usual for comics narratives to span over large periods of time (sometimes even decades), thus involving different generations of creative teams (authors and graphic artist first and foremost, but also inkers or letterers) working on the same story. Also, numerous comics series in the past have been the object of changes of ownership, with strategies of new publishers seeing characters "repackaging" (moving them from one magazine to another), awarding them their own issues (e.g., "Superman" – see Figure 3.7 or "Spider-Man")



Thanks to Ken and Flo Jacobs, Ernie Gehr, Paul Pavel, Louise Fili, Steven Heller,
Deborah Karl, and Mala Spiegelman, whose appreciation and support
have helped bring this book into the world.

And thanks to Françoise Mouly for her intelligence and integrity,
for her editorial skills, and for her love.

Maus, Volume I copyright © 1973, 1980, 1981, 1982, 1983, 1984, 1985, 1986
by Art Spiegelman
Maus, Volume II copyright © 1986, 1989, 1990, 1991 by Art Spiegelman

All rights reserved under international and Pan-American Copyright Conventions. Published in the United States by Pantheon Books, a division of Random House, Inc., New York, and simultaneously in Canada by Random House of Canada Limited, Toronto.

Chapters one through six of *Maus, Volume I* and chapters one through four of *Maus, Volume II* first appeared, in a somewhat different form, in *Raw* magazine between 1980 and 1991.
"Prisoner of the Hell Planet" originally appeared in *Short Order Comix #1*, 1973.

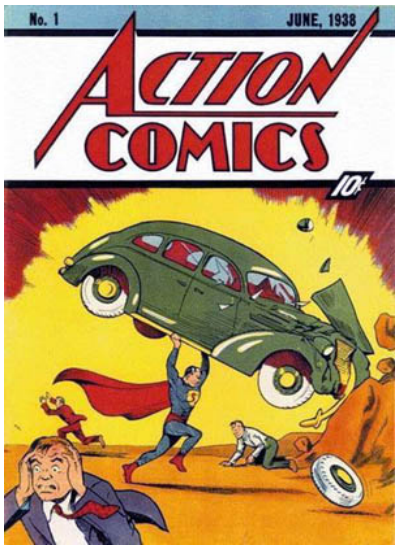
Library of Congress Cataloging-in-Publication Data
Spiegelman, Art.
Maus : a survivor's tale / Art Spiegelman.
p. cm.
Contents: My fathers bleeds history — And here my troubles began.
ISBN 978-679-40641-9
1. Spiegelman, Vladek—Comic books, strips, etc.
2. Holocaust, Jewish (1939-1945)—Poland—Biography—Comic books, strips, etc.
3. Holocaust survivors—United States—Biography—Comic books, strips, etc.
4. Spiegelman, Art—Comic books, strips, etc.
5. Children of Holocaust survivors—United States—Biography—Comic books, strips, etc. I. Title.
DS135.P63S68 1997
940.53'18'0922—dc20 [B] 96-32796 CIP

25th anniversary edition, October, 2011
Printed in China

1 3 5 7 9 C 8 6 4 2 0

Figure 3.6.: Facsimile of the title page of the 2011 hard cover edition of *Maus*. Notice the bibliographic particularities of this publication, i.e. serialisation, progressive publishing and composition (blue), multiple copyright (green). In the bottom half, notice the multiple data and classifications made by the US Library of Congress Cataloguing (blue).
Source: *Maus*, Art Spiegelman [Spiegelman, 2011]

(a)



(b)

(c)

(d)

Figure 3.7.: "Superman" stories moved from newspaper strips (a), to series of stories in comic books (b), to the creation of the *Superman* comic book (c). More so, the magazine has been renamed and renumbered on several occasions (d) and special issues have been inserted inside the original series.
a The first tier of the first Superman Sunday strip, Jerry Siegel (story) and Joe Shuster (art) (Nov. 5, 1939). The Superman strips ran from 1939-1966 and at their peak were published in almost 400 newspapers, with a readership of approximately 20 millions. Source: The Speeding Bullet, Fair use.
b Cover of *Action Comics #1* (June 1938, DC Comics). Art by Joe Shuster (art), and Jack Adler (colour). *Action Comics* is the first comic book to host a series of "Superman" stories. Source: "Action Comics #1" at The Grand Comics Database. Retrieved October 31, 2006. (similar file if not the one originally uploaded.), Fair use.
c Cover of *Superman #1* (Summer, 1939). Art by Joe Shuster. Following their success in *Action Comics*, "Superman" stories started featuring in their own title comic book starting with 1939. Source: Grand Comics Database, Fair use.
d Cover of the last issue of *The Adventures of Superman #649* (Apr., 2006). Art by Ivan Reis. Between 1986-2006, *Superman* magazine was renamed *The Adventures of Superman*, with its "first" issue continuing from #424, while a brand new series used the *Superman* title. In May 2006, after 228 monthly issues, it was returned to its initial title and numbering. Source: Wikipedia, Fair use.

or simply reinitialising a series' count, as a marketing move to attract attention on the refreshed title.

Serialisation issues can get particularly complicated in the case of cartoons and comics published by early days comic strip syndicates that sought to place their authors' work in as many newspapers or magazines as possible. Alternatively, DC's staple reboot technique saw the revival of a number of earlier superhero characters (e.g., The Flash, Green Lantern, Hawkman, The Atom, Wonder Woman, Batman, Superman), sometimes in renumbered series, adding more complexity to the work of those following the continuity of such characters. This is the case of Marvel's *Fantastic Four*, which appears to have no less than three #1 issues, in 1961, 1982 and 1996. Despite their more recent beginnings, graphic novels too have come with their fair share of bibliographic complications, as we have previously seen in the case of *Maus'* publishing saga.

Batman is another interesting case, rather because of the character's complex origins. Batman itself is not a pure character, being based on previous Golden Age pulp heroes (e.g., The Shadow, Doc Savage, The Spider, The Black Bat). Its official début, by the creative duo made of Bob Kane and Bill Finger, was in *Detective Comics #27* (1939); the character was rebooted in 1960 within the *Justice League of America*, with several artists developing different aspects of its art and narrative before its great 1986 revival, in Frank Miller's *Dark Knight Returns*. Just like before, Batman's successful "revival" led to DC developing subsequent stories during the 1990s and 2000s under the direction of several different artistic teams and published in different formats: ongoing series, limited series, graphic novels and spin-offs.

These are just a few examples supporting the necessity of technical solutions adapted to the bibliographic complexities of comic books publishing. It is important that such solutions be able to keep track of the entire metadata time-line of a comic book title and as we have seen in some cases, that of its characters. In addition, such solutions should foresee the requirements imposed by the development of digital comics publications. In the following section, we introduce one such standard and discuss its possible implementation into future technical formats for comics encoding.

Overview of bibliographic standards

We carry on with a short overview of several of the most relevant bibliographic standards and formats, from standard identifier registries (ISBN, ISSN, ISTC) and actionable identifiers (DOI) to rich metadata schemas (ONIX) and authority files for the characterisation of a publication's content (GND, VIAF, LCSH).

ISBN, ISSN, ISTC

International Standard Book Number (ISBN)[10] is a standard number assigned to each edition and variation of a book, meaning that a paperback, a hardcover and an electronic version of the same book will each have different ISBNs. The number consists of a 10 or 13 digit sequence, usually represented in the form of an EAN bar code and has the following structure:

1. prefix (3 digits): used to identify the "book" (978-) according to the EAN schema[11] and is pre-pended to ISBN-10 codes to obtain the ISBN-13[12];
2. registration group element (1-5 digits): identifies the language of a publication, within the 978- prefix. For example, 0 and 1 corresponds to English-speaking countries, 2 to French, 3 to German, 4 to Japanese etc. For a more efficient use of the number space, longer sequences are assigned to books published in rare languages (e.g., 99936 corresponds to Bhutan);
3. the registrant and the publication elements (4-8 digits): a subsequence of digits designates a publisher, which then uses the remaining subsequence to identify its publications;
4. checksum (1 digit): a control digit (or "X"), obtained from all the other digits using modulo arithmetic.

International Standard Serial Number (ISSN)[13] is an 8-digit[14] code that can be used in conjunction with ISBN to uniquely identify serial publications, however holding no information pertaining to location, language or publisher and a new number is generated when a publication suffers modifications. The standard does distinguish, however between the publication of the same serial in different media (e.g., print and digital) for which it requires the generation of different codes. While potentially useful before the 1990s, this requirement is needless at best today, rendering the standard too variable with respect to the same content, essentially ISSN's main limitation. This has been addressed as of 2007 with the introduction of ISSN-L, a method for linking the same series across different media.

[10] https://www.isbn-international.org/

[11] EAN is a geographically based schema with the first 3 digits corresponding to a geo-location code. For the sake of consistency, since EAN identifies books irrespective of their location, the 978- prefix identifies the fictive "country" of "Bookland".

[12] This follows the adoption of ISBN by EAN, as the latter wished to avoid the creation of their own redundant code.

[13] http://www.issn.org/

[14] For EAN compliance, similar to the ISBN prefix (978-), the 977- "country" prefix is reserved for ISSN codes.

International Standard Text Code (ISTC)[15] is a unique identifier used for text-based works, which can easily be used in conjunction with ISBN, typically as an attribute identifying the content of a publication. It is useful to identify similar content published under different titles, group together different, but related content or to separate different content, published under the same title.

DOI

Digital Object Identifier (DOI)[16] is a persistent identifier designed to uniquely identify objects (physical or digital)[17] and link to their associated metadata. It is structurally composed of a suffix and a prefix, separated by a slash:

1. prefix: has the form `10.NNNN`, where `NNNN` is a series of at least 4 digits (at least `1000`), which can be further subdivided; it identifies the registrant of the name (e.g., `10.1007` identifies "Springer New York".)

2. suffix: can be any Unicode character and is used by the registrant according to its own needs (e.g., it can also be an ISBN, in which case it achieves precisely that function).

DOI has several important advantages, including its capability of storing volatile information (i.e. resource locations, ownership) as a permanent unique identifier, independent of any object properties (physical or logical). Thus, an object can change properties by updating its related metadata and without changing its DOI, resulting in a more robust identification scheme. Another useful advantage of DOI is its capability of encoding objects of and at different levels of granularity (e.g., books, chapters, articles, tables, figures etc.). In the case of digital comics, this could be applied for encoding the different parts of a comic book (e.g., series, issue, page, panel, character, balloon etc.).

ONIX

Online Information Exchange (ONIX) is a set of three *XML*-based formats (ONIX for Books[18], for Serials and for Publication Licenses) designed for capturing and exchanging a wide array of bibliographic data related to print and digital books. Maintained by the Association of American Publishers and *EDItEUR*, the format is not meant to characterise the content of a publication, but rather to provide with a rich set of metadata about the book as a product (authorship, publishing, physical properties and even sales data).

GND, VIAF, LCSH

Integrated Authority File (de. Gemeinsame Normdatei) (GND)[19] is an international authority file managed by the German National Library *Deutsche Nationalbibliothek*),

[15] http://www.istc-international.org/
[16] https://www.doi.org/
[17] DOI means "digital identifier of an object" rather than "identifier of a digital object".
[18] http://www.editeur.org/83/Overview/
[19] http://www.dnb.de/EN/Standardisierung/GND/gnd.html

enabling documentation tasks in libraries, archives and museums. GND provides with a high-level hierarchy of entities (person, name, corporate body, event, work, topical term, geographical place name), as well as an *RDF*[20] format ontology.

Virtual International Authority File (VIAF)[21] is an international project hosted by the Online Computer Library Center (OCLC) with the objective of linking and thus increasing the utility of different national authority files[22]. VIAF contains bibliographic information on: corporate names, geographic names, personal names, works and expressions.

One of the most widespread used bibliographic standards is the Library of Congress Subject Headings (LCSH)[23], an comprehensive controlled vocabulary of subject headings, that allows users to query a publication's collection beyond common bibliographic metadata (i.e. title, author, publisher). LCSH gathers several collections of headings covering every subject used to characterise the content of a publication.

All these standards have long proven their efficiency in matters of identification and description of various aspects of print and digital publishing. Earlier in this section, we gave several examples of bibliographic particularities of comic book publications, with which none of the standards presented so far deals in a complete and satisfactory manner. Consequently, we dedicate the following section to an extensive presentation of one of the most complete and useful standards rooted in bibliographic and library practises.

Functional Requirements for Bibliographic References

For the past 50 years, the International Federation of Library Associations and Institutions (IFLA) has dedicated its efforts towards standardising a large array of library practises. One of their numerous resulting standards, Functional Requirements for Bibliographic Records (FRBR) [FRBR, 1998] was developed as a set of user-centric requirements, rooted in the need to fully understand the role of bibliographic references for librarian, archiving and cataloguing activities and the way these can assist their users in achieving their tasks.

One would argue the very technical nature of this set of requirements, but it is exactly the exhaustiveness of *FRBR* that provides with answers or at least solid leads for many of the bibliographic challenges of comics publications. In *FRBR* users interact with an Entity-Relationship Model, to achieve four main tasks [FRBR, 1998]:

[20] Resource Description Format
[21] http://viaf.org/
[22] Amongst them the US Library of Congress, the German National Library (*Deutsche Nationalbibliothek*), the British Library, the French National Library (Bibliothèque Nationale de France), Vatican Library (*Bibliotheca Apostolica Vaticana* etc.
[23] http://id.loc.gov/authorities/subjects.html

1. find an entity: the user states a series of search criteria, based on entity attributes or relationships between entity; the model is able to return one or more matching entities;
2. identify an entity: the user confirms that the entity found is the one they were seeking; in case of more than one matching entities, the user chooses the best one;
3. select an entity: users select an entity according to their specific needs (i.e. content, format, etc.);
4. acquire an entity: through purchase, loan, online access etc.

Entities

The *FRBR* model defines entities as key objects of interest to users of bibliographic data and separates them in three groups.

Group 1 Entities define "the different aspects of user interest in the products of intellectual or artistic endeavour". Figure 3.8 below illustrates the four group 1 entities: Work, Expression, Manifestation and Item and their primary relationships with one another.

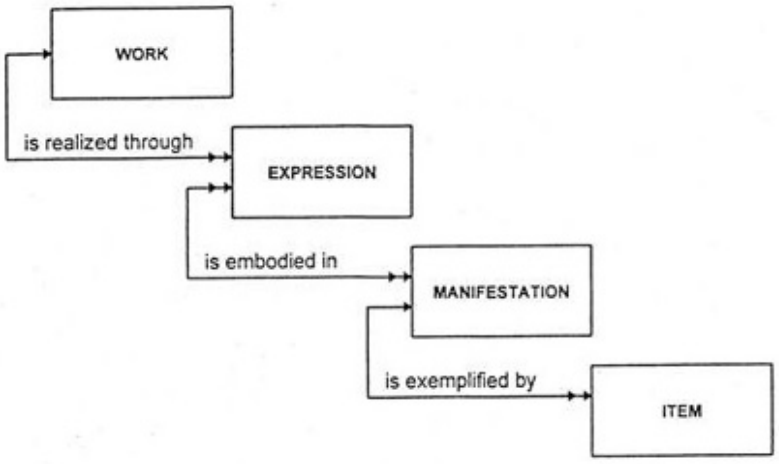


Figure 3.8.: Group 1 Entities in the FRBR model.
Source: IFLA.

Work is the most abstract entity of the model and identifies a distinct intellectual or artistic creation. The Work is defining the commonality of its Expressions.

Expression is "the intellectual or artistic realisation of a Work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or

any combination of such forms" [FRBR, 1998]. Several particular cases need to be mentioned here, as they are very often encountered in comics:

- a change in the form of an Expression leads to the creation of a new Expression of the same Work. This is the case of revisions, updates or augmentations (with illustrations, notes or glosses).
  Example:

  w Art Spiegelman's "Maus"

  e1 the first complete edition of the graphic novel

  e2 the second complete edition of the graphic novel, with revised drawings

  e3 the last complete edition of the graphic novel, together with interviews

- translations from one language to another generate a new Expression *Example*:

  w Scott McCloud's "Understanding Comics"

  e1 the original English language edition

  e2 *L'art invisible* – the French translation

In essence, an Expression should capture the intellectual or artistic differences within the content of the same Work; at the same time it is a common link for different Manifestations.

Manifestation is the physical embodiment of an Expression. It refers to all objects bearing the same characteristics, in terms of both intellectual content and physical support. This means that any kind of change impacting the physical form generates a new Manifestation within the current Expression. This is a very common case in the comics industry where different publishing and marketing strategies see comics repackaged (e.g., bundles – several comics sold together, deluxe editions etc.), digitised, reprinted in different formats etc.
Example:

w Art Spiegelman's Maus

e1 the complete original English version

m1 the first soft-cover edition printed by Pantheon

m2 the first hard-cover edition printed by Pantheon

m3 a scanned version packaged as a *CBR* archive

Item is the single exemplar of a Manifestation. This is the most concrete entity in Group 1 and, in the cases of comics can be a book, a flip-flop magazine, the digital files on

which the content is stored etc. Thinking about it in terms of intellectual content, an Item is one and the same thing as its Manifestation. The difference is made in terms of physical properties, an Item being simply one of many (most of the time) physical or digital instances of a Manifestation.
Example:

w Art Spiegelman's Maus

e1 the complete original English version

m1 the first soft-cover edition printed by Pantheon

i1 copy autographed by the author

i2 copy held by the Local Library

Group 2 Entities are responsible for the creation, production, dissemination and custodianship of the intellectual content. There are two such entities:

Person. Defines the person or persons, dead or alive, who were involved in the creation or realisation of the Work. Alternatively, the model allows for this entity to identify a person who is the subject of a Work. In the case of comic books, this entity would be used to identify those responsible with the creation of a work (authors, writers, graphic artists, pencilers, inkers, letterers), those involved in its publishing and distribution (editors, publishers, booksellers), but it may very well identify characters in a comic book (Superman, Batman, Vladek – from *Maus*, Ally – from *Fun Home* etc.)

Corporate Body. Describes an organisation, group of individuals or organisations involved in the creation or realisation of a Work, acting as a unit and identifiable by a name. An example pertinent to the case of comic book is that of a publishing house (e.g., Pantheon, Glénat), a distributor (e.g., Hachette, Amazon) or a digital platform (e.g., ComiXology, iBooks or our partner's very own Sequencity). On top of this, the specification allows for this same entity to describe organised gatherings and events. This is tremendously useful for a full and accurate description of a comics ecosystem in which alternate means of engagement with comics, such as fan-mail, "fan-zines" (magazines with fan-created content) and comic conventions have historically played an important part.

Finally, the **Group 3 Entities** of the model are an additional set of entities that allow for a more complete description of the subject of a Work. Here is how the *FRBR* standard describes each of them [FRBR, 1998, Section 3.2]:

Concept. Refers to an abstract notion or an idea, such as fields of knowledge, disciplines, schools of thought (philosophies, religions, political ideologies, etc.), theories, processes, techniques, practises, etc.

Object. Refers to a material thing, usually describing animate and inanimate objects occurring in nature, fixed, movable, and moving objects that are the product of human creation or objects that no longer exist.

Events. Refer to a comprehensive range of actions and occurrences that may be the subject of a work: historical events, epochs, periods of time, etc.

Place. Refers to a diversity of locations – terrestrial and extra-terrestrial, historical and contemporary, geographic features and geopolitical jurisdictions.

Besides the obvious advantages of using these latter entities for a rich description of bibliographic metadata pertaining to a Work, they have also been tremendously helpful in formally confirming the results of our effort – based largely on scholar research about comics – of identifying the most important elements within, amenable to crowdsourced annotation.

Attributes

Each of the previous entities exhibit a set of attributes allowing for an accurate description of both related metadata and their roles in the model. Table 3.1 below lists the attributes most relevant to comics for each of the Group 1 and 2 Entities. For a full description of all attributes of all entities, please refer to the full *FRBR* specification [FRBR, 1998, Chapter 4].

Relationships

Relationships between entities are means of assisting the user to "navigate" the universe that is represented in a bibliography, catalogue, or bibliographic database. Beyond this basic purpose, the relationships in the *FRBR* model are meant to standardise the very diverse and rather non-uniform language used by professionals in the publishing industry.

FRBR defines two big classes of relationships:

Relationships in high level diagrams refer to the main relationships between Work, Expression, Manifestation and Item (Group 1 Entities), such as they were depicted in Figure 3.8 earlier in this section. These relationships should be viewed as a logical flow describing the intellectual content from the most abstract (Work) to its physical embodiment (Item). But other such relationships can be established between these same Group 1 Entities and Group 2 Entities (Persons and Corporate Bodies). In this case, the relationships describe authorship, intellectual property or the distribution process. Finally, relationships linking to a Person or a Corporate Body can also describe the role of these entities as subjects featured within the content (i.e. characters).

A second category of relationships are those that link two Group 1 Entities of the same kind (Work-to-Work, Expression-to-Expression, Manifestation-to-Manifestation, Item-

Entities	Attributes
Work	title, form, date, other distinguishing characteristics, intended termination, intended audience, context
Expression	title, form, date, language, other distinguishing characteristics, extensibility, revisability, extent, summarization, context, critical response, use restrictions, sequencing pattern (serial), expected frequency (serial), technique
Manifestation	title, statement of responsibility, edition/issue designation, place of publication/distribution, publisher/distributor, date of publication/distribution, fabricator/manufacturer, series statement, form of carrier, extent of carrier, physical medium, capture mode, dimensions of the carrier, identifier, source for acquisition, terms of availability, access restrictions, typeface, type size, foliation (hand-printed book), collation (hand-printed book), publication status (serial), numbering (serial), system requirements (electronic resource), file characteristics (electronic resource), mode of access (electronic resource), access address (electronic resource)
Item	item, fingerprint, provenance, marks/inscriptions, exhibition history, condition, treatment history, scheduled treatment, access restrictions
Person	name, date, title, other associated designations
Corporate Body	name, associated number, associated place, associated date, other associated designations

Table 3.1.: A selection of comic book-specific attributes corresponding to Group 1 (upper) and 2 (lower) Entities in the *FRBR* model [FRBR, 1998, Chapter 4].

to-Item) or of different kinds (Expression-to-Work, Manifestation-to-Item). These relationships cover a large number of use cases, some of them accurately serving the needs of comic books publishers and readers:

- Work-to-Work and Expression-to-Work: successor and supplement relationships to describe the serialisation of a work (e.g., the volumes of a graphic novel, such as *Maus* or *The Dark Knight Returns*); adaptation (e.g., translations), transformation (e.g., the screenplay based on Batman which is later transformed into a movie) or imitation (e.g. parodies to superhero comics);
- Expression-to-Expression: of the same Work (e.g., abridgement, revision or translation) or between Expressions of different Works. This is just an "instantiation" of the Work-to-Work relationships allowing for the serialisation or transformation of Expressions of comics and graphic novels.

- Manifestation-to-Manifestation relationships refer to either reproductions (e.g., a reprint of the original *Binky Brown* with the author's notes and comments in facsimile) or alternate releases.
- Manifestation-to-Item and Item-to-Item relationships describe those situations related to the physical or electronic support of the publication, such as reconfigurations (e.g., different bindings of a comic book as a flip-flop or as a hard-cover) and reproductions.

Alternatively, each Entity is equipped with a Whole/Part-Of relationship, allowing for compositions/decompositions at different levels: a Work composed of chapters, sections, volumes, issues; Expressions into table of contents, volumes, issues, illustrations; the physical components or bindings of an Item.

Now that we have seen all the different Entities and Relationships composing the *FRBR* model, let us take a closer look at Figure 3.9 and Figure 3.10, illustrating the model's use to represent a bibliographically complex comic book – Art Spiegelman's, *Maus* [Spiegelman, 2011].

Figure 3.9 presents a description of Spiegelman's story "Prisoner on the Hell Planet", which was initially published in an underground "comix" magazine (1973) before it was included in the same form inside *Maus*' chapter 5 (1983). FRBR will not differentiate between these two instances at abstract levels (work, expression), but will do so at the more concrete manifestation level. The *wSOC, eSOC* refer to the entire work and expression of *Short Order Comix #1*, the magazine that first hosted the story.

Figure 3.9.: A description of Art Spiegelman's "Prisoner on Planet Hell" using FRBR.

Several remarks need to be made for a better understanding of this example of applying FRBR to one of the most bibliographically complex works of literature, an extreme case – if you will – of all the particularities that can arise in comic books:

1. the first ten chapters of the general work (*w0*) initially appeared on a per-chapter basis, published as an insert story within the yearly issues of *Raw Magazine Vol.1 #2 – Vol.2 #3* (1980-1991). Hence, it was necessary to identify each chapter-story

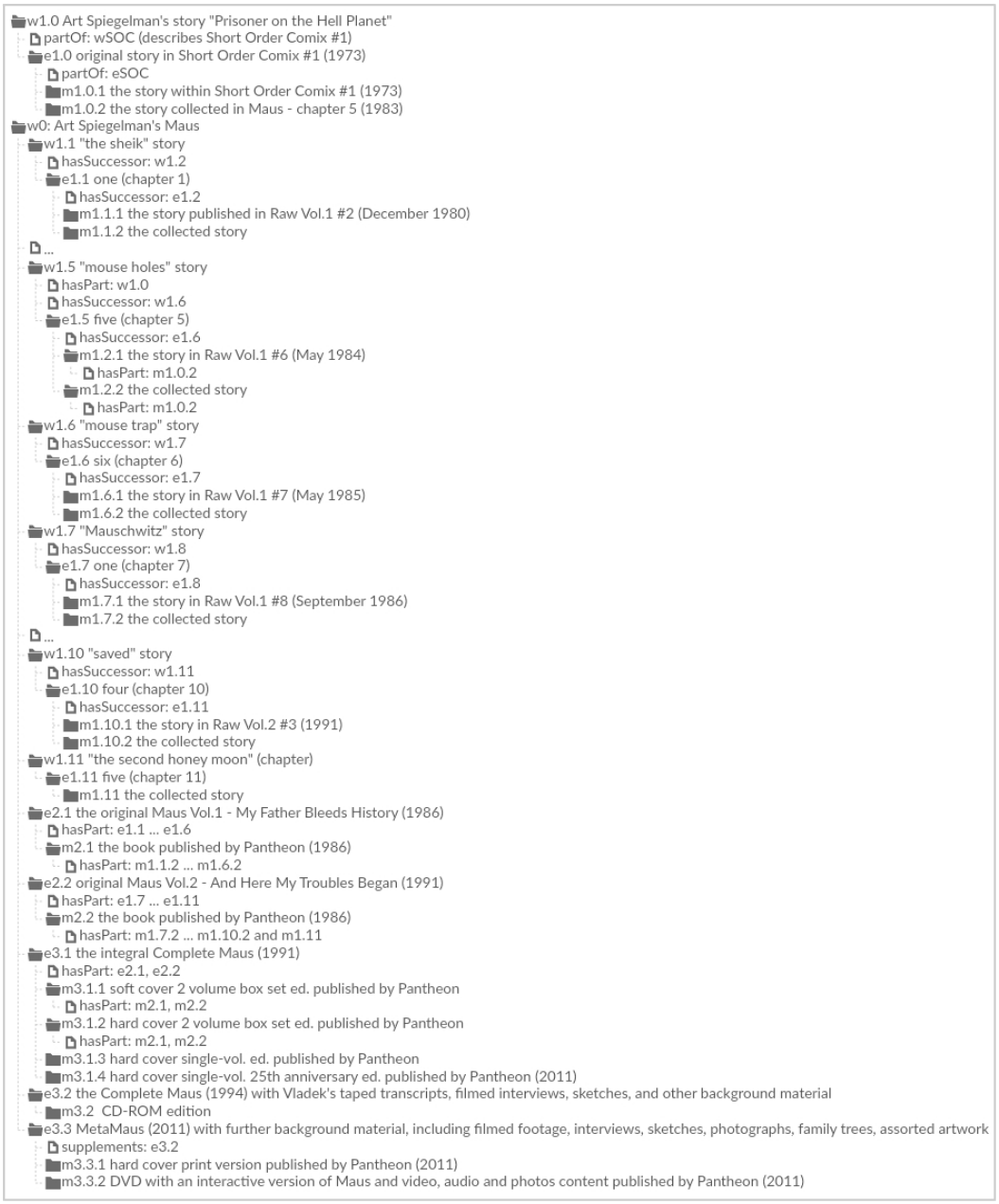


Figure 3.10.: An extended bibliographic description of Art Spiegelman's *Maus* using FRBR.

as its own work (*w1.1* to *w1.10*) which are linked to the main work through the Work-to-Work *hasPart* relationship. The work in each chapter is embodied in a corresponding expression (*e1* to *e10*);

2. the first six chapters were collected in *Maus Vol.1 – My Father Bleeds History* (1986). This generates a new expression, corresponding to this volume (*e2.1*). This expression is composed by the expressions corresponding to the first six chapters (*e1.1* to *e1.6*), to which it is linked by an Expression-to-Expression *hasPart* relationship;

3. chapter 11, the last one, was not previously serialised in any other publication. Its first appearance was in the collected *Maus Vol.2 – And Here My Troubles Began* (1991). Its expression (*e2.2*) includes those corresponding to chapters seven to eleven (*e1.7* to *e1.11*);

4. each expression of a chapter has two manifestations (embodiments): the first one, as a story in each issue of *Raw Magazine* (*m1.X.1*) and the second as part of one of the two volumes of the collected work (*m1.X.2*). The only exception is chapter 11, which, as we have seen, has only one manifestation (m.1.11) within *Maus – Vol.2*;

5. chapter five presents another particularity, as it encloses "Prisoner on the Hell Planet", a separate story of Spiegelman, which had appeared in an underground "comix" magazine in 1973. Earlier in this example we have seen how the expression of this story is embodied by two manifestations: one for the text which appeared in the magazine and one for the different collections of the text. The latter (*m1.0.2*) is connected to chapter's five both manifestations in *Raw Magazine* (*m1.5.1*) and within *Maus* (*m1.5.2*) by a Manifestation-to-Manifestation *hasPart* relationship;

6. the expressions corresponding to the two separate volumes (*e2.1* and *e2.2*) are derived from the root *Maus* work (*w0*). Since they are just parts of the same work, according to the FRBR model, their simple division in distinct volumes doesn't require us to model two new works. In the same way, the integral collections are represented as expressions (*e3.1*, *e3.2* and *e3.3*) of the same root work;

7. the integral *Maus* has been published by Pantheon (*e3.1*) in different variations (soft and hard cover two volumes, one-cover single volume), each of these corresponding to a manifestation;

8. the integral *Maus* published by Voyager in 1994 is embodied by only one CD-ROM manifestation. Due to the additional content within (Vladek's taped transcripts, filmed interviews, sketches) it required its inclusion into its own expression (*e3.2*), as opposed to being just a manifestation of *e3.1*;

9. *MetaMaus* is a multimedia type of publication, hence the creation of a separate expression (*e3.3*), holding a print (*m3.3.1*) and a DVD manifestation (*m3.3.2*), each of which featuring sizable additional content.

3.3. Putting everything together

This chapter has covered the diversity of standards and formats needed for a full specification of a comic book publication, from the publishing, semantic and bibliographic perspectives. While these standards are currently used to address problems within the scope of their specific domains, the open nature of (digital) content creation and its large scale use in the recent years lead us to believe that a new, integrated view over the creation, consumption and exploitation of digital publications has to be considered. In particular, it is our belief that complex, user-oriented standards such as *FRBR* should have a bigger influence over the future evolutions of packaging standards (i.e. *EPUB Open Container Format*). At the same time, semantic formats such as *CBML* should be able to enhance content standards (i.e. *EPUB Content Documents*) and support use cases specific to bibliographic standards (*FRBR*). Equally important, such bibliographic standards should be able to better synchronise with publication metadata standards, such as *EPUB Publications*.

In the following chapters, we will discuss possible integrations of these stand-alone standards and formats, hopefully opening up opportunities for the future development of digital publications.

Chapter 4.

New standards for comics' digital publishing

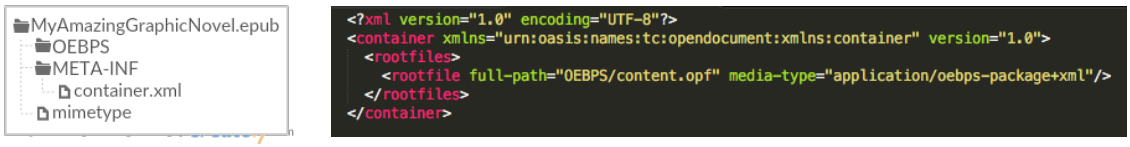
This chapter describes a new open standard for the digital publishing of comics, taking into account the limitations of the different standards and formats discussed in the previous chapter. In it, we saw the emergence of *EPUB* as the most complete open source standard supporting the packaging, presentation and distribution of image-based publications. These allow us to place *EPUB* at the centre of our development of new open standards for the digital publishing of comics. We have also seen how *CBML* – the format for semantic encoding of comics – should be able to enhance *EPUB*'s presentation capabilities, while bibliographic standards – such as *FRBR* – help untangle some of the complexities related to comic books bibliographies and offer new development directions in terms of packaging and descriptive metadata.

4.1. EPUB

This section briefly discusses the technicalities of what an *EPUB* package really is. Such a package may seem confusing and overwhelming at first. However, with the help of a detailed example we will tear it down bit by bit, explain the role of each part, then put everything back together and see how all parts function together as a system.

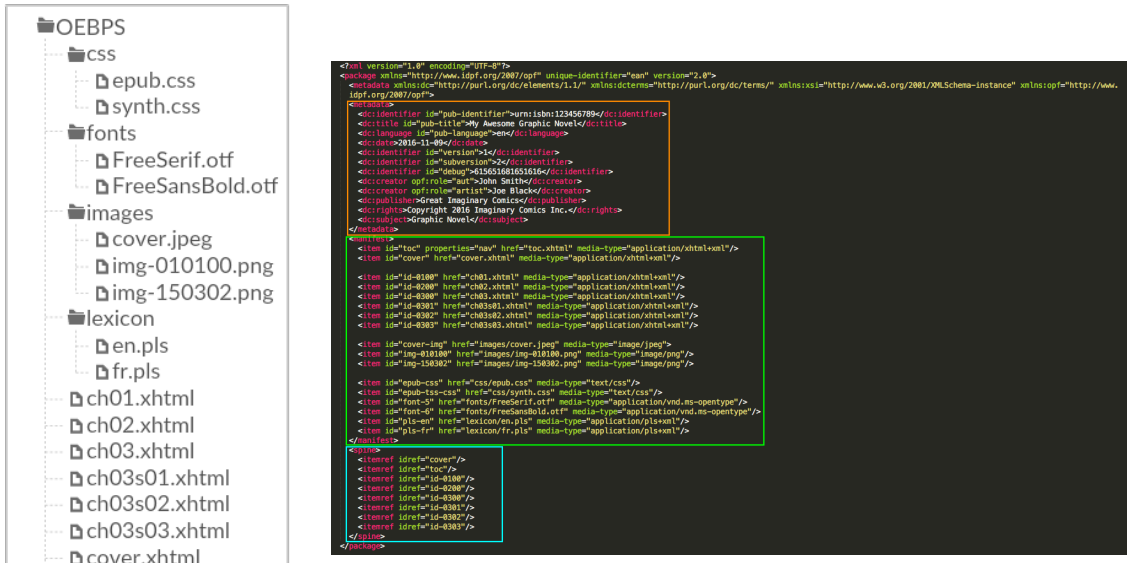
We recall the overview of the *EPUB* family of standards from the previous chapter, defining the three main parts of the standard: Publications [EPUB-Publications, 2014], Content Documents [EPUB-ContentDocs, 2014] and Open Container Format [EPUB-OCF, 2014]. How does it all come together, from the creation of the files, their packaging, all the way to the rendering of the file on our computer, tablet or mobile phone?

The **Content Format** specification, tells us that an *EPUB* file is nothing more than a *ZIP* package, containing a series of content documents. This is done through a simple line in the package-specific `mimetype` file, placed at the root level of the container (see



(a) (b)

Figure 4.1.: The high-level structure of a typical EPUB package (a) and of the `META-INf/container.xml` document (b). Within, notice the reference to `OEBPS/content.opf`, the Open Package document.



(a) (b)

Figure 4.2.: The placement of the Content Documents within the EPUB Package (a). It contains all *HTML* files, images, stylesheets, fonts and auxiliary files necessary for the rendering of the *EPUB* file.
All these elements come together in the `OEBPS/package.opf` document (b), in which we distinguish three parts: the metadata (orange), the manifest (green – a list of all the content documents in the package) and the spine (blue – the rendering order of the final content documents).

Figure 4.1). At the same level, the `META-INF/container.xml` file is the place where the path to the root file of the content documents is specified.

At the root file location, Figure 4.2 gives an overview of the typical **content documents**, which are usually zipped inside an *EPUB* package. The **Content Documents** specification defines these documents as subsets of the most common web-based documents, namely *XHTML, CSS* and *SVG*. Two additional specifications enhance the content documents with capabilities for synchronising text and audio (Media Overlays [EPUB-MediaOverlays, 2014]), as well as referencing arbitrary content within the publication (Canonical Fragment Identifiers (CFI) [EPUB-CFI, 2016]).

At the very same root level, one can finally notice the `package.opf` file. This mandatory file is defined according to the **Publications** specification and is composed of three main parts:

1. the **metadata section** contains a variety of publication and package metadata, most of it rooted in the *Dublin Core* schema[1]. Some of the most commonly used are: `dc:identifier` (the publication's identifier – e.g., ISBN, ISSN, DOI etc.), `dc:title` (the publication's title; it can also be refined to allow for subtitles), `dc:language` (the publication's language), `dc:date` (allowing for example, for different release dates), as well as editorial metadata (creator, contributor, editor, with the possibility of refining the contributor's role to, for example in the case of comics, main author, drawer, inker, translator);

2. the **manifest section** is a list-style enumeration of all the composing resources (content documents). The correct rendition of each of these items depends on the correct specification of a series of attributes: the resource's `id`, its relative path (`href`) and its `media-type`.

3. the **spine section** is an enumeration of the items in the section above, based on their defined `id`'s, this time in the order in which they will be rendered (e.g., chapters in a book or a graphic novel, pages in a comic book etc.).

As for the reading systems, due to the its open-source, democratic nature, the standard doesn't impose any requirement on how the reading system should handle the complexity of *EPUB* packages. Instead, it expects that *EPUB*-compliant reading systems be able to perform a certain sequence of operations:

1. identify the type of package (according to its `mimetype`);

2. unzip the *EPUB* archive, if detected;

3. read the `META-INF/container.xml` file and identify the root file path inside;

4. go to the root file path and process each section from the `package.opf` file;

5. render the content documents based on the `manifest` section and in the order described in the `spine` section.

With the basics laid down, the following sections will explore the evolutions of the *EPUB* standard, some of which have already acquired specification status, while others are still in the design or mitigation stages within *IDPF*'s different working groups.

[1] http://dublincore.org/

4.2. Development of the new *EPUB* standards for comics

In this section we present two major developments of the standard, with respect to image-based publications:

- the set of standards developed under the coordination of the EPUB Advanced and Hybrid Layouts (AHL) working group. Our contributions to AHL saw us taking part in the collaborative development of most of its deliverables. In particular, we played a central role in the design of the EPUB Structural Semantic Vocabulary for comics;
- the creation, together with our industrial partner (Actialuna), of the EPUB Sequences specification for transitions in comics and other image-based publications.

4.2.1. Development context – Advanced and Hybrid Layouts

EPUB Advanced and Hybrid Layouts (AHL) was an *IDPF* working group, active between November 2012 and September 2015. Its activity was dedicated to the elaboration of a set of specifications extending the core *EPUB 3* functionality, "to better meet the full needs of the ecosystem with regards to user and context adaptation and reading of image-based publications such as comics and magazines" [EPUB-AHL, 2012].

The core term here is **image-based publications**. As the group's Charter [EPUB-AHL, 2012] states, the project stems in those requirements which weren't addressed by the original EPUB Fixed-Layout Documents specification [Kanai et al., 2012]. Most of these requirements were referring to the context of EPUB publications "where the main content is represented using image data", i.e. image-based publications.

The scope of the group's work stated the creation of three major deliverables:

- a mechanism for automatic and interactive **rendition selection**, generally meeting the use cases of multiple rendition publications (e.g., multi-language books, multi-edition books packaged together);
- a mechanism for **inter-rendition mapping**, to create correspondence points between different renditions of the same publication;
- a mechanism for **sub-page region mapping and navigation**, meeting sub-page level navigation requirements (e.g., panel-to-panel navigation in comics).

As a full member of the working group, I took part in its activities, which included two extended face-to-face meetings, in addition to the weekly conference calls. The work load consisted of individual or collaborative weekly assignments, ranging from research, requirements elicitation, to the elaboration of the final set of specifications. During this

time, we worked together with representatives of some of the most important stakeholders in the digital and publishing industry. Among them, we particularly mention our colleagues from Google, Barnes & Nobles, Fujifilm, Hachette Livre, Sony, Feedbooks, Aquafadas, Actialuna and the group's chair Murata Makoto (Japan Electronic Publishing Association).

The group concluded its activity with the publication of three main specifications: EPUB Multiple Rendition Publications, EPUB Region Based Navigation and EPUB Previews.

4.2.2. AHL new standard specification #1: EPUB Multiple Rendition Publications

Today's digital publishing products exhibit great variation in terms of rendition characteristics, in particular language and layout. This requires a digital publication to adapt to just as many parameters. Depending on the locale, a publication can be rendered in different languages or impose a different reading direction. The reading device (computer, tablet, smartphone, audio device) often imposes physical constraints, while development platforms and technologies (web, app, mobile) push their own specific storage requirements. Thus, for only one publication, one can have a multitude of publishable instances, with each one suitable in different contexts. For example, a Japanese *manga* issue will be rendered in Japanese, with a right-to-left reading order, in its mobile or web versions; alternatively, the same *manga* can also be rendered in English or French, with a left-to-right reading order, in a mobile or web version. Each of these specific embodiments of a publication and their exact context, is what EPUB is calling an **EPUB rendition**.

There is, of course, the possibility of packaging each rendition in its own EPUB package, but there are a few problems with this approach. First, it is not elegant; even though the intellectual content of different renditions would be very different (the case of multiple languages), most of the styling and formatting would still apply across all renditions. In this case, multiple packages would involve an unnecessary duplication of resources. Second, in the case of image-based publications, much of the content is very much the same even throughout different renditions. Back to the previous example, a Japanese *manga* will exhibit the same graphical content when it will be translated into English or French and it is the textual fraction of the content that is going to adapt to the language. In this case, most of the content will be common for the two renditions.

Besides multiple languages, this specification is also concerned with issues related to different possible layouts of the same publication. This use case originates from the needs of magazine publishers to have their specific content (also a blend of text and images, but in different proportions) adapt to different screen layouts. Thus, an *EPUB* package could hold two renditions of the same issue of a magazine, corresponding to the fixed-layout (print replica), for tablet size screens and the reflowable layout, suitable for smaller size mobile phone screens.

From a technical standpoint, the specification describes a way for the reading system to determine that an *EPUB* package contains multiple renditions, how to select the default rendition and how to choose between different available renditions.

Rendition Selection

The Rendition Selection involves the specification of a series of attributes, attached to the `rootfile` elements, inside the container file (*OCF*):

- `rendition:media` – identifies the media features (i.e. screen layout) for which the given rendition is most suitable (Figure 4.3);

```
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container"
           xmlns:rendition="http://www.idpf.org/2013/rendition"
           version="1.0">
   <rootfiles>
      <rootfile full-path="Sandman.opf"
           media-type="application/oebps-package+xml"/>
      <rootfile full-path="Sandman-large.opf"
           media-type="application/oebps-package+xml"
            rendition:media="(min-width: 1920px)"/>
   </rootfiles>
</container>
```

Figure 4.3.: The following example shows two Renditions of *The Sandman* bundled in the same container, one optimised for screens 1920 pixels or wider. The Default Rendition will be used for screen sizes smaller than 1920 pixels by default.
Credit: [Conboy et al., 2015b]

- `rendition:layout` – indicates whether the given rendition is reflowable or pre-paginated (Figure 4.4);

```
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container"
           xmlns:rendition="http://www.idpf.org/2013/rendition"
           version="1.0">
   <rootfiles>
      <rootfile full-path="EPUB/reflow/magazine.opf"
           media-type="application/oebps-package+xml"/>
      <rootfile full-path="EPUB/fxl/magazine.opf"
           media-type="application/oebps-package+xml"
           rendition:layout="pre-paginated"/>
    </rootfiles>
</container>
```

Figure 4.4.: The following example shows two Renditions of a magazine bundled in the same container. Note that it is not necessary to state that the Default Rendition is reflowable, since Renditions are reflowable by default.
Credit: [Conboy et al., 2015b]

- `rendition:language` – indicates that the given rendition is specified for a certain language (Figure 4.5). This attribute is more precise than the `dc:language` element

present at package level, the latter indicating the predominant language used in the prose of a publication.

```
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container"
           xmlns:rendition="http://www.idpf.org/2013/rendition"
           version="1.0">
   <rootfiles>
      <rootfile full-path="EPUB/en/package.opf"
         media-type="application/oebps-package+xml"
         rendition:language="en"/>
      <rootfile full-path="EPUB/fr/package.opf"
         media-type="application/oebps-package+xml"
         rendition:language="fr"/>
      <rootfile full-path="EPUB/es/package.opf"
         media-type="application/oebps-package+xml"
         rendition:language="es"/>
   </rootfiles>
</container>
```

Figure 4.5.: The following example shows a multilingual EPUB Publication, with English, French and Spanish Renditions of the content.
Credit: [Conboy et al., 2015b]

- `rendition:accessMode` – identifies the way in which the intellectual content is communicated with this rendition. The possible values are: `auditory`, `tactile`, `textual` or `visual` (Figure 4.6). This attribute describes the primary mode in which the content is being rendered, but doesn't exclude a mixture of modes. For example, the average comic book contains a larger proportion of images, compared to text; in this case, the value of the attribute will be `visual`. Same applies if the content is a scanned version of a print source (regardless of it being image-based or largely textual);

```
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container"
           xmlns:rendition="http://www.idpf.org/2013/rendition"
           version="1.0">
   <rootfiles>
      <rootfile full-path="EPUB/comic/package.opf"
         media-type="application/oebps-package+xml"
         rendition:accessMode="visual"/>
      <rootfile full-path="EPUB/novel/package.opf"
         media-type="application/oebps-package+xml"
         rendition:accessMode="textual"/>
   </rootfiles>
</container>
```

Figure 4.6.: The following example shows an EPUB Publication with an image-based Rendition and a text-based serialisation available.
Credit: [Conboy et al., 2015b]

- `rendition:label` – provides a way to annotate the current rendition with a human-readable name (Figure 4.7). This is especially useful for manually selecting between available renditions.

```
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container"
           xmlns:rendition="http://www.idpf.org/2013/rendition"
           version="1.0"
           xml:lang="en">
   <rootfiles>
      …
      <rootfile full-path="EPUB/package.opf"
           media-type="application/oebps-package+xml"
           rendition:media="color, min-width: 1024"
           rendition:layout="pre-paginated"
           rendition:label="Color-optimized print replica"/>
   </rootfiles>
</container>
```

Figure 4.7.: The following example shows the `rendition:label` attribute being used to provide a human-readable name for a Rendition.
Credit: [Conboy et al., 2015b]

Rendition selection **should** occur on the first rendering of the *EPUB* publication and automatically react to all changes in the user's environment (e.g., rotating the reading device, resizing the viewport etc.). Alternatively, the process should be triggered when rendition selection is done manually. The processing model is straightforward. For a given rendition, it evaluates each of the existing attributes (from the list above). If all of them evaluate to `true`, the current rendition is selected, the process stops and the corresponding content is rendered. If at least one of them evaluates to `false`, the process moves to the preceding `rootfile`, corresponding to the next rendition above. All unrecognised `rootfile` attributes are discarded. If none of the alternate renditions is selected, the default rendition (first one in the list) is chosen and the process will terminate.

Rendition Mapping

Rendition Mapping is complementary to Rendition Selection and establishes connection points throughout the different renditions of a multiple renditions publication. This enables users to stay at approximately similar points along the content in the case of switching between renditions.

Imagine one would like to switch between the French and English renditions of *The Adventures of Tintin*. While doing so, it would be undesirable for the reader to have to scroll from the beginning of the content all the way back to the point where they were when the language switch occurred. The same applies, in case the rendition is automatically switched following a change in the reader's environment (e.g., tilting or rotating the reading device). The nuisance this would cause if one would rotate their device halfway through Dostoevsky's *War and Peace*!

The mapping mechanism involves the introduction of an additional content document – the Rendition Mapping Document (Figure 4.8). This is essentially an *XHTML* file, containing one or more unordered lists under the document's `<nav>` element. Each of these

lists represents a common point across all renditions (e.g., a chapter, page, paragraph, image etc.); with each element in the list being a link to the location of the corresponding content document of a particular rendition.

```
<html xmlns="http://www.w3.org/1999/xhtml">
    <head>
        <meta charset="utf-8"/>
    </head>
    <body>
        <nav epub:type="resource-map">
            <ul>
                <!-- Chapter 1 -->
                <li><a href = "../../en/en.opf#epubcfi(/6/2)" /></li>
                <li><a href = "../../fr/fr.opf#epubcfi(/6/4)" /></li>
                <li><a href = "../../de/de.opf#epubcfi(/6/2)" /></li>
                <li><a href = "../../es/es.opf#epubcfi(/6/6)" /></li>
                <li><a href = "../../it/it.opf#epubcfi(/6/2)" /></li>
            </ul>
            <ul>
                <!-- Chapter 2 -->
                <li><a href = "../../en/en.opf#epubcfi(/6/4)" /></li>
                <li><a href = "../../fr/fr.opf#epubcfi(/6/6)" /></li>
                <li><a href = "../../de/de.opf#epubcfi(/6/4)" /></li>
                <li><a href = "../../es/es.opf#epubcfi(/6/8)" /></li>
                <li><a href = "../../it/it.opf#epubcfi(/6/6)" /></li>
            </ul>
            <ul>
                <!-- Chapter 3 -->
                <li><a href = "../../en/en.opf#epubcfi(/6/6)" /></li>
                <li><a href = "../../fr/fr.opf#epubcfi(/6/8)" /></li>
                <li><a href = "../../de/de.opf#epubcfi(/6/6)" /></li>
                <li><a href = "../../es/es.opf#epubcfi(/6/10)" /></li>
                <li><a href = "../../it/it.opf#epubcfi(/6/8)" /></li>
            </ul>
            <!-- …  could be even more fine grained for better results … -->
        </nav>
    </body>
</html>
```

Figure 4.8.: The following example shows a multilingual EPUB Publication with each language in a separate Rendition. [Conboy et al., 2015b, Section 4]

The Multiple Rendition specification has generated extended debate during AHL's working group meetings. Solutions have generally toggled between monolithic and modular approaches. We have seen the limits of the former, with duplication of content becoming a big issue especially in the case of "heavy" image-based publications. Despite the inconvenience of having to alter the previous structure of *EPUB* packages by introducing the Rendition Mapping Document, the modular solution has generally found more traction with the group's members. This also constitutes an important step in the direction of a visionary approach to what one could call *CloudPUB* – a lightweight EPUB package, which instead of holding different renditions and their respective content documents and resources, it simply references and serves them from the "cloud".

4.2.3. AHL new standard specification #2: EPUB Region Based Navigation

The period following the appearance of tablets and smartphones saw an increasing trend in the consumption of publications on small screen devices. In the beginning, most content came as fixed layout (*PDF* or *HTML*) or, in most of the image-based publications cases, simple scans of the print version. The user experience was far from being optimal, with simple user-environment changes, such as rotating or tilting the device, or resizing the viewport, often making the content difficult to consume. Addressing these issues quickly became of paramount importance for the successful development of digital magazines and the fast growing sector of digital comics.

Let's consider the typical case of a comic book page, as illustrated in Figure 4.9.

Figure 4.9.: The desired segmentation of a page into composing panels.
Source: *Clues Tome 1 – Sur les traces du passé*, Mara (2008), Copyright ©Akileos, Retrieved from [Conboy et al., 2015b, Section 4]

Displaying this kind of content on a small screen device would lead to one of two compromises: (1) give up on the quality of the image, while rendering it entirely on the screen, thus largely reducing its size and making its consumption by the reader difficult or (2) allow to zoom in on the different panels of the screen, at the cost of forcing the user to make a lot of on-screen gestures to freely navigate from panel to panel. Both of

these solutions contradict basic recommendations regarding user experience design and ergonomics principles applying to mobile devices.

To address this, the solution for Region Based Navigation proposes the segmentation of the content into sub-page level regions, called **EPUB Region of Interest (RoI)**, annotating them with minimal semantic metadata and leave it up to the reading system developers to decide how a device is supposed to render them in a sequence. For the page above, one possible segmentation is the one in Figure 4.10 below:

From a technical standpoint, this requires the introduction of another content document (essentially an *XHTML*) in the EPUB package. The Data Navigation Document thus defines a series of `<nav>` elements, each carrying the attribute-value pair `epub:type=''region-based''`. Each of these elements hosts precisely one ordered list element (`<ol>`), whose items (`li`) contain anchor references (`<a href>`) pointing to the RoI inside a fixed-layout content document.

The example below (Figure 4.11) illustrates the solution for our comic book page example. Notice that the list items of the ordered list impose the order in which each region should be rendered by the reading system. The model also allows for the description of sub-regions, by specifying them as `<ol>` children elements of the `<li>` element corresponding to the encompassing region.

The examples above generally use basic Media Fragments for describing the content and geometry of different RoI's. However, the current specification allows for three such ways to declare regions of interests:

- Media Fragments: useful for the definition of rectangular, circular or polygonal shapes;
- Synthetic regions: allowing creators to declare content spanning across pages (double-spread panels / pages), a very common use case in comics (Figure 4.12 below);
- SVG regions: simply specifying that a certain SVG component should be treated as a RoI.

Finally, the specification encourages the use of semantic metadata in conjunction with the definition of each RoI. To this end, EPUB defines a number of specific structural semantic vocabularies [EPUB-SemVoc, 2016], adapted to different types of publications: magazines, comics, education. We conclude this section with a brief presentation of the **Structural Vocabulary for Comics** – our main individual contribution within AHL – which defines a **minimal** group of terms allowing for the semantic enrichment of comics-related regions of interest:

panel : an individual frame or drawing (in the case of border-less frames);

panel-group : a group of panels. Useful in describing the artistic intent of grouping certain panels together. In this case, the reading device would have the option of rendering them together, rather than in a sequence;

Figure 4.10.: Rendering the page in Figure 4.9 on a cellphone screen, using a semi-transparent letterboxing effect to obscure the surrounding areas. Source: [Conboy et al., 2015b, Section 4].

```
<nav epub:type="region-based">
   <ol>
      <li><a href="page1.xhtml#xywh=percent:5,5,66,33">
        <meta property="ahl:character-name" content="Emily"/>
      </a></li>
      <li><a href="page1.xhtml#xywh=percent:68,5,33,33"></a></li>
      <li><a href="page1.xhtml#xywh=percent:5,40,50,33"></a></li>
      ...
   </ol>
</nav>
```

Figure 4.11.: The highest level of the `region-based nav` element inside the Data Navigation Document corresponding to the previous example. Notice the use of the `<meta property>` element to describe the characters inside a panel. Six such elements can be found in the EPUB Structural Semantics Vocabulary for comics.

balloon : an area in a comic panel that contains the words – spoken or thought – of a character. This could be very useful in triggering certain effects by the reading system (e.g., a short musical background for a thought, a sharp noise if a character is screaming);

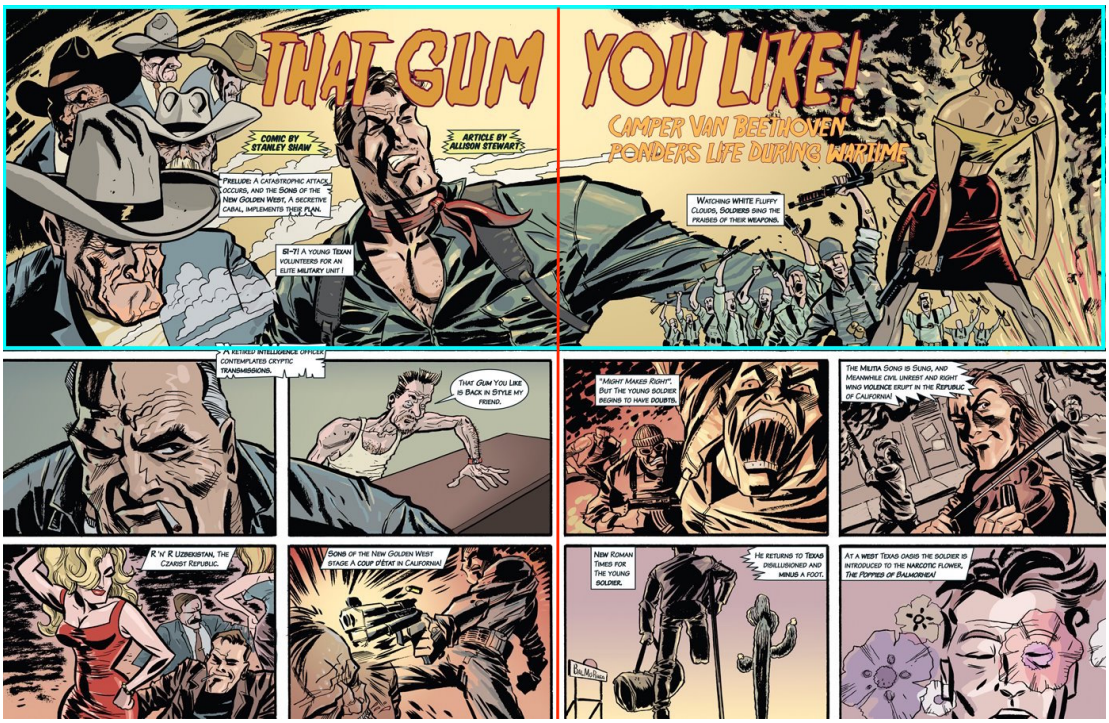
text-area : an area of text in a comic panel. Used to represent titles, narrative text, character dialogue (inside a balloon or not). The reading device could react in different ways to this kind of text (e.g., a background music could accompany narrative text);

sound-area : an area of text in a comic panel that represents a sound. This is a mean of isolating onomatopoeia in comics and enabling the reading device to have a particular behaviour – haptic, visual or sound effects, when rendering this type of RoI (e.g., the device could vibrate if a panel would contain the *BOOM!* onomatopoeia, it could get briefly lit-up if a gun-shot onomatopoeia is detected or it could generate a specific sound if the rain falling onomatopoeia is detected).

4.2.4. New standard specification for image-based publications: EPUB Sequences

In this section we introduce EPUB Sequences, an extension we developed together with our industrial partner (Actialuna), completely outside the scope of AHL. EPUB Sequences primarily addresses AHL's limitations concerning the rendering of image-based publications, and in particular the specification of transitions and navigation effects.

A retrospective of the history of digital comics in the past 30 years takes us from the first experiments with comics through introductory scenes to video games, over to CD ROMs, the first webcomics and all the way to the blogs, apps and turbo media. The common point of all these experiments is their effort of preserving the linearity of the reading act. To clarify: when we talk about conserving the linearity of the story we are not referring to the progression of the narration (which could, obviously, evoke a future or a past tense,



(a)

```
<li>
    <span>
        <a href="page22.xhtml#xywh=percent:0,0,100,50"></a>   // left region
        <a href="page23.xhtml#xywh=percent:0,0,100,50"></a>    // right region
    </span>
</li>
```

(b)

Figure 4.12.: A typical double spread (a) and the corresponding EPUB snippet (b). Notice how both regions (left and right) are starting from the top-left corner of their pages (0, 0) and extend 100% of the width and 50% of the height of each page. Credit: Stan Shaw (Nov. 17, 2014), for *Non Depression Magazine*, Retrieved: Draw Stanley

i.e. a flashback – before resuming its present tense); one shouldn't confuse the reading experience as it is offered to the reader (which should be linear) with the artistic richness of the literary work.

At this point, it is worth remembering one of those trivial facts about digital books that is often forgotten. The concept of "book" covers two aspects: the book as an object on one side and, on the other side, the work that is digitally divided between the reading device (dedicated software included) and the file which contains it. Since comics involve both a written and a graphical part, it is of paramount importance for creators to master the visual experience of the reader. And this is equally valid for other publishing areas that are artistically employing the form of the written work.

Reflowable EPUBs are displayed by reading systems, the latter imposing graphical transitions while progressing through the reading. However, these transitions are chosen by the system developer (e.g., iBook's 3D page turn). In comics, any kind of imposed transitions (those of the reading system included) become a real problem from the author's standpoint. The purpose of EPUB Sequences is thus to bridge the gap between the EPUB standard and the authors' need for tools enabling them to fully master the visual experience that they wish to offer to their readers. Therefore, the two main working directions of Sequences are:

- transitions between reading instances (whether these are pages as those inherited from the paper print or in a broader sense, as it is implied by digital technologies);
- triggering an event, using the same gesture as that dedicated to "turning the page" or progressing with the reading of the publication.

Terminology

To support the effort of creating a model for describing sequences in EPUB, we begin by defining the core concepts of such a model:

Navigation units generally identify either pages or regions of interest inside an EPUB publication. Navigation in an EPUB publication is of two types:

- Intra-unit navigation: navigation inside a navigation unit (i.e. the user explores a page or a RoI, using an action such as scroll, without going to the next navigation unit);
- Inter-unit navigation: navigation between navigation units (i.e. the user navigates between two consecutive navigation units; turning a page or moving from RoI to RoI, for example).

A user-performed action is the interaction between the user and the reading device resulting in progression toward an adjacent navigation unit. Actions can be gestures (swipe / tap screen, tilt / rotate / shake device), mouse clicks or pressing of physical buttons.

A navigation effect is occurring as a result of a user-performed action and can be a transformation or a transition.

Any combination between these three elements (navigation units, user performed actions and navigation effects) can occur and their interplay is entirely up to the creator of digital content (Figure 4.13). Additionally, in order to establish a clear minimal terminology for the specification of sequences in EPUB the following definitions are added.

A transformation is any kind of transform applied to a navigation unit or to an element within; it results in changing that element's original position or structural property (shape, colour, size). In the defined terminology, geometric transforms (translation, rotation, scale, skew) are considered transformations.

A transition is the navigation from one navigation unit to the next one. For example, the commonly used page-turn effects are considered transitions. Enhanced transformations include such effects as fade, cut, dissolve.

Animation. When we consider page-to-page navigation, the difference between animations and transitions can also be seen from the perspective of their application to pages: a transition is applied to one or both consecutive pages; an animation is simply inserted between two consecutive pages (to this extent, it can be seen as "a page" itself).

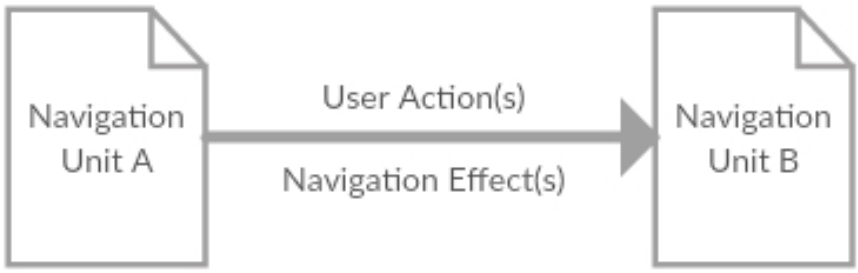


Figure 4.13.: Illustration of the basic principle behind EPUB Sequences: the navigation between two adjacent navigation units is triggered by one or more user-performed actions and results in at least one (i.e. the page turn) navigation effect.

Use Cases

The specification draft carries on with the description of a number of use cases from the digital publishing sector, that motivate the design of a model for specifying sequences in EPUB.

In the following, we shortly discuss a subset of those cases, that we consider most challenging or simply most interesting. We also strongly encourage our readers, to take their time and consult the on-line examples referenced by each of the use cases. This will give you a more visual idea, resulting in a better understanding of the discussed use cases.

Page to page navigation, augmented by a navigation effect. During the navigation between two pages, a transition effect is applied to one or both pages. The best illustrations of this are "infinite comics".

- *Wolverine: Japan's Most Wanted Infinite Comic*;
- *Infinity Against the Tide Infinite Comic*;
- *Ultimate Spider Man Infinite Comic*;
- "Newton" on Le blog á Malec[2](2013);
- "Mes anges", in *Professeur Cyclope*[3] (2014), by Loïc Sécheresse;
- Slide based presentations: "animations" or "transition effects".

Guided navigation (RoI to RoI transition), with a navigation effect. This use case describes the guided navigation for a document that has been fragmented in RoIs. However, when transitioning from one RoI to another an artistic effect (either a transformation or an animation) is occurring. Imagine for instance that in a publication, when we transition from RoI#1 to RoI#2 a certain fade effect can occur until the first region disappears while the second one appears.

- The digital comic *CIA:Operation Ajax* [Burwen and de Seve, 2012] uses many instances of RoI to RoI navigation, in which the current panel is fading in, while the previous one(s) are fading out and translating to occupy a final position on the page; in this way each page is composed panel by panel, but focus stays on the latest panel;
- in presentation documents, this is somehow similar to the effects available in software like *Prezi*.

Animations on RoIs, controlled by a gesture. Animations are applied to elements inside the page. As opposed to the previous cases, the animation is synchronised with the gesture performed by the user on the device. The animation can progress forward and/or backwards.
Comment: based on the assumption that reading is a linear act, we have to take into account that the animation is part of the artistic intent and is thus meant to be played (i.e. consumed by the reader). However, in order to avoid conditioning the reader to play all animations (and potentially blocking the reading process for whatever reason), the recommendation is to include metadata stating the role of the animation in the reading sequence.

[2] http://leblogamalec.blogspot.fr/2013/11/newton.html
[3] http://profxz3kdev.professeurcyclope.fr/revues/demo/#chapter/mes_anges

- Every Last Drop[4]. The scrolling is synchronised with a set of actions on elements of the scene (e.g. the bed moving in the scene, the shower curtain being dragged, the car moving on the street etc.) as well as narration elements (e.g. appearance / disappearance of text areas or text-bubbles);
- Zivot Jedne Kuchyne[5];
- the examples achieved with Super Scrollorama[6].

Wide horizontal or long vertical scroll pages, with animations triggered by a gesture. The navigation is done on a potentially infinite horizontal or vertical scrollable page. Animations are occurring at certain points during the navigation and are triggered and / or synchronised with the progression gestures.

- *Bongcheon-Dong Ghost* [English version][7]. This digital comic showcases the use of this type of small animations triggered at certain points, by the reader's scroll (e.g., the ghost turning its head, the ghost running towards the girl);
- editorialised pages[8].

Changing reading direction. This is considered to be an extreme use case (with respect to its frequency in the publications we have examined): there are certain publications in which the reading direction / navigation changes; however this happens without compromising the sequential nature of the reading act.

- Hugues Micol in *du9 - l'autre bande dessinée*[9]. In this interview from a French alternative *bande dessinée*, the navigation is initially achieved through vertical scroll. Then, at certain points, the navigation becomes horizontal, before switching back to vertical.

Additionally, the EPUB Sequences draft isolates and discuss a sensitive use case, that of the artistic parallax. The **parallax** consists in simulating graphic depth by moving different layers at different speeds, relative to a reference layer (the layer readers interact with). This technique is regularly used by creators of digital publications to achieve high quality graphic effects. We distinguish two important subcases posing different technical problems:

- intra-unit parallax: using the parallax within the reading unit, in which the movement is free. This is a commonly used technique, but poses some ergonomic problems (e.g., *Hobo Lobo of Hamelin*[10]);

[4] http://everylastdrop.co.uk/
[5] http://www.zivotjednekuchyne.cz/
[6] http://janpaepke.github.io/ScrollMagic/
[7] http://comic.naver.com/webtoon/detail.nhn?titleId=350217&no=31&weekday=tue
[8] http://www.polygon.com/a/ps4-review/conclusion
[9] http://www.du9.org/entretien/hugues-micol/
[10] http://hobolobo.net/

- transition with parallax: using the parallax for transitioning between two reading units. It is preferable from an ergonomic standpoint, as its quality is to fix the reader's view on a reading instance. However, it is more difficult to solve since it involves solving the problem of graphical elements persisting between two reading units and assigning a transformation to each of them (i.e. a transformation in the case of a transition).

4.2.5. AHL new standard specification #4: EPUB Previews

This third specification developed by the AHL working group, describes how preview content can be included in an *EPUB* publication. It also identifies two models for creating publication previews: Preview Publication and Embedded Preview.

Preview Publication is a model in which the preview is packaged as a standalone *EPUB* publication. The [EPUB-Previews, 2015] specification mentions several situations in which this would be desirable: "to begin marketing a book before the full manuscript is complete, to limit the content available when not using DRM [digital rights management] schemes, to minimise download sizes for media-rich works, and so on". When distributing such a preview, the standard requires that a user is made aware that the type of publication he or she is consuming is a preview (Figure 4.14).

```
<package ... unique-identifier="pub-id">
   ...
   <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:title>Norwegian Wood</dc:title>
      <dc:language>en</dc:language>
      <dc:type>preview</dc:type>

      <dc:identifier id="pub-id">urn:uuid:A1B0D67E-2E81-4DF5-9E67-A64CBE366809</dc:identifier>
      <meta property="dcterms:modified">2011-01-01T12:00:00Z</meta>
      <dc:source>urn:isbn:9781448103706</dc:source>

      <link href="http://example.org/book/9781448103706"
            rel="acquire"
            media-type="text/html" />
      <link href="http://example.org/book/9781448103706.atom"
            rel="acquire"
            media-type="application/atom+xml;type=entry;profile=opds-catalog" />
   </metadata>
   ...
</package>
```

Figure 4.14.: Preview Publication – The following example shows a minimally complete metadata section, where the Publication is identified as a preview, metadata links back to the source publication and a link to acquire it has also been included. [EPUB-Previews, 2015]

Embedded Preview is a model in which the *EPUB* publication identifies a subset of its content documents and resources which will be presented in a preview context, as well. This model provides content creators with control over which parts of the intellectual content is available as a preview. In addition, when the full publication is accessed, the user will not have to download a new *EPUB* package, corresponding to the full length publication, but simply unlock the rest of the content from the existing package. To

specify an Embedded Preview, an *EPUB* publication will need to include a package document (*OPF*) that contains a `<collection role=''preview''>` element, stating its role as a preview. In turn, this will contain a manifest child collection (`<collection role=''manifest''>`). Nested within this collection, a series of links will point to the content documents and resources (stylesheets, images etc.) composing the embedded preview. The order of these links is the order in which the content documents will be previewed (Figure 4.15).

```
<collection role="preview">
    <collection role="manifest">
        <link href="chapter1.xhtml"
              media-type="application/xhtml+xml" />
        <link href="chapter2.xhtml"
              media-type="application/xhtml+xml" />
        <link href="chapter3.xhtml"
              media-type="application/xhtml+xml" />
        <link href="stylesheet.css"
              media-type="text/css" />
        <link href="image1.jpg"
              media-type="image/jpeg" />
    </collection>
    <link href="chapter1.xhtml"/>
    <link href="chapter2.xhtml"/>
    <link href="chapter3.xhtml#preview"/>
</collection>
```

Figure 4.15.: Embedded Preview – The following example shows how a publication can provide a collection that indicates which EPUB Content Documents are included in a preview (here, only part of chapter 3), as well as a manifest of the resources necessary for this preview. [EPUB-Previews, 2015]

In both these cases, the specification requires that the user has a way to access or unlock the original publication from which the preview has been derived.

4.3. New standards for publication annotations

At the conclusion of the extensive work done by the AHL group, another working group started its activity within the IDPF. Functioning between November 2013 and May 2014, Open Annotations in EPUB (OA) set off "to define how to apply Open Annotations in the context of EPUB, thereby enabling users and service providers to utilize a non-proprietary representation of annotation data" [11].

Much more compact than the AHL group, OA hosted a number of industry specialists and scholars working in the field of digital publications and open annotations. The working

[11] OA in EPUB GitHub: https://github.com/IDPF/epub-revision/wiki/Open-Annotation-in-EPUB

methodology, based on weekly conference calls, resembled largely to that of AHL, with the observation that no face-to-face meetings were organised for OA. As opposed to AHL, my role within the OA group was rather observational and was motivated by future technical and research needs of our industrial partner, Actialuna.

The final specification, [EPUB-OA, 2014], which became an EPUB recommendation early in 2016, defines "a profile of the W3C Open Annotation specification for the creation, distribution and rendering of annotations for EPUB Publications".

Annotation Models

The format for serialising open annotations in EPUB is *JSON*; the specification refers to this as the Annotation Document. Much like in the case of Previews, there are two models for packaging and distributing open annotations in EPUB.

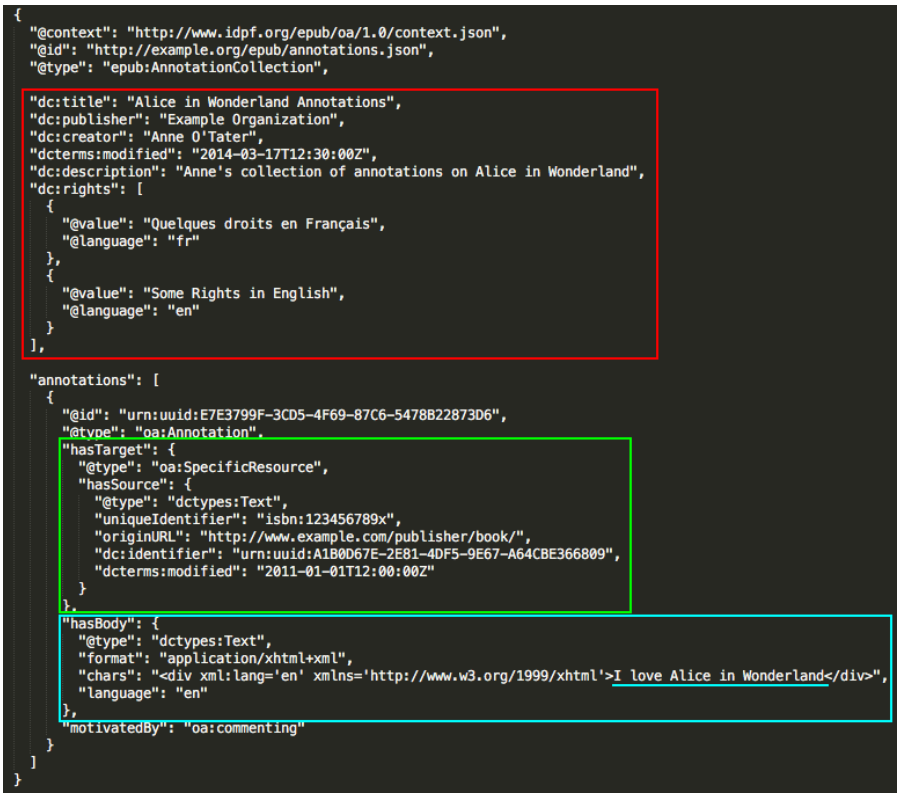
- **Standalone Collections** refer to the situation in which annotations are packaged separately from the EPUB publication. Several use case scenarios are met by this model: when publishers desire to produce and distribute the annotations separately, as an addition to the publication itself or when the reading system is expected to update the current set of annotations with a fresh one (e.g., an update from the publisher, users exchanging notes etc.). If the Annotation Document doesn't reference any local resource, then it can be distributed as it is. In any other case, it must add the referenced resources and package everything as a *ZIP* file.
- **Embedded Collections** refer to the case in which the annotations are delivered together with the resources they are referencing. Such a collection may thus be included in the rendition it is targeting or, if it targets multiple renditions, it will be included in the default rendition. In this case, the annotations are NOT distributed as a *ZIP* package. Instead, they are listed in the package document manifest (*OPF* file) and included in the EPUB container just like any other content document.

Collections

An Annotation Document must contain exactly one object, representing the collection of annotations. The collection is described using a series of attributes (`@context, @id, @type`) and the most important, `annotations`. The latter is nothing more than a *JSON* array containing the annotations. Descriptive metadata for the collection can also be added, with *Dublin Core (DCMI)* being the recommended namespace for metadata terms (`dc:creator`, `dc:created`, `dc:description`, `dc:language`, `dcterms:modified`, `dc:publisher`, `dc:rights`, `dc:title`).

Annotations

Each annotation in a collection is defined by encapsulating three main parts: the metadata, its body and the target object (Figure 4.16)[12].



```
{
  "@context": "http://www.idpf.org/epub/oa/1.0/context.json",
  "@id": "http://example.org/epub/annotations.json",
  "@type": "epub:AnnotationCollection",

  "dc:title": "Alice in Wonderland Annotations",
  "dc:publisher": "Example Organization",
  "dc:creator": "Anne O'Tater",
  "dcterms:modified": "2014-03-17T12:30:00Z",
  "dc:description": "Anne's collection of annotations on Alice in Wonderland",
  "dc:rights": [
    {
      "@value": "Quelques droits en Français",
      "@language": "fr"
    },
    {
      "@value": "Some Rights in English",
      "@language": "en"
    }
  ],

  "annotations": [
    {
      "@id": "urn:uuid:E7E3799F-3CD5-4F69-87C6-5478B22873D6",
      "@type": "oa:Annotation",
      "hasTarget": {
        "@type": "oa:SpecificResource",
        "hasSource": {
          "@type": "dctypes:Text",
          "uniqueIdentifier": "isbn:123456789x",
          "originURL": "http://www.example.com/publisher/book/",
          "dc:identifier": "urn:uuid:A1B0D67E-2E81-4DF5-9E67-A64CBE366809",
          "dcterms:modified": "2011-01-01T12:00:00Z"
        }
      },
      "hasBody": {
        "@type": "dctypes:Text",
        "format": "application/xhtml+xml",
        "chars": "<div xml:lang='en' xmlns='http://www.w3.org/1999/xhtml'>I love Alice in Wonderland</div>",
        "language": "en"
      },
      "motivatedBy": "oa:commenting"
    }
  ]
}
```


Figure 4.16.: A comprehensive example of using Open Annotations in EPUB. Notice the three distinct sections of an annotation: the metadata (red – publisher, creator, description etc.), the annotation target (green – identifies the type and the annotated publication) and the annotation body (blue – holds the actual content of the annotation). [EPUB-OA, 2014]

Several metadata are used to describe an annotation: the compulsory set of attributes (`@id, @type, @target`) uniquely identify each annotation within the collection; additionally, other metadata can be included to describe the provenance and motivation of any given annotation. The latter can be created according to W3C Open Annotations schemas.

The body of an annotation refers to its actual content and is stored within the `hasBody` child property. The specification requires this to be the fragment of an *XHTML* document, with any desired styling being possibly specified using exclusively *HTML5*'s `style` attribute.

The target object of an annotation, identified by its `hasTarget` property, references an EPUB publication or a fragment within a publication (using EPUB Common Fragment

[12] An extended compilation of a variety of use cases is available in the Open Annotations in EPUB specification draft [EPUB-OA, 2014].

Identifiers). When referencing a publication, the `specificityLevel` attribute can be useful in making clear the exact scope of the publication (`release`, `publication` or `work`).

Finally, it is worth mentioning the additional case of an annotation pointing to another annotation, forming content similar to a discussion thread (e.g., the discussion between a teacher and a student, the discussions of several people collaboratively reading a book etc.).

4.4. A look into EPUB 3.1

The EPUB 3.1 will be the first major revision of the standard since EPUB 3. This revision of the core EPUB specifications is mainly focused on consolidating, simplifying and adding a select set of new features. A dedicated working group started its activity in September 2015 and released its latest editor's draft [EPUB-31, 2016] in April 2016, having in sight the end of the same year for a public member's review. Due to the considerable amount of issues the group set up to solve, the work was divided between a number of subgroups.

Metadata deals with requests by authors for the inclusion of new metadata in the *OPF* core specification (e.g., series and accessibility metadata). In order to avoid bloating the *OPF* document with expressions that might become obsolete in future evolutions of the standard, they will not be included in the 3.1 specification and instead, their usage patterns will be tracked until they will stabilise. A second important change, concerns the replacement of the `@refines` attribute with *Dublin Core* custom attributes (for expressing roles such as creator, publisher, contributor etc.).

Browser-Friendly Format . Despite being a web-based format, *EPUB* doesn't fare very well when rendered in browsers. This group's aim is to update the 3.1 specification to the point at which it allows web developers to either display an EPUB package in an unzipped ("exploded") form or to allow for a *JSON* serialisation of the `container.xml` file and the *OPF* package document.

CSS Profile is one of the most sensitive parts of the current EPUB specification. Many users of the format have complained with respect to the direction taken by the EPUB Style Sheets [EPUB-ContentDocs, 2014, Section 3], which seem to be acting as a mandatory subset of *CSS*, rather than the intended *"minimum required subset of CSS"*. The latter, not only it would be in alignment with the philosophy of the Open Web Platform (OWP), but would mean that the minimum mandatory *CSS* could be easily enriched with custom expressions as per the author's necessities.

Scripting Components. This has always been one of the most technically sensitive aspects of EPUB standards, due to the need of balancing the democratic nature of the standard (minimum demands imposed on reading systems developers), the desire of providing safe content (scripting has always been linked to security issues) and the quality rendering that can be achieved with their use. The current iteration of the

standard aims at providing "clearer rules for scripting support in Reading Systems and content", such as "the ability for content to query reading systems on which scripting features it supports/allows, and models for the passing of events between content and reading systems".

Accessibility has traditionally been one of the format's strong points. Since early versions, *EPUB* publications were able to support different accessibility (**a11y**) use cases, mainly through the use of Media Overlays. This iteration of the standard is set to meet the ecosystem's needs for ways to evaluate, discover and certify accessibility features within *EPUB* publications. In addition, the current update defines metadata for classifying the accessibility degree of the content (i.e., discovery-only, accessible, optimised).

Consolidation. This subgroup is mainly in charge with studying the integration opportunities of the newly developed specifications into *EPUB*'s core specifications. For example, it is currently dealing with the integration of the three AHL specifications (Multiple Renditions, Region-Based Navigation and Previews) in EPUB 3.1.

Serialisation. With EPUB currently allowing exclusively *XHTML* serialisations of the *HTML*, this subgroup is set to investigating a problem that is redundant at every milestone of the standard: what is the impact of allowing *HTML* serialisations, with a minimum impact on the rest of the ecosystem (i.e. authoring tools, reading systems development, distribution and retail chain systems etc.)?

4.4.1. Limits of the standard and future directions

Finally, we acknowledge two particular subgroups, **restructuring** and **external resources**, for their pivotal role in the near future development of EPUB as an open standard.

Restructuring discussions revolve around the intrinsic complexity of the current three-plus-two specifications at the core of the standard. There are concerns about the increasing difficulty in navigating the standards, which can generate a large overhead when creating EPUB-compliant content, authoring tools or reading systems. Moreover, the complexity of the standard (recall the EPUB dissection we did in Section 4.1) is only bound to increase with the addition of new specification modules. Our personal belief (shared by other working group participants) is that all these problems originate from an insufficient understanding of the "book" object in general, thus leading to a poor layering of the technical solution. Unfortunately, it feels like EPUB's earlier iterations, those which in fact defined its core, failed to involve all the major stakeholders of the publishing industry, with librarians and every-day users being mostly disregarded. And while the organisms working on subsequent iterations have become much more aware of user needs, use cases originating from libraries, museums and collections still seem to be disregarded. In Section 3.2.3 we have explored *FRBR*, a set of requirements centred on the needs of the user-librarian binomial, providing a

potentially useful paradigm for looking at the **"book"** at different abstraction levels (work, expression, manifestation, item). We strongly believe that such an approach could have yielded a different solution to the problem of Multiple Renditions and opened up the path towards lighter EPUB versions sooner. Alternatively, semantic formats like *CBML* (see Section 3.2.2) should be considered for integration with future versions of *EPUB* and the current Structural Vocabulary for Comics is already largely inspired from *CBML*. Instead of creating a new (reduced and soon insufficient) schema, enhanced versions of EPUB could make use of those of *CBML* and *TEI* (much richer, covering many more practical cases).

Which leads us to the work done on the topic of external resources. External resources are those resources that are not packaged with the bulk *EPUB* publication. At the moment the standard is only allowing external resources which are media files, but with *EPUB* publications evolving according to authors' and users' needs, more flexible and interactive publications will become more frequent. Moreover, with offline consumption still being a frequent use case, all this will generate an increasing pressure for developing more lightweight digital publications. This, in fact, is one of the most strenuous demands on image-based publications, which are only bound to get "heavier" in their quest for enhanced content rendering.

Therefore, discussions about different serialisations of the format (*JSON*, for example) are welcome sooner, rather than later. Alternatively, the idea of creating what we have called *CloudPUB* – by completely moving resources and content documents in the cloud – should be given more thought as it has the potential to address several of the above concerns. In such a scenario, the role of the package document (potentially serialised as a *JSON* file) would be that of a collection of external references and composition rules.

However, all these proposals are not depending exclusively on the technical skills of the different IDPF members and are frequently influenced by the organisational strategies of IDPF, particularly their positioning with respect to the World Wide Web Consortium (W3C). This is to say, that any future development of EPUB is subject to the "political" decisions inherent in standards organisations.

4.5. Enhancements to *CBML* – the semantic format for comics

In this section we propose solutions to the limitations of the *CBML* format, which we briefly presented at the end of Section 3.2.2. The reader will recall that most of these limitations were related to the format's capability of encoding more particular graphic elements of the comics medium, such as: objects, visual cues, types of panel borders, splash pages, etc.

Objects

In Section 2.2.1 we discuss the elements of the comics medium and the way in which they function together to give the reader the sense of time and movement. We have also seen the importance of "framing the action" within the constraints of a panel, meaning that the author will choose to render only a limited space and a particular moment of a scene. With the page imposing its own physical constraints on this process, it is clear that an artist can afford to illustrate only the most relevant graphic aspects of a scene. Thus, in comics more than in other literary media, objects are particularly important and their proper encoding could be of great help to all those working with the medium.

Thus, one of *CBML*'s most strenuous constraints is the lack of any element specification for the description of objects in the scene. More so, *TEI* also fails to provide with such an element and lack of consensus concerning its design, has put any solution implementation on hold since 2012. Until a generic object-specific element is developed by the TEI community, we give our own approach to this problem (Figure 4.17).

```
<cbml:panel>
    <!-- Method 1: introducing new cbml elements-->
    <cbml:objectList>
        <cbml:object who="#capamerica">shield</cbml:object>
        ...
    <cbml:objectList>

    <!-- Method 2: using a typed tei note element-->
    <note type="object">Captain America's shield</note>
</cbmlPanel>
```

Figure 4.17.: Two *CBML* extensions for annotating objects on a comic book page. The first solution we propose is based on the introduction of a new `<cbml:object>` element, whereas the second one uses a typed version of the existing *TEI* `<note>` element.

Solution 1. The less intrusive solution that we propose, is to specify the presence of an object inside a panel, by using the *TEI*-specific element `<note type=''object''>` (notice the custom value that the `@type` attribute receives) as a child of a `<cbml:panel>` element. Alternatively, `<note>` could be used within any other *CBML*-specific element as well as in a number of *TEI*-stock elements.

Solution 2. The creation of a custom `<cbml:object>` element. This solution should be envisaged only in case TEI renounces to work towards the creation of its own object-encoding element. We have sketched a potential schema specification, which sees such an element included at different levels:

- connected to other elements of the scene, within *TEI*-specific elements such as `<person>` (i.e. possession), `<event>` (i.e. accessory), `<place>` (i.e. presence) or within another `<cbml:object>` (i.e. composition, part-of);

- connected directly to the scene itself, by specifying any number of `cbml:object` elements within a custom `<cbml:objectList>` element. The latter functions in the same way as the *TEI*-specific `eventList` or `placeList` and in our context should occur as a child of one of the *CBML*-specific elements. Such an element should be able to encode (if desired) physical attributes of an object (shape, size, colour, texture etc.) as well as its relationship to other elements in the scene (placement, possession, use etc.). Some of these are present in the form of *TEI* attributes (e.g. `@who`, `@where`), while others need to be defined.

Visual cues

In Section 2.2.1 we discuss the importance of visual cues as a graphic means for the artist to involve the reader's other senses: equilibrium, motion (see Figure 2.36), smell, hearing etc. Much like in the case of object encoding, there are two technical solutions (Figure 4.18):

Solution 1. The less intrusive solution that we propose, just like before, is to use the `<note type=''visualCue''>` mechanism, allowing the inclusion of this element as a child of any other *CBML* or comics-related *TEI* element.

Solution 2. The creation of a custom `<cbml:visualCue>` element, which could be attached either to different elements of the scene (`<person>`, `<event>`, `<place>`) or to the entirety of the scene (`<cbml:panel>`). Additionally, this solution would give the possibility of encoding more information about visual cues, namely: (1) what kind of action does it support (i.e. the smell coming from a boiling pot, motion lines indicating a speeding car, via the *TEI*-specific `@rend` attribute); (2) the sense it involves (i.e. smell, visual, audio, motion, impact etc.); (3) in the case of motion, what is the perspective (see Figure 2.36) from which it is perceived (i.e. fixed camera, mobile camera, "driver seat").

Panel borders, loose pages, double-spread pages

While *CBML* doesn't provide with a direct way of describing the different types of panel borders, such as they were discussed in Section 2.2.1 and illustrated in Figure 2.20, the `<interpretation>` element allows us to define a border typology and use it within the `@ana` attribute of the `<cbml:panel>` element (Figure 4.19). An alternative solution could see the definition of an additional attribute (e.g., `@border-type`) specific to the `<cbml:panel>` element.

In what concerns loose page structures (splash pages and meta panels), *CBML*'s current capabilities allow for the definition of panels containing other panels, which would allow for the definition of meta panels. At the same time, the use of the `@ana`/`<interpretation>` mechanism we have seen before could allow for the definition of logical identifiers for splash pages and meta panels.

```
<!-- inside the <teiHeader> -->
<interpGrp resp="#mtufis" type="senseType">
    <desc>The <gi>interp</gi> elements below describe the different
        senses that can be engaged through the use of visual cues.
    </desc>
    <interp xml:id="smell"/>
    <interp xml:id="motion"/>
    <interp xml:id="audio"/>
    <interp xml:id="visual"/>
    <interp xml:id="impact"/>
</interpGrp>

<!-- inside the <teiHeader> -->
<interpGrp resp="#mtufis" type="motionPerspective">
    <desc>The <gi>interp</gi> elements below describe different
        types of perspective suggested by the artist together
        with the use of motion lines. This follows the definitions
        given as they were defined by Scott McCloud
        in his <bibl><title>Understanding Comics</title>
        <biblScope type="pp">(113-114)</biblScope></bibl>.</desc>
    <interp xml:id="fixed-camera"/>
    <interp xml:id="moving-camera"/>
    <interp xml:id="driver-seat"/>
</interpGrp>

<cbml:panel>
    <!-- Method 1: introducing a new cbml element-->
    <cbml:visualCue sense="#motion" perspective="#mobile-cam">
        car passing at high speed
    </cbml:visualCue>

    <!-- Method 2: using a typed tei note element-->
    <note type="visualCue" ana="#motion #mobile-cam">
        car passing at high speed
    </note>
</cbml:panel>
```

Figure 4.18.: Two methods of describing visual cues in *CBML*. The first method we propose is based on the introduction of a new `<cbml:visualCue>` element, whereas the second one makes use of a double-typed *TEI*-stock `<note>` element. Notice how both of them rely on two typologies (senses described by the visual cues and types of motion perspectives) defined with the help of `<interpGrp>` elements and included in the `<teiHeader>`.

```
<!-- inside the <teiHeader> -->
<interpGrp resp="#mtufis" type="panelBorderType">
    <desc>The <gi>interp</gi> elements below describe a typology
        of panel borders as they were defined by Will Eisner in his
        <bibl><title>Comics and Sequential Art</title>
        <biblScope type="pp">(48-49)</biblScope></bibl>.</desc>
    <interp xml:id="regular-border"/>
    <interp xml:id="irregular-border"/>
    <interp xml:id="no-frame"/>
    <interp xml:id="portal-border"/>
    <interp xml:id="wavy-border"/>
</interpGrp>


<cbml:panel n=3 @ana="action-to-action #regular-border">
    ...
</cbml:panel>
```

Figure 4.19.: *CBML* encoding of the border type using on the `@ana` attribute of the current `<cbml:panel>` element. The values of this attribute rely on the typology defined in the `<interpGrp>` elements specified in the `<teiHeader>`. Notice how the `@ana` attribute can hold multiple such interpretation values separated by spaces.

Finally, in order to deal with *CBML*'s lack of specification for double spread pages (see Figure 4.12b), a very straightforward solution (Figure 4.20) is to rely once more on the `@ana`/`<interpretation>` mechanism which allows us to define a panel's inclusion in a double- or multi-spread. Previously, a `<div type=''panelGrp''>` should be used in order to mark the logical grouping of the two (or more) panels into one scene.

```
<!-- inside the <teiHeader> -->
<interpGrp resp="#mtufis" type="motionPerspective">
    <desc>The <gi>interp</gi> elements below help identify scenes
        spreading over two or multiple panels. A normal panel will
        not be identified by any of these values.
    </desc>
    <interp xml:id="double-spread"/>
    <interp xml:id="multi-spread"/>
</interpGrp>


<div type="panelGrp" ana="#double-spread">
    <cbml:panel n=1 xml:id="panel_001_left">
        ...
    </cbml:panel>

    <cbml:panel n=2 xml:id="panel_001_right">
        ...
    </cbml:panel>
</div>
```

Figure 4.20.: *CBML* encoding for scenes spreading two or more pages. The method makes use of the `@ana`/`interpretation` mechanism applied to a `<div type=''panelGrp''>` element. The values of the `@ana` attribute rely on the typology defined in the `<interpGrp>` elements specified in the `<teiHeader>`.

In this chapter – which concludes the second part of our work – we covered our contributions to the development and improvement of different standards and formats for comics specification. We presented our contributions in IDPF's different working groups (AHL, OA), dedicated to the most recent evolutions of the EPUB format. In particular, we stress our role in the specification of the Structural Semantic Vocabulary for comics, in connection to the Region Based Navigation specification, and in the creation of the EPUB Sequences specification, which formalises sequences and transitions in image based publications. Finally, we introduced important additions to CBML – the semantic format which addresses the multimodal nature of comics.

Part III.

Crowdsourcing comics annotations

Chapter 5.

Crowdsourcing – literature review

In this chapter we review the body of work that has been dedicated to the study of crowdsourcing, since the first use of the term in 2006, marking the first academic research done on the topic starting with 2008 and some of the most recent developments in what is now a proper problem solving approach, data collection methodology and means to tap into collective knowledge resources. Across this literature review we will gradually try to shed some light over the differences in meaning between some of the buzzwords that have certainly flooded academia and media in recent years – outsourcing, open source, citizen science, crowdfunding, crowd economy, social engagement – and establish their relationship, similarities and differences with respect to crowdsourcing. We will begin by presenting research that deal with crowdsourcing on a more general basis, before we zoom in on a body of work that is concerned with different aspects and components of crowdsourcing, such as motivations, cost and incentive mechanisms, human factors, data collection and quality assurance, as well as several applications. Our inclusion of these reports in any of the previous categories doesn't entail that the said work deals exclusively with that specific component of crowdsourcing, but rather that the main topic of the work we report on or its relation to our own work is rather viewed from that particular perspective.

5.1. A review of crowdsourcing initiatives

In this section we give a very synthetic review of a number of crowdsourcing platforms and projects we have come across over the course of the past three years. This list is obviously far from being exhaustive, given there are hundreds of both active and inactive crowdsourcing projects on the web and their number is only set to increase as the practice is becoming more and more popular. The projects featured on this list are there because: they are of great relevance to our own crowdsourcing effort, or they are rare examples of certain crowdsourcing typologies, or they are successful examples of crowdsourcing or... we simply loved them!

Threadless is a t-shirt design company founded in 2000, one of the most successful American small businesses ($ 30 million revenue in 2008, at a 30% profit margin). *Threadless* uses sources the t-shirt design process to an online community, which votes the best designs over the course of one week. The winning designs, get to be printed and sold on the platform at prices between $ 15-20. As of 2009, winning designs are rewarded with $ 2,000 each and a $ 500 gift card for purchases on the platform. *Threadless* also stands out due to its successful use of non-monetary mechanisms, as a result of building a vibrant, tight community around the brand, sometimes borderlining addiction, as reported in a qualitative study by [Brabham, 2010].

iStockPhoto is a web-based company founded in 2000, that sells royalty free photography largely taken by an amateur crowd. Clients purchase credit points which they can then use to acquire stock photos, with price varying according to size and resolution between $ 1-5, although over-sized pictures and even videos can be acquired for prices up to $ 50. *iStockPhoto* offers its contributors 20% of each purchase, whereas those who chose to get further involved with the platform, get 40%. *iStockPhoto* is an excellent case of how crowdsourcing contributors skill acquisition becomes one of the main motivators for their participation. In general, participants on this platform are both amateur and professional, but by developing their expertise we are witnessing the transformation of the former into "professional amateurs".

NextStopDesign is a project put in place by the *Utah Transit Authority*, calling for participants to submit their design for revamping a bus shelter in the city of Salt Lake City, Utah. Initially targeting local, state-wide and to a lesser extent national publics, the project got international exposure after being mentioned on several professional design blogs. This attracted a worldwide audience and with it professional design submissions. *NextStopDesign* is a fine example of dealing with unexpected and sometimes unwanted crowds.

Crowdsourced advertisement . Marketing strategists couldn't have missed on the golden opportunity of tapping into the creative potential of target crowds. Since 2010 a considerable number of advertisement campaigns[1] made use of crowdsourcing. The expectation in such campaigns is that nobody knows what "the crowd" likes, better than "the crowd" itself. Companies like JetBlue, Sony, Chrysler or Toyota are only a few of those which have successfully used crowdsourcing in designing their advertisement campaigns. An interesting phenomenon in creative crowdsourcing campaigns is what has been called "crowdslapping", essentially another instance of crowds behaving in an unexpected way. This is the case of Chevrolet's 2007 crowdsourced advertisement campaign for Chevy Tahoe, during which participants took it upon themselves to reinterpret the requester's call and instead turned in a

[1] An interactive time line of crowdsourcing advertisement campaigns by global brands: http://www.tiki-toki.com/timeline/entry/52997/Crowdsourcing-by-Worlds-Best-Global-Brands/

sizable amount of parody contributions on themes such as SUVs, the environment or American car industry.

InnoCentive is a crowdsourcing platform which "enables companies to tap into the talents of a global scientific community for innovative solutions to tough R&D problems". The rewards for winning submissions (usually technical specifications and written reports) range between $ 10,000-100,000 and in rare cases go as high as $1 million. While open to any member of the public, upon registration on the platform participants need to give specify their obtained degrees, if any.

Goldcorp Challenge was a crowdsourcing competition organised in 2000, by Canadian gold mining company Goldcorp. The company released geologic data and asked the public to identify where to direct their next national mining operations. Incoming submissions helped the company confirm supposed deposits and identify new ones and the top 25 contributors received a total reward of $ 500,000. In 2001, Goldcorp launched the worldwide challenge, promising a reward of $ 2 million.

White House SAVE Award . This competition is a prime example of applying crowdsourcing in government. Specifically, the competition consisted in a call for Federal employees to submit ideas that would lead to cost cuts in public administration. By directing the call to a closed community, the requester relies on the crowd's familiarity with the task at hand, in what it is essentially a form of community-sourcing. In a second stage, the evaluation of these ideas is also crowdsourced, this time to the general public, citizens and government employees alike.

Peer To Patent Community Patent Review is a government project by the US Patent and Trademark Office, which seeks to address the increasing number of patent requests by enlisting members of the general public in the process of reviewing existing patterns and thus enabling the Patent Office to be more efficient in reviews of current applications.

SeeClickFix is a platform for citizen engagement which crowdsources small infrastructure problems at local level (e.g., a pot hole in the road, a fallen tree, nonfunctional street illumination etc.) which are more easy for citizens to identify and can be quickly addressed by local administrations.

mcollect is a citizen engagement platform, which contributors can use to crowdsource prices from local agricultural markets in Burkina Faso, Mali or Senegal. Similarly, the *Community Knowledge Worker* project enables citizens from rural areas in Uganda to crowdsource agricultural information from farmers, establishing links between farmers, the government and other agricultural entities.

Ushahidi is a geolocalisation crowdsourcing platform that has been successfully deployed in numerous disaster relief efforts worldwide. The project's most famous instances saw it used during the violence following the 2007 Kenyan elections and in the aftermath

of the 2010 devastating earthquake in Haiti and Chile, or as a crisis mapping tool by the United Nations during the military operations in Lybia in 2011.

BeMyEyes is a mobile platform application which visually-impaired people can use to crowdsource help with everyday situations, from reading the expiry date on food products to navigating unknown surroundings. The requester uploads a video, which is then watched and described by a willing respondent. Focused on providing situational assistance, *BeMyEyes* falls in the same category of project as *Ushahidi*.

British Library Georeferencer . The purpose of the project was to "geo-enable" historic maps, by annotating them with geographic metadata. Managed by the British Library, it hosted 752 digitised historical maps which were presented to a crowd of passionate annotators. Surprisingly, the project had a remarkably fast execution time, with all maps being fully annotated in only four days. Useful, of course, but the project managers realised that most of the infrastructure set up aside the project was essentially useless. Despite reaching its annotation goal, *BLG* didn't generate any community around it.

Letters of 1916 . A remarkable project that looks into a massive collection of letters held by different Irish institutions, focusing on daily and major events in Irish society from a century ago (the Easter Rising, World War I). The project is fascinating in the way in which it mobilised people to explore and discover parts of Irish and sometimes family history. This intimate link seems to be a powerful, yet rare motivator for crowd engagement and the only other project where we have observed it is *Measuring the ANZACs* (annotating war diaries about Australian and New Zealand troops taking part in World War I).

Transcribe Bentham is a fascinating project dedicated to annotating 40,000 folios of unpublished work by English philosopher Jeremy Bentham, which were later subsequently published in a collected edition. It is a classic example of using crowdsourcing for transcription tasks in humanities.

Old Weather is a *Zooniverse* project dedicated to annotating diaries of old ships from the 19th and 20th century. The initial goal of the project was to record weather data, which would advance climate research. Building a strong community around it, the motivation of the crowd diverged from the designed goal and even though its objectives were attained, participants seemed more motivated by the historical nature of the data.

TROVE used crowdsourcing to correct and validate OCR digitisations of Australian newspapers beginning from 1803 onwards, a corpus simply too large for the Australian National Archives to process on their own. The real benefit of this massive annotation effort was the creation of a fully indexed and searchable digital collection.

Prism is an interesting example of collaborative tagging of literary texts. In *Prism*, participants are asked to provide their own interpretations of a text, which in turn are not judged against each other to establish the "correct" contribution. Researchers are rather interested in overlaying different points of view of the same text.

Scottish Words Place Names is a rare example of crowdsourcing project dedicated to recording ephemera, in this case to establish a vocabulary of Scots[2] words currently in use and their meaning. The project also benefited from *GLOW* (the intranet of Scottish schools), which facilitated the development of a community of followers and contributors among schoolchildren and teachers. Just like *Letters of 1916*, the rather intimate cultural link might also have played an important role in the success of the project.

Johny Cash project is a beautiful crowdsourcing project[3] set to create a collaborative music video for Johny Cash's for the title song of the singer's second posthumous album, "Ain't no grave". Participants from around the world drew sketches for each frame of the music video. Participants can then vote to promote their favourite sketch corresponding to each frame. Alternatively, the platform classifies the sketches according to the drawing styles, resulting in different renditions of the same video.

Peer-to-peer social services . Refer to the large number of platforms which over the past 10 years have been steadily building what we now call "crowd or shared economy". These services usually cater for a certain need in the general population and offer a competitive price compared to long existing solutions. Amongst them we mention successful ride-sharing platform *blablacar*, controversial alternative "taxi" service *uber*, or flat-sharing solution *airbnb*, which has established itself as a serious competitor to hotel business around the world. In the crowd economy, there is a solution for (almost) everything, from affordable interior design (*houzz*) to renting a boat for the holiday (*Antlos*) or hiring a "pet-sitter" (*PetMe*).

5.2. General

Brabham, D.C. (2008). Crowdsourcing as a Model for Problem Solving

Darren Brabham has the merit of being the first researcher to investigate the crowdsourcing phenomenon from a research perspective. In this article, Brabham defines crowdsourcing as a "distributed, problem solving model", a definition rooted in his analysis of several such projects. This definition has the quality of moving away from the original self-evident definition crowdsourcing given by [Howe, 2006] in 2006 and to provide the seed

[2]Scots is not a distinct language in the way that, for example, Gaelic is, but it has a community of people who believe strongly in its cultural and linguistic importance [Dunn et al., 2012].

[3]http://www.thejohnnycashproject.com/

for future theoretical work in crowdsourcing. This is indeed the other main contribution of [Brabham, 2008], who in his paper proposes the hypothesis of adapting crowdsourcing from a successful business model, into a model suitable for non-profit applications. Almost then years after his suggestion, crowdsourcing is currently succesfuly applied in research, innovation, health, social and government sectors.

By analysing some of the most successful crowdsourcing initiatives (Threadless, iStockPhoto, InnoCentive, GoldCorp Challenge, White House Prize etc.), the author reinforces the problem solving potential of an approach that uses a large number of potentially non-expert members of the "crowd". To better understand this phenomenon, Brabham turns to "the wisdom of crowds" a notion extensively explored by [Surowiecki, 2004] whose finds that "under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them". The discussion thus shifts on creating these so-called "right circumstances" which today might be precisely the technologies offered by the world wide web. The author follows up with a solid argumentation on why crowdsourcing should be grounded in the web – a decentralised, global and open environment where culturally diverse individuals, who are able to move freely around the world, contribute and exchange information. In it, humans are not just consumers of information, but find themselves as active participants *in* this medium as actors, creators and innovators.

One interesting point made by the author is the difference between open source and crowdsourcing. Open source is a philosophy for product development, where all the pieces are openly available for desiring contributors to improve the final product. There is no material reward in open source and investigations into open source incentives have revealed that participants to such projects are motivated by "the thrill of finding a solution" or the social capital among fellow hobbyists. However, Brabham considers it is hard to use such motivations in crowdsourcing settings, since participants might be reluctant into donating their free time to "produce" information of products that are used by others. Nevertheless, based on the success of crowdsourced citizen science platforms such as Zooniverse or Crowd4U, we argue that such motivations that were until now associated to open source participants can now be (carefully!) extended into crowdsourcing scenarios.

In the final part of his work, the author discusses advantages and disadvantages of crowdsourcing. One of the most controversial facets of paid crowdsourcing is the danger of creating a "slave economy", as the wages paid to participants, even in settings that differ from micro-paid micro-tasks are substantially smaller than those on the labour market. For instance, a photographer for iStockPhoto receives only 20% of the platform's income for their photo, while a successful Threadless designer can potentially earn $ 2000, a fraction of the company's $ 18 million gross revenue in 2006. On the positive side, a certain type of crowdsourcing allows participants to develop new skills, cultivate their hobby or even launch into a new career. What is most important is that in creative settings (such as the case of Threadless), crowdsourcing diminishes the impact of business giant, by reintroducing the eventual consumer into the design process. It is also worth to contemplate on Brabham's observations of the entrepreneurial dynamics within the crowds, for in the author's view,

with crowdsourcing the real "winners" are not those entrepreneurial minds that surface from the crowds and tap into their freshly discovered talent, but the creators of crowdsourcing opportunities, those who develop platforms and bring the crowds together in the first place.

Finally, the author makes the case for careful "crowd design", especially from a diversity standpoint. As the power of crowdsourcing resides in the crowd, it is its diversity (age, gender, ethnicity, culture) that enables a crowd to converge towards an acceptable (generalisable) solution. In the same way, the requester needs to insure its quality assurance mechanisms are not culturally or demographically biased, as such a bias would do nothing else but sink down contributions coming from disagreeing environments.

The author concludes with an agenda for crowdsourcing research, highlighting several research directions, some of which have already started to being addressed: matching the right set of skills to a given task, strategies for attracting and retaining crowds (especially given the constant development of social media technologies), deploying crowdsourcing in environmental, governmental, urban planning, emergency planning or cultural activities. At the same time, with crowdsourcing being increasingly used by companies, governments, health and research organisations, there is an increasing need for discussions about the legal and ethical implications, including one very stringent question: should crowdsourcing be subject to government regulations?

Chittilappilly, A.I., Chen L., and Amer-Yahia S. A Survey of General Purpose Crowdsourcing Techniques (2016)

This is (one of) the most recent and certainly most comprehensive general survey of crowdsourcing techniques, a state of the art review in its own right, that provides an extensive coverage of these first eight years of crowdsourcing research and analyses the domain from three main perspectives: incentive design, task assignment and quality control.

Looking at crowdsourcing as a process, the study first identifies its three main stages and their sub-processes that they include:

1. Initialising, which involves the sub-stages of task design, task breakdown and incentive design;
2. Implementation, which involves finding a crowd and assigning tasks to the crowd;
3. Finalising, which involves quality control and aggregation.

The authors recognise four main recurrent challenges that occur in general crowdsourcing settings: designing proper incentives in order to retain workers, proper recruitment of workers taking into account the required tasks and both sub-processes in the finalising stage, namely assuring the good quality of the workers' contributions and aggregating their contributions according to the necessities of the requester.

Incentives In crowdsourcing settings, incentives are both financial and non-financial. A large body of research has stressed the importance of financial incentives in attracting and retaining participants on crowdsourcing platforms. A 2009 survey of AMT workers showed that 91% of them preferred financial incentives, whereas 25% of American and 13% of Indian *turkers* declared AMT as being their primary source of income. Concerning non-monetary incentives, the study provides with a detailed overview of such motivations. Fun and entertainment play an important role in attracting and retaining participants and should be considered when designing crowdsourcing tasks. Such incentive is used in "gamified" approaches to data collection, such as *games with a purpose*. Other settings choose to design a competition setting, where contributions are rewarded according to their usefulness or quality. Personal development is another strong motivator for online participation and some of the best showcases to this respect are knowledge production systems such as *Wikipedia*, *Quora* or *Duolingo*, where people develop skills while contributing data to the respective platforms. Reputation mechanisms have also proven valuable by motivating participants to improve their status in a given reputation scheme. At the same time, it has been observed that participants might lose interest and be less committed once their status drops. A related incentive scheme is that of material (yet non-financial!) rewards, such as points, credits, vouchers or other products offered as a reward for participating contributors.

One thing surfaces from the study of various monetary incentive methods – the tensions that usually appear between requesters and workers: if the work is under-priced, there will be little involvement with the task, whereas overpriced work leads to data quality issues, especially in micro-task settings, such as AMT. With this in mind and based on the forms of crowd motivation previously presented, the authors identify and discuss several types of monetary incentive models:

- Reputation-based model. In this type of model, more tasks are assigned to workers with higher reputation level, which in turn is being reassessed periodically. However, such models assume that all tasks are the same, which may lead to performance problems if the right task is not matched to the right worker.
- Rating and reward-dividing model. This model assumes workers have different levels of expertise with respect to different tasks, which in turn impact the task assignment and reputation protocols, respectively. In the example cited by [Chittilappilly et al., 2016], a service administrator is involved, whose role is to divide the reward between those workers whose contributions have been accepted by the requester.
- Relevance-based model. This model doesn't take into account workers' reputation. Instead, workers are being informed that their contributions are to be checked against those of an expert. Three payment schemes were subsequently be devised:
  - positive incentive: a bonus is given if the worker and the expert agree over an assessment;

  - negative incentive: a penalty is applied to the base payment if the expert disagrees with the worker's assessment;
  - combined incentive: a bonus is given for agreement and a penalty is applied in case of disagreement between the expert and a worker.

Task Decomposition As simple and intuitive as it may seem, the application of the principle of task decomposition is paramount to the success of a crowdsourcing project. In their survey, [Chittilappilly et al., 2016] identify three types of decomposition strategies:

- sequential: this involves the serial execution of sub-tasks, in situations where the execution of later sub-tasks is dependent on the output of previous sub-tasks;
- parallel: these kind of tasks usually involve the division of the subject matter into smaller, independent pieces, each corresponding to a sub-task that can be performed separately;
- divide-and-conquer: in this setting, larger tasks are recursively separated into sub-tasks, up to a certain level of atomicity. At this stage they are individually resolved by workers and the solutions recursively put back together;
- macro-tasks: not really a decomposition technique, but it is worth noting that due to their nature, certain tasks are simply unsuitable for decomposition (e.g. hiring tasks, t-shirt design etc.).

Task Assignment Two important questions need to be taken into account when designing task assignment protocols for crowdsourcing: what is the expected result (task requirements) and what is achievable by the crowd (human attributes). The authors present two main classes of task assignment algorithms:

- In **offline** task assignment, the requester knows in advance all information about workers expertise and their bid;
- In **online** task assignment no prior information about workers is known by the requester. Instead, such information will be available a-posteriori and it will be up to the requester to identify the workers' level of expertise based on the collected data.

We encourage our reader to refer to the survey of [Chittilappilly et al., 2016] for more details and a synthetic analysis of the pros and cons of these algorithms.

Quality Control The quality control discussion in the survey is based on the authors' statement that "the main objective of a crowdsourcing system is to maximise productivity at minimal cost". From the analysis of different crowdsourcing settings, it appears that quality control can be done before or following the execution of the tasks. Quality control prior to task assignment and execution can be achieved in several ways. First, a requester needs to make sure that the proper incentives are in place in order to attract and retain the right crowd. Next, tasks need to be carefully planned and designed; clear tasks will be more

easily understood by workers, who will be more likely to provide better contributions. Also, at this point, it would be a good idea to estimate the size of the crowd that is being needed, as both too few or too many participants could have a negative impact on the quality of the data. At the same time, we know that participants' expertise can vary with different types of tasks. Thus, based on the learning effect noticed in crowdsourcing settings, some research have highlighted the importance of training workers in order to increase their expertise with the required task. Finally, there are techniques of estimating workers' skill based on prior work (however, not all platforms can provide this information) or, more commonly through qualification tests, aimed at determining their level of expertise with different proposed tasks. Post-task execution quality control is achieved through strategies such as worker selection, workers with specific objectives or aggregation. In **worker selection**, participants are selected based on their expertise calculated either as a trust score based on their previous contributions or as an estimate following a learning phase. The second strategy – **workers with specific objectives** – is based on the assumption that groups of workers assigned to sub-tasks perform better than all workers working on all tasks. Specifically, a subset of contributors would work on the main task, while their results would be verified by another subset of workers acting as a control group; then, the assessments of the control group are sent back to the crowd for an additional verification step, thus aiming to diminish the impact of cheaters in either group. Finally, **aggregation** methods are rely on the idea that multiple contributions for the same task typically converge towards a reduced (or unique) solution. Iterative aggregation methods are usually composed of two steps: first, all answers are aggregated, taking into account the expertise of each contributor; then, the expertise of each contributor is updated based on the answer they provided. On the other hand, most common non-iterative aggregation methods are improved variants of the well known majority voting rule. Another quality control strategy worth noting is the use of gold standard questions inserted during the execution of the tasks. These questions, whose answers are previously known by the requester, are presented as typical questions and are part of the task being executed. Failing to provide correct answers to such questions should have a negative influence on the reputation profile of the worker in question.

The study of [Chittilappilly et al., 2016] was of great importance to our own research. Not only that it provided us with a wide view over the main aspects of crowdsourcing, it also allowed us to position our work with respect to the most recent body of work dedicated to crowdsourcing research. With respect to the four methods used to attract participants (volunteering, authority, payment and pay-to-use) our project is currently focusing on the volunteering component, promoting itself as a citizen science project dedicated to supporting research in comics and sequential art. One other point made by the survey is that participants are more motivated to participate by seeing the impact of their contributions. To this end we are constantly engaging with our community through social media channels, by providing updates of our data collection effort and all of its applications. Concerning task design, our tasks can be executed both as a serial sequence and in parallel, depending on each participant's skill and preferences. Finally, in what concerns the quality

control, our approach proposes the use of a non-iterative aggregation method based on agreement measures between multiple annotators; alternatively there is a possibility of extending our approach to an iterative process that also involves modelling the expertise of our participants.

Saxton, G.D., Oh, O., Kishore, R. (2013). Rules of Crowdsourcing: Models, Issues, and Systems of Control

The objectives of the work of [Saxton et al., 2013] are to improve Jeff Howe's original definition of crowdsourcing, building a taxonomy of nine types of crowdsourcing models and perform an analysis of several important issues arising in crowdsourcing systems (e.g., collaboration, task types and managerial control).

The authors sets to improve the original definition of crowdsourcing, by incorporating the notions of "crowds", outsourcing and social web technologies, in a way that it distinguishes crowdsourcing from e-business, e-commerce, outsourcing, Web 2.0 and social media. The main motivation for an updated definition is to provide academics with a clear conceptual base for their future research in the field. The authors also interpret the definition given by [Brabham, 2008] (a "problem-solving model") as limiting crowdsourcing exclusively to its innovation potential. The crowdsourcing definition that they provide with is the following:

> Crowdsourcing is a sourcing model in which organisations use predominantly advanced Internet technologies to harness the efforts of a virtual crowd to perform specific organisational tasks.

[Saxton et al., 2013] agree with [Brabham, 2008] on the fact that crowdsourcing and outsourcing achieve their objectives by sourcing business needs to an outside entity. However, while in outsourcing these needs are achieved through third-party entities, crowdsourcing relies on an "undefined, non-professional, heterogeneous online crowd", with the expectation that its contributions can be more valuable than those of a handful of professionals. The last key element in this definition is the role fast, easy-to-use, over-reaching Internet technologies in engaging with a wide public, tapping into individuals' skills and aggregating them into the desired form of knowledge.

The second important contribution of [Saxton et al., 2013] is the creation of a taxonomy of crowdsourcing models, following an empirical analysis of over 100 crowdsourcing projects (their projects sample includes English language, crowdsourcing-centred initiatives, thus excluding one-time idea competitions or small-scale data collection projects) from four key perspectives: the type of service/product being outsourced, the role of community users, the level of collaboration and the type of managerial control systems.

1. Intermediary Model. Online participants are a virtual workforce which are typically engaged in a find-execute-earn kind of workflow. This is the classic case of AMT or *InnoCentive*.

2. Citizen Media Production Model. Revenue is generated through user-created content (video, news, commercials). Examples of this model are citizen journalism platforms.

3. Collaborative Software Development Model. The typical processes in software development and software products marketing are sourced to the "crowd".

4. Digital Goods Sales Model. The creation or transformation (e.g., through tagging, metadata enrichment etc.) is outsourced to an online community, with the platform addressing potential intellectual property issues. One of the most well known examples is that of *iStockPhoto*.

5. Product Design Model. The design of products (e.g., t-shirts, mugs, shoes etc.) is outsourced to an online community, with selected designs receiving some sort of financial reward. This is the case of *Threadless*, *Zazzle* or *Fluevog*.

6. Peer-to-Peer Social Financing Model. Bypasses traditional financial institutions and directly connects lenders with borrowers. An observation needs to be made with respect to this model, which can normally be extended to all peer-to-peer social services that have appeared in recent years. Thus, noteworthy instances of this model are platforms such as *airbnb* or *blablacar*.

7. Consumer Report Model. The review process is outsourced to online consumers, through online platforms such as *Epinions.com*.

8. Knowledge Base Building Model. In this model, specific knowledge is acquired from an online community, whose contributions are usually subject to aggregation methods or other quality control mechanisms.

9. Collaborative Science Project Model. Large online crowds take part in certain stages of scientific research, usually helping out in areas where computer algorithms are still under-performing. This is essentially the case of citizen science platforms (e.g., *Zooniverse*) where participants are annotating or transcribing hand written texts, counting animals, analysing galaxy shapes etc.

The authors subsequently identify three key dimensions along which their taxonomy stands: the product/service being outsourced, the level of collaboration and managerial control systems. A great variety of product and services that are being outsourced through crowdsourcing platform, result from their empirical analysis. What is truly fascinating is that the low cost and easy-to-use social media technologies empower crowdsourcing to transform existing product and service markets (*airbnb*, *uber*) and even to create new ones (*blablacar*, *houzz*). In what concerns collaboration, the authors note that in crowdsourcing it tends to occur in various degrees. However, collaboration is not a staple of most crowdsourcing projects. What does seem to stand out is the capacity of a platform to attract a community and leverage its various skills to elicit the desired outcome. This idea is also at the foundation of managerial control systems for crowdsourcing. While in classical work settings, the management focuses on enhancing workers' skills, the strategy

in crowdsourcing is to leverage whatever skills are already available and manage them in such a way as to achieve the desired outcome. The authors focus on three ways of achieving this:

- compensation schemes: half of the platforms in the analysis provide monetary or equivalent rewards (of different types and in a wide range of values), while some non-financial rewards involve open recognition of high quality contributors;
- trust-building systems;
- voting and commenting: take into account community voting or ratings into the decision-making process or reward mechanisms.

Brabham, D.C. (2012). Managing Unexpected Publics Online

In this subsequent article, crowdsourcing scholar Darren Brabham presents a case study focused on engagement and management strategies of "crowds" in the context of a crowdsourced design competition. He argues that the wide-reacing web-based tools can complicate and sometimes even undermine the efforts of PR practitioners in reaching target publics, especially in local contexts. The findings of the case study contradict our expectations of how "crowds" engage with their tasks and reveal unexpected interactions within.

The case study concerns "Next Stop Design" a crowdsourcing competition organised by the *Utah Transit Authority (UTA)*, with the aim of designing the next bus stop shelter at a very busy location in Salt Like City. The project was destined particularly to commuters from Salt Lake City or the Utah state in general, with the possibility of reaching participants from neighbouring states as well as the rest of the USA. Thus, initial promotion efforts were focused on local and state media, mailing lists of urban planning, design, communication and architecture programs, social media campaigns and it also featured on the White House blog. A few weeks later, a New York City design blog picked up the news and soon other design and architecture websites (with large followings) reported it themselves. Once Google SketchUp featured it on their blog, the story reached international audience.

As a consequence, despite efforts to focus the project towards a local crowd only a bit more than half of the website visitors (53.7%) and less than half (45.4%) of the registered users came from the United States. The author admits that most of the public on *Next Stop Design* was unexpected and to some extent unwanted, since the input of someone half-way around the planet on a public matter concerning Utah locals could be ill-informed (are all non-Utah contributors familiar to the particularities of the local climate, for instance?) and could even be considered as "undemocratic".

Due to the featuring of the project on different professional architecture and urban design websites, most contributors ended up being professional designers. While some of the first designs included a lot of "pencil sketches, cleanly drawn and effective at conveying

an idea, but amateurish", as soon as the first 3-D models were submitted, amateur sketches quickly ceased to be submitted. Moreover, in a series of qualitative interviews following the project, Brabham noted that some of the professional designers who had submitted declared that amateurs shouldn't have even been allowed to vote for the winning design.

The author underlines that unexpected publics are not always be detrimental to a crowdsourcing campaign and admits that without the diversity of worldwide participants or the professional input of trained architects and designers, the quality of the submissions would not have been so high. On the other hand, this kind of participation was far from being desired as it broke the basic principles of public participation.

Finally, the author concludes with a series of online community management recommendations.

- Do not assume your public, as offline communities do not exactly map to those online;
- Use offline tools for online projects, especially when targeting local publics;
- Use filters to give different privileges and weight to different publics;
- Track and respond, using different analytic tools;
- Make a plan for handling varying degrees of the unexpected;
- Have an online community manager, as their workload at different stages of a campaign can be quite important.

As for our crowdsourcing project, these guidelines proved to be useful in managing our own community of participants.

Sharma, A. (2010). Crowdsourcing Critical Success Factor Model. Strategies to harness the collective intelligence of the crowd

This paper provides an adaptation of a critical success factor model from outsourcing to crowdsourcing. The model assumes **motive alignment** – "the extent to which the crowd is able to associate with the long term objective of the crowdsourcing initiative" – as the primary factor for crowdsourcing success and identifies five peripheral factors that influence crowd motivation:

- Vision and strategy. A properly presented vision statement might attract the support of important stakeholders, including government.
- Human capital. Virtually the crowd size and their skills. If needed and possible, these skills should be developed through proper training.

- Infrastructure should be easily accessible, reliable and provide a good quality of service (i.e. good telecommunication or Internet services). A critical aspect here is the initial capital investment in developing the crowdsourcing infrastructure.
- Linkages and trust. The degree to which a crowdsourcing project manages to engage with concerned communities or with other successful ventures.
- External environment. The macroeconomic environment in which a crowdsourcing project is active; it comprises of the living environment, government, businesses and other economic agents.

This paper is a qualitative study from a more business-oriented perspective, thus provides a different view on crowdsourcing motivation factors. It is still unclear what is considered "successful" crowdsourcing and from what we were able to understand this is determined by looking at the crowd size. Even under this assumption, a better definition on **how** to estimate crowd **size** would still be needed: is it everyone who logged into the platform, is it all who have contributed at least once, is it daily active users? While our intuition tells us that these outsourcing-specific factors should easily extend to crowdsourcing as well, in absence of a quantitative analysis and for the purpose of our work we will consider them as being strictly referential.

Difallah, D.E., Demartini, G., and Cudré-Mauroux, P. (2014). Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms

This paper reviews current techniques for detecting different kinds of spammers and malicious workers as well as their limitations, by describing various attacks on crowdsourcing tasks. The work of [Diffallah et al., 2012] deals exclusively with monetary incentives, which have been shown to raise problems with the quality of the crowdsourced data. Cheaters, bots or humans, are acting towards maximising their reward by completing tasks as fast as possible, in a random or semi-random fashion, which leads to a drop in the quality of the collected data. The authors thus claim that current quality control techniques are insufficient to counter organised groups of cheaters, working exclusively towards maximising their reward, since these techniques are not generalisable across a large variety of task types, in particular on subjective tasks. Controlling the reward has also been proven ineffective as previous research has shown that this only improves dwell time, with negative impact on the quality of the contributions.

We have previously seen, in the general survey of [Chittilappilly et al., 2016], that quality control techniques are applied before or after a task is executed. Task design is usually employed before executing a task and with cheaters being more discouraged to take part in new tasks or tasks of creative nature. After the execution of the tasks, typical quality control methods are aggregation, test questions and machine learning techniques.

Finally, the authors present a number of potential attacks against the quality control techniques currently in use:

- In **individual attacks**, cheaters usually try to complete tasks as fast as possible in order maximise their reward by performing as many tasks as possible. Typical strategies see attackers provide random, automatic or semi-automatic answers. Random answers are easily detectable by looking at completion times (often very low) and are typically addressed by collecting multiple answers on the same task. Automated answers come in the form of random or low level artificial intelligence bots, but are usually dealt with by introducing test questions within the workflow. Semi-automated bots try to parse the available answers and subsequently query a knowledge base or ask a human.
- **Group attacks** are more even more organised and exploit the repeatability of a task in order to build a Shared Question-Answer Dictionary. Typical strategies involve:
  - major agreement on one answer: the first cheater to perform a task gives a random answer, which is then stored in the dictionary and used by the following cheaters. This tactic is very effective against the majority voting rule;
  - answer sharing: this tactic is aimed at tricking test questions, by recognising either their redundancy (gold standard questions) or their difficulty (Turing questions, such as CAPTCHA). In the first case, once detected, cheaters agree to provide an honest answer, which is then stored in the dictionary for future reuse; in the second case, cheaters are alerted that such a task can be performed only by a human, in which case they briefly take over, complete the Turing task and then resume the automated cheating strategy.
  - artificial clones: in this kind of attack, a malicious worker honestly completes all available tasks and stores the answers in a dictionary, then creates multiple automated bots that are able to replicate the honest behaviour based on this dictionary. The authors recognise the difficulty in spotting this kind of attack which can easily result in increased financial gains for the cheater and inconclusive data for the requester.

As crowdsourcing becomes increasingly popular and more organisations will fund such data collection campaigns, we can expect that cheating strategies will also improve and become more elaborate. [Diffallah et al., 2012] draw a signal concerning the naivety and lack of robustness of current quality control mechanisms. They recommend that future research focuses on discouraging cheaters (worker analysis systems, non-financial incentives), rather than a-posteriori data filtering.

Kosiński, M., Bachrach, Y. (2012). Crowd IQ: Measuring the Intelligence of Crowdsourcing Platforms

[Kosinski and Bachrach, 2012] crowdsource an IQ-test on Amazon Mechanical Turk (AMT) and analyse the performance of the participating crowd under different conditions: different payment levels, impact on reputation, number of workers per task.

The authors took a non-trivial IQ-test and transformed each of its questions into a Human Intelligence Task (HIT). In this way, the experiment was focused on measuring quality rather than quantity of work and at the same time it was designed to measure the IQ of the "crowd" rather than that of each individual.

In the rejection experiment, the authors designed two conditions: in both of them, participants were informed that only correct answers will be rewarded (fixed payment of $ 0.10), but in one of the conditions incorrect answers would be rejected, which on AMT leads to a drop of the reputation level. Results showed that the reputation-fearing crowd scored better than the control crowd (138 vs. 119 IQ points); at the same time, the reputation condition "scared" some of the participants, which chose not to take part at all, thus resulting in a lower crowd turnaround and a longer completion time.

In the payment experiment, the crowd is divided into four payment conditions ($ 0.01, $ 0.05, $ 0.10, $0.20) and the time to complete the test in each condition was then recorded. No rejection rule was used, hence there was no impact on participants' reputation. Consistent with other research in crowdsourcing and behavioural economics, there is a payment "sweet-spot" (in this case at $ 0.05) after which any further increase in payment doesn't lead to an increase in the general IQ of the crowd. There is however a monotonic relationship between the payment and the IQ increase rate (and completion time): turkers in the $ 0.20 condition finished all questions in only 15 minutes at an IQ rate of points-per-minute 14 times higher than turkers in the $ 0.01 condition, who completed the test in 18 hours! The authors thus conclude that if speed is not a priority, it is recommended to pay less and wait longer for the solution.

In the reputation experiment, authors studied the impact of different reputation levels over the quality of the work. While in previous experiments, only workers with high reputation (98% acceptance rate) were accepted, this experiment allowed workers with at least 80% acceptance rate to take part, at a fixed payment level of $ 0.05. It comes at no surprise that groups of workers with higher reputation (i.e. acceptance rate) yield to better crowd IQ scores.

Finally, in the aggregation experiment, the authors test the impact of combining an increasing number of responses. The interesting finding here, is a significant increase in the crowd's IQ following its initial augmentation in size; afterwards, no matter how many workers contribute, the IQ increase is only marginal. This also confirm previous findings about the existence of a population size "sweet spot". The authors report even better results

for small crowd once a dynamic approach – assigning more workers only to problematic tasks – is put in place.

At 122 IQ-points, the result obtained for an assignment of one worker-per-task, the crowd performs better on the IQ test than 91% of individuals in the general population. Despite its limitations (focusing on a specific type of task, setting the experiment on only one platform, using very simple aggregation techniques), this paper shows the strength and efficiency of crowdsourcing, as long as the right designs are put in place, particularly concerning the tasks, the crowd and the financial and reputation incentives.

5.3. Applications

von Ahn, Luis and Dalbish, Laura (2004). Labelling Images with a Computer Game

In their work [von Ahn and Dabbish, 2004] start from the premise that games are fun and that having fun is a good enough motivator for participants to a crowdsourced annotation task.

In their annotation game, the "fun" is provided by the incentive of continuously gathering bonus points when two players agree on an image. What differentiates their initiative from others of the kind is the questions design. Thus every player is asked **"to think at what the others would think of"** when seeing an image (as opposed to the straightforward "what do you see in the image?").

The game is based on what the authors call "taboo words", words which are the set of words having already been identified as good labels for an image. Each image starts with no taboo words and will be annotated with a maximum of 6. An annotation becomes a taboo word the moment a minimum number of agreements over it has been reached between the players.

A round consists of 15 images being shown to 2 players, who are then given 30 seconds to annotate the image. The time limit is used as a primary mechanism to prevent cheating by collusion between players. von Ahn and Dalbish also introduce a mechanism for **massive agreement prevention**: collusion is detected by measuring the average time when large agreements occur; to address this they introduce pre-recorded players (previously played games) while cheating is still happening.

Whilst the premise of their work ("fun makes tedious tasks easier") is clear, it is not obvious how exactly they plan to "bring the fun" in their image labelling game, though some features as well as design choices seem to support their endeavour:

- selecting the images. The game uses a corpus of 350 000 images. In order to keep the game attractive and avoid that the users get bored, the images are randomly selected from different sources, covering different topics;
- spelling. To alleviate the input task for the players and to keep the collected data as clean as possible (e.g. no misspellings), the game server is also equipped with an online dictionary;
- theme "rooms". Some players may be interested in using their expertise on a specific theme. The authors suggest the introduction of so called theme "rooms" where only images related to a certain theme are presented to players. The advantage is two-fold: the players keep having fun working on what they want, the system benefits of the expertise in a certain field of a benevolent annotator.

In what concerns the **evaluation** of the collected data, the authors propose three experiments, all of the yielding good results in terms of annotation quality:

1. using taboo words as search queries. 10 random frequent taboo words (more than 8 occurrences) were chose and used as search queries in an image search engine. Then the top image results were assessed for relevancy. All 10 words yielded relevant results;
2. assessment by a non-playing annotators group. 15 annotators were presented with the same 20 images (chosen randomly out of those images which had at least 5 taboo words assigned by the game). They were then asked to list 6 words which **they feel** best describe the images. These were then compared to the taboo words for each image. The results showed that for all the 20 images, at least 5 of the 6 words had been already suggested during the game;
3. relevance assessment. 15 subjects were presented with the same 20 images (randomly chosen from those having at least 5 taboo words assigned by the game). For each image they were also presented with the first 6 taboo words assigned during the game. They were then asked how many of these words **they would use** to describe the image and how many have nothing to do with the image. The subjects found that on average 5.105 words ($\sigma = 1.04$) were relevant according to their preferences, while 0.105 words ($\sigma = 0.25$) are not.

The method proposed by the authors is certainly very attractive. The strength of their method seems to reside in the "fun" approach (in the form of a game) in an otherwise tedious job even for a big crowd. Their cheating detection mechanisms as well as the small innovations meant to keep the game as fun as possible differentiate it from typical crowdsourcing image annotation projects. Moreover, making use of the element of "fun" may constitute both a good enough incentive for participants (suitable for those projects which cannot propose monetary rewards for the crowd) as well as a potentially efficient method to select those users which are motivated by the task and not by the reward.

The one downside seems to be their evaluation protocol. In the case where several participants recognise the scenario, the whole experiment could be biased towards selecting the relevant results or the contrary. We aim to address this problem, by proposing a masked, continuous quality assessment integrated in the annotation process itself.

Eagle, Nathan (2009). txteagle: Mobile Crowdsourcing

The idea for txteagle [Eagle, 2009] originated while the author was working in a hospital in Kenya, facing recurring blood supplies shortages. A first version of (SMS Bloodbank) was asking for the nursing staff in all the hospitals nearby to update their hospital's current blood supply by sending a simple text message. While virtually every targeted nurse had access to a mobile phone they were extremely reluctant in using their own money (phone credit) for the task, therefore the system was initially a failure. A first change to the system following a deal with a local phone operator allowed the participants to receive airtime or phone credit when they were providing the blood level updates. This small fine tune on the motivational component of the crowdsourcing setting turned the initiative into a successful one, which is now used throughout Kenya.

Furthermore, the project expanded the task diversity and is now targeting the whole Kenyan mobile phone users. The workflow is very simple: a participant is receiving an SMS with the micro-task. Upon completion, he replies with the solution. Tasks are as different as:

- **transcription** – participants are asked to transcribe medical files. With a 98% accuracy rate, the reported savings in terms of transcription costs are around 60% in the worse case scenario;
- **software localisation** – with Kenya being a cultural melting pot, where over 60 different languages are largely spoken, writing locales for software solutions can be a tedious jobs. *txteagle* has proposed the translation of a phrase book of the most frequently used words and by 2009 the participants had managed to translate them in 15 languages;
- **citizen journalism** – this is probably the most innovative type of tasks participants can take part at. It addresses the need for people to be content creators and it provides decent incentives for aspiring journalists;
- **search relevancy**
- **blog / news translations and sentiment** - evaluating the sentiment or opinion of participants with respect to different products on the market;
- **survey and market research** – making use of the large penetration rate of GSM technology in Kenya to reach even the most remote areas and include more people in SMS-based surveys;

- **voice tasks** – gathering audio recordings from users to be used as training corpus for speech recognition algorithms.

txteagle uses a simple maximum likelihood estimator to estimate the the task difficulty and the user's expertise. Thus, each participant receives tasks calibrated to their expertise, while the quality is evaluated according to the difficulty of the task. While the approach is fairly simplistic and takes into account only the difficulty as a factor influencing the response quality, it is nevertheless one of the best examples of how simple and efficient a crowdsourcing data collection initiatives can be.

Finally, it's worth noting that the project has gotten a fair amount of media exposure in Kenya and world wide [Marwaha, 2009] and has now developed as a Kenyan start-up company – Jana.

Brabham, D.C. (2010). Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application

Crowdsourcing scholar Darren Brabham pushes forward the discussion about motivators behind crowds taking part in crowdsourcing processes, by presenting yet another case study, that of t-shirt design platform *Threadless*. This qualitative study was undertaken in the form of interviews with members of the mentioned community.

Threadless is a crowdsourcing platform on which members of a sizable community take part in a weekly t-shirt design competition. The best designs are selected by the community during a one-week voting period and are subsequently produced and sold back to the crowd by *Threadless* for prices between $ 15-20. The company, one of the most successful small enterprises in United States, had revenues projected at around $ 30 million in 2008, at a 30% profit margin. This success, Brabham acknowledges, lies in the diversity of the crowd the platform attracts, which translates into a diversity of ideas, a vital element in creative endeavours.

The author then summarises different motivators for crowd participation in similar settings. First, the possibilities of today's web technologies have transformed crowds from simple user/consumers into participants and producers of content – by the crowd and for the crowd. And with web technology being the main infrastructure for crowdsourcing, there seems to be no better way to harness this crowd engagement and put it to good use. Second, by referring to open source projects, which have a great deal in common with crowdsourcing, Brabham reminds that contributors to such projects are usually motivated by the pleasure of doing what they enjoy. Other projects managed to surface intrinsic motivators, such as the thrill of solving a problem in the case of *InnoCentive* or altruistic, such as participants enjoying to share their skills and taking part in the creation of crowdsourced movie *Star Wreck*. However, in settings such as *iStockPhoto*, the desire to earn money or improve

one's skills trumped the desire to network with similarly interested peers or other altruistic motivations.

The study consisted of interviews with 17 members of the *Threadless* community, seven of which had submitted at least one design, while the other ten were actively involved in other ways: voting designs, discussing on the forum or shopping on the website. Five themes emerged concerning participants' motivations:

1. Making money. This was a recurring theme for all interview participants who had already submitted at least one design, whereas only one respondent in the non-designer group mentioned it as a motivator in case they would ever submit.

2. Improving creative skills. For many at participants at *Threadless*, design is merely a hobby, but having a supportive community, able to provide feedback, helped some of them become better at it.

3. Freelance opportunities. For some of the members, the experience acquired by actively participating at *Threadless* turned their hobby into a professional opportunity, either by moving into full time design jobs or starting their own successful ventures.

4. The sense of community. Seems to have been a strong motivator for retaining participants on the platform. All but two respondents mentioned their attachment to the community and appreciated the humorous dialogue on the platform's forum as a reason for constantly returning. To this respect *Threadless* is a fine example of how successful crowdsourcing platforms are able to create a brand community, which in turn feeds back into the success of the platform.

5. Addiction. This was perhaps the most interesting emerging theme, as 11 of the respondents used language associated to addiction when talking about their experience and interaction with *Threadless*. While for some it may be the case of hyperbolic language to express their attachment to the platform, others have reported even being more careless with their daily commitments (work, school) in order to spend more time on *Threadless*.

[Brabham, 2010] provides with a very helpful study in what are crowd motivators in creative crowdsourcing. And while some of these motivations seem to emerge in other crowdsourcing projects as well – the financial incentives are important at *iStockPhoto* and *InnoCentive*, developing skills and freelance opportunities at *iStockPhoto* – there seems to be no set of motivators applying to all types of crowdsourcing initiatives. Brabham's work is central to our own, as it helped us to take more time to better understand the crowd we were going to target so as to carefully design our platform and social media campaign. At the same time, the idea of "addiction" has been linked with comics fandom; thus, part of our target public are comic fans and scholars, people whose comics addiction and willingness to share their comics related "skills" we want to harness.

Dunn, S., Hedges, M. (2012). Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research

This extensive study is part of a research project aimed at determining the possibility of using crowdsourcing in digital humanities research, by assessing the impact that crowdsourcing could have in the study of humanities, the motivations of the communities that choose to take part and to put forward a typology of existing typologies.

Crowdsourcing is relatively new in academic research. Even newer in humanities, it is employed in different activities, such as content digitisation, content creation or processing, content enrichment, relevance assessment, to name just a few. Just as much as in any other field, crowdsourcing in digital humanities helps researchers in areas where automated processing and algorithms are still outperformed by human agents.

The project surveyed 54 academic publications, as well as 51 projects, platforms and individual activities in the field of digital humanities and which make use of crowdsourcing to attain their goals. In addition to that, the project hosted two workshops, one dedicated to researchers managing crowdsourcing projects in humanities and the other to members of the communities participating in such projects.

Following these workshops, the authors have identified four factors (the list is certainly not exhaustive!) that characterise crowdsourcing for the humanities:

1. The existence of a clearly defined humanities direction and/or research question.
2. The potential for an open group to transform, add value or interpret a given primary material. A distinction is made here between crowdsourcing and the narrower term community-sourcing.
3. The possibility of defining a task or to break it into a workflow of replicable, definable tasks.
4. The activity should be scalable, both in terms of data volume and crowd size.

When discussing the different levels of engagement, the authors make a clear distinction between social engagement and crowdsourcing. In the first model, participants take advantage of existing technologies in order to contribute with their time, skill and knowledge towards a goal that may not always be clearly defined (e.g., uploading a video, tagging images on social media, enriching data on a map etc.), without the expectation of financial reward. On the other hand, in crowdsourcing, participants answer to an open call and work towards a well defined goal and in many instances receive some form of material reward. The authors note that crowdsourcing may involve more work and a deeper commitment to volunteer free time and from this perspective it is closer to citizen science, an analogous term which defines members of the public engaging in small tasks, but working towards achieving a large-scale scientific goal (e.g., bird counting, reporting on insect population,

observing stars etc.). We sum up the discussion about levels of engagements with an observation made by digital humanities practitioner Trevor Owens:

> Most successful crowdsourcing projects are not about large anonymous masses of people. They are not about crowds. They are about inviting participation from interested and engaged members of the public.

The body of work we reviewed so far awarded a great deal of attention to the motivational component in different types of crowdsourcing settings. In what concerns academic crowdsourcing studies acknowledge three motivators for crowd engagement: personal interest in the subject area, the opportunity of exercising the interest and the opportunity of engaging with others that share the same interest. This is certainly different from findings in materially motivated crowdsourcing, in which we recall that the main motivations were the financial reward and the possibility to improve one's skills.

Other approaches mention gamification as a means to make crowdsourcing more "fun" and thus attract and retain participants. Nevertheless, authors draw attention on the risks of gamification, which might be perceived as a trivialisation of participants' contributions and might be off-putting, especially for the most skilled of them. Competition is another important motivator and certain projects integrate it by using points system and leader boards acknowledging the participants' contribution. Just like in the case of gamification, platform designers should proceed with care, as they run the risk of driving away valuable users who simply don't have a competitive nature and might feel demoralised once their contributions seem not to be properly acknowledged via rankings. Another possibility is to break the "main" competition into several smaller competitions corresponding to different types of tasks and have separate rankings for each of them.

An important aspect brought up by [Dunn et al., 2012] are the different roles assumed by members in the crowd. Depending on the model of engagement proposed by the platform, participants can evolve from one role to another after certain achievements (much like changing status when acquiring more experience), assume these roles on their own (depending on their set of skills) or have these roles emerge, based on their contributions (for instance, in "long-tail" type of projects just a few participants – the so-called "super-contributors" – contribute with the largest amount of data).

The are various reasons for which academics choose to make use of crowdsourcing and in most cases it is used to fill a gap or accelerate a certain stage in the research process. For institutions in the Galleries, Libraries, Archives, Museums (GLAM) sector, crowdsourcing is a means to acquire information about assets that are difficult to access and in doing so, to stand up to their mission of engaging the public. In other cases, crowdsourcing is used as a way to build communities around research and scholar activities and, similar to GLAM institutions, this is particularly helpful at a time when humanities research units need to quantify their social impact.

The authors dedicate the second part of this study to the description of a typology of crowdsourcing for humanities. Due to the complexity of crowdsourcing and its composing elements, they propose a fluid, multi-faceted categorisation, built around four key typological areas: asset type, process type, task type and output type.

1. Processes:

    - Collaborative tagging. Tags are based on controlled vocabularies or free input. They are growing as the crowd contributes new terms and are used to index the annotated corpus and make it searchable. It has been found that in GLAM, tags freely provided by contributors can differ substantially from those in a controlled vocabulary supplied by a curator. In certain projects, which involve a certain degree of analysis and synthesis (much like *Comics++*), the goal might not be to find "the right tag", but rather to see how aggregated tasks give meaning to assets;
    - Linking;
    - Correcting/modifying content;
    - Transcribing is currently one of the most widespread crowdsourcing activities related to the field of humanities. It is tightly connected to correcting and modifying content as it usually involves participants transcribing handwritten text or text that is difficult to process through Optical Character Recognition (OCR) techniques;
    - Recording and creating content. May apply to recording ephemera and intangible culture. The contributions are usually subject to editorial control and they are either disseminated in publication form or are used for further analyses;
    - Commenting, critical responses and stating preferences. These might be collected in case they would make the object of a very specific research question (e.g., analysing the global reactions to a literary text);
    - Categorising;
    - Cataloguing. This is a very time-consuming, yet very common activity in GLAM institutions and some of them have tried to use crowdsourcing as a cost-effective alternative;
    - Contextualisation;
    - Mapping. Refers to creating a spatial representation of a specific asset (i.e. a building, a street, a neighbourhood etc.);
    - Geo-referencing. Consists in indicating the geographical position of an asset;
    - Translating.

2. Assets:
    - Geospatial;
    - Text is the most common type of asset in crowdsourcing projects;
    - Numerical or statistical information. These assets are sometimes difficult to differentiate from textual, as they either appear in textual form or are part of larger textual assets;
    - Sound;
    - Image. A common type of asset as well, usually subject to tagging, categorising, contextualisation or geo-referencing;
    - Video. One of the lesser used assets in crowdsourcing. Some projects have seen the creation of films based on video assets recorded or created by participants;
    - Ephemera and intangible cultural heritage (e.g. oral history). These are perhaps the most intriguing and research-promising type of assets, certainly amenable to crowdsourcing, but at the same time the most underdeveloped from a methodological standpoint.
3. Task:
    - Mechanical: processing discrete units of information;
    - Configurational: identifying patterns or "configurations" in information;
    - Editorial;
    - Synthetic. These tasks require analytical skills and are very likely to enable the contributor to develop expertise;
    - Investigative. Involves seeking for information that is latent or implicit in corpora;
    - Creative.
4. Output: Text (original, transcribed, corrected, enhanced), Transcribed music, Metadata (descriptive or preservation), Structured data, Knowledge/awareness, Funding, Synthesis, Composite digital collections (structured or unstructured)

Probably the main contribution of the extensive report of [Dunn et al., 2012], these four facets offer a clearer and more comprehensive means of analysing crowdsourcing projects for humanities research. More so, this report shows us that in the case of humanities research, crowdsourcing distances itself from the lucrative spectrum under which it was first promoted ten years ago.

Samson, C., Fiesler, C., Kane, S.K. (2016). Crowdsourcing Comic Book Transcriptions

Despite this project being in its early stages, we consider it important for the current state of the art review, since it is, to the best of our knowledge, the only example of crowdsourcing used for annotating comics, except for our own *Comics++*.

In this exploratory study, [Samson et al., 2016] recruited 60 workers on AMT, which were presented with the same two pages of a comic book and were asked to provide a description of the actions within. The crowd was divided in two conditions: a group that was being informed that these transcriptions would be used in a text-to-speech system for visually impaired readers and a control group, which was not given any further detail. Two other questions, aiming to establish participants' familiarity with the given comic book and their capacity to recognise characters and their alter-identities, were also included in the questionnaire.

Since there is no gold standard to compare to, for the purpose of this study the authors chose a word count metric, based on the assumption that longer descriptions are associated with transcribers who are willing to put more effort into the task and are thus more reliable. Subsequently, they performed an analysis of variance (ANOVA) using word count as a dependant variable and the impact of the three factors (blind condition, expertise, character recognition).

An interesting observation with respect to the level of engagement of the turkers is the varied nature of their contributions, from very concise interpretations of the entire page, to accurate panel-by-panel descriptions. The findings of the study reveal that participants in the blind condition produce longer descriptions than those in the control group. Second, there is encouraging evidence that participants reporting familiarity with comics (based on their answers to the two additional questions) provided with more detailed descriptions. However, more investigations need to be made, since the results of the test didn't reach significance level.

The experiment of [Samson et al., 2016] yields encouraging results in what concerns the possibility of using crowdsourcing to annotate comic book scenes. It looks into motivations specific to readership and more particular to comic book fans, suggesting – just like us – that the passion of comic book readers can be harnessed to produce usable annotations. The authors admit to one important limitation – the lack of a gold standard for evaluating the quality of their collected annotations. This is one of the reasons for which we believe that the extent of *Comics++* can provide such comics research projects with a usable corpus.

Bongard, J.C., Hines, P.D.H., Conger, D., Hurd, P., Lu, Z. (2013). Crowdsourcing Predictors of Behavioural Outcomes

This paper presents a novel approach in machine science, in which the task of choosing features for behavioural outcomes – typically based on the expertise of a domain specialist – is instead achieved by outsourcing it to a crowd of non-experts. In this setting, crowdsourcing is used to both create a prediction model and to acquire the data needed to test it.

The proposed approach is based on a three fixed workflows, corresponding to the investigator, the participant and the predictive model. First, the investigator defines a human behavioural outcome and a set of seed questions, known to correlate well with the outcome. Participants initially arriving on the platform answer the seed questions, thus providing initial data. At the same time, they are able to suggest new questions of their own, which upon validation by the investigator become new predictors of the outcome. Subsequent participants will have a larger pool of questions to answer to. Finally, as the number of predictive variable changes, new models are generated and the data that is gradually collected is used to build and test different iterations of different predictive models. An interesting design choice, is to combine feedback with the idea of competition (two important motivators, as we previously saw), by showing each participant, once they provided all answers, how did theirs compared to others'.

This workflows can be adapted to different experimental settings, as [Bongard et al., 2013] did for two different experiments of predicting human behavioural outcomes: the energy efficiency of a household and the body-mass index (BMI) respectively. The authors then analyse the predictive power of different factors as they compose different models. In the first experiment, predicting energy efficiency, it is found, under satisfying level of significance, that the three factors with the highest predictive power are user-generated. While encouraging, the authors are still cautious about these results, mainly due to the small crowd size, which they attribute to the potential inconvenient of completing this task (i.e. searching for electrical bills). They account for such problems in the second experiment, in which they seek to predict the BMI, an information that the crowd should access much easier. The experiment lasted shortly under one week, during which 64 *reddit* users took part and provided over 2,000 answers to 57 questions. Every 5 minutes, a regression model was generated. The results show that there is no relationship between the predictive power of a question and the time when it is asked or how many responses it attracts. However, the predictive power of the questions seems to be distributed according to the power law, meaning that when considered as unique predictors, most questions have a very small predictive impact. This is not the case for the number of responses, which means that it apparently cannot explain the power law distribution of the most predictive questions. The authors warn on the fact that the number of questions is comparable to that of participants, with model over-fitting representing a potential limitation of this second experiment.

The authors conclude with a few remarks about crowd engagement. First they notice the motivation and potential of the crowd to uncover features of human behaviour outcomes;

at the same time, given that participants arriving at a later stage contribute with less data, questions of task design arise either to retain and motivate participants to provide more data or to direct them towards those small sets of questions which seem to be the most promising.

This approach is one of the first to address the concern expressed by [Dunn et al., 2012], who noticed that most crowdsourcing approaches are designed based on the need of the requester and, to a very small extent based on the characteristics of the "crowd", mainly due to the difficulty in predicting it. Such an approach can provide with new directions in one of the most researched fields of machine science, namely feature selection. Besides that, the strength of using crowdsourcing in this setting allows us to acquire, at the same time, the data needed to test the crowdsourced models.

5.4. Evaluation

Kilgariff, A. (1998). Gold standard datasets for evaluating word sense disambiguation programs

In his journal article dating back all the way to 1998, [Kilgariff, 1998] discusses the limits of the human tagging process, virtually what is today known as annotation. Considering the state of the art at the time the author discusses the limits of what we currently classify as expert annotation process. By underlining the high cost and high demand in terms of work intensity of human expert annotation he reiterates on one of the reasons which has probably lead to the need of non-expert annotations; added to this, the always recurrent need of already (manually) annotated corpus. Moreover, expert annotations seem to fall short on meeting what the author calls "adequate standards of replicability", meaning basically that two annotators working on the same task are far too frequently not assigning the same label.

Though his work concerns word sense disambiguation tasks, it is easy to extrapolate on what we are today calling human intelligence task. In what concerns evaluation protocols and based on the work of [Gale et al., 1992] the performance of every system should fall between a lower and an upper bound, with the lower bound being the performance that a naïve algorithm would achieve on the task, while the upper bound is the amount of time people would need to reach agreement. Although not practical from an implementation perspective, this view summarises the role of the human factor in the annotation process – human agreement is in fact the best achievable result.

The author then introduces the opportunity of using measures of inter-annotator agreement (IAA) and their use in conjunction with gold standard. Based on his remark that "without high IAA, a gold standard is fool's gold, our approach will seek to avoid the use of such techniques. As the author states, we don't only believe they're time- and

cost-expensive, but also highly static for a process (crowdsourcing) that we see as inherently dynamic and adaptive. We do however stop over the three potential reasons he identifies for annotation agreement failure and some solutions.

1. **irreducible indeterminacy of ambiguity in the data**. The author's suggestion is to provide with an explicit annotating scheme;
2. **poorly defined labels**. The solution in this case is very simple: make the labels as explicitly as possible!
3. **mistakes by annotators**. The author's solution involves using experts of the domain in conjunction with proper motivations. However, in current crowdsourcing settings as far as expertise goes, recommendations go in the direction of providing basic training to non-experts or simply counting on them getting acquainted to tasks.

Feng, D., Besana, S. and Zajac, R. (2009). Acquiring High Quality Non-Expert Knowledge from On-demand Workforce

[Feng et al., 2009] propose a practical framework for acquiring high quality non-expert annotations using Amazon Mechanical Turk (AMT) [4]. Particularly, the authors are interested in detecting bad annotators, improving data labels and achieving the right balance between cost and data quality using AMT.

Considering **data quality, cost** and **scale** as the key parameters in a crowdsourcing project, the authors propose a 2 phase methodology for the data collection effort using AMT. In the initial *validation phase*, a first basic questions should be asked whether AMT is appropriate for a certain task and if it is what are the parameters for obtaining high quality data. Specifically what is the right cost for a task and what is the appropriate number of labels to be gathered. Since the observed agreement is subject to chance agreement, the authors use K-statistics [Artstein and Poesio, 2008] to evaluate the agreement between non-expert annotators. Finally, they propose an intuitive algorithm for detecting outlier annotators: for each label of each annotator the agreement is computed with respect to the aggregated annotations (majority voting) of the other annotators. The distribution of the agreement measure is then analysed to identify and treat the outlier annotators. At the end of the first phase a certain cost / annotator and number of annotators will be identified as the input parameters for the second phase - **large scale submission**, when the tasks will be available to a larger number of turkers.

The experiments proposed by [Feng et al., 2009] show some interesting results. They propose a simple classification task asking the participants whether a given query refers to a local business or not. Two groups (in-house experts and non-expert turkers) take part at the validation phase and are assigned to 4 price conditions (1 ¢, 2 ¢, 5 ¢, 10 ¢).

[4] https://www.mturk.com/mturk/welcome

A first result shows that increasing the payment decreases the total number of turkers needed to complete the task, decreases the outliers and decreases the total turnaround time. Simply put, the task is done faster, with less participants which apparently agree more. However the agreement analysis shows a very interesting effect. **The 5 ¢condition seems to outperform all other conditions, including the 10 ¢one**. This result is against our intuition that if we increase the reward we also increase quality of the work, but completely in line with a large corpus of research done in the field of behavioural economics [Ariely et al., 2005]. This notion of sweet spot seems to replicate when it comes to the number of turkers assigned per task, with 7 being the optimal number of non-experts needed to replicate the agreement between the group of experts.

As for handling the outliers, the authors suggest two approaches:

- removing all the queries with outlier labels. This seems too extreme and involves losing all the good labels associated to a query which might have a minority of bad ones;
- remove only the labels assigned by the outliers. It's a less radical approach, in line with current recommendations for data pre-processing.

On the whole, [Feng et al., 2009] introduce a simple, yet useful methodology for parametrising crowdsourcing projects using AMT. Focusing on what they believe to be key aspects in such a project (data quality, cost, scale) may be a good approach to gradually identify all the factors influencing the success of a crowdsourcing data collection effort. Since their methodology was evaluated exclusively for query relevance assessment tasks, we plan to test if it applies to our controlled image annotation tasks.

Perhaps the best contribution of this paper lies in the cost and population analysis. The detection of the price and population "sweet spots" suggests that crowdsourcing projects do have optimal parameters. We would like to combine this knowledge with the approach suggested by [Hsueh et al., 2009] for detecting the right amount of financial motivation.

Snow, R., O'Connor, B., Jurafsky, D., Ng, A. (2008). Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks

In their study [Snow et al., 2008] try to put in perspective the advantages (cheap and fast) and disadvantages (unreliability) of crowdsourcing annotations for a number of different tasks, through task design. They also attempt to answer an over-aching question in the field: "how many non-expert annotators are there needed to reach the same quality of an expert?". Finally, their experiments have helped them draw preliminary conclusions on the behaviour of non-experts in different crowdsourcing settings.

Their experiments [5] are based on two very simple premises regarding task design – tasks descriptions should be succinct and followed by examples – and the size of the collected data – 10 annotations per item per task.

As previously stated, the most relevant contribution of their study is trying to answer the question every crowdsourcing project designer has in mind: **how many non-expert annotators are there needed to equate for the quality of an expert one**? To generalise their findings they propose a varied type of crowdsourceable tasks and compare their non-experts results with the existing literature on expert annotations.

- **Affective text-analysis**. Participants are asked to rate a news headline from 0 to 100 for each of the six emotions: anger, disgust, fear, joy, sadness and surprise, before giving an overall rating between -100 and 100. Experiments have shown that it takes on average 4 aggregated non-experts to obtain at least the same performance of an expert.
- **Word similarity**. Participants in this experiment were presented with different pairs of words and were asked to grade on a scale from 0 to 10 the similarity between the terms in a pair. This is a task which typically yields a very high inter-annotator agreement, as it has been replicated in the current study. Perhaps the surprising result was the very high speed of completion of this task. However, regarding the research question, it seems it takes around 10 non-experts to get to the performance of an expert annotator.
- **Recognising textual entailment**. Participants are presented with 2 sentences and asked whether the second sentence can be inferred from the first. The results show again that an aggregation of 10 non-experts achieve *almost* the same precision as an expert.
- **Event annotation**. In this task, each participant is presented with pairs of events and asked to judge which one occurs first in a temporal ordering. And again, the results show that 10 non-expert annotators are enough to equate for the performance of one expert.
- **Word sense disambiguation**. This is a task on which machine learning algorithms are known to perform very well. Participants have to choose what is the meaning of selected terms from phrases they are presented with. However, a comparison to the expert agreement could not be made since it was not reported in the initial study.

As can be seen, there's a recurring result suggesting that for most experiments around 10 non-expert annotators should have an aggregated performance equal to that of an expert.

This is even more interesting, if we take into account that the study of [Feng et al., 2009] suggests that there's a sweet spot of 7 non-expert labels per item, that realise an optimal

[5] The experiments are fully available at this URL: https://sites.google.com/site/nlpannotations/

balance between cost and performance. This makes us wonder of course, whether the number of non-experts needed to perform a task varies a lot with the task (no such study has been done, to our knowledge, for image annotation tasks) or stays around the values reported in literature for different other types of tasks.

Furthermore, the research of [Snow et al., 2008] gives us very interesting insights about the behaviour of non-expert annotators for the six given tasks:

- a small number of non-experts do most of the work. This comes to show that a certain, small number of participants in crowdsourcing projects are probably more motivated by the task, rather than by the financial incentive.
- a small few give very noisy annotations. Probably, there is a small number of annotators which have really bad expertise for accomplishing the task or are simply cheaters.
- some non-experts are focused, others are simply careless. This could also explain the previous conclusion. This conclusion seems a bit , it is difficult to assess the attention of an uncontrolled, fully online crowd.

Finally, taking into account their experiments, the authors point to a few guidelines for keeping data quality high, despite errors from non-experts:

- use more workers. This is in line with the general principles of crowdsourcing. A diverse crowd will give a diverse result which converges towards agreement;
- use compensation mechanisms to direct payments to those who work better (and not more, as was the case of AMT);
- model reliability and biases of individual workers and correct them.

The last point is one of the research lanes we are pursuing in the current thesis. In the following we will try to present a model which tracks both annotator and annotation quality through the entire crowdsourced image annotation process.

Grady, C. and Lease, M. (2010). Crowdsourcing Document Relevance Assessment with Mechanical Turk

[Grady and Lease, 2010] propose an experimental methodology for determining what are the relationships between the design of Human Intelligence Task (HIT) [6] and accuracy, dwell time and cost. The experiments are performed for a relevance assessment task.

The experimental variables used in the study are:

[6] HITs are just the way AMT names its tasks.

- **the query** – the authors make 2 hypothesis about the query to be assessed. First, they hypothesise that the longer the query description with respect to the query title, the longer the dwell time of a participant (she needs more time to read the description) and thus the bigger its involvement or dedication. The second hypothesis is that a very brief description usually leads to longer dwell time (the participants need more time to figure out what is asked of them) which could also be indicative of a very dedicated worker.

- **terminology** – the authors argue between 2 types of title terminologies: a technical one or a simple "yes/no" approach. It seems that specificity attracts workers more knowledgeable to a certain domain, whereas mentioning a research purpose can attract those people who are more animated by the idea of contributing to a collective effort.

- **payment** – the authors use 2 price conditions (1 ¢and 2 ¢[7]). They noticed that high rates seem to attract both a higher number of legitimate workers, but with them a higher number of spammers which introduce noise in the annotation. Referring back to the conclusion of [Feng et al., 2009] (reviewed in section 5.4) about the existence of a sweet-spot in terms of cost for query relevance tasks in AMT, as well as to the research in behavioural economics about optimal motivation [Ariely et al., 2005] this conclusion suggests that the starting rate is influencing the behaviour of the participants. Furthermore it shows that crowdsourcing projects don't have to go over their heads in trying to offer attractive financial incentives and that a big incentive can usually do more damage than expected.

- **bonus** – the authors use 2 conditions: no bonus and 2 ¢. In this setting bonuses are offered only after human evaluations of the qualitative explanations participants give over a task. This is particularly interesting, since alternative research suggests that qualitative input from users is associated with better better annotation quality. The authors seek to find whether a-priori knowledge about the existence of a bonus is affecting the quality of the workers and whether actually receiving a bonus is changing the performance of a user on future HITs.

The constants in this experiment design are:

- **description** – not to be confused with what the authors call "terminology" (which is an experimental variable). This simply refers to the steps that a participant has to take to fulfil a task (e.g. decide on the relevancy - click "yes" or "no" - click "submit");

- **keywords** – used to index tasks for search purposes. An interesting idea, currently unexplored is to see which are those keywords workers look for when searching for tasks;

- **task duration** – how long is a task available;

[7] AMT's anecdotal base rate

- **approval rate** – how many items per task were approved;
- **HIT approval** – it shouldn't be immediate; this will suggest that someone is actually taking time to assess the input from the participants;
- feedback to workers – proved to be useful in order to build a community around the project; shows involvement from the management side and may imply a good reaction from the participants.

The corpus was made of *XML* documents from the Text REtrieval Conference (TREC) TIPSTER [8] collection of news articles.

We conclude on the work of [Grady and Lease, 2010] by presenting some behavioural observations about the participants. First, a recurring observation – **there are some workers which complete significantly more tasks than others**. This exactly the same conclusion which [Snow et al., 2008] reached during their experiments and it is a perfect examples of one of those less obvious facts that we can still learn about crowdsourcing. An immediate research direction would be to study what is the exact distribution of task resolution by participants and see if this is sensitive to the type of task.

Considering those tasks with a high number of participants, according to the authors, this can mean one of two things: (the good) either the task was attractive, either (the bad) the incentives for the task were not enough to encourage only "reliable few" (which we previously talked about) to perform more tasks.

From a design standpoint, the use of query descriptions didn't seem to improve the accuracy. This conclusion seems a bit too bold and not so well formulated. We remind that in this experimental design, the query description refers simply at an optional display of the steps to be followed in order to resolve a task (e.g. Think - Choose - Submit). Since this is a very basic description which doesn't bring any additional information for a participant to solve a task, it was probably ignored by most respondents.

Further, an interesting effect suggests that as the number of workers decreases, the accuracy increases. Though counter intuitive to the current perception of crowdsourcing (that its power resides in large populations of participants) this conclusion seems to confirm that of [Feng et al., 2009] (reviewed in section 5.4) about the existence of a payment "sweet spot" after which any increase in the financial reward would not increase the performance of the annotators.

Finally, qualitative comments about the task seem to be indicative of the quality of a participant's work. We intend to use this finding in our proposal, as a proxy for measuring the quality of each annotator.

To conclude this review, a small note on **automatic** bonuses. While they seem to be cost efficient and stimulate workers to provide quality work, they could have the exact

[8] http://www.itl.nist.gov/iaui/894.02/related_projects/tipster/

opposite effect and encourage workers to simply input random comments. However, we believe that basic text analysis solutions could automatically detect and eliminate the cheaters.

Hsueh, P.Y., Melville, P. and Sindhvani, V. (2009). Data Quality from Crowdsourcing: A Study of Annotation Selection Criteria

[Hsueh et al., 2009] have conducted an empirical study evaluating, amongst others, the utility of annotation selection on classification accuracy and efficiency. Their study concerns sentiment analysis of political blog snippets and uses a sample of 3 expert annotators and 25 non-expert turkers.

To this end 2 types of annotations are collected:

- sentiment class annotation – annotators are asked what is the general feel of a snippet (the message it transmits about the subject: positive, negative, both, neutral);
- alignment annotation – annotators are asked to state whether the snippet is supporting or opposing the candidate in question. For the unclear snippets, annotators can tag them as neutral.

Subsequently the authors define 2 types of annotation quality measures:

- annotator-level noise – is defined as the deviation between the annotations received from a certain annotator and the gold standard, corrected by a chance agreement factor. Then using a threshold, the noisier annotators could be filtered out to improve the IAA;
- snippet-level sentiment ambiguity – the method they propose takes into account two dimensions: the strength of the annotator's judgements (otherwise put, the annotator's quality) and the polarity of the annotations.

The authors finally introduce confusion, as an integrated measure of both the the previous measures.

We are not getting into the details of the empirical evaluations here. Since they have been conducted against a gold standard they are of less importance to our review. What presents interest for us are the conclusions the authors draw at the end of their study. First of all, their results confirm the usefulness of having multiple noisy annotators as a substitute for experts. They then identify three dimensions to take into account when using non-expert annotators for labelling and classification tasks, the most relevant from a crowdsourcing standpoint being the annotator's noise level. Of course this means that the data quality can be improved by either eliminating noisy annotators or ambiguous annotations. The last conclusion should be carefully considered. It is strikingly similar to the one of [Feng et al., 2009] (Section 5.4) and implies the same discussion about how to

treat noisy annotators. While their complete exclusion might lead to the loss of a lot of data (some annotators may be bad only on some tasks), excluding only the noisy annotations might just be enough to keep a diversity of annotators without sacrificing too much of the quality of their aggregated effort.

Sheng, V., Provost, F., Ipeiritos, P. (2008). Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labellers

Considering the setting of supervised learning algorithms, [Sheng et al., 2008] investigate if, when and for which data points multiple training labels are needed and once acquired what to do with them. Although we are not necessarily interested in supervised learning algorithms, our project is nevertheless part a label collection initiative. Moreover, we find ourselves precisely in the repeated labelling setting described by the authors of the cited paper.

The premise of their work is that the performance of supervised learning models is dependant on both the quality of the data as well as the size of the data. We believe that this applies to crowdsourcing models as well and that the complex interplay of the 2 factors influences the success of such a project.

Repeated labelling is the process of collecting another label for the same data point. Repeated labels can alleviate the danger of having noisy annotators through aggregation of multiple labels.

Throughout this section, we will use the same notations as the authors, as following: x_i – features of a data point y_i – expected label for a data point y_{ij} – the label assigned to feature i, by annotator j p_j – quality of annotator j $Pr(y_{ij} = y_i)$

The **first assumption** considers label quality as being independent of the specific data point ($Pr(y_{ij} = y_i|x_i) = Pr(y_{ij} = y_i) = p_j$). What this means is that under the current assumption all data points (tasks) are considered equally difficult.

$\hat{y}_i$ – the majority voting result for votes y_{ij} $q_i = Pr(y_i = \hat{y}_i)$ – the integrated quality of an annotator

The authors then study two different scenarios, for the case of **majority voting**. First, considering a **uniform label quality** for all annotators ($p_j = p, \forall j$) it turns out (not unexpectedly) that the integrated quality q is larger than the individual label quality p, only when $p > 0.5$. This means that adding more (uniform) annotators does help. However, for values of $p < 0.5$, repeated labelling doesn't improve the integrated quality. Relaxing the last assumption, labels with **different label quality** are taken into account. For simplicity the authors consider the case of 3 annotators with qualities $p - d, p$ and $p + d$. The analysis shows that for certain values of d with respect to p (please refer to the dependency curve between parameters d and p in the cited article of [Sheng et al., 2008]) multiple annotators

do improve the integrated quality. The largest allowed variation is $d = 0.1$ for $p = 0.8$. For any other values of d and p using the single highest quality annotators is sufficient. This analysis is offering us a way of factoring variations in the quality of annotators before deciding whether the number of annotations is satisfactory for a quality result or we need more. However, the assumptions underlining the dependency between the 2 parameters should be remembered before generalizing this conclusion. In most crowdsourcing tasks we will probably use more than 3 annotators and the variation in quality between them will most likely be not as ideal as in this setting. Nevertheless, the initial study provided by the authors suggests a pattern which allows us for more variation as the quality of annotators increases up to a certain point. From that point beyond, it is intuitive that as we have better annotators we can only allow less qualitative variation. In the case of the perfect annotator ($p = 1$) no variation whatsoever should be allowed.

One drawback with the majority voting approach presented above is losing information about the uncertainty of labels. To address this [Sheng et al., 2008] introduce a third method – **uncertainty preserving labelling**. For each set of features, all assigned labels are preserved (thus becoming "soft labels"). This means that for each label y_{ij} a separate entry is created mapping the annotated features set x_i and $y_{ij}, \forall j$. Additionally, each of these mappings receives the same weight, equal to a fraction of the total number of occurrences of a certain label. Further, the authors suggests using these weights for different learning algorithms for automated labelling. As far as we are concerned, the challenge in a crowdsourcing scenario consists in the finding the right aggregation of annotations and outputting the best quality annotations. We believe that the uncertainty preserving labelling method that they present, though initial has potential to inspire our future work in terms of integrating both the annotators and labels quality.

Based on the conclusions above [Sheng et al., 2008] carry on with a cost analysis of repeated annotations. The problem is approached from three perspectives: i) how cheap can repeated labelling be, ii) for a given cost how much better is repeated labelling and iii) can we select specific data points to improve performance? The cost components considered by the authors are:

- C_U – the cost of getting a new unlabelled data point
- C_L – the cost of getting a new label (for an already labelled data point)

Thus, at any point a choice will be made between acquiring a new unlabelled example and labelling it ($C_U + C_L$), and labelling an already acquired example (C_L).

The two strategies explored are the generalised round robin and selective repeated labelling. The empirical evaluation is done for classification problems applied to datasets from the UCI repository for Machine Learning databases [9]. The selection strategy for **generalised round robin** is as simple as choosing the next example with the fewest labels. In the extreme case in which the labelling costs are considerably larger than those

[9] http://www.ics.uci.edu/~mlearn/MLRepository.html

of actually acquiring the unlabelled example ($C_U << C_L$) repeated labelling (with majority voting) performs better than single labelling. Apart from this case, a look over the general costs reveals that for small values of p (low quality annotators), as the ratio ($\rho = \frac{C_U}{C_L}$) increases (i.e. the labelling cost becomes smaller compared to the cost of acquiring new data points), the best performing strategy is the repeated labelling with preserving uncertainty. Alternatively, for a fixed value of ρ, an increase in the quality of annotators (p), either flavour of round robin repeated labelling (majority voting or label uncertainty preserving) perform equally well and outperform the single labelling.

The second strategy – **selective repeated labelling** – aims at finding those data points whose labelling will improve the integrated quality. Given a set of labels for a certain data point, the initial temptation would be to use some entropy based measure and ask for additional labels for those data points whose labels are not homogeneous enough. This could, however prove to be the wrong decision since the quality of labels should also be taken into account. Thus, a strategy is needed to avoid such cases as those when a data point with low quality labels is not deemed suitable for repeated labelling, just because his current set of labels is homogeneous enough. To address the authors compute a Bayesian estimate for the uncertainty in the class of the example (i.e. the uncertainty that the true label of an example is the outcome of the majority vote for the currently available labels). The reported results show that modelling labels uncertainty performs significantly better than general cost round robin repeated labelling (previously shown to give good results).

To synthesise, [Sheng et al., 2008] show that under certain assumptions, repeated labelling improves the quality of the labelled data. Additionally, when the annotators' quality is low preserving the labels uncertainty can improve the quality of the annotation.

Several suggestions of future work are particularly interesting for us, given the crowdsourcing setting in which we would like to annotate our comics corpus. First, having good estimates on the quality of individual annotators, based on the labels they've already assigned might improve the repeated labelling process. Second, adapting the incentive (as reflected by the C_L and C_U costs) such that annotators are rewarded according to the level of noise they introduce in the data [Morrison and Cohen, 2005]. Third, there's the obvious need of relaxing the initial assumption that all tasks are equally difficult. This is clearly not the case when using non-expert annotations, who might have trouble with an inherently difficult task or due to their lack of expertise with some tasks. Another assumption which should be relaxed is that of considering the true label. In almost all "real life" scenarios, we are dealing with **noisy annotators** over **uncertain labels**. Finally, the best annotation schemes could be hybrid ones, which according to the context (i.e. a cost vs. accuracy evaluation) will decide for either the single annotators strategy or a repeated labelling strategy.

Donmez, P., Carbonell G., J. and Schneider, J. (2009). Efficiently Learning the Accuracy of Labelling Sources for Selective Sampling

The work of [Donmez et al., 2009] is a critique to the model proposed by [Sheng et al., 2008] and reviewed in the the previous subsection. While appreciative of the analytical quality of the latter study, they underline several reasons for its practical inapplicability:

- high dependency on the different cost models – the practical value of repeated labelling is highly dependent on different cost models (i.e. when the cost of pre-processing data is non-negligible) and varies with the quality of the annotator
- over-simplifying assumptions – as we've also noted in the previous subsection there is simply no practical use for the assumption considering equal qualities for all annotators. Simply put, that means that all the annotators in the system have the same probability of making an annotation error on all tasks.

In real data collection settings not only that different annotators will perform differently on different tasks, but we won't have any a-priori information about their quality. As our approach will suggest, we will have to infer it and adjust it along the way.

Back to [Donmez et al., 2009], they propose a solution for selecting the most accurate annotators to provide with repeated labels for those data points where this is required. This solution is based on the Interval Estimation algorithm [Moore and Schneider, 1995] – at each point the action yielding the maximal result is selected. In this case, the sought action is choosing the best annotators, whereas the measure for doing so is provided by her performance on previous tasks. Also, in the absence of having a-priori information about the right label, the latter needs to be estimated. The authors argue that majority voting provides with a good enough approximation of the expected label. And since bad annotators will cause the algorithm to converge slower, a threshold technique (IEThresh) is used to filter out those annotators with inferior qualities.

Eight benchmark data sets are chosen for the experimental evaluation, six of which are classification data sets. All of them were either originally binary or were converted to binary values. *IEThresh* is compared against 2 other methods: i) all annotators are asked to provide with a new label for a given data point (the *Repeated* condition) and ii) random annotators are asked to provide with a new label (the *Random* condition).

Results show that though a slow starter, *IEThresh* outperforms both *Random* and *Repeated* after approximately 200 queries. An interesting observation is that as more queries are performed, IEThresh filters out the unreliable annotators and selects more and more the reliable ones, until eventually the algorithm converges. Even more remarkable is that the algorithm converges even for more heterogeneous distributions of annotators quality. The authors explain this by comparing the results with another flavour of the Interval Estimation algorithm, one which takes into account only the mean rewards of the sample. This variant performs worse even than the baseline algorithms, suggesting that

using variance, in addition to the mean of the sample improves the exploration capability. This is desired especially in those cases where there are many unreliable annotators and the few reliable ones need to be fast discovered.

The authors conclude with potential improvements to their proposal, some of which we will likely include in our work. Insightful observations regarding the behaviour of annotators could be made by tracking the variation with time of their annotations. The authors suggest that such observations could be associated to other factors influencing the annotation process, such as fatigue or attention [Snow et al., 2008]. Then, there is the issue of task difficulty inherited from the critiqued work of [Sheng et al., 2008]. This is in fact one of the hypothesis of [Sheng et al., 2008] which they fail to relax. As previously discussed, it is fairly unrealistic to consider that all tasks are equally difficult for all annotators. Also, as noted earlier, both task difficulty and annotator's familiarity with the topic are amongst the factors influencing the quality of the annotations. Last, since every data collection initiative incurs certain costs, the authors suggest the inclusion of a cost component with their model. In its most simple form this should take into account the accuracy of the annotator and the difficulty of the task; but we will discuss more about this in the sections to come.

Nowak, S. and Lukashevich, H. (2009). Multilabel Classification Evaluation using Ontology Information

Citing the difficulty in evaluating multilabel classification using ontology information [Nowak and Lukashevich, 2009] introduce a novel evaluation measure, ontology score (OS). In fact, the reviewed research introduces the measurement tool for the extended study by [Nowak and Rüger, 2010] about the reliability of crowdsourced annotations for images.

Based on the premises of the hierarchical organisation of concepts in ontologies, the authors suggest the use of the same principle for the evaluation stage. Upon review of the literature concerning hierarchical measures, the authors give several suggestions as to what a good measure of this kind might be. The basic premise is that concepts which are closer in a hierarchy are judged as being more similar than concepts which are further apart. Inspired by their application to uni-label annotations, the authors stress the importance of two types of hierarchical measures:

- depth-independent distance based misclassification cost – in this case each relationship in the hierarchy has the same weight
- depth-**dependent** distance based misclassification cost – in this case, an classification mistake at a lower level of the hierarchy is less penalised than one at a higher level

At this point it is worth noting that in another related work, [Blockheel et al., 2002] propose an evaluation measure for hierarchical **multi-classification**, by extrapolating the distances between individual labels to distances between sets of labels. Finally [Cesa-Bianchi et al., 2006] propose a hierarchical loss function; the first misclassified in-

stance on a path adds up to this function, while all children of that node are not explored any more.

Back to the reviewed work of [Nowak and Lukashevich, 2009], they identify a series of problems that hierarchical evaluation measures should take into account. According to them, none of the previously presented measures comply to all of them:

1. a measure should consider partial matches;
2. establishing the ground-truth set is not an easy task – the authors suggest the empirical computation of the agreement rate for each concept and using it as a factor for the cost function;
3. take into account the types of relationships in the underlying ontology (i.e. disjoint relationships, pre-conditions etc.).

Nowak, S. and Rüger S. (2010). How reliable are annotations via crowdsourcing? A study about inter-annotator agreement for multi-label image annotation

One of the most extensive studies on the applicability of crowdsourcing since the wide spread of the term is the work of [Nowak and Rüger, 2010]. It bears additional relevancy for us since it makes extensive use of work done on i) multiple and repeated labelling, ii) inter-annotator agreement measures, and iii) creating non-expert based gold standards and seeks to apply it to a topic closer to ours – image based annotations.

The experiment is somehow similar to the ImageCLEF 2009 Photo Annotation competition [Nowak and Dunker, 2010] during which a subset of 18000 images from the MIR Flickr dataset [Huiskes and Lew, 2008] were annotated with 53 concepts by 19 team expert teams in 74 configuration runs.

The case of [Nowak and Rüger, 2010] is virtually a downsized repetition of the experiment, during which 99 images from the same corpus were annotated by a total of 11 experts and a crowd of AMT non-experts, also referred to as turkers.

Annotations were provided at a per-image basis. Annotators were asked to use the 53 concepts previously identified during ImageCLEF 2009, which were a-priori categorised as following:

- mutually exclusive concepts (e.g. for the category *Place* one could choose exactly one of *Indoor*, *Outdoor* or *No Visual Place*)
- optional concepts – one or more concepts could be chosen for categories such as *Landscape Elements*

The **annotation tools** were a customised desktop application with an easy to use GUI, to be used by the experts and HIT deployed on AMT, for the non-experts. The latter was designed in the form of questionnaires with a GUI as close as possible to that of the desktop application. Explanations and examples accompanied each concept (a good way of insuring both help and continuous training for annotators). To keep it simple though, instructions were more simple for non-expert annotators. Finally, the image was continuously displayed for all annotators throughout the completion of the task. This design choice was most likely taken to avoid that the annotators loose focus on their task by switching between the image environment and the task user interface.

Moving to the experimental design, let us take a look at a summary of the **measures** used in the evaluation process. First of all, in image based annotations one can always assess accuracy either between sets of multiple images or between sets of annotations on a per-image basis.

For the purpose of their evaluations, the authors make use of three classes of measures:

- image annotation evaluation measures: ontology score (OS), equal error rate (EER) and area under the curve (AUC);
- rank correlation measures: Kendall's τ and Kolmogorov-Smirnoff's D;
- the Kappa statistics.

From the first category, OS [Nowak and Lukashevich, 2009] makes use of an ontology of concepts on which it measures partial matches between a ground truth and the current output. The score itself is based on i) structural information (the distance between concepts in the domain ontology) ii) relationships between concepts in the ontology, and iii) the per-concept agreement between annotators. More details about this measure will be give in the review of [Nowak and Lukashevich, 2009] in the following subsection.

EER and the AUC are both based on the receiver operator curves (ROC). EER is the point on the ROC in which the false acceptance rate is equal to the false rejection rate (i.e. false positives are equal to false negatives), while AUC is the summing up of the area under the ROC. In the given setting they were both computed on a per-concept basis and then averaged.

Kendall's τ [Kendall, 1938] is a statistical measure used to compare two given rankings. It takes into account the number of swaps needed to turn one ranking into the other. Values range between -1 (opposite rankings) and +1 (identical rankings), with 0 being the correlation between random rankings. *Kolmogorov-Smirnoff's D* [Kolmogorov, 1933] assumes that two rankings are similar and is sensitive to the degree of disorder between rankings.

The experimental design of [Nowak and Rüger, 2010] is supporting four types of analyses:

1. analysis of agreement amongst experts. Each expert annotator labels the same set of images and accuracy is computed with each of the annotators as ground-truth. Additionally the "merged annotator" is computed using the majority voting rule over all annotators.

2. influence of different sets of expert annotators on ranking the performance of the system. Each expert is considered as a ground truth and all the others are compared against her. The systems are then ranked according to different scores: OS, EER or AUC. Results in this experiment show a high level of agreement between experts. In what concerns the purpose of the experiment, there seems to be no significant difference in terms of system ranking as to which expert we choose as ground truth.

3. analysis of the agreement between experts and non-experts. Since in this case not every image has the same number of annotators, correlation cannot be measured. Therefore inter-annotator agreement (IAA), based on Kappa statistics is computed for labels either at concept-level or image-level. Finally the comparison is made between the "merged expert" annotator and the "merged non-expert" annotator. Results show that the "merged non-expert" and the "the merged expert" perform similarly in terms of accuracy. We remind however that accuracy doesn't account for chance agreement, therefore a better picture might be given by the Kappa statistics. According to these some concepts exhibit significant IAA differences between experts and non-experts. It also seems that the majority voting rule seems to be enough to filter out most of the noise of the non-experts, for a large number of concepts, according to the same Kappa statistics.

4. influence of average expert annotators compared to average non-expert annotators on system rankings. In this experiment, the "merged non-expert" annotator is used as ground-truth. The expert annotators are then compared and ranked against it, according to different scores: OS, EER or AUC.

Their experiments of [Nowak and Rüger, 2010] both confirm some expected results and reveal some interesting findings, all of them supporting the utility of non-expert based crowdsourced data collection initiatives. First, they confirm that experts generally show high consistency in their annotations as well as good agreement, according to the K-statistics. It is therefore without surprise that there seems to be no apparent difference between the performance of systems considering different experts as ground-truth. This means that repeated expert annotation on the data is not necessary as long as rules are clearly defined from the beginning. The authors recommend however that IAA is verified on a small sample of data as it can fluctuate according to the concepts (probably as an influence of the difficulty of annotating some concepts).

Considering the non-experts, they perform equally well in terms of accuracy. However, since this measure doesn't take into account chance agreement the study of Kappa statistics seems more suitable. The latter thus reveals lower (but still acceptable) rates of IAA between non-experts.

Despite the thoroughness of the reviewed work several open questions still remain, constituting promising lines of research some of which we will tackle in our work. First, the question is raised with respect to the sensitivity of the existing evaluation measures to the particularities of image based annotation processes. This is coupled with a natural question whether the encouraging findings by [Nowak and Rüger, 2010] are scaling to bigger data sets. Last, a question which has already received a great deal of attention: how many non-experts are there needed to account for the quality of an expert.

Aroyo, L., Welty, C. (2013). Crowd Truth: Harnessing disagreement in crowdsourcing a relation extraction gold standard

This paper challenges the classical approach of using ground truth for web data analytics and instead employing a new, richer and more diverse type of ground truth, the *crowd truth*. Thus, the authors of this study start from the simple, yet radical premise that *annotator disagreement is not noise, but signal*. They choose to interpret it as a source of information about the difficulty, ambiguity or some other property of the concerned task.

Classic annotation approaches focus on achieving a high inter-annotator agreement (IAA) and to this end restrict annotators to follow only one precise annotation. As soon as the corpus presents ambiguity, this form of controlled annotation fails and leads to disagreement. The authors worry that while focusing to achieve high IAA, task design might have to suffer up to the point where it can become artificial, or biased, we might add. The main claim of [Aroyo and Welty, 2013] is that in crowdsourcing ground truth data, there is no one correct answer and several interpretations for a data point might apply.

The experiment took over on the *CrowdFlower* platform and participants were presented with sentences from the medical domain, in which two entities were marked and were asked to choose all relationships between them, from a given set. Two additional relationship options were added: none and other. Contributions were stored as binary vectors of dimension $n + 2$, where n is the total number of relations. A sentence vector is then obtained by summing all vectors corresponding to a contribution. Cosine similarity measures are then used to study the quality of a certain contributor, the clarity of a certain sentence and the relationships between sentences.

The work of [Aroyo and Welty, 2013] presents particular interest to our work. In our approach, we choose to annotate potentially ambiguous information such as event descriptions. Contributions of various participants which are semantically close, might be expressed differently in terms of content (i.e. different words), causing IAA measures not to take into account this kind of agreement. Just like in the work presented here, we wish to harness apparent disagreement and interpret it as diversity of language/opinion, or in some cases, as a measure of task difficulty.

Chapter 6.

Comics++: a new system for crowdsourcing comics annotations

"No matter who you are, most of the smartest people work for someone else."

— Bill Joy, co-founder of Sun Microsystems

Part II of this thesis explored the landscape of standards and formats for print and digital publications, with an emphasis on the role of *EPUB* as an open format for packaging, rendering and distributing digital publications. In Chapter 4 we presented in detail the evolutions of the format and our thoughts about possible enhancements.

We expect *EPUB* to deliver better content and for that the standard needs a stronger semantic layer to back its rendition-oriented capabilities. To this end, we have already identified a rich, flexible schema – *CBML* – which through its *TEI* inheritance has already proven its usefulness, albeit largely in academic research. Nevertheless, *CBML*'s richness comes at a price and that is the difficulty of fully populating such a file, considering the diversity and complexity that comics exhibit in both structure and content. Therefore, it is important that we clearly understand our data requirements, based on which we design a suitable data collection strategy.

In this chapter we introduce **Comics++** – a crowdsourcing platform for annotating comics. We begin by understanding the data requirements, the "crowd" we wish to engage, and then proceed to the practical aspects of developing and implementing our crowdsourcing platform for citizen science. We argue our technical choices and take our reader through the development and deployment details, discussing the challenges we faced and how we surmounted them. We hope that the present chapter will serve as a useful guide to all those who will dedicate their time, knowledge and energy to similar undertakings.

6.1. Data requirements for the description of comics

So what do we wish to extract from a comic book? At the time this chapter is being edited and to the best of our knowledge, no similar approaches to encoding data related to sequential art are in place. We have thus expanded the range of our search to include projects dedicated to all kinds of publications and one of them stood out in particular.

Small Demons

Small Demons launched in 2011 and set to collect, catalogue and index details from both fiction and non-fiction books. Valla Vakili – CEO and co-founder of Small Demons explains:

> "The problem was that the wine, the music, the streets of foreign cities found inside of books generally remained just there - on the page. I wanted them accessible to me, gathered together in one place where I could see everything meaningful inside a book, and everywhere that could take me. We founded Small Demons to do just that. To look through the pages of everyone's favourite authors, for all the people, places and things worth exploring further. For every book, we create a visual index that puts all those details at anyone's fingertips."

At its core, the platform had an analysis engine that mined text in digital books in search for people, places, organisations, but also things (i.e., drinks, food, works of art etc.), or mentions of other books. Interestingly enough, instead of public domain books, the platform stroke important partnerships and indexed more than 5,000 titles from catalogues of world renowned publishers (Random House, Simon & Schuster, Hachette, Harper Collins and Penguin). The final product? A genuine universe of everything one would find in books, with multiple cross-overs between books, real life and fiction. Or, as Valla Vakili explains, the so-called *"storyverse"* created by Small Demons:

> "So now you can see, in one place, the music in Nick Hornbys *High Fidelity*, the tech and gadgets in *Steve Jobs*, the places from *The Girl With the Dragon Tattoo*, historical figures in *John Adams*, the comedians, actors and actresses in Steve Martins *Born Standing Up*."

Unfortunately, due to lack of funding, the company ceased to exist in late 2015 and their discovery platform is offline ever since. Despite this, the idea not only that inspired us at the time, but also provided us with a useful typology (see Figure 6.1 below) for the data we could extract to build our *"comic-verse"*.

Comics-related data

To better understand our data requirements, we turn back to Chapter 1 and the motivations behind our project. At that time, we had identified three main areas that could largely benefit from our work – digital publications, digital humanities research and the GLAM sector – each of which come equipped not only with their own use cases, but also with

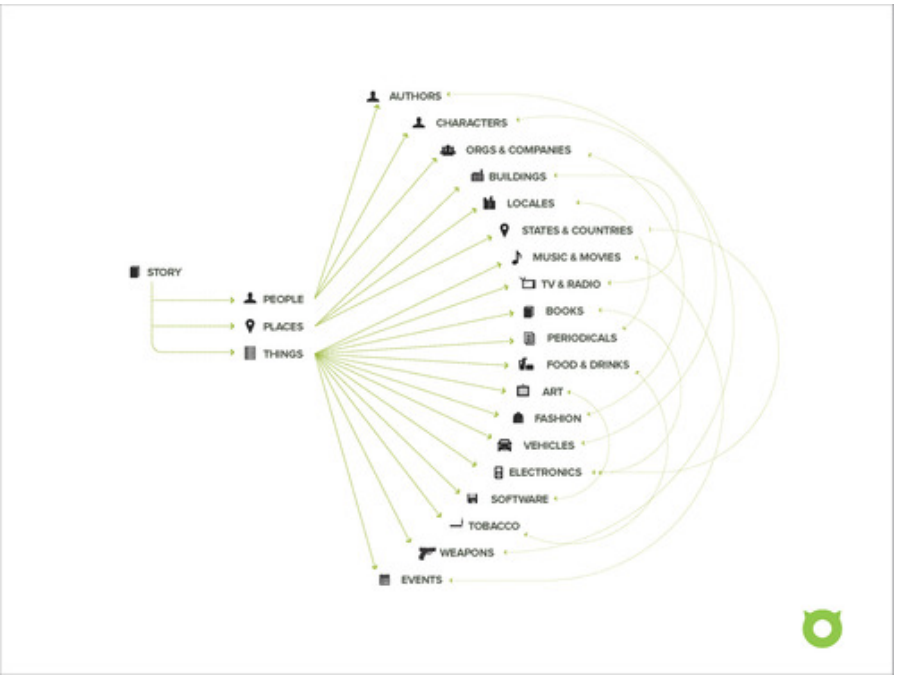


Figure 6.1.: The elements composing Small Demons' "story-verse". Participants on our platform will create a "comic-verse" by finding at least a subset of these elements: people, places, events and as many relevant objects as possible.

the standards and formats that treat these use cases, and that have been discussed in Chapter 3. In Section 3.2.2 we argued in favour of *CBML* as the suitable file format for encoding comics, and given our desire to push it as a complementary solution to publishing standards, it seemed obvious that this format should also be one of the basic constructs of our data requirements. The format's core encoding capabilities revolve around three main elements: `cbml:panel`, `cbml:balloon` and `cbml:caption`, allowing for the easy encoding of these particular structures. Moreover, as we recall from *CBML*'s rich schema, each of these elements is equipped with a variety of attributes allowing for the detailed encoding of other comic book elements: characters, places, events, onomatopoeia, transition types between panels etc., themselves aligned with key concepts from the comic book theory that we discussed in Section 2.2.1.

Bibliographic metadata

Finally, we investigated whether library and bibliographic standards can go beyond descriptive or provenance metadata schemas (see Section 3.2.3) and can provide us with additional clues about the kind of data to collect. We thus identified potential solutions within *FRBR*, the library-based, user-oriented specification, and particularly its Group 2 (Person, Corporate Body) and Group 3 Entities (Object, Event, Place)

Synthesising from these three sources (content enriching projects, comics-specific data and bibliographic metadata), we give below the primary data[1] that we set off to collect:

- structural data:
  - panels (shape, border style, groups of panels);
  - page type (i.e., whether a page contains a double-spread or whether it is a title splash page);
- data about the graphic content:
  - characters (names and depicted emotions);
  - places (named and unnamed);
  - events;
  - objects (with a focus on the recurring ones or those perceived as important);
  - visual cues for the non-visual (e.g., movement lines, lines depicting smell etc.);
- textual content data:
  - balloons' transcriptions;
  - onomatopoeia (textual transcription and interpretation).

6.2. *Comics++*: Crowdsourcing system overview

Given the richness and the variety of this data and the typology of the supporting tasks, it is clear that we cannot rely on the usual data collection approaches – in-house experts or trained in-house annotators. We rule out these approaches due to their known limitations in high scale data collection efforts: experts are hard to train, expensive and their use doesn't completely rule out agreement problems, when working on tasks with a certain degree of interpretability. Moreover, these groups tend to be rather small and homogeneous, translating into high turnaround time, difficult to accept in the case of projects involving large corpora, such as ours is supposed to. Thus, when in need of fast, inexpensive and diverse data, tapping into "the wisdom of crowds" is the the right choice.

Taking into account its novel, interdisciplinary, and largely unexplored nature, there is still much debate on how to define crowdsourcing and how the different crowd-based initiatives fall within one of the existing typologies. Crowdsourcing itself is too "young"

[1] We call it primary data, since this data is collected directly from participants, as opposed to secondary data, which is collected by tracking their interaction with the platform (i.e., task accomplishment time, number of contributions etc.).

and research about it is still inconclusive with scholars still working on understanding its underlying formal model. In the context of our work, we consider crowdsourcing to be nothing more than a data collection methodology, one that we choose for the advantages discussed before. Subsequently, our efforts focus on the design of a data collection platform suitable for the desired data, as well as on the evaluation and aggregation of the crowdsourced data.

Figure 6.2 gives an overview of our crowdsourcing system for annotating comics. The diagram focuses on the composing processing engines (annotation collection, annotation aggregation), the different databases (annotations, comics library) and storage units (*CBML*, *EPUB*), as well as the users (readers, annotators) and other concerned stakeholders (Digital Humanities scholars, museums, digital comics publishers).

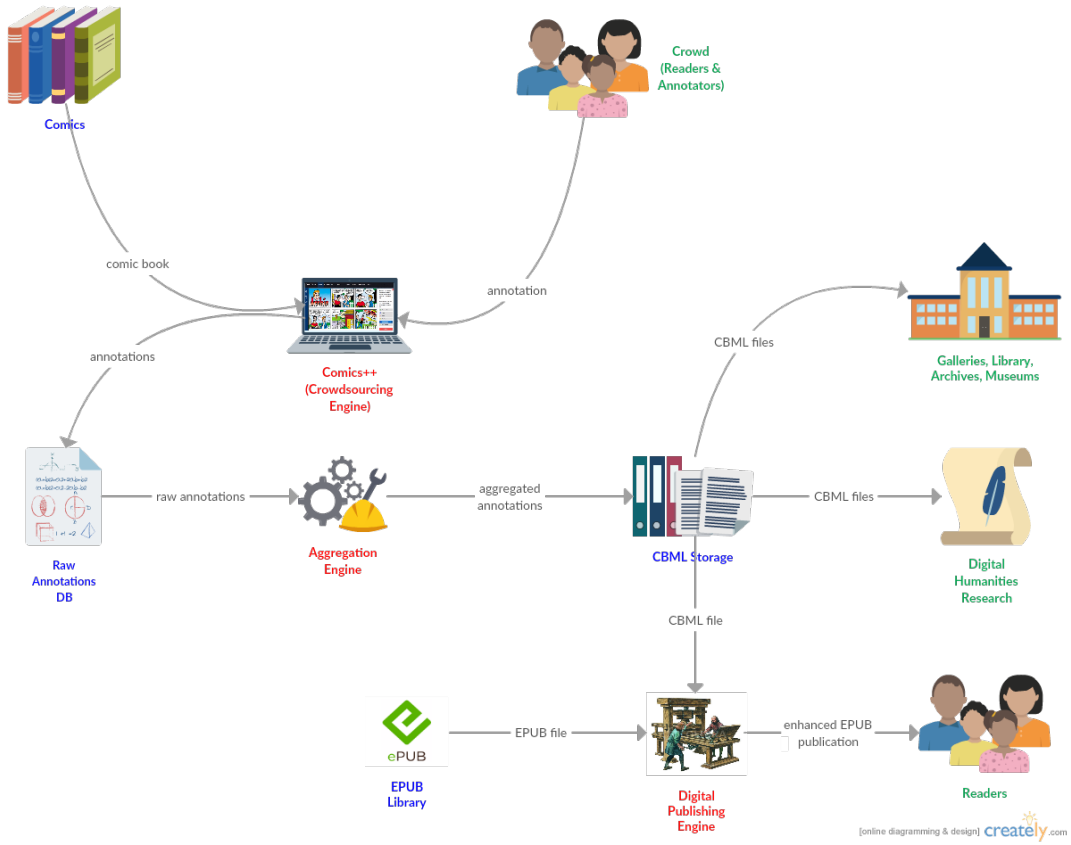


Figure 6.2.: Animated data flow diagram, giving an overview of the crowdsourcing annotation system. It highlights its main components: the data processing engines (in red), the data storage units (in blue), and the users and stakeholders (in green). See Figure 6.4 for a more detailed view of the *Comics++* crowdsourcing platform.

The design for our crowdsourcing platforms is first and foremost based on the principles formulated by [Snow et al., 2008] and reiterated by [Alonso and Mizzaro, 2012]. Both of these studies insist on the importance of the task description being as concise as possible

and followed up by illustrative examples of what is expected from the annotator. Training the annotators either through dummy tasks or accompanying examples is acknowledged by [Hirth et al., 2013] as an important factor in decreasing some of the cost components of the annotation effort.

We obviously sought inspiration in similar participative science platforms, such as The Public Catalogue Foundation's Tagger [2] (tagging names of people, places or events in paintings), Zooniverse's Measuring the ANZACs [3] (transcribing military records from World War I) or Old Weather [4] (transcribing meteorological data from ship logs). These pioneer projects were tremendously helpful during the prototyping stage, providing us with valuable good practises and examples of bad designs.

Based on design recommendations by [Alonso and Mizzaro, 2012], we opted for the separation of our tasks into two large types: marking and transcribing. In **marking tasks** participants are asked to simply point and click at comic book elements that they identify on the screen. This allows us to reduce the nuisance of asking a distinct question for each of the many elements we were interested in. The tasks are mechanically simple and require the participants to read the page (as they would normally would in an off-platform setting) and then simply point and click in the region of the page where a given element of the comic medium is found. A typical example would have a reader identify Batman jumping from a building, and simply placing a mark in the approximate area where this event and this character are occurring on the page. Other marking tasks require participants to answer single or multiple choice questions about the structure of each page.

In **transcribe tasks**, we display previously acquired marks and ask our participants to simply transcribe them. In the case of the example before, our annotators are expected to simply input the name of "Batman" or, in case the marked element is the event, simply state "Batman jumping from building". There is an additional role which transcription tasks are playing, especially those that could get a bit more tedious (transcribing the narrative associated to a character, for instance): [Grady and Lease, 2010] and [Kazai and Milic-Frayling, 2009] noticed that participants' willingness to contribute with qualitative assessments (e.g., answering open questions) is related to their intrinsic motivation, translating into a lesser likelihood of them contributing with intentional noise. In other words, if an annotator who gave a "wrong" answer has also taken the time to contribute to a qualitative task, then this annotator is less susceptible of being a "bad annotator" overall.

By separating our tasks, we seek to engage our participants in different ways, both in terms of intellectual and mechanical processes required for them to complete. We expect them to choose the tasks that they feel more comfortable with, thus hoping for a longer engagement with our platform (during the current session, but also as returning users).

[2] http://artuk.org/participate/tag-artworks
[3] https://www.measuringtheanzacs.org/
[4] https://www.oldweather.org/

According to the same design recommendations of [Alonso and Mizzaro, 2012], each of the tasks are deployed together with an accompanying help module. The design principle in our case is to keep these help sections as short and informative as possible, equipping them with clear examples; in this way we hope to achieve two important objectives: first, to train our participants to the tasks at hand and second, to offer them an educational incentive, that of finding out more about the interesting world of comic books.

We will continue this discussion in Section 6.4.2, where we will go into the implementation details of each of the tasks, their related help content and their composition into workflows.

6.3. *Comics++*: Choosing the "crowd"

Building the "crowd" is essential to the success of a crowdsourcing campaign. Obviously one cannot spell "crowdsourcing" without "crowd", but this blatant reality can turn into one of the most painful experiences when rolling out a crowdsourcing platform. The literature often reports about the advantages of using "crowds" (large, heterogeneous, inexpensive), but in reality these "crowds" are easily engaged almost exclusively by platforms like AMT, which are tempting due to their micro-payments mechanisms. The bulk of existing projects and platforms, especially those promoting citizen science attract different participants, most of them having primary motivations of a less material nature. A recurring theme in the body of work we reviewed in Chapter 5 mention the difficulty of predicting the crowd, when designing the infrastructure of a crowdsourcing system. Most of the time, these platforms are developed based on the data needs of the requester, who relies on task and incentive design and a subsequent media campaign to attract and retain participants. Sometimes, the infrastructure can prove to be insufficient (such was the case of *GalaxyZoo*, which was overwhelmed by contributions in the first days after launching), whereas sometimes infrastructure might be insufficient – as in the case of the *British Library Georeferencer* which had its 750 maps transcribed in only four days, rendering the rest of the projected interface completely useless. There seem to be two important aspects to look at, when designing for a "crowd": its motivation and the cost of "employing" it. Consequently, two questions are typically asked: what is the target "crowd" and how do we engage with it?

Much like in other citizen science projects we rely on a crowd of passionate participants, in our case about reading and collecting comics. A particularity that sets our "crowd" apart is the enthusiasm of comic book readers, fans and collectors. In Section 2.2.2 we discussed how comics can engage their public in ways that go beyond the content itself, we have seen the role of fan mail in the comic books industry and the fans' level of engagement through alternative content publications (*fan-zines*) or social events reaching unprecedented popularity (the *comic-cons*). These people are our primary target "crowd" and because of their expertise and passion for comics, we expect them to deliver in terms of both quantity and quality.

Second, we target curious people who get motivated by the idea of contributing to a research project. [Dunn et al., 2012] report on how the possibility of actively contributing to ongoing research, being part of a community or simply discovering something new rank particularly high on a list of non-financial incentives in projects engaging the crowds with humanities. It is one of the reasons for which our platform also acts as an educator for those who are not passionate about comics... or at least not yet. As we have previously stated, our help sections and tutorials aim at being as informative, yet concentrated as possible. We do not wish to bore our target public, nor do we want them to interpret what they are doing as "work"; instead we expect them to engage with the platform for as long as they find it "fun" (another motivator cited by the report of [Dunn et al., 2012]), whether this comes from showcasing their knowledge or learning something new.

In this context and without relying on an already existing user base, such as that of the AMT, most projects find it difficult to reach to their desired participants, a problem encountered even by better known public participation platforms, such as Crowdflower[5] or Crowdcrafting[6]. GalaxyZoo[7] (Zooniverse's initial project) was somewhat of an exception attracting more than the initially desired "crowd", whereas other projects, such as Next Stop Design [8] or Letters of 1916 [9] benefited greatly from both traditional and new media campaigns.

Currently lacking any possibility to advertise our projects through traditional media campaigns, we focused our efforts into continuously building our online "crowd" with the help of new media. We launched calls for participation through the Twitter platform and reached towards potentially interested communities (crowdsourcing professionals, comics scholars, fan sites etc.). Alternatively, we broadcast an email campaign targeting primarily our closest research community (members of the LIP6, DH researchers at Sorbonne and partners from related projects). In the future, we hope that the potential involvement of GLAM actors, such as the CIBD Angoulême, will draw awareness amongst visitors of these institutions.

Finally, we hope that upon successful resolution of currently sensitive copyright issues, future iterations of our platform will be able to involve our industrial partner, online comics library Sequencity[10]. This would allow us to draw a part of their comics passionate crowd onto our platform and give us the opportunity to offer them material incentives, such as unlocking an issue of a comic book after a certain number of valid contributions. Despite it not being a fully financial reward, it would be interesting to see if the payment sweet spot effect noticed by [Feng et al., 2009] and [Grady and Lease, 2010] applies to lesser monetised incentives and more specialised crowds.

[5] https://www.crowdflower.com/
[6] https://crowdcrafting.org/
[7] https://www.galaxyzoo.org/
[8] http://nextstopdesign.com
[9] http://www.letters1916.ie
[10] https://www.sequencity.com/

6.4. *Comics++*: System development and implementation

We dedicated our resources to the development of three crowdsourcing platforms for collecting annotations:

Comics++ – a platform[11] for reading and annotating structural and content elements in comics and other forms of sequential art.

(Dis)Similitudes – a platform for discovering and annotating comparisons and similes in prose poetry from the 19th and early 20th centuries. Its two versions boast an array of English[12] and French[13] authors respectively, and are meant to support one of our Digital Humanities research projects aiming at the automatic identification of such structures [M'Pouli Njanga Seh, 2016];

The following deals primarily with *Comics++*, since this particular platform sustains the main objective of our work – the characterisation of sequential art to support the evolution of digital publication formats for comics. Nevertheless, the same principles were largely used in the development and deployment of our other platform, *(Dis)Similitudes*.

Like most crowdsourcing systems, *Comics++* displays a number of mandatory features:

1. the database: for storing the subjects, the annotators and their classifications (see Figure 6.3)
2. the back-end: the logic of the crowdsourcing engine implementing methods and rules for serving subjects to the users. These can differ according to the purpose of the project and the constraints imposed on the collection effort. For *Comics++*, these constraints are:
   - an annotator should not classify a subject more than one time;
   - when working through the annotation steps of a subject, an annotator cannot return to a previous step;
   - a subject is withdrawn after a certain number of annotations has been provided;
3. the front-end: provides with a simple, user-friendly interface for presenting each annotator with the subject to annotate; additionally, the user interface should provide with non-intrusive help sections, on-demand tutorials and separate sections dedicated to the description of the project;

[11] http://dissimilitudes.lip6.fr:8182
[12] http://dissimilitudes.lip6.fr:8180
[13] http://dissimilitudes.lip6.fr:8181

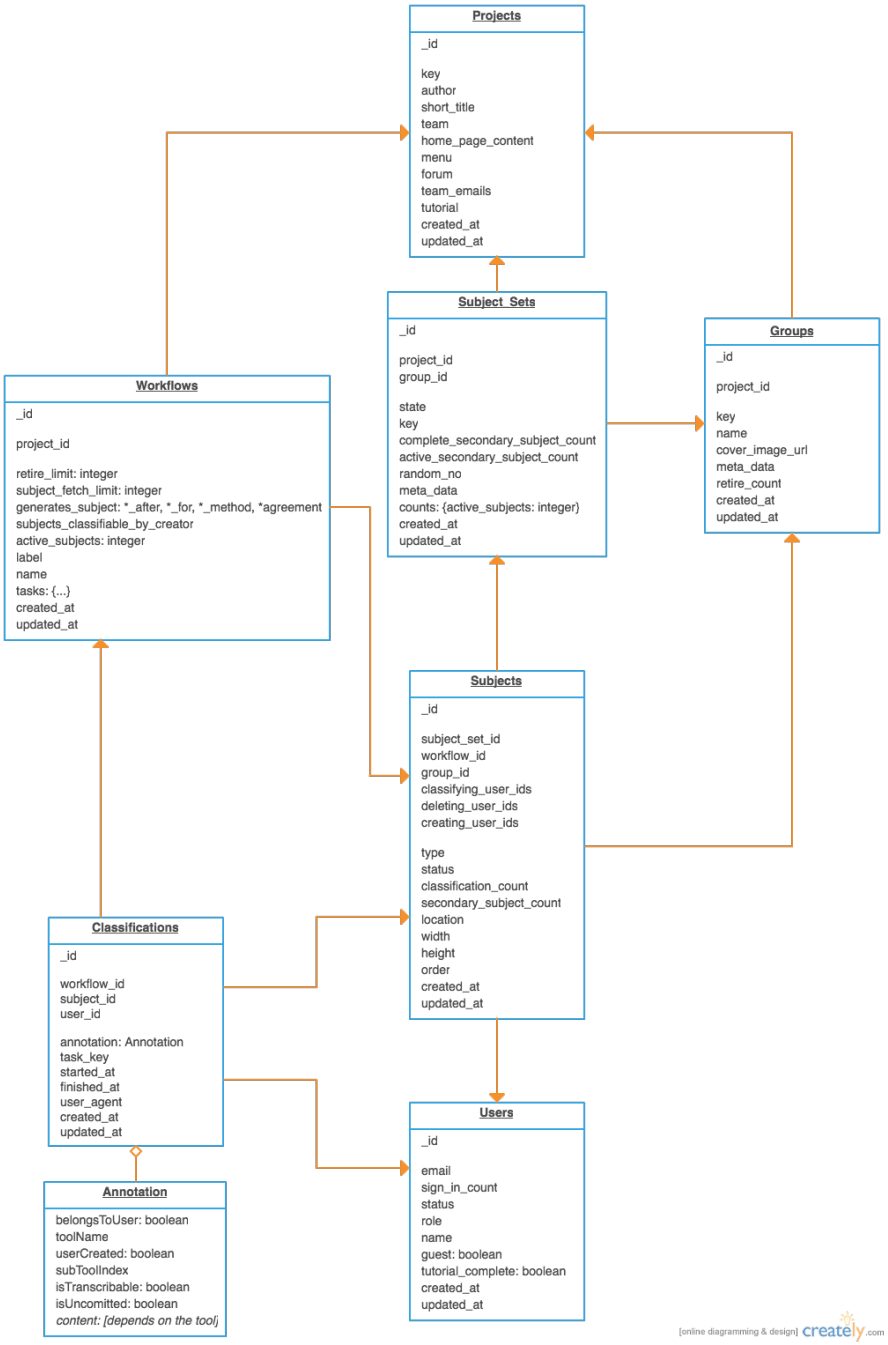


Figure 6.3.: UML diagram of the *scribe* database, illustrating the collections used in a typical *scribe* crowdsourcing platform.

4. a discussion forum: offers participants the possibility to engage with both the developers (in case any problems are detected) and their fellow participants, thus building a sense of community around the project.

6.4.1. Technical choices

At the time we started the implementation phase for *Comics++*, AMT represented more or less the norm for crowdsourcing platforms. However, its overly-simplistic user interface and its predominantly financially-motivated user base, were strong enough reasons for ruling it out as a technical solution for our data collection effort.

Alternatively, investigations about similar crowdsourcing platforms, in particular the report by [Brabham, 2013] about the use of crowdsourcing in government and the scoping study of [Dunn et al., 2012] concerning the engagement of crowds with humanities through crowdsourcing, provided with a comprehensive overview of the problems and solutions proposed by other teams in their effort to develop similar platforms and run similar campaigns. Amongst these, a recurrent issue is the difficulty of developing a completely new crowdsourcing infrastructure (i.e. the back-end in the list above) and without the financial or human resources of the said projects, we focused our efforts into seeking alternative solutions.

Our efforts yielded only a couple of potential candidates suitable to our needs, both of which were platforms designed for hosting crowdsourcing campaigns. CrowdFlower[14] had the advantage of a ready-made back-end, adaptable to the flexible types of tasks (and task constraints) desired by the data collector. But just as the AMT, its main drawback was that the crowd was mostly motivated by financial rewards. With CrowdCrafting[15] however, it seemed we had finally discovered our solution. Just as we desired, the platform was coming with a well designed back-end for managing tasks, subjects and classifications and with an easy to configure user interface based on *HTML* and *Javascript*. On top of this, it had the advantage of being developed on top of an open source API (PyBossa[16]). Its main backdrop was related to one important issue in the back-end which affected the configuration of the user interface. More specifically, CrowdCrafting allows for the upload of a set of subject images, which it can only serve to the users as a continuous stream. Considering our subjects are pages of comic books, we needed a way to serve these pages grouped in their original context, that of pages of separate comic books. However, with no other option available and with time becoming an issue, we decided to temporarily concede on this point and carry on with the development of a first CrowdCrafting prototype.

After a couple of months and a first review of the prototype, the spring of 2015 brought a new and exciting development. Delivering on a Google grant received in late 2013, Zooniverse[17], the citizen science platform behind the successful GalaxyZoo project, unveiled their new platform. With an appealing, slick, new interface on top of an improved back-end, its main feature, *Panoptes*, is an easy to use module allowing for the fast building of highly customisable citizen science projects. Moreover, due to integrating and hosting the resulting

[14] https://www.crowdflower.com
[15] http://crowdcrafting.org
[16] http://pybossa.com/
[17] https://www.zooniverse.org/

projects straight on Zooniverse's home page, deployment issues are automatically cast aside and projects are instantly ready to run. At the same time, a project is also automatically exposed to Zooniverse's user base, much larger and certainly not financially incentivised. We immediately reacted and quickly designed two new prototypes (one for *(Dis)Similitudes* and *Comics++*, each) and made them available for in-house testing and review. In the meantime, more good news was coming, with Zooniverse "secretly" developing *scribe* [18], an open API addressing *Panoptes*' major limitation – the inability to define textual transcription tasks.

6.4.2. Implementing the crowdsourcing platform

Developed with *Ruby on Rails*, the *scribe* Application Program Interface (API) exposes the components necessary for building a crowdsourcing platform and stores the respective data in a *Mongo* database. In the following we present the anatomy of a *scribe* project, namely the implementation details of our own platform – *Comics++*.

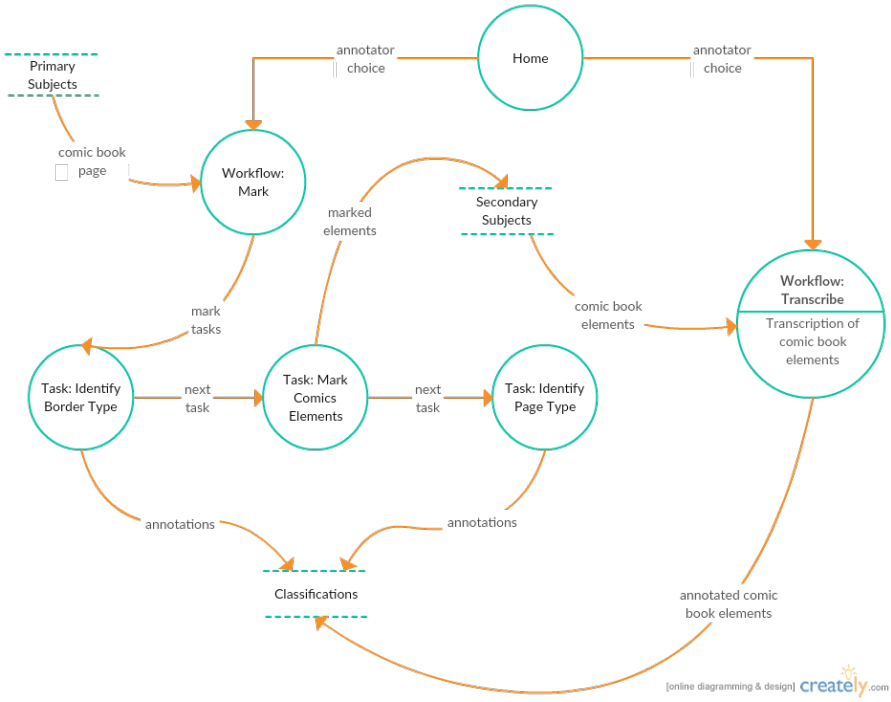


Figure 6.4.: Data flow diagram of the *Comics++* crowdsourcing platform, illustrating the interplay between workflows, tasks, subjects and generated classifications (annotations).

A typical *scribe* project is based on the following building blocks, as illustrated in Figure 6.4. Notice that this diagram is a more detailed view of the **Crowdsourcing Engine** component from the system diagram in Figure 6.2:

[18] https://github.com/zooniverse/scribeAPI

Subjects. A **subject** is a uniquely identifiable image a user performs a task on. In *Comics++*, a subject is a comic book page that our participants are asked to annotate. These initial images are called **primary subjects**. Primary subjects can be organised in **groups**, a useful feature in all our projects: in *Comics++* it allows us to group pages by comic book issues, whereas in *(Dis)Similitudes* it allows the grouping of poems by author.

Tasks. A **task** is an action which a user performs on a subject. Typical actions in *Comics++* require annotators to indicate and transcribe elements they recognise in a page or answer multiple-choice questions.

Workflows. A **workflow** represents a series of ordered or unordered tasks that are expected to be performed on a set of subjects. In *Comics++* we use two types of workflows: **Mark** and **Transcribe**. Tasks in a marking workflow are applied to primary subjects and their output – the **secondary subjects** – become the input to tasks of the transcription workflow. *scribe* also allows for the generation of so-called **tertiary subjects**, which can result from transcription tasks and become the input of a third type of workflow – **Verify**. None of our projects are making use of Verify type of tasks.

Classifications. A classification is a record of a user's response to a given task performed on a subject. In *Comics++* these are the set of marks, transcriptions and answers resulting from completing the proposed tasks.

The elements above are organised according to a file structure similar to that in Figure 6.5 below.

The presence of the `project.json` file is compulsory; this file contains basic information about the project which is used by the back-end at compile time. An example of configuration file corresponding to *Comics++* is illustrated in Figure 6.6 below. The file's fields are specified as hash-maps, with some of the most important being:

- `title` and `short_description`, which are compiled and then featured on the project's home page;
- `forum`, which links towards the URL where the discussion board is hosted (in our case, a Zooniverse Talk page);
- `terms_map`, for defining a project-specific vocabulary for consistent use throughout the platform (in our case, it identifies the generic *scribe* subject to as a comic book page and a group of such subjects as an issue).

After this file has been parametrised, the rest of the `project` folder, with the exception of the `subjects` and `workflows` subfolders – which we will cover in the next sections – hold the actual front-end content of the platform.

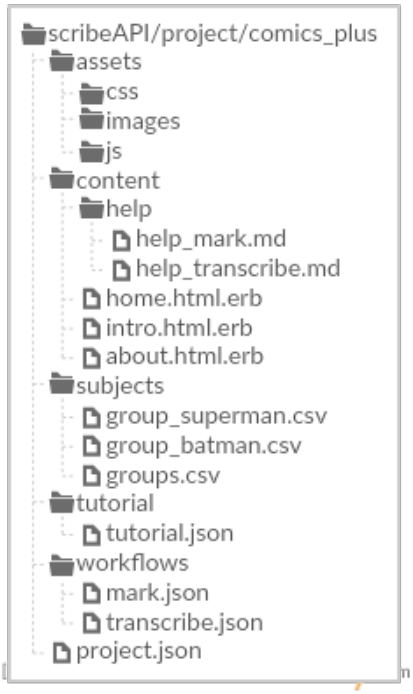


Figure 6.5.: The file structure of the *Comics++* crowdsourcing platform.

- `content`. This subfolder holds *HTML* and *Markdown* content, which at compile time places the main menu tabs in the top section of every page of the platform. This folder can additionally hold the `help` subfolder, which also contains a series of *HTML* and *Markdown* documents supporting the help and tutorial sections of the project.
- `tutorial`. This subfolder holds the `tutorial.json` file, which describes the sequential series of documents placed in `content/help`; these are the documents mentioned above, holding the content for the help and tutorial sections.
- `images`. This subfolder is a local container for the images used in the project, such as background images or pictures supporting the tutorials.

Workflows and tasks implementations

Tasks are specified in workflow-specific *JSON* files, which are placed within the `workflows` subfolder. Each of these files has a specific structure that describes a task sequence, the content of each task, the type of output data and the corresponding tutorial or help files, if any. Because of an API limitation, the workflow files can only bare the actual name of the workflow (i.e. `mark.json, transcribe.json, verify.json`).

```
{
  "title": "Comics++",
  "short_title": "Comics++",
  "summary": "Annotating difficult aspects of comics as a medium",
  "admin_email": "mihnea.tufis@lip6.fr",
  "team_emails": [],
  "terms_map": {
    "group": "issue",
    "subject set": "issue",
    "subject": "page"
  },
  "team": [...
  ],

  "organizations": [
    {
     "title": "LIP6",
     "url": "http://lip6.fr"
    },
    {
     "title": "Actialuna",
     "url": "http://actialuna.fr"
    }
  ],
  "forum": {
    "type": "talk",
    "account_username": "acasa",
    "project_name": "en-comics"
  },
  "discuss_url": "https://www.zooniverse.org/projects/acasa/en-comics/talk",
  "feedback_form_url": "http://goo.gl/forms/rrSAue1uuG",
  "metadata_search": {
    "fields": [
      {
        "label": "comic",
        "field": "name"
      }
    ],
    "feature_on_homepage": true
  }
}
```


Figure 6.6.: A typical configuration of the `project.json` file, in which the most important fields are `title`, `forum` and `terms_map`.

The Mark workflow and its tasks

A typical `mark.json` file, such as that used for *Comics++* is composed of a header (Figure 6.7) and its `tasks` section. The header contains a series of properties, which allow

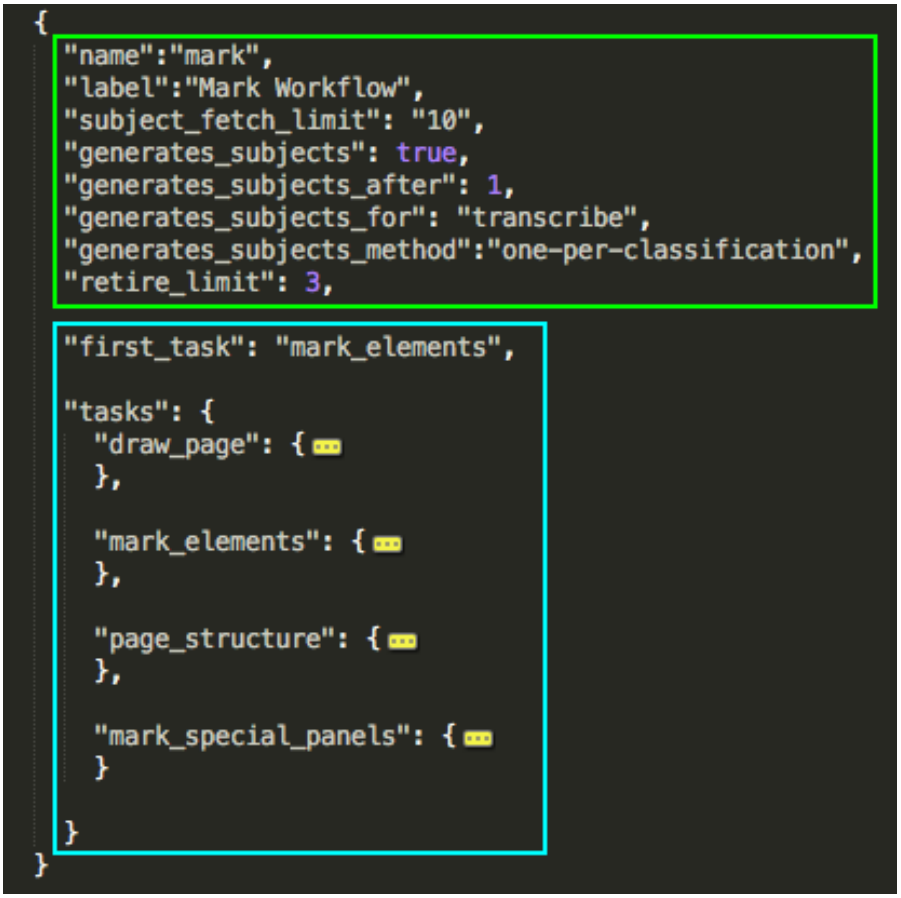


Figure 6.7.: The outline of the `mark.json` file, composed of a header (green) and the tasks section (blue).

for the identification of the workflow within the project and the specification of its link to other workflows:

- `name` and `label` identify the workflow with a name and a user-readable alias;
- `generates_subject`: if `true` signals that secondary subjects are generated for the next workflow;
- `generates_subject_for`: the workflow for which secondary subjects are generated; in our case, `transcribe`;
- `generates_subject_method`: the method for generating secondary subjects; in *Comics++*, each mark classification generates a different secondary-subject for transcribing;

- `retire_limit`: how many times annotators have to say there is nothing left to mark before a subject is retired from this workflow.

Next, the body of the **mark workflow** comprises the description of the composing tasks and their succession in the workflow. The four tasks in *Comics++* make use of only four of the so called marking tools. The first prototypes were developed in order to cover as many data requirements as possible. However, as we rolled these prototypes out for in-house testing, the feedback we received reiterated the idea that the annotation process can occasionally be off-putting, as it takes a bit too long of the participants' time. Thus, some tasks were eliminated, while others were grouped together in a more compact workflow, which in its current iteration has only three steps: one single-answer question and two point-and-click marking tasks:

`mark_elements` is the most engaging task, asking participants to pick and mark the different comic book elements (*Characters, Places, Events, Objects, Onomatopoeia, Visual Metaphors*). It uses the `pickOneMarkOne` tool, configured (`tool_config`) for each distinct type of comic book element included in the annotation:

- `pointTool` is the geometric `type` of the marking tool used to mark each specific element;
- `label` and `color` allow for the individualisation and styling of the tools corresponding to each comic book element;
- `generates_subject_type` is referring to the name of the transcription task for which each specific marking will become a secondary subject;
- `help` allows for the creation of contextual help, specific to each element.

The result of putting all this together is depicted in Figure 6.8 below:

`mark_special_panels` is yet another task built with the `pickOneMarkOne` tool. It asks participants to pick and mark (if existing) panels that exhibit special border shapes or styling, from a given list (*Wavy, No Frame, Portal, Irregular, Other*). As in the case of the first task, a tool configuration is necessary, with the supporting `rectangleTool` being used on this occasion for marking the often polygonal shapes of panels (see Figure 6.9 below);

`page_structure` is a simple task asking the users whether the current subject-page contains any special structure (i.e. is a splash page or holds a double-spread). The tool supporting this is the `pickMany` allowing for the choice of none, one or both possible answers (see Figure 6.10 below).

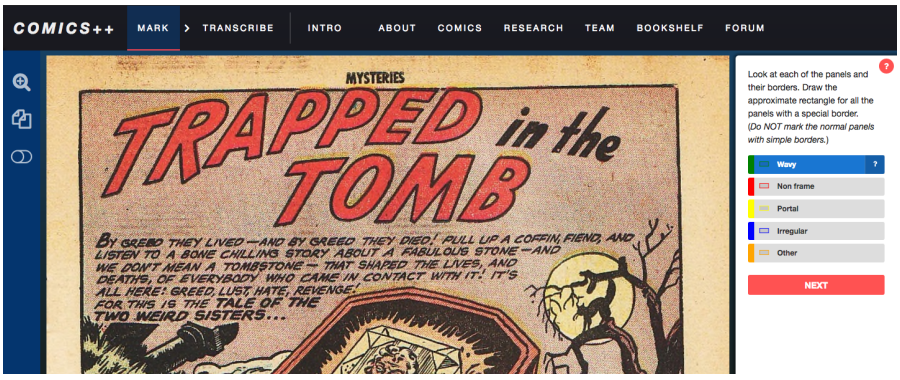
The `next_task` hash contains the name reference to the following task in the workflow, while a value of `null` marks its end.

(a)

```
"mark_elements": {
  "generates_subjects": true,
  "tool": "pickOneMarkOne",
  "instruction":"Read this page. Are there any missing **narrative elements**? Mark them by placing the mark close to their position in each panel.<br><br>Stop bef
  "tool_config": {
    "options": [
      {"type": "pointTool", "label": "Characters", "color": "green", "generates_subject_type": "transcribe_characters", "help": {"file": "m_record_char"}},
      {"type": "pointTool", "label": "Places", "color": "blue", "generates_subject_type": "transcribe_places", "help": {"file": "m_record_places"}},
      {"type": "pointTool", "label": "Events", "color": "red", "generates_subject_type": "transcribe_events", "help": {"file": "m_record_events"}},
      {"type": "pointTool", "label": "Objects", "color": "yellow", "generates_subject_type": "transcribe_objects", "help": {"file": "m_record_objects"}},
      {"type": "pointTool", "label": "Onomatopoeia", "color": "purple", "generates_subject_type": "transcribe_ono", "help": {"file": "m_record_ono"}},
      {"type": "pointTool", "label": "Visual Metaphors", "color": "orange", "generates_subject_type": "transcribe_visuals", "help": {"file": "m_record_visuals"}}
    ]
  },
  "help": {"file": "m_howto_elements"},
  "next_task": "mark_special_panels"
},
```

(b)

Figure 6.8.: The task requiring a participant to mark the elements of a comic book page (a); the corresponding code (b).

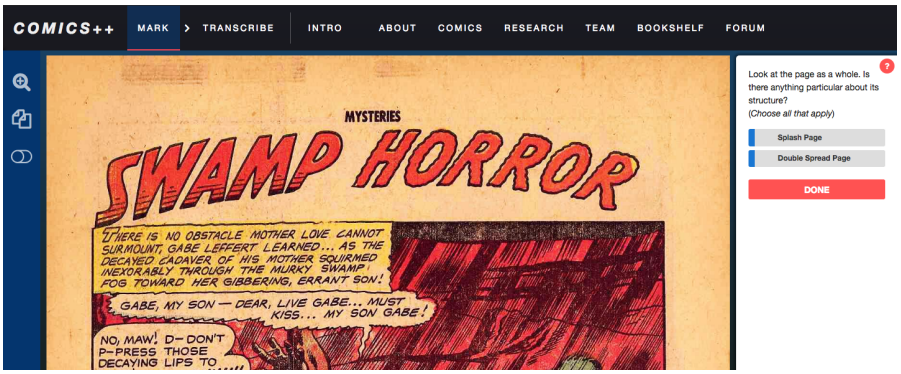


(a)

```
"page_structure": {
  "tool": "pickMany",
  "generates_subjects": true,
  "instruction": "Look at the page as a whole. Is there anything particular about its structure?<br>(*Choose all that apply*)",
  "tool_config": {
    "options": [
      {"value":"splash-page", "label": "Splash Page", "generates_subject_type": "no_type"},
      {"value":"double-page", "label": "Double Spread Page", "generates_subject_type": "no_type"}
    ]
  },
  "help": {"file": "m_record_structure"},
  "next_task": null
},
```

(b)

Figure 6.9.: The task requiring a participant to identify panels with special borders (a) and the corresponding code (b).



(a)

```
"mark_special_panels": {
  "generates_subjects": true,
  "tool": "pickOneMarkOne",
  "instruction":"Look at each of the panels and their borders. Draw the approximate rectangle for all the panels with a special border.<br>(*Do NOT mark
  "tool_config": {
    "options": [
      {"type": "rectangleTool", "label": "Wavy", "color": "green", "generates_subject_type": "no_type_1", "help": {"file": "m_border_wavy"}},
      {"type": "rectangleTool", "label": "Non frame", "color": "red", "generates_subject_type": "no_type_2", "help": {"file": "m_border_noframe"}},
      {"type": "rectangleTool", "label": "Portal", "color": "yellow", "generates_subject_type": "no_type_3", "help": {"file": "m_border_portal"}},
      {"type": "rectangleTool", "label": "Irregular", "color": "blue", "generates_subject_type": "no_type_4", "help": {"file": "m_border_irregular"}},
      {"type": "rectangleTool", "label": "Other", "color": "orange", "generates_subject_type": "no_type_5", "help": {"file": "m_border_other"}}
    ]
  },
  "help": {"file": "m_howto_specials"},
  "next_task": "page_structure"
}
```

(b)

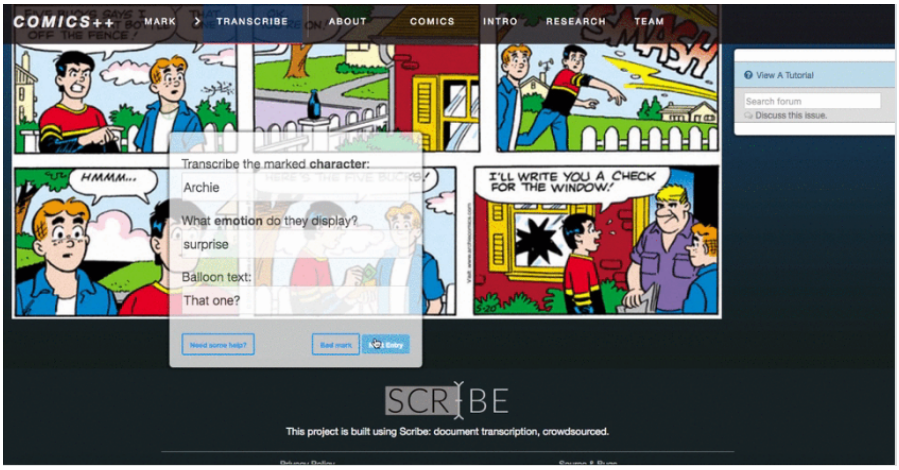
Figure 6.10.: The task requiring a participant to identify special types of pages (a) and the corresponding code (b).

The Transcribe workflow and its tasks

In *scribe*, the transcription tasks are directly influenced by the secondary subjects generated from the **mark workflow**. Thus, *Comics++* allows for the transcription of the comic book elements marked in the first task (i.e. characters, places, events, objects, onomatopoeia, visual metaphors). However, each of these elements have different particularities and thus impose slightly different task configurations.

`transcribe_characters`. This is achieved by using a `compositeTool` comprising of three different `tools`, each of them of type `textTool`, but with different individual configurations. The task requires the encoding of the following three aspects related to a character. Figure 6.11 illustrates the tool configurations and the resulting design:

- the character's name, or description – in case we are referring to an unnamed or collective character;
- the character's facial expression or body posture, chosen from a set of such expressions: *anger*, *fear*, *joy*, *surprise*, *deviousness*, *threat*, *power* (see Figure 2.32);
- the narrative contribution of a character (i.e. the text inside a balloon associated to the character);



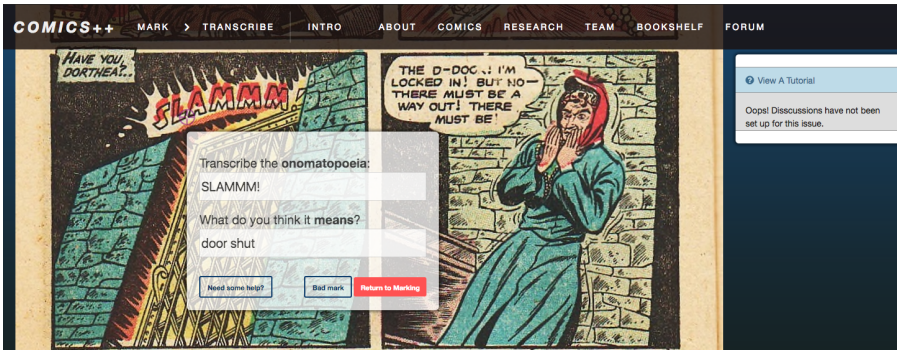
(a)

```
"transcribe_characters": {
    "tool": "compositeTool",
    "tool_config": {
        "tools": {
            "char_transcript": {
                "tool": "textTool",
                "tool_config": {},
                "label": "Transcribe the marked **character**:"
            },
            "expression_transcript": {
                "tool": "textTool",
                "tool_config": {"suggest": ["anger", "fear", "joy", "surprise", "deviousness", "threat", "power"]},
                "label": "What **emotion** do they display?<br><font size=2>*Click an arrow key for suggestions!*</font>",
                "value": "Placeholder"
            },
            "balloon": {
                "tool": "textTool",
                "tool_config": {"limit": 150},
                "label": "Balloon text:"
            }
        }
    },
    "generates_subjects_type": "transcribed_character",
    "help": {"file": "m_record_char"},
    "next_task": null
},
```

(b)

Figure 6.11.: The task requiring a participant to transcribe the characters (a) and the corresponding code (b).

`transcribe_ono` is the task facilitating the transcription of onomatopoeia and their perceived meaning. Just like for the task before, a `compositeTool` of two `textTools` is needed (see Figure 6.12b below);



(a)

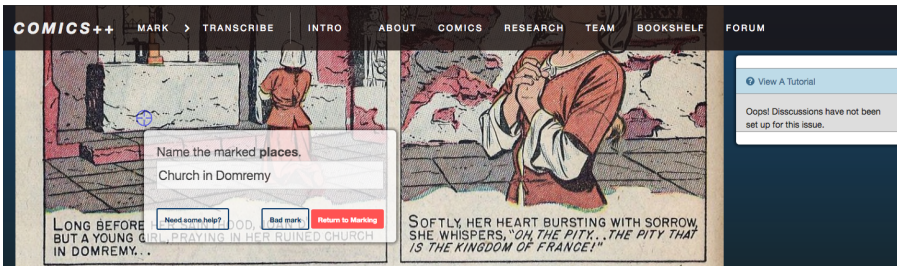
```
"transcribe_ono": {
    "tool": "compositeTool",
    "tool_config": {
        "tools": {
            "ono_transcript": {
                "tool": "textTool",
                "tool_config": {"suggest": "common"},
                "label": "Transcribe the **onomatopoeia**:"
            },
            "ono_meaning": {
                "tool": "textTool",
                "tool_config": {"suggest": "common"},
                "label": "What do you think it **means**?"
            }
        }
    },
    "instruction": "",
    "generates_subjects_type":"transcribed_ono",
    "help": {"file": "m_record_ono"},
    "next_task": null
},
```

(b)

Figure 6.12.: The task requiring a participant to transcribe onomatopoeia (a) and the corresponding code (b).

`transcribe_places`, `transcribe_events`, `transcribe_objects`, `transcribe_visuals`. The rest of the transcription tasks are simple `textTools`, allowing for the straight description of the respective element. Figure 6.13 below illustrates the implementation of the place transcription task and its visual result.

scribeAPI allows for the configuration of a much larger number of mark, transcribe and even verify tasks and tools and even grants a developer the possibility of coding tools tailored to their own needs. As it has been the case for *Comics++*, we couldn't avoid



(a)

```
"transcribe_places": {
    "tool": "textTool",
    "tool_config": {"suggest": "common"},
    "instruction": "Name the marked **places**.",
    "help": {"file": "m_record_places"},
    "next_task": null
},
```

(b)

Figure 6.13.: The task requiring a participant to transcribe places (a) and the corresponding code (b).

diving deeper into the core code of *scribe* and adapt it to respond to some of our particular necessities.

The subjects

The `subjects` folder of a *scribe* project typically holds a number of *CSV* files, in which one defines sets of primary subjects. Remember that in *scribe* the secondary and tertiary subjects (those used by the **transcribe** and **verify workflows**) are automatically generated from primary subjects. The header of these files is based on a subset of fixed columns, which can be extended according to the project's needs.

- `order` – an *integer* giving the order of the subject inside its containing set;
- `filepath` – the URL path to the full subject file;
- `thumbnails` – the URL path to the thumbnail image; if missing, the full image is adjusted;
- the `width` and `height` of the image; if missing, *scribe* tries to determine them automatically.

Figure 6.14) below illustrates the *CSV* file containing the subject-set corresponding to a comic book issue in *Comics++*.

```
order,file_path,set_key
1,/images/comics/Mysteries_Weird_and_Strange_002/page_0003.jpg,mws-002
2,/images/comics/Mysteries_Weird_and_Strange_002/page_0003.jpg,mws-002
3,/images/comics/Mysteries_Weird_and_Strange_002/page_0003.jpg,mws-002
4,/images/comics/Mysteries_Weird_and_Strange_002/page_0003.jpg,mws-002
...
```

Figure 6.14.: A snippet of a typical subjects *CSV* file, listing the individual subjects (i.e. pages) of a subjects group (i.e. comic book). Observe the last column (`set_key`), which identifies a set of subjects as members of the same group (i.e. `mws-02`).

We briefly digress to draw the attention on an important development issue: the file path and image dimension columns are tightly linked to probably the most vicious bug we corrected in the *scribe* API. During our internal tests, we discovered that any image which wasn't stored at an *HTTP* accessible location was virtually inaccessible due to the way the path parsing was done. With all our assets stored in Rails' `public` folder (hence, not an *HTTP* location), this meant the URI towards them was incorrectly parsed, leading to the impossibility of establishing the size of each subject and therefore breaking the whole *ReactJS* layout. However, since this didn't result in a compile time error, the problem was occurring only after the *Rails* server had been launched and a request for subjects had been made.

The subjects are finally loaded from the `groups.csv` file, which aggregates the references to all *CSV* files which contain the subject sets. The header of the `groups.csv` file features the following columns:

- `order` – an *integer* giving the order of each group;
- `key` – the key in the file name of the subject set (e.g. a key `eisner` refers to the subject set in `group_eisner.csv`);
- `description` – a textual description of the group;
- `cover_image_url` – the URL referencing the image that will feature on the main page as a cover for the entire group;
- `external_url` – the URL referencing additional information for the group;

Besides these standard columns, developers can also define custom columns, serving the purpose of an individual project. In the case of *Comics++* some of these columns are:

- `issue` – the name of the featured comic book issue;
- `date` – the publishing date;
- `author` and `publisher`
- `copyright` – the license under which these images are being presented.

The corpus

Building a corpus of subjects for *Comics++* has been one of the most important challenges in our project. With comics publishing becoming such an important segment of the market, the financial objectives of major publishers are hampering most academic and research initiatives with respect to comics, particularly those that require their sharing with the large public. Under these conditions, we turned our attention to copyright free comic outlets, such as the Digital Comics Museum[19] or the Internet Archive for Comic Books and Graphic Novels[20]. The former hosts an impressive collection of public domain comics from the Golden Age (1940s-1960s), including several issues from Will Eisner's famous *"Spirit"*. There are, however, some inconveniences in using these comics outlets. First, the quality of the available scanned version is in some cases far from ideal, meaning a decreased reading experience and an increased risk of frustration and loss of motivation for the annotators. Second, the corpora hosted by these platforms don't exhibit enough cultural variance. With much of the content adhering to the structural and narrative style specific to the American Golden Age, the results of any empirical study concerning this corpora would be affected by this cultural bias.

To tackle this, we are considering the possibility of integrating our annotation platform with Sequencity[21], the online comics reading engine developed by our industrial partner, Actialuna. This would grant us access to their own publication catalogue, which features a larger variety of comics, including *manga* and *bandes dessinnées* collections, thus satisfying the criteria for cultural diversity. At the same time, this would also solve two of our crowdsourcing related issues: reaching to a larger public of dedicated readers and potential annotators, as well as the possibility to provide material rewards (e.g., reading vouchers on Sequencity) for our best contributors.

Finally, a potential collaboration with the *Cité Internationale de la Bande Dessinnée et de l'Image (CIBD) – Angoulême*, which would see *CIBD* open parts of their own Catalogue to *Comics++* annotators, would not only meet our corpus demands, but would also enable *CIBD* to acquire fast and complete annotations, leading to a better indexing of their Catalogue content, an excellent showcase of how citizen science research can be applied in the GLAM sector.

User interface and interaction

This section briefly introduces the user interface as well as the user interaction principles behind it. Aiming for a simple, yet slick interface, *Comics++* and *(Dis)Similitudes* are built by making the best of two popular design frameworks – Material Design[22] and CSS

[19] http://digitalcomicmuseum.com/
[20] https://archive.org/details/comics
[21] https://www.sequencity.com/
[22] https://material.google.com/

Bootstrap[23]. Developed by Google in 2014, the Material Design language creates a visual experience based on the physical properties of paper and ink. The material metaphor is perfectly summarised by its creators:

> "Surfaces and edges of the material provide visual cues that are grounded in reality. The use of familiar tactile attributes helps users quickly understand affordances. Yet the flexibility of the material creates new affordances that supercede those in the physical world, without breaking the rules of physics".

Since then, Material has been successfully applied to YouTube and most of Google's applications (GMail, Google Docs, Sheets and Slides, Google Drive, and Google Maps). Perhaps the most well known application of Material was to the UI design of recent Android platforms.

Bootstrap on the other hand, needs little introduction, since it is, at the moment, "the most popular web-based framework for developing mobile-first web applications". In *Comics++* and *(Dis)Similitudes* we combine the layout design principles of Bootstrap with the visual design principles of Material. Luckily, much of the *scribeAPI* generated layout is designed according to Bootstrap. For example, the layout features a fixed navigation bar in the top part of the screen, from where a user can access any of the platform's section, as well as a fixed footer, hosting sections dedicated to privacy policy and acknowledgements.

The platform's home page (Figure 6.15) features a simple design, with the name of the project and the host organisation taking the centre of the page and tries to encourage participation with a short, proactive message. The other element featured on this page is a large Bootstrap-style button.

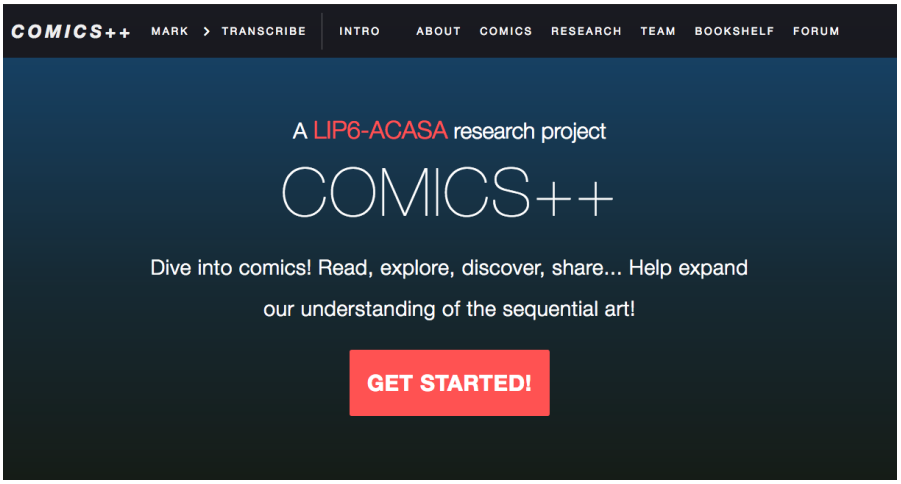


Figure 6.15.: The home screen of the *Comics++* crowdsourcing platform, featuring a simple design based on concepts drawn from Material and Bootstrap.

[23] http://getbootstrap.com/

From here, the page links to the introduction tutorial, designed to quickly get participants acquainted with the whole platform. According to crowdsourcing design principles, it shortly states the purpose of the project, the tasks that the annotators are expected to help with and gives clear, visual examples of how exactly to perform them. From here, participants can embark on working on either marking or transcription tasks.

The marking page is made up of three distinct interface elements, illustrated in Figure 6.16 below:

- the content pane, in the centre, taking the largest part of the screen;
- the toggle bar, to the left, featuring the thumbnails of each of the subjects belonging to a set;
- the question pane, to the right, containing a series of elements of its own:
  - the task help button (in the upper right corner of the pane);
  - the task itself (in the central area of the pane);
  - the extras area (in the lower part of the pane), comprising a series of quick links towards the forum, the general tutorial, the details page for the current comic book issue, the **transcribe workflow** and social media.

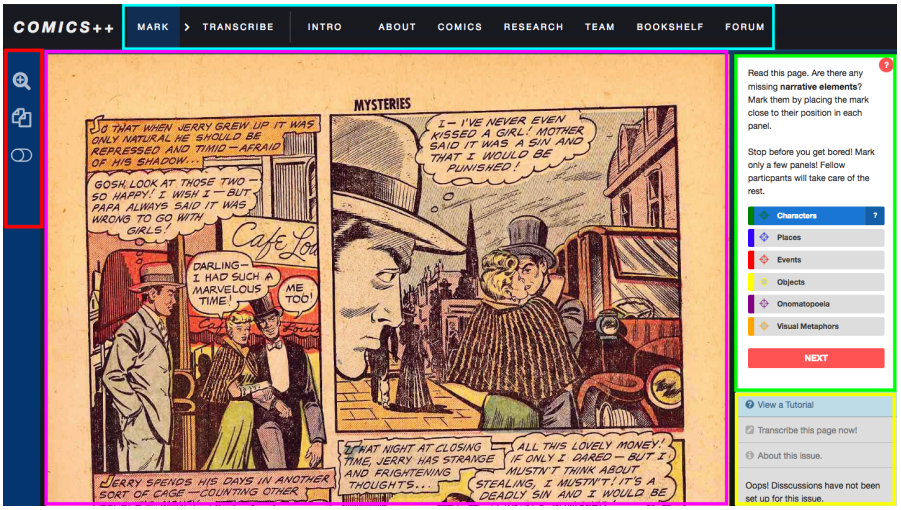


Figure 6.16.: A typical screen of the **mark workflow** in *Comics++*. The main areas are: the main menu (blue), the content pane (purple), the toggle bar (red), the question pane (green), the extras (yellow).

Alternatively, the transcribe page comprises only a couple of interface elements, as illustrated in Figure 6.17 below:

- the annotation pane, displaying the annotation task;
- the central pane, displaying the current comic book page and the existing marks;

- the extras area, displaying only two quick links, one towards the general transcription tutorial and the second to the Forum.

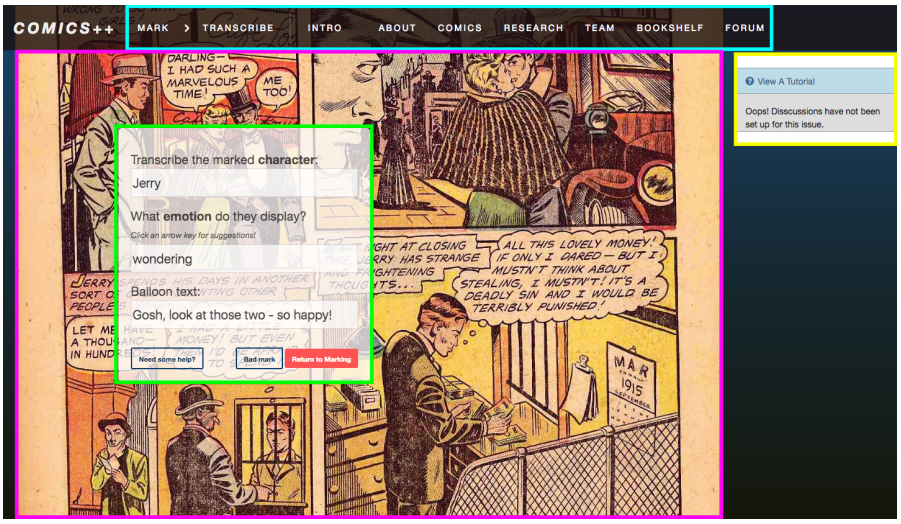


Figure 6.17.: A typical screen of the **mark workflow** in *Comics++*. The main areas are: the main menu (blue), the central pane (purple), the annotation pane (green), the extras (yellow).

In what concerns the styling and the colour scheme in particular, we adhered to recommendations from Material.

> "Material takes cues from contemporary architecture, road signs, pavement marking tape, and athletic courts. Color should be unexpected and vibrant."

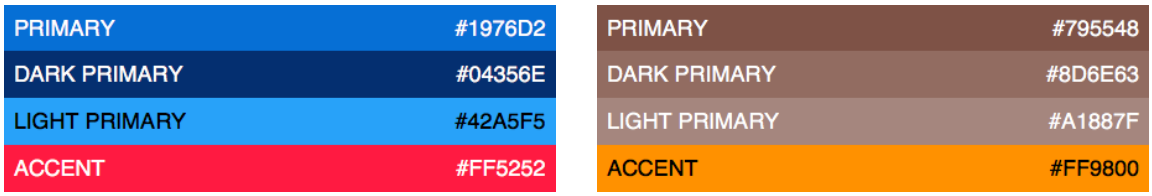
And according to these principles, the user interface should make use of a consistent colour scheme comprising of:

- Primary colour – the most widely used across all screens and components.
- Secondary colour – used to indicate a related action or information. The secondary colour may be a darker or lighter variation of the primary colour.
- Accent colour – used for the buttons and interactive elements, such as: text fields and cursors, progress bars, links etc.

The selection of the application's colour palette has to take into account the visual identity of the platform or organisation behind it. Thus, *(Dis)Similitudes'* primary colour is a darker hue of brown derived from its association with the *Observatory for Literary Life (LabEx OBVIL)*. For *Comics++* on the other had, finding a visual identity was a bit more problematic, since neither the main organisation behind the project, nor our industrial partner are promoting a particular colour identity. In this circumstances, we set off to creating a brand new chromatic identity for our platform by investigating what are the most frequent colour combinations in comic books. In the absence of relevant research,

one of the few analysis on the topic[24] reports on the most common colour combinations in American superhero comics: red and blue (Wonder Woman, Superman, Spiderman, Captain America), blue and yellow (X-Men), red and yellow (Ironman, Daredevil), green and purple (usually for antagonistic characters). Combining them with Material principles, our personal preference leaned towards blue (as a more sober, primary colour) and red (as the accent colour).

The complete colour palettes for the two projects can be visualised in Figure 6.18.



(a) **(b)**

Figure 6.18.: The colour palettes of our crowdsourcing platforms *Comics++* (a) and *(Dis)Similitudes* (b), based on Material Design principles.

[24] [COLOURlovers, 2011]

Chapter 7.

Data Analysis of Comics++

"With most things, the average is mediocrity. With decision making, it's often excellence. You could say it's as if we've been programmed to be collectively smart."

— James Surowiecki, author of *Wisdom of Crowds* (2004)

The previous chapter dealt with the design and implementation of *Comics++*, our crowdsourcing platform for annotating comics. In Section 6.1 we analysed the nature and type of data we need to collect (also called "primary data"), whereas in Section 6.2 we presented an overview of how this can be achieved by crowdsourcing the data collection process.

A secondary objective of our work is to investigate and improve our understanding of the factors underlying the success of crowdsourcing campaigns, expanding from the work of [Sharma, 2010] concerning "critical success factors" and more recent research of [Amer-Yahia and Senjuti Basu, 2016] with respect to "human factors" in crowdsourcing projects. In fact, by internalising all the corpus of research presented in Chapter 5, much like [Grady and Lease, 2010] we focus on a subset of these factors, whose interplay and influence on the quality of our crowdsourced data we would like to understand. To this end, we record and analyse data related to the behaviour of our participants ("secondary data"):

- task difficulty:
  - time to finish a task;
  - page type identification (sequential art theory suggests that pages with loose structures and border-less panels are likely to be more difficult to grasp by the reader);
- user expertise in a given context (i.e. task):
  - time to finish a task;

- task design:
  - exposing the annotator to different types of workflows and questions;
- motivation (intrinsic and external motivators):
  - time spent on the platform (intrinsic motivators);
  - number of annotations (intrinsic motivators);
  - unlocked items on our industrial partner's platform (external reward);
- attention:
  - time to finish a task

In the following we present an experiment setup designed to analyse three different aspects of our crowdsourcing system:

- the interaction between participants and the platform;
- the design of the two workflows: mark and transcribe;
- the content of their contributions.

7.1. Experiment design

The first of *Comics++* made available a number of 10 entire comic books, each consisting of 4-5 stories of different genres, with lengths ranging between 4-12 pages, for a total of 335 pages. After a couple of weeks, only a small fraction of this corpus was annotated (mostly marked), with most participants engaging with the problem only once, following social media or targeted e-mail campaigns.

To address the issue of reaching the crowd, we designed an experiment in which a downsized crowd annotates a downsized corpus, consisting of only 4 stories (from 3 different comic books), with lengths ranging between 4-8 pages, for a total of 24 pages. Despite the reduced corpus, we aimed at gathering a large number of annotations, which would allow us to perform an extensive data analysis about *Comics++*. In order to adapt to the desired diversity of the crowd, the stories were selected such that they cover a diverse range of representative genres and topics (sports, superhero, comedy, horror), and include most of the comic elements that we wish to annotate (i.e., characters, places, events, onomatopoeia, visual cues, special panels, splash pages):

- "Dash Dillon", in *Daredevil #2* (November 1941), by C. Biro, R. Briefer, J. Robinson and D. Wood (4 pages);

- "Nightro", in *Daredevil #2* (November 1941), by C. Biro, R. Briefer, J. Robinson and D. Wood (5 pages);
- "Truant Toy", in *Funnyman #1* (January 1948), by Jerry Siegel and Joe Shuster (8 pages);
- "Shrinking Horror", in *Mysteries #3* (September 1953), N/A (7 pages).

The experiment consisted of two stages:

1. Mark – participants visualised the story page-by-page and were asked to identify and mark elements of the comic medium. In particular, they were instructed not to provide markings for those elements that they consider to be correctly marked by other participants. We expected this step to be concluded very fast and offer sufficient data for the next stage;
2. Transcribe – participants were asked to transcribe a series of information referring to each of the elements that were marked in the previous stage. In case of disagreement with a certain marking, participants had the option to flag it appropriately.

To build the community, we started from a "seed crowd" of 80 members directly connected to our team (members of our research group, members of different professional groups, friends, family, etc.), an approach somewhat inspired by the "closed crowd" experiment described by [Nallapati et al., 2012]. We engaged our crowd through a targeted email campaign, explaining the context of our research and the structure of our experiment. Each member of the "seed crowd" was then asked to forward our call for participation to at least one other person that could be interested in taking part to the experiment. For the first stage (marking), the crowd was divided into two groups, each marking one short and one long story (in total, 12 pages). For the second stage (transcription), the crowd was divided into four groups, each group working on an individual story. However, in both stages, participants were encouraged to "help" their peers working on other stories.

In parallel, we started a social media campaign on the Twitter platform, involving daily tweets with updates about the progress of our experiments. This attracted the attention of several comic fans and GLAM institutions with an interest in comics and comics studies.

7.2. Experimental results and analysis of Comics++

The main goal of our experiment is to study whether crowdsourcing and *Comics++* are suitable for the purpose they are designed: gathering quality annotations for comics in a cheap and timely manner.

A crowd of 45 contributors participated in the two stages of our experiment (marking and transcribing), which ran over the course of 14 days. The transcription was launched five

days into the campaign, without explicitly putting an end to the marking phase. However, most necessary marks were created in the first three days of the experiment.

Participants spent a total effective time of 3,132 seconds (less than a day) to provide most of the annotations in the first stage (marking), confirming our expectations of the fast turnaround time for this workflow. As for the second stage, participants spent a total effective time of 64,600 seconds transcribing the different elements of comic books. This is equal to approximately 18 hours, meaning that the tasks were effectively performed in less than a day.

Both these fast turnaround times confirm the power of crowdsourcing and its suitability as a data collection methodology for annotating comics. In the following, we present a number of specific analyses, focused on:

- studying the difficulty of annotating different types of elements;
- studying the difficulty of annotating different types of comic stories;
- studying the crowd engagement with each of the two workflows;
- studying the capacity of the crowd to identify the main narrative elements in each story;
- studying the capacity of the crowd to identify subtle elements of the comic medium (e.g., splash pages, onomatopoeia, visual cues);
- modalities of identifying bad annotators.

7.2.1. Analysis of special panels

This analysis concerns the second task in the **mark workflow**, the identification and annotation of special panels, such as they were defined in Section 2.2.1 and illustrated in Figure 2.20. On the platform the task is depicted as following:
Look at each of the panels and their borders. Draw the approximate rectangle for all the panels with a special border. (Do NOT mark the normal panels with simple borders.).

Participants choose from a list of special border types (wavy, no-frame, portal, irregular, other) and then draw the approximate rectangle that encompasses the respective panel.

The purpose of this analysis is to assess the capacity of the crowd to recognise specific elements of the comic medium, in this case panels with different types of borders. These are particular notions, first introduced in comics theory by [Eisner, 1985], meaning that participants are very unlikely to be familiar with these concepts; therefore, we propose that before they proceed, they go through a tutorial about these issues.

9 annotators provided 69 panel markings, considering that there were actually 18 special panels (5 of type *no-frame* and 13 of type *other*) in the entire corpus. Table 7.1 and Table 7.2 summarise the number of panels classified as special (**Annotated**), by story and by panel type, respectively, then compare them to their real count (**Actual**) and to how many of these were correctly identified (**Correct**).

	Dash Dillon	**Nightro**	**Shrinking Horror**	**Truant Toy**
Actual	6	0	5	7
Annotated	12	20	25	2
Correct	0 (1*)	0	4	0 (2*)
Precision	0 (0.08*)	0	0.16	0 (1*)
Recall	(0.16*)	NA	0.8	0.28
F-score	0 (0.11*)	NA	0.27	0 (0.44*)

Table 7.1.: Special panels by story – annotated vs. actual vs. correct.
(*) indicates special panels whose positions were correctly identified, but were assigned to the wrong type.

	Wavy	**No frame**	**Portal**	**Irregular**	**Other**
Actual	0	5	0	0	13
Annotated	17	20	4	16	2
Correct	0	4	0	0	0 (3*)
Precision	0	0.2	0	0	0 (0.6*)
Recall	NA	0.8	NA	NA	0 (0.23*)
F-score	NA	0.32	NA	NA	0 (0.33*)

Table 7.2.: Special panels by type – annotated vs. actual vs. correct.
(*) indicates special panels whose positions were correctly identified, but were assigned to the wrong type.

Provided a good understanding of what was required and the clear drawing style used by the authors of the four stories, our expectations with respect to this task were high. However, from the results presented in Table 7.1 and Table 7.2 we observe very low precision rates, with a direct impact on the reported F-score. What was really surprising was the identification of types of panels that don't exist in the corpus (wavy, portal, irregular). Following a visual investigation of these misclassified panels, we observed one frequent mistake: annotators looked at the shape of the text balloon comprised by a panel and annotated the whole panel based on this. This explains the large number of normal panels, that were misclassified as wavy or irregular, only due to the presence of a thought- or scream-balloon within.

One such example occurred in "Shrinking Horror", where several thought balloons, which were drawn slightly outside their encompassing panel, lead annotators to classify them as *wavy* instead of simply considering them as normal.

Three other annotations, two from "Truant Toy" and one from "Dash Dillon" identified the correct positions of the special panels, but wrongly classified them as *irregular* instead of *other*[1]. Also in "Truant Toy", in two different instances, the authors employ "complementary" panels – a long panel that is divided into irregular smaller panels; in this case, participants noticed only one of the two "complementary" panels.

Finally, "Nightro" proved to be the most problematic story. Despite no special panels actually existing in this story, participants misclassified a large number of normal panels and even text balloons.

Despite our expectations, the data collected with respect to each type of special panel suggest a lot of confusion in our participants' contributions and lead us to believe that either they didn't read the necessary contextual tutorial, or that this was poorly designed. The small data sample is another limitation of this experiment, one that we wish to correct in a future iteration, and considering a different experimental setup.

7.2.2. Analysis of page structure

This analysis concerns the last question of the **mark workflow**:
Look at the page as a whole. Is there anything particular about its structure? (Choose all that apply).
Participants have two options to choose from: *splash-page* and *double-page*. A page can also be a "splash-double-page" (in which case participants select both) or a normal page (participants don't select any option and proceed to the next page).

Just like in the previous section, the purpose of this analysis is to assess the capacity of participants to understand specific concepts of the comic book medium (i.e., splash- and double-page) and indetify instances of their occurrence.

Our corpus contains 4 splash pages (corresponding to the first page in each story), 20 normal pages and no double page. 15 different participants answered this question, providing a total number of 30 answers (23 splash-pages and 7 double-pages). Table 7.3 summarises the distribution of the 23 answers identifying splash-pages, baring in mind that only the first page in each story (first row of the table) is an actual splash-page.

Most of these answers (17) correctly identified the splash pages (true positives), while 6 of them misclassified normal pages as splash (false positives). This translates to a precision

[1]This was the case of round panels, which under our classifications are considered of type *other*.

Page #	Dash Dillon	Nightro	Shrinking Horror	Truant Toy
1	1	11	2	3
2	1	2	0	0
3	1	1	0	0
4	1	0	0	0
5	-	0	0	0
6	-	-	0	0
7	-	-	0	0
8	-	-	-	0

Table 7.3.: Page structure markings for pages in each comic.

value of 0.74 an $F - score = 0.85$, both acceptable values considering the limited size of the collected data.

We draw attention on the case of "Nightro", for which all four classifications, one for each page, came from the same user. In addition, 6 of the 7 double-page misclassifications[2] occurred on this same story. Our only explanation for the recurrent misclassifications on "Nightro" is that this was the first story that participants engaged with. This would be consistent with the learning effect reported by [Kazai and Milic-Frayling, 2009], who noticed that annotators become better as they repeatedly perform the same task.

Alternatively, based on similar experiments by [Grady and Lease, 2010], we analyse the average time that participants needed to accomplish this task and. Based on the assumption that bad annotators have a shorter dwell time, we compare the average time between correct and incorrect classifications.

Thus, the average accomplishment time for the page structure task was of approximately 87 seconds, with participants spending on average more time to correctly identify splash-pages (92 seconds), as opposed to those who misclassified splash-pages (79 seconds). Figure 7.1 summarises this point and also gives an overview of the dwell time data. Unfortunately, this data is insufficient for establishing any connection between dwell time and the correct identification of splash-pages[3].

[2]We recall that there are NO double pages in all our corpus

[3]There are two problems: first, the distribution of the dwell time is strongly skewed to the left and second, the t-test we performed resulted in an unsatisfactory p-value. Therefore, more data is needed to test this hypothesis.

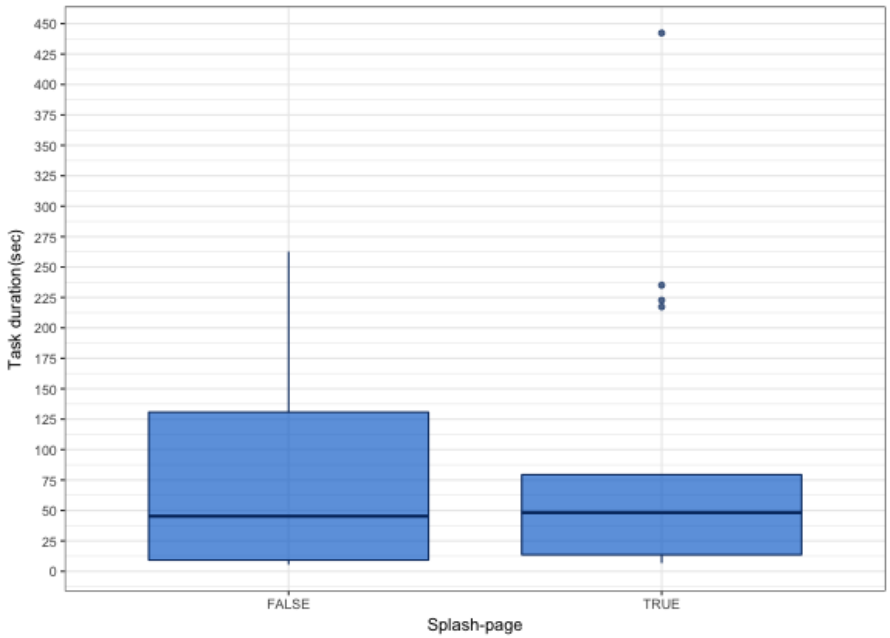


Figure 7.1.: Page structure - Distribution of task accomplishment times, for correctly and misclassified splash-pages.

Analysis of marked elements

This analysis concerns the first task in the **mark workflow**, requiring participants to identify different types of graphic elements specific to the comics medium (i.e., characters, places, events, onomatopoeia, visual cues). The exact requirement for this task is the following:
Read this page. Find the narrative elements and place a mark close to their position in each panel. Skip those that seem to be correctly marked by others.
This setting seemed suitable for three reasons:

- we offer participants a choice between taking part in the fast, marking tasks or the more detailed oriented (hence, slower) transcription tasks;
- it is a good mechanism to avoid duplication of correct markings;
- we achieve a form of early validation, since marks that are not deleted can be considered validated by the crowd.

The purpose of this analysis is to understand the difficulty of identifying the different elements of the comics medium using non-experts, assessing whether the type of story has any influence on the crowd's participation in the task, as well as having an overview of the crowd's engagement with the fast marking tasks.

17 different annotators identified a total of 594 graphic elements specific to the comics medium, according to our instructions. In Figure 7.2 we give an overview of the different types of elements identified by the crowd throughout the entire corpus.

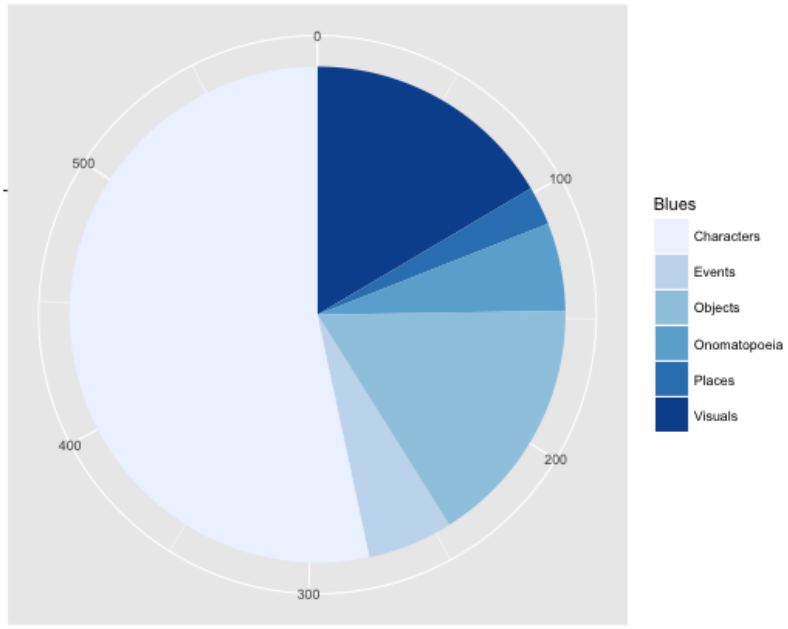


Figure 7.2.: Total counts for each type of marked element.

Not surprisingly, characters account for more than half of the marked elements, most likely due to the ease of identifying them and their role as a narrative vehicle in comics. While numerous in our corpus, places and events account for only a small proportion (8%) of the collected marks. Since in the instructions, we asked participants to identify only those places or events that they judge to be relevant to the narration, a possible explanation for the low count of these elements is that annotators had difficulties in assessing their relevance. This is not the case for marking objects, which account for approximately 16% of the marked elements, likely due to their clearer presence. An unexpected outcome is the large number of visual cues (e.g., movement lines), which represent 16% of the marked elements. We hypothesise[4] that this is due to a piqued interest by participants in this subtle, yet powerful element of the comics medium.

Alternatively, Table 7.4 summarises the total count of all marked elements within each story, as well as their average per-page. One of the questions we asked was whether there is any particular engagement or lack thereof, with any of these stories. To answer this, we performed a pairwise t-test between the Poison counts of the average number of marks per page for each of the stories. With one exception, the test indicates no significant difference between the average page counts at 5%. The exception is the comparison between "Dash

[4]We explore these two hypotheses in a subsequent survey concerning participation on *Comics++*.

Dillon" and "Shrinking Horror", with an average count of marks per page significantly larger in favour of the former, at the 5% level (p-value=0.011). Nevertheless, since this difference is isolated, it is safe to conclude that there are no significant differences in the way participants marked the proposed stories.

	Dash Dillon	Nightro	Shrinking Horror	Truant Toy
Count	133	141	120	200
Avg. per page	33.2	28.2	17.1	25

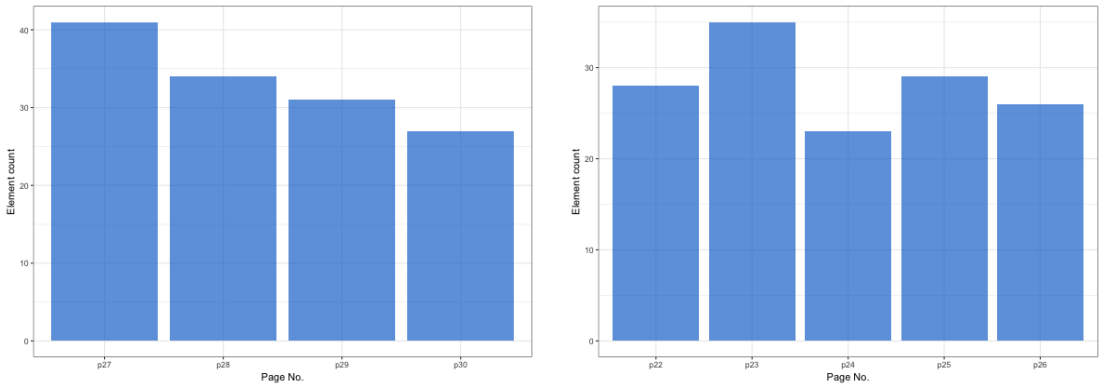
Table 7.4.: Distribution of marked elements for each story.

In Figure 7.3 we take a look at the distribution per page of the marked elements, in each of the four stories. These distributions could be indicative of several things: difficulty in reading a certain part of a story, a particular page structure (e.g., changes in the grid layout, the presence of a splash page) or the level of annotators' engagement (e.g., attention, fatigue). Overall, the maximum number of marks per page is of 41 and the minimum of 4 (corresponding to the full-page splash on the first page of "Truant Toy").

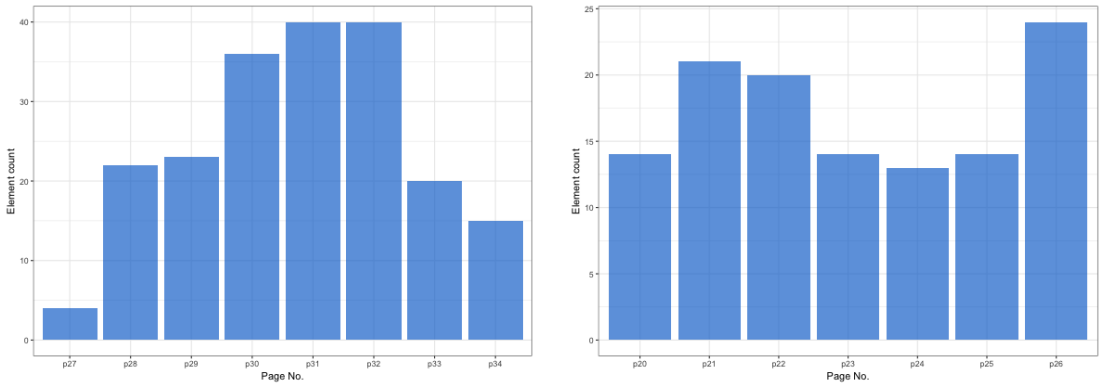
The marks appear to be uniformly distributed over the pages of the short stories (Figure 7.3a and Figure 7.3b). Given the fact that all pages within a story exhibit the same layout, these distributions suggest that participants held the same level of engagement throughout the marking of the short stories.

This changes in the case of the long stories (Figure 7.3c and Figure 7.3d), for which marks seem to follow a normal distribution over the pages. And while the left tail of these distributions can be explained by the low count of elements in the first pages of each story (large splash pages), we have difficulties in explaining the right tail of these distributions, particularly in the case of "Truant Toy". Given the uniform structure of the pages throughout the story, our hypothesis is that this distribution can only be explained by participants' level of engagement. While interesting, these hypotheses should be tested on a larger number of stories, before drawing any conclusion.

Finally, in Figure 7.4 we analyse the overall participation in the marking workflow, by plotting the daily cumulative number of contributed marks. We observe the majority of marks (530 of the total 594) were submitted during the first day of the campaign, followed by a slow increase over the next couple of days and only feeble marking activity during the rest of the campaign. This is also a consequence of our decision to move the experiment into the second phase (transcription) after just three days, due to the sufficient number of initial contributions. At this point, while markings were still allowed, we had asked our participants to focus primarily on transcriptions.



(a) (b)



(c) (d)

Figure 7.3.: The page distribution of marked elements, for each of the stories: a) Dash Dillon, b) Nightro, c) Truant Toy, d) Shrinking Horror.

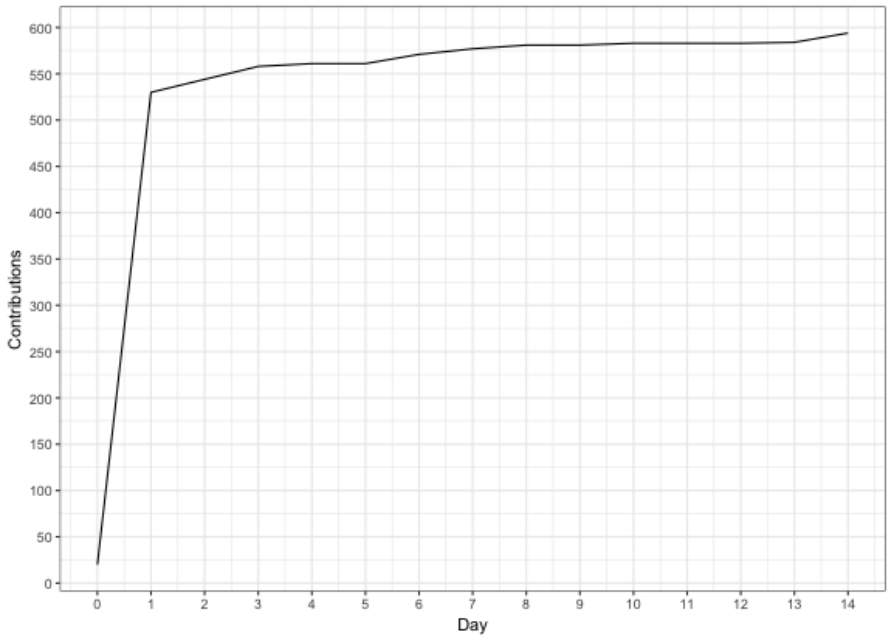


Figure 7.4.: Daily cumulative number of contributed marks.

7.2.3. Analysis of transcriptions

This analysis concerns the set of tasks which correspond to the **transcription workflow**. Participants are presented with the set of marks resulting from the **mark workflow**, in which they had identified different types of elements from the comics medium and are asked to give a descriptive transcription of these.

The goal is to determine whether the proposed approach is suitable for identifying the main narrative elements of comic book stories (characters, places, events, objects, onomatopoeia, visual cues) and, just like before, to determine the level of engagement of participants with the platform and the previously marked content.

23 different participants contributed with 1,476 transcriptions corresponding to the 594 elements previously marked, resulting in an average of approximately 2.5 transcriptions per mark.

Table 7.5 synthesises the distribution of transcriptions corresponding to each type of element (also see Figure 7.5). Based on this and looking at the average number of transcriptions provided for each marked element, we try to understand whether participants show preference for some type of elements in particular.

While characters did attract the largest number of transcriptions (also due to the large number of marks of this type), it is not them, but the objects that generated the highest

Element	Count	Transcriptions-per-Mark
Characters	812	2.56
Events	79	2.39
Objects	252	2.60
Onomatopoeia	41	1.21
Places	23	1.53
Visuals	213	2.17

Table 7.5.: Distribution of transcriptions and per-mark ratios for each type of element.

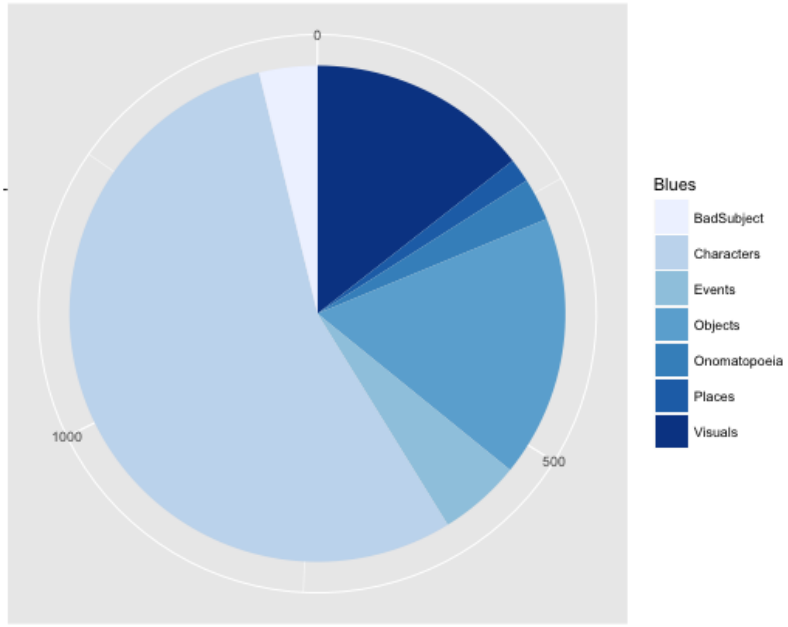


Figure 7.5.: Distribution of transcribed elements.

transcriptions-per-mark ratio. Once again, just like during the marking tasks, we notice an unexpectedly high level of transcriptions of visual cues. Nevertheless, there are no significant differences between the average number of transcriptions-per-marking in the case of characters, events, objects or visual cues. At the other end, the lowest such rates occurred in the case of places and onomatopoeia.

Next, we wanted to investigate the distribution of transcriptions at the page level, in each of the four stories. Just like in the case of marking, such data can be indicative of difficulties in understanding the page or the degree of participants' engagement (e.g., attention, fatigue). Observing Figure 7.6, these distributions seem completely aligned with

	Dash Dillon	Nightro	Shrinking Horror	Truant Toy
Count	370	361	365	380
Avg. per-page	92.5	72.2	52.1	47.5
Avg. per-element	2.8	2.6	3.0	1.9

Table 7.6.: Distribution of transcribed elements for each story.

the data from the **mark workflow** (Figure 7.3), confirming the uniform distributions of transcriptions-per-marking across all stories, just as we can observe it on the last line of Table 7.6.

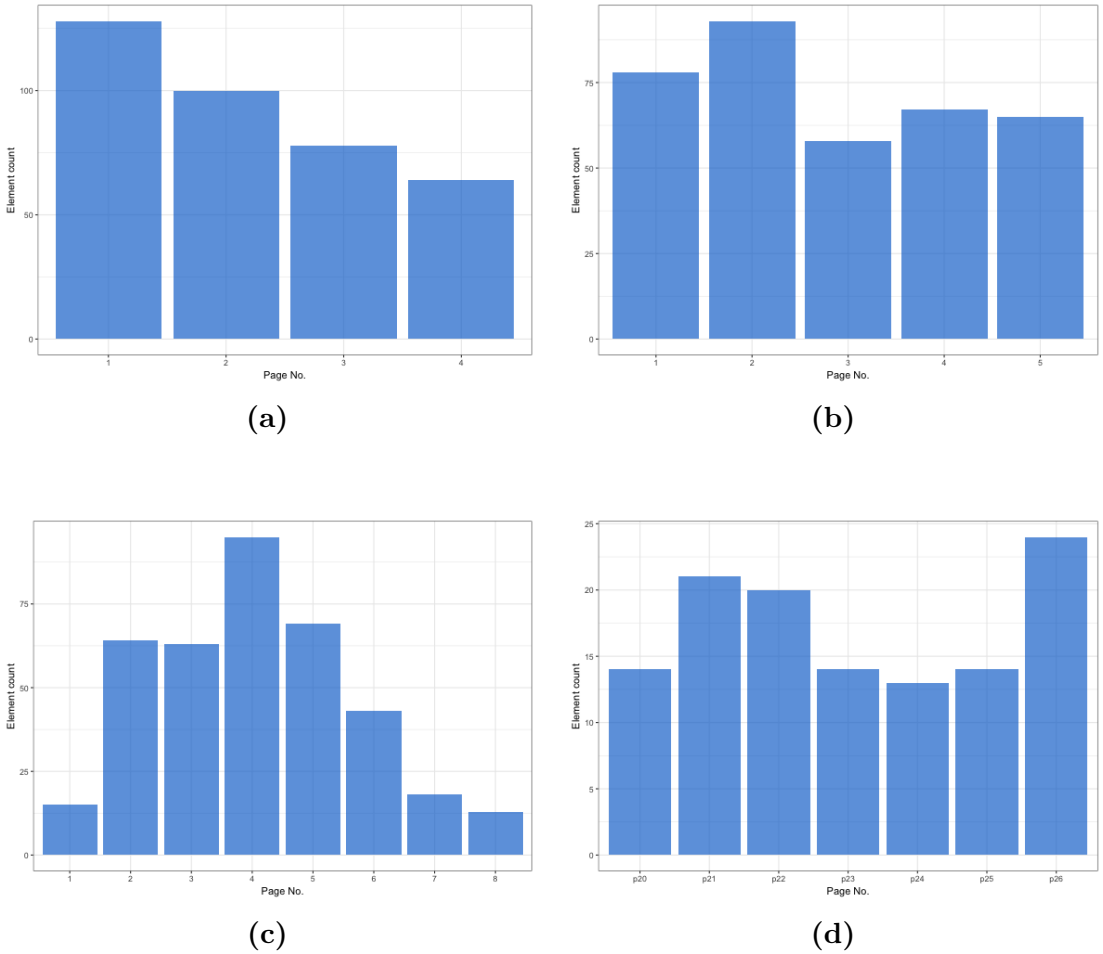


Figure 7.6.: The page distribution of transcribed elements, for each of the stories: a) Dash Dillon, b) Nightro, c) Truant Toy, d) Shrinking Horror.

Last, we try to understand participants' involvement in the transcribe workflow, by observing the daily evolution of the number of collected transcriptions in Figure 7.7. This

allows us to visualise different variations in the number of contributions and to link them to the external events that influenced our experiment. Thus, we observe several activity peaks, which correspond to:

- day 6: the day following the launch of the transcription phase of our experiment;
- day 11: the day we sent a reminder about our experiment to the initial "seed crowd";
- day 14: two experimented users contributed with an important number of transcriptions.

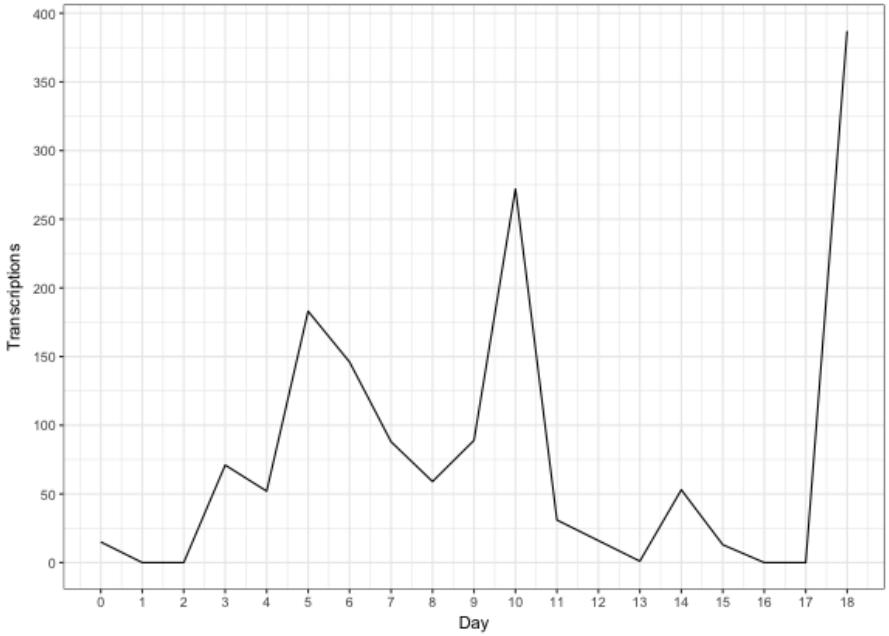


Figure 7.7.: Daily numbers of contributed transcriptions.

7.2.4. Analysis of element types

In this section we take a look at the actual transcription data, that we collected for each type of element – characters, onomatopoeia, places, events, objects and visual cues. We begin with an overview of this data, in Table 7.7 where we report, for each element, the number of completed transcriptions, the number of transcriptions flagged as bad and the average completion time. These reports provide us with additional insight into participants' preferences and difficulties while transcribing elements of each type. For instance, the number of transcribed elements (and their per-mark ratio) can be indicative of participants' willingness to engage with these. On the other hand, bad flag counts are direct identifications of misclassifications, whereas the duration of a task has already been shown to be an indicator of its difficulty (difficult tasks take longer to accomplish) or the

quality of the corresponding contribution (short dwell times are characteristic to spammers contributing with random answers or to a contributor's lack of focus).

	Completed	Bad Flag	Avg. Time (sec)
Characters	559	108	59
Places	18	1	14
Events	61	7	22
Objects	216	10	27
Onomatopoeia	20	12	21
Visuals	151	19	26

Table 7.7.: Transcribed elements: completed vs. bad flags

Based on this, we now carry on by looking into the transcription data corresponding to each type of element.

Characters

The value reported in Table 7.7 (559) corresponds to all character transcriptions, for which at least the name of the character has been provided. The second column (108) is the number of character transcriptions which were flagged as potentially bad at least once. Currently, in our implementation, a subject is withdrawn after receiving bad flags from three different annotators. We also notice the large average completion time (59 seconds) compared to the rest of the elements: this can be explained by the composite nature of the character transcription task (i.e., name transcription, face expression transcription, corresponding balloon transcription), but can also be indicative of participants' good level of engagement.

Next, we assess the potential of our platform to extract the characters from our corpus of stories. Table 7.8 summarises the most frequent character identifying annotations, as well as their number in each story[5].

We are now able to draw several conclusions:

- the shorter stories attracted more participants, which translated into a larger number of character annotations. Even more, there seems to be a statistically significant

[5]The numbers on the second row include all attempted character transcriptions, irrespective of the completed sub-task – name, expression or balloon, as well as flagged marks

Dash Dillon	**Nightro**	**Shrinking Horror**	**Truant Toy**
229	206	193	184
Dash Dillon (25)	Nightro (27)	Dan Turk (41)	Funnyman (29)
Lynn Taylor (21)	Hugh Goddard (23)	dan (21)	June Farrell (19)
Hale player (11)	Tollini (23)	Dan (20)	funnyman (11)
ball player (11)	Hoag (13)	Max (16)	thief (11)
coach Taylor (7)	nightro (8)	policeman (6)	Thief (10)
Lynn (6)	tollini (8)	guard (5)	Larry Davis (9)
Dash (3)	goddard (3)	max (4)	larry davis (5)
Lynn's father (3)	Goddard (3)	watchman (4)	woman (4)

Table 7.8.: Examples of annotated characters.

difference between the total number of identified characters in short stories, compared to those in the long stories[6], which suggests a better engagement with shorter stories;

- for all stories, our method is able to accurately identify the main characters (Dash Dillon, Nightro, Dan Turk, Funnyman), the supporting characters (Lynn Taylor, Tollini and Hoag, max, June Farrell) and even less important ones (ball player, guard, policeman, woman);
- in the case of superhero comics, the method is also able to accurately identify alternate persona of the same character (Nightro/Hugh Goddard, Funnyman/Larry Davis);
- the method needs to be further refined, such that it is able to merge together annotations referring to the same character, irrespective of the way in which they are described (Lynn/lynn/Lynn Taylor) or to identify different roles of the same character (coach Taylor/Lynn's father);
- it is also worth mentioning that some participants provided longer answers which describe the annotated character in more detail, sometimes going as far as describing the character in the course of an action ("Dash getting ready to bat", "Max approaching Dan", "Dan running on the roof of the train"). Future iterations of our method should be able to identify and extract the valuable information from such contributions.

As we previously mentioned, the character transcription task involves two other subtasks: the description of the expression conveyed by the character and the corresponding text balloon. Thus, 157 balloons transcriptions have been submitted with the average length for the transcribed text of 56 characters. Recommendations from comic book theory

[6]According to the Poisson count t-test, at the 95% confidence interval – p-value=0.02.

([Eisner, 1985, Ch. 8]) state that a balloon should contain about 30 words (roughly 150 characters), meaning that our participants, while willing to engage with this rather daunting task, prefer to do so only for shorter text balloons.

Last, we can report that only 116 character transcriptions were fully complete, in the sense that they featured a response for all three sub-tasks.

Onomatopoeia

Just like the character transcription, onomatopoeia were also collected in the form of composite tasks, asking participants to i) simply transcribe the onomatopoeia, and ii) provide a short description of its meaning. From the data we collected, only 20 onomatopoeia markings were finally transcribed, 18 of which fully. 12 others were flagged as badly marked, a rather large proportion of those initially marked. In fact, in most of these cases it seems participants mistook visual effects, such as movement lines or lines portraying loud noises (a ball being hit, a hard impact), for onomatopoeia. In the end, our method allowed for the identification of 14 different onomatopoeia, out of a total of 36 in the entire corpus. In some cases, participants chose to transcribe the description and neglect the rather simple onomatopoeia ("OAAAA" was simply annotated as "yell").

Places

Despite the fact that a very small number of places were identified in all four comics, their transcription yielded encouraging results. From the 18 complete place transcriptions, only one was flagged as bad (and even that refers to a well identified place, with one annotator simply deeming it not important for the narration). Our method showed that we are able to identify both named (Yarvard baseball field) and unnamed places (street, sidewalk, cafe, baseball field, court, forest in the mountains). It remains to be seen how this task can be better designed such that it better engages with the participants. Also, as in the case of character transcriptions, aggregation methods should take into account the different ways of referencing the same place (cafe/Cafe, road/road by the house, Yarvard baseball field/baseball field, street/sidewalk, woods/forest in the mountains).

Events

Another delicate task, because of potential ambiguities resulting either from its design or the way it is interpreted by participants is the transcription of events. A satisfactory amount of 61 event annotations, all providing brief event descriptions, were gathered through our platform. Moreover, a better look at the 7 flagged events gives us reason to believe that there is much potential for this type of task:

- two of the 7 flagged annotations come from "Truant Toy". We have carefully inspected them and from what we can tell, these designate legitimate actions in the story, so the flags are likely to be incorrectly raised; in any case, these annotations have not been deleted from the database, as this occurs only after 3 such flags are raised;
- the remaining five come from "Shrinking Horror", where one participant seems to have accidentally used the events tool to annotate objects.

Another curious observation is that none of the participants in the **mark workflow** have identified events in "Nightro" – a superhero story packed with action scenes[7].

Objects

There was a considerable amount of object transcriptions, following the set of objects marked in the first stage. In total, 216 object-related transcriptions were collected and only 10 of them were flagged as potentially bad. Table 7.9 summarises the most frequently identified objects in each story[8]. What stands out is the particularly large number of objects annotated in "Shrinking Horror", indeed a story abundant in different sorts of objects, both accessories as well as decorative.

Dash Dillon	**Nightro**	**Shrinking Horror**	**Truant Toy**
23	56	123	50
apple (5)	door (5)	door (5)	toy kangaroo (8)
fruit (3)	car (4)	needle (5)	gun (6)
baseball bats (2)	film roll (3)	Florence flask (3)	kangaroo toy (5)
hydrant (2)	microscope (2)	armchair (3)	mirror (3)
	reel (2)	toy train (6)	boxing gloves (2)
	stick (1)	Big needle (2)	diamond ring (2)
			pistol (2)

Table 7.9.: Examples of objects transcriptions in each story.

In fact, as in the case of characters, our method appears to be perfectly capable of identifying both the most important elements in comics stories (apple, baseball bats, needle, toy train, toy kangaroo, diamond ring), as well as decorative and referential elements (hydrant, car, film roll, mirror). However, in contrast with characters annotations, important objects don't necessarily appear multiple times in the content. In fact, in some

[7]We will try to better understand this through our follow-up questionnaire.
[8]The distribution includes all attempted object transcriptions, as well as flagged objects

cases (diamond ring), they might appear only briefly in a certain part of the story (e.g., at the end). Thus, it is difficult to identify the importance of an object simply by looking at its frequency of occurrence.

Additionally, here too, proper aggregation techniques should be put in place to address such issues as generality (apple/fruit), specificity (needle/Big needle), synonyms (film roll/reel) or words order (toy kangaroo/kangaroo toy).

Visual cues

Another task which enjoyed a surprising degree of popularity on our platform was the transcription of visual cues for non-visual phenomena. Our participants submitted 151 such contributions, and another 19 of them were flagged as bad. Most of these cues are classical movement, speed and impact lines, and to a lesser extent, smoke and odour. Our participants most commonly described them by using action verbs, such as: move, swing, fall, throw, smash, hit etc. As we previously discussed, this task stirred a particular interest with our participants and this despite the potential ambiguity of open responses. In this case, it seems that the fear of being wrong (which we think occurred when annotating places or events) faded away in the light of a challenging task during which most participants also learnt something new about comics.

7.2.5. Crowd analysis

This section is dedicated to further analysing the user engagement with our platform. Based on suggestions from other crowdsourcing projects in the humanities ([Dunn et al., 2012, Das Gupta et al., 2016, Wieneke et al., 2016]), we proceed to the identification of super-annotators. Usually, they are found by plotting the histogram of participants' contributions, in decreasing order. At the same time, we are also interested in finding the worse contributors, based on their answers to the easier tasks, and understand the ways in which they behaved during our campaign.

A total of 45 participants engaged with *Comics++* during the two stages of our two weeks experiment. In the introduction to both stages they were specifically asked to log in to our platform using one of the three available authentication modalities: Facebook, Google and Zooniverse. Unfortunately, almost half of the participants didn't follow this advice (see Table 7.10), meaning that we don't have a complete tracking of the actual number of unique participants to our platform. However, since most contributors engaged in one-time working sessions, we believe that the actual number of unique participants is close to the reported value of 45.

Facebook	Google	Zooniverse	Guest
9	12	3	21

Table 7.10.: Participants' choice of authentication modalities.

We proceed to identifying the super-annotators in the mark workflow, by looking at Figure 7.8, which illustrates the total number of elements marked by each user, in decreasing order. The graph follows a power law, a good indicative for the presence of such super-users. We identify two super-annotators (participants `58ad78d4766972687e220200` and `58ad6309766972687e4d0000`), who contributed with 85.85% of the annotations in the **mark workflow**.

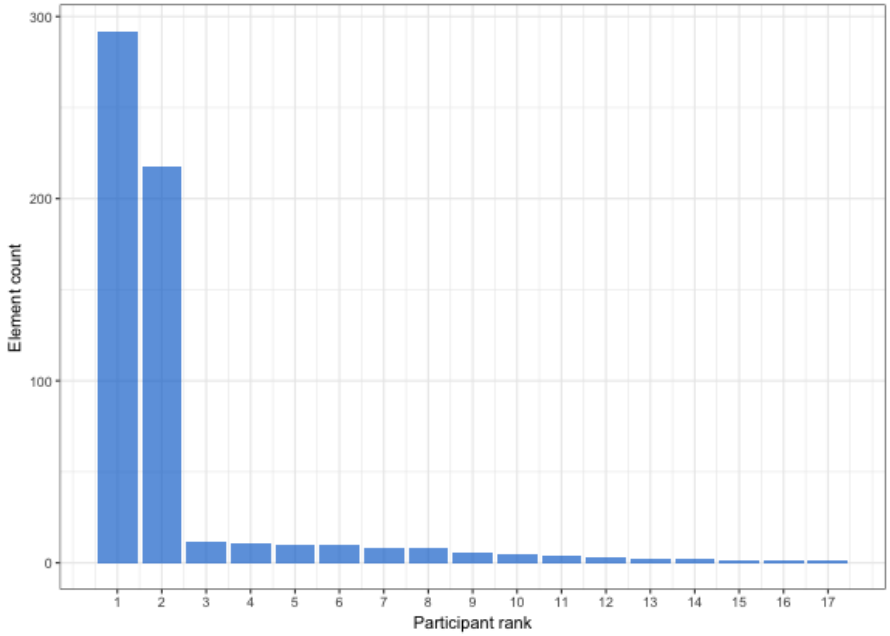


Figure 7.8.: Distribution of marked elements.

We perform the same analysis for the **transcription workflow** (Figure 7.9) and, despite the absence of a clear power law distribution, we are still able to identify two super-transcribers (participants `58b2cea9766972687eea0500`, `58c4272f766972687e850b00`), who account for 34.6% of the total transcriptions. They are, however, closely followed by the next three participants in this workflow, with the top-3 accounting for 45.4%, the top-4 for 53.9% and the top-5 for no less than 61% of the transcriptions.

Finally, we perform a similar analysis considering the total time that participants spent on the platform (Figure 7.10). Once again, there is no power law characteristic

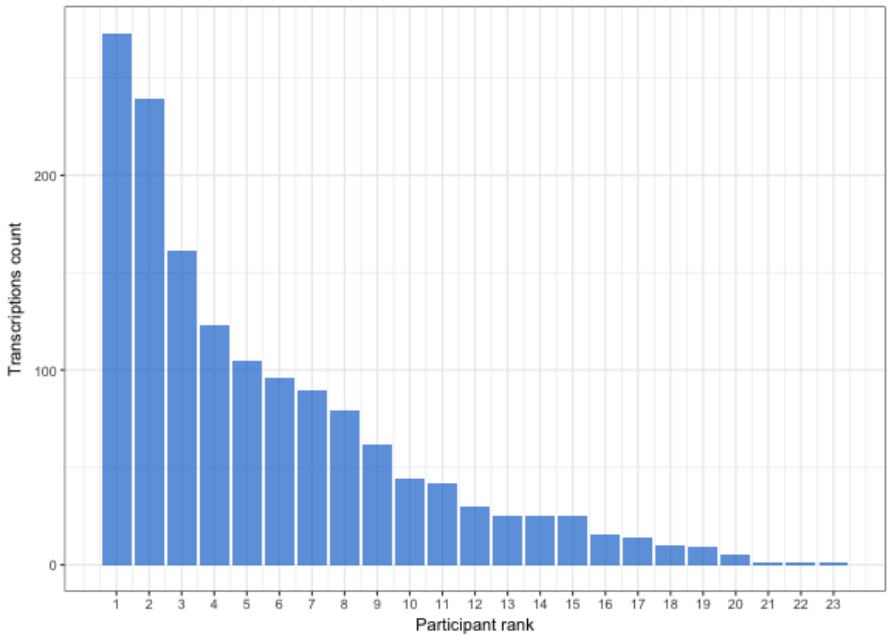


Figure 7.9.: Distribution of transcribed elements.

in our transcription data, but we are able to identify what looks like a super-transcriber (participant `58b33ba4766972687e590600`).

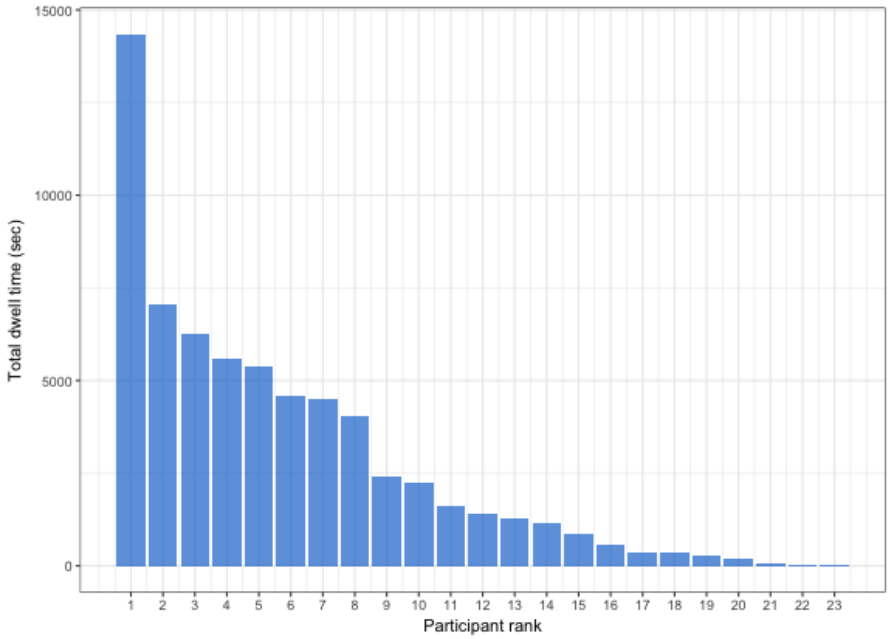


Figure 7.10.: Distribution of transcribed elements.

At the same time, the nature as well as the results obtained in several other tasks, allowed us to use them to better understand the behaviour of our crowd. The tasks we refer to are mainly those for which we could easily create corresponding ground truth, thus rendering them ideal as gold standards tasks for assessing annotator quality ([Chittilappilly et al., 2016]): identifying special panels, identifying splash pages and transcribing onomatopoeia.

Concerning the identification of special panels, the very few correct identifications of positions and types of special panels were provided by only one participant (`58ad6309766972687e4d0000`), who happens to be one of the two super-contributors to the **mark workflow** that we have previously identified. At the same time, looking at the number of incorrect answers provided by the rest of participants, two of them stood out (`58b2cea9766972687eea0500` – 27 misclassifications, and `58b3cc02766972687e9f0600` – 13 misclassifications).

The splash page identification task seemed much easier for our participants, with 13 of them correctly identifying such structures, with our two marking super-contributors being part of this group. The interesting result however, concerns those who didn't manage to correctly accomplish this simple task, with two of the six misclassifying participants also being the two worse performers on the special panels task before.

Finally, based on recommendations for testing data quality ([Grady and Lease, 2010]), we analyse participants' involvement with tasks requiring analytical and interpretative skills. In this case, we are interested in finding out who are the participants that engaged with such tasks as that requiring to explain the meaning of an identified onomatopoeia. We found that from the 10 respondents, all but two were in the top-10 contributors to the **transcription workflow**, with the other two ranking very close as well (positions 12 and 15).

These last four results lead us to believe that a certain polarisation of the crowd could have appeared during our crowdsourcing experiment: on one side, a very involved crowd, constantly providing with good answers, possibly slowly turning into experts; on the other, a less invested crowd, which provided wrong answers even for the most simple tasks. This result isn't necessarily surprising, with similar phenomena being previously reported by [Brabham, 2012] in his *NextStopDesign* crowdsourcing experiment. However, just like in the case of [Brabham, 2010], our crowd is perhaps too small to generalise any results and more data would be needed to perform more complex, inferential analysis.

Chapter 8.

Conclusions

The growing popularity of comics and their development as a medium at the intersection between literature and graphic art supported their expansion into increasingly diverse fields of practice. In publishing, comics continue a century-long tradition and, like the rest of the industry, are currently dealing with the challenges and opportunities of their transgression into the rapidly growing world of digital (and mobile digital). In this respect, one of the most stringent issues concerns the possibility of rendering comics content on the wide variety of existing technological platforms. The emergence of open publication standards partially solved basic rendering issues, but current standards still fail to address more sophisticated use cases, such as pro-actively using the technological capabilities of reading devices to further push the limits of digital comics. Today, such issues are highly dependant on a semantic layer that is either non-existing or extremely low coupled with the available presentation layers. Thus, it is ever so clear that semantic layers, such as that offered through the Comic Book Markup Language (CBML), need to be interfaced with presentation layers, offered by standards like EPUB.

At the same time, in recent years comics have become the focus of activities of different academic communities. Once deeply rooted in art and literature, today, comics studies benefit from the research interest of various scientific communities, which through their respective methodologies help cast new light and further our understanding of the sequential art. This is the case of cognitive science, which through specific experiments (eye-tracking or fMRI) helps us understand reader behaviour and reading patters. In digital humanities, computational approaches to text are now being adapted for studying text stylistics in comics and other forms of sequential art, while research based in media studies focuses on the multi-modal aspect of the comics environment.

In the past 30 years, following the social impact of the "comix" underground movement and their recognition as a literary form through the great productions of "the 1986 *trifecta*" and those of the early 1990s, comics have caught the attention of cultural institutions – Galleries, Libraries, Archives, Museums (GLAM). Today, a worldwide network of museums, libraries, media and research centres are dedicated to the valorisation of comics as objects of cultural heritage. Interestingly, once perceived as opposing the authority, comics have now

become part of (academic) authority, through their critical recognition and their adoption by scholar communities.

In our present work we have shown that the success of all these diverse, comics-centred initiatives (digital publishing, academic research, cultural heritage) and their subsumed activities are highly dependant on the existence of a diverse and richly annotated corpus of comics. While expert annotation can be slow, tedious and sometimes generate unsatisfactory results at a great cost, comics have turned to the computer science research community in search of methods for automatic extraction of graphic and textual content. Currently, these techniques range form simple OCR text extraction, to the identification and extraction of structural elements of the page (i.e., balloons and panels), while most recent machine- and deep-learning techniques seek to identify highly specific content, such as characters. However, we have seen that such techniques still have their limitations in identifying highly specific graphic content (objects, places, events, visual effects and some onomatopoeia) and in some cases, suffer themselves due to the lack of gold standard annotated corpora.

With the advent of crowdsourcing as a problem solving methodology, data collection is in turn directed towards crowds of non-experts. By virtue of its qualities – fast, cheap, making use of large, diverse crowds – we have shown that crowdsourcing is the suitable approach for creating a corpus of annotated comics.

Our work set to provide a characterisation of comics and sequential art for supporting the evolution of digital formats for comics. To this end, our contributions were two-fold and consisted in:

- the development of new open standards for rendering and semantically encoding comics;
- the development of a crowdsourcing platform for annotating comics, with the purpose of generating semantically rich comics encodings.

In the first part of the thesis, we provided the reader with an introduction in the captivating, yet complicated world of comics. We unveiled the cultural background of comics, underlining their quality as a sequential art and their potential for developing complex graphic narratives. We subsequently presented the vocabulary of the medium and argued that due to its structural complexity and their content richness, comics are difficult to encode for any of the purposes previously discussed.

The second part of our work was dedicated to the development of new open standards for comics' publishing. In it, we covered the state of the art of publishing standards and formats, as well as previous attempts to encode comics. We discussed each of their limitations and chose EPUB as the core format for our future development of open standards for comics. The choice was based on EPUB's wide adoption in the industry of digital publishing, as well as its potential for future development and easy coupling with semantic standards. We followed up with a presentation of our contributions to EPUB's different initiatives for

improving the rendering of image-based publications: region-based navigation, rendition mapping and rendition selection. Our main contributions saw us propose the Structural Semantic Vocabulary for comics, a metadata schema that allows for improved region-based navigation on small screen devices. We subsequently addressed one of the main limitations of the set of standards proposed by the EPUB Advanced and Hybrid Layouts working group, by developing an entire specification – EPUB Sequences – for the description of transitions and sequences in image-based publications.

The third part of the thesis was dedicated to the design and development of *Comics++*, a unique system for crowdsourcing comics annotations. We argued our data requirements for extensive annotations of comics, as well as the technical choice of developing the platform using Zooniverse's *scribeAPI*. In the end, we decided to focus the annotation effort on those elements of the comics medium that are difficult to extract automatically: characters, places, events, objects, onomatopoeia, visual cues, special panel shapes and special types of pages. With respect to the four methods used to attract online participants (volunteering, authority, payment and pay-to-use), our project is currently focused on the volunteering component, promoting itself as a citizen science project dedicated to supporting research in comics and sequential art. Alternatively, together with our industrial partner (Actialuna) we considered putting in place a pay-to-use scheme, but these plans are still on-hold due to commercial and intellectual property reasons.

The task design of *Comics++* sees participants engage in two types of activities:

- mark: participants are asked to simply point and click at comic book elements that they identify upon reading a page;
- transcribe: we display previously acquired marks and ask our participants to simply transcribe them.

We concluded our work with the presentation of an experimental study, in which a reduced crowd used the *Comics++* platform to mark and transcribe four copyright-free comic stories, with lengths ranging between 4 – 8 pages. The experiment took place over two weeks, and showed several interesting results. First, the effective task accomplishment time was of less than one hour for the marking tasks and of about 18 hours for the transcription tasks. This confirmed our hypothesis that crowdsourcing can be successfully applied to comics content and, as in other instances, it is a fast and cheap data collection method. The analysis of the crowdsourced data revealed a satisfactory level of quality, as we were able to easily identify all main characters from all four stories, as well as supporting and sometimes referential characters. The same applies to onomatopoeia, which were largely identified despite a smaller engagement by participants with this task. An interesting result was the good level of engagement that we observed on the tasks concerning the identification and transcription of visual cues. Since these elements are highly specific, we feared that participants would be put off by this task; on the contrary, it attracted an important crowd and the quality of its contributions showed us that there is potential to motivate participants to take part in tasks that allow them to acquire new knowledge.

On the contrary, places and events transcriptions were not so successful and we believe that this is due to participants getting confused by our requirement of identifying only those which they judged to be relevant. While this seemed like a good approach, to limit the number of such annotations and over-working our crowd, the effect might have been that participants felt overwhelmed by the idea of making relevance assessments and thus deciding what information is further processed and what is excluded.

One of the most surprising results, concerns the identification of special panels and splash pages. An unexpected number of contributors misclassified normal panels for the different types of special ones, usually based on the border shape of the text balloon comprised in the respective panels. This could be indicative of either lack of attention by participants or of poor task design. Even more surprising is the classification of some pages as double-pages, given that such structures are easily identified by readers and that our corpus didn't feature any.

Nevertheless, we tried to turn these results to our advantage, by using these tasks as a filter for discovering potentially bad annotators. In general, by investigating the total number of contributions and the total dwell time of participants on platform, we were able to isolate a group of several super-contributors who, it turns out, were amongst the few to perform well on the two problematic tasks previously described. On the other hand, participants who contributed with bad data for the same two tasks, found themselves in the tail of these distributions.

8.1. Future work

Unfortunately, one of the limitations of our experimental study is that the data we collected was insufficient to perform any statistically significant inference test. In the following period, the platform will continue to be active and several campaigns to engage a wider audience will be launched, with the hope of acquiring more data.

The results concerning the annotations of comics specific elements are indeed encouraging. We have seen that our crowdsourced method is able to identify all main characters, a large number of the supporting characters and even referential ones; the same applies for identifying some of the main objects in the text as well as places and visual cues. However, aggregation techniques, possibly inspired from named entity recognition and natural language processing are needed to account for the various input we receive for the same data point (e.g., cafe/Cafe, Lynn Taylor/Lynn, Lynn's father/baseball coach, etc.). At the same time, if more data would be provided, we need to assume that not all would have the high quality it currently does. A more diverse crowd could attract more spammers and consequently, in the future we will need to apply methods for detecting and addressing bad annotations and annotators. Some of the work we started in this direction is based on adapting inter-annotator agreement measures to our scenario.

Some of the poor results in our experiments come from the identification of special panels and splash pages. We have previously seen how these tasks were used in our analysis to identify poor annotators. While more data is needed to determine whether the bad results on these tasks are connected to poor design or participants' lack of attention, it is clear that in future designs, such questions could be used as gold standard questions (much like in the case of *CAPTCHA* tasks), allowing us to easily identify potential spammers.

From a design standpoint, we are also investigating the possibility of separating *Comics++* into different annotation and transcription projects. This would see very specific types of tasks (e.g., identification of special panels, splash-pages, double-pages, onomatopoeia or visual cues) be deployed each in individual projects. Thus, we would try to focus the attention of participants on only one specific type of question, with the expectation of better quality annotations and a faster turnaround time.

Currently, we are also waiting to begin future collaborations with different GLAM partner institutions, such as the *Cité Internationale de la Bande Dessinnée et de l'Image* Angoulême. Such an arrangement could see *Comics++* deployed within the centre's online and on-site infrastructure, thus opening the project to CIBD's different visitors. Such a collaboration would benefit the institution from at least two perspectives: on one hand, it could use the collected data to index and enrich its comics catalogue and on the other, it could demonstrate a different level of engagement with its public, at a time when GLAM institutions are pressured to quantify their social impact [Dunn et al., 2012].

One of the most exciting research directions consists in integrating *Comics++* with *Sequencity*, the comics reading and recommendation engine developed by our industrial partner (Actialuna). This would allow us to tap into a larger, more diverse crowd, primarily motivated by the opportunity of reading comics and contributing with their expertise to further understanding the object of their interest. Provided all intellectual property issues are sorted out, such an integration would greatly diversify our corpus and enable us to perform more diverse analysis, taking into account different genres of comics. In this case, participants on the *Sequencity/Comics++* platform could be subject to more complex types of incentives, such as competitions and points systems, allowing them to unlock other comics on the platform, according to their contributions.

Finally, we hope that the collected data will support machine- and deep-learning based approaches for segmentation of comic book pages and extracting features therein. We believe that the rich, annotated corpus of comics resulting from our crowdsourcing approach can prove its usefulness in the form of ground truth for advancements in panel extraction [Arai and Tolle, 2010, Rigaud et al., 2013], identification of special structures, events and scene annotation [Samson et al., 2016], or in the form of training sets for deep-learning approaches to characters extraction and identification [Nguyen et al., 2017] or named entity recognition of characters and places.

Appendix A.

Appendix

A.1. FRBR

The specification of *Maus* [Spiegelman, 2011], a bibliographically complex graphic novel in FRBR. This example demonstrates the various capabilities of the format.

w1.0 Art Spiegelman's story Prisoner on the Hell Planet

partOf: wSOC

e1.0 original story in Short Order Comix #1 (1973)

partOf: eSOC

m1.0.1 the story within Short Order Comix #1 (1973)

m1.0.2 the story collected in Maus – chapter 5 (1983)

w0 Art Spiegelman's *Maus*

hasPart: w1.1, . . . , w1.11

w1.1 "the sheik" story

hasSuccessor: w1.2

e1.1 one (chapter 1)

hasSuccessor: e1.2

m1.1.1 the story published in Raw Vol.1 #2 (December 1980)

m1.1.2 the collected story

...

w1.5 "mouse holes" story

hasPart: w1.0

hasSuccessor: w1.6

e1.5 five (chapter 5)

hasSuccessor: e1.6

m1.5.1 the story in *Raw Vol.1 #6* (May 1984)

hasPart: m1.0.2

m1.5.2 the collected story

hasPart: m1.0.2

w1.6 "mouse trap" story

hasSuccessor: w1.7

e1.6 six (chapter 6)

hasSuccessor: e1.7

m1.6.1 the story in *Raw Vol.1 #7* (May 1985)

m1.6.2 the collected story

w1.7 "Mauschwitz" story

hasSuccessor: w1.8

e1.7 one (chapter 7)

hasSuccessor: e1.8

m1.7.1 the story in *Raw Vol.1 #8* (September 1986)

m1.7.2 the collected story

...

w1.10 "saved" story

hasSuccessor: w1.11

e1.10 four (chapter 10)

hasSuccessor: e1.11

m1.10.1 the story in *Raw Vol.2 #3* (1991)

m1.10.2 the collected story

w1.11 "the second honey moon" (chapter)

e1.11 five (chapter 11)

m1.11 the collected story

e2.1 the original *Maus Vol.1 – My Father Bleeds History* (1986)

hasPart: e1.1, ..., e1.6

m2.1 the book published by Pantheon (1986)

hasPart: m1.1.2, ..., m1.6.2

e2.2 original *Maus Vol.2 – And Here My Troubles Began* (1991)

hasPart: e1.7, ..., e1.11

m2.2 the book published by Pantheon (1991)

hasPart: m1.7.2, ..., m1.10.2 and m1.11

e3.1 the integral *Complete Maus* (1991)

hasPart: e2.1, e2.2

m3.1.1 soft cover 2 volume box set ed. published by Pantheon

hasPart: m2.1, m2.2

m3.1.2 hard cover 2 volume box set ed. published by Pantheon

hasPart: m2.1, m2.2

m3.1.3 hard cover single-vol. ed. published by Pantheon

m3.1.4 hard cover single-vol. 25th anniversary ed. published by Pantheon (2011)

e3.2 the *Complete Maus* (1994) with Vladek's taped transcripts, filmed interviews, sketches, and other background material

m3.2 the CD-ROM edition

e3.3 *MetaMaus* (2011) with further background material, including filmed footage of Vladek, an interview of Art Spiegelman conducted by Hillary Chute, interviews with Spiegelman's wife and children, sketches, photographs, family trees, assorted artwork

supplements: e3.2

m3.3.1 hard cover print version published by Pantheon (2011)

m3.3.2 DVD with an interactive version of *Maus* and video, audio and photos content published by Pantheon (2011)

A.2. EPUB Sequences – Charter Draft

A.3. Comics++: Platform deployment

A.3.1. Deploying the platform

The final step following the implementation and correct configuration of the *scribe* crowdsourcing core is the deployment of the resulting platform. Several sequential steps are needed in order to achieve this:

1. **Configuring the Mongo database.** *Mongo* usually defaults on a `test` database. However, we recommend the configuration of a brand new database, dedicated exclusively to the project. *scribe* is developed using the *mongoid* database connection gem,

allowing for fast interfacing between *Ruby, Rails* and *Mongo* databases. Its standard configuration requires the creation of specific profiles (i.e. development, test, production) in the corresponding `mongoid.yml` file. An example of the respective configuration for the production environment of *Comics++* is given in Figure A.1 below.

```
# mongoid
#
# As discussed in: http://blog.mongolab.com/2014/02/mongodb-driver-tips-tricks-mongoid-3
#
production:
  options:

    raise_not_found_error: false

  sessions:
    default:

      # Use environment variables or a config file to keep your
      # credentials safe.
      database: <%= ENV['MONGO_DB'] || 'scribe_comics_dev_db' %>

      hosts:
        - localhost:27017

      options:

        max_retries: 30
        retry_interval: 1
        timeout: 15
        refresh_interval: 10

test:
  sessions:
    default:
      database: scribe_comics_test_db
      hosts:
        - localhost:27017
      options:
        read: primary
        # In the test environment we lower the retries and retry interval to
        # low amounts for fast failures.
        max_retries: 1
        retry_interval: 0
```

Figure A.1.: Configuration of the "production" and "test" environments of the `mongoid.yml` file for the *Comics++ scribeAPI* project.

2. **Configuring the deployment environment.** The project is deployed on a *Linux* virtual machine functioning on LIP6's internal infrastructure, configured according to *scribe*'s prerequisites for the deployment environment: *Ruby 2.1.5*, *Node JS* and a series *Ruby gems* to be bundled.

3. **Launching the platform.** Prior to launching, the ruby called is compiled with the following command:
   `MONGO_DB=<db_name> rake project:load[comics++] RAILS_ENV=''production''`
   This command essentially tells the ruby compiler to compile all assets, according to

the "production" configurations (within `mongoid.yml`) and connecting to the database referred through the `MONGO_DB` environment variable.

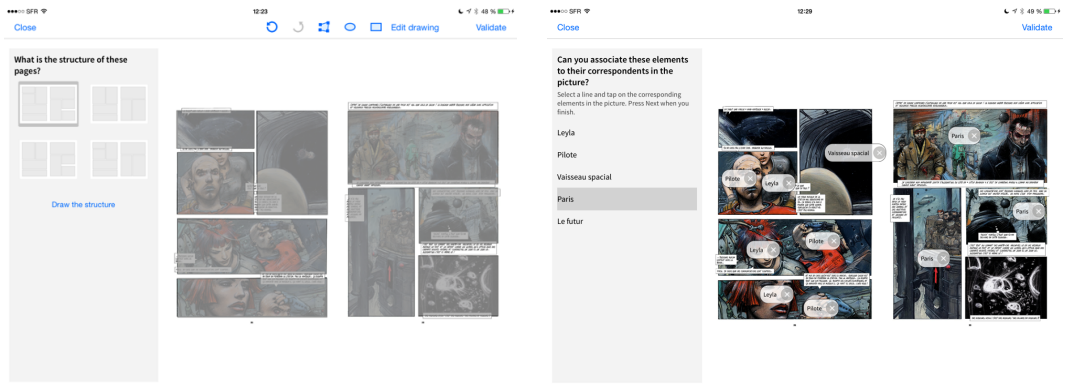
4. **Starting the server.** We are now ready to launch the *Rails* server:
   `MONGO_DB=<db_name> rails s -p <port_no> -e ''production''`
   This launches the rails sever on port number `port_no`, connecting to the database referred through the `MONGO_DB` environment variable and according to the "production" environment configurations.

A.3.2. Comics++: The mobile prototype

One of the first prototypes of our crowdsourcing platform, was a mobile version for iPad, developed by Actialuna. Although it was never rolled in for testing, this tool was great prototype for illustrating the proof of concept of our crowdsourced annotation effort, in the early development stages.

It also focuses on a sequence of tasks, with a few user experience enhancements allowing the annotator to focus on either the whole page or one panel at a time.

Figure A.2 below gives a few examples form the prototyped interface.



(a) (b)

Figure A.2.: Two stages of the Actialuna mobile prototype of *Comics++*: a) page structure annotation step and b narrative elements annotation.

A.4. Comics Code Authority

CODE OF THE COMICS MAGAZINE ASSOCIATION OF AMERICA, INC.[1]

Adopted October 26, 1954

PREAMBLE

The comic-book medium, having come of age on the American cultural scene, must measure up to its responsibilities.

Constantly improving techniques and higher standards go hand in hand with these responsibilities.

To make a positive contribution to contemporary life, the industry must seek new areas for developing sound, wholesome entertainment. The people responsible for writing, drawing, printing, publishing, and selling comic books have done a commendable job in the past, and have been striving toward this goal.

Their record of progress and continuing improvement compares favorably with other media in the communications industry. An outstanding example is the development of comic books as a unique and effective tool for instruction and education. Comic books have also made their contribution in the field of letters and criticism of contemporary life.

In keeping with the American tradition, the members of this industry will and must continue to work together in the future.

In the same tradition, members of the industry must see to it that gains made in this medium are not lost and that violations of standards of good taste, which might tend toward corruption of the comic book as an instructive and wholesome form of entertainment, will be eliminated.

Therefore, the Comics Magazine Association of America, Inc. has adopted this code, and placed strong powers of enforcement in the hands of an independent code authority.

Further, members of the association have endorsed the purpose and spirit of this code as a vital instrument to the growth of the industry.

[1]Retrieved from: https://en.wikisource.org/wiki/Comic_book_code_of_1954

To this end, they have pledged themselves to conscientiously adhere to its principles and to abide by all decisions based on the code made by the administrator.

They are confident that this positive and forthright statement will provide an effective bulwark for the protection and enhancement of the American reading public, and that it will become a landmark in the history of self-regulation for the entire communications industry.

CODE FOR EDITORIAL MATTER

General standards - Part A

1. Crimes shall never be presented in such a way as to create sympathy for the criminal, to promote distrust of the forces of law and justice, or to inspire others with a desire to imitate criminals.
2. No comics shall explicitly present the unique details and methods of a crime.
3. Policemen, judges, Government officials and respected institutions shall never be presented in such a way as to create disrespect for established authority.
4. If crime is depicted it shall be as a sordid and unpleasant activity.
5. Criminals shall not be presented so as to be rendered glamorous or to occupy a position which creates a desire for emulation.
6. In every instance good shall triumph over evil and the criminal punished for his misdeeds.
7. Scenes of excessive violence shall be prohibited. Scenes of brutal torture, excessive and unnecessary knife and gunplay, physical agony, gory and gruesome crime shall be eliminated.
8. No unique or unusual methods of concealing weapons shall be shown.
9. Instances of law-enforcement officers dying as a result of a criminals activities should be discouraged.
10. The crime of kidnapping shall never be portrayed in any detail, nor shall any profit accrue to the abductor or kidnaper. The criminal or the kidnaper must be punished in every case.

11. The letters of the word crime on a comics-magazine cover shall never be appreciably greater in dimension than the other words contained in the title. The word crime shall never appear alone on a cover.

12. Restraint in the use of the word crime in titles or subtitles shall be exercised.

General standards - Part B

1. No comic magazine shall use the word horror or terror in its title.

2. All scenes of horror, excessive bloodshed, gory or gruesome crimes, depravity, lust, sadism, masochism shall not be permitted.

3. All lurid, unsavory, gruesome illustrations shall be eliminated.

4. Inclusion of stories dealing with evil shall be used or shall be published only where the intent is to illustrate a moral issue and in no case shall evil be presented alluringly, nor so as to injure the sensibilities of the reader.

5. Scenes dealing with, or instruments associated with walking dead, torture, vampires and vampirism, ghouls, cannibalism, and werewolfism are prohibited.

General standards - Part C

All elements or techniques not specifically mentioned herein, but which are contrary to the spirit and intent of the code, and are considered violations of good taste or decency, shall be prohibited.

Dialogue

1. Profanity, obscenity, smut, vulgarity, or words or symbols which have acquired undesirable meanings are forbidden.

2. Special precautions to avoid references to physical afflictions or deformities shall be taken.

3. Although slang and colloquialisms are acceptable, excessive use should be discouraged and, wherever possible, good grammar shall be employed.

Religion

1. Ridicule or attack on any religious or racial group is never permissible.

Costume

1. Nudity in any form is prohibited, as is indecent or undue exposure.
2. Suggestive and salacious illustration or suggestive posture is unacceptable.
3. All characters shall be depicted in dress reasonably acceptable to society.
4. Females shall be drawn realistically without exaggeration of any physical qualities.

NOTE. It should be recognized that all prohibitions dealing with costume, dialogue, or artwork applies as specifically to the cover of a comic magazine as they do to the contents.

Marriage and sex

1. Divorce shall not be treated humorously nor represented as desirable.
2. Illicit sex relations are neither to be hinted at nor portrayed. Violent love scenes as well as sexual abnormalities are unacceptable.
3. Respect for parents, the moral code, and for honorable behavior shall be fostered. A sympathetic understanding of the problems of love is not a license for morbid distortion.
4. The treatment of live-romance stories shall emphasize the value of the home and the sanctity of marriage.
5. Passion or romantic interest shall never be treated in such a way as to stimulate the lower and baser emotions.
6. Seduction and rape shall never be shown or suggested.
7. Sex perversion or any inference to same is strictly forbidden.

CODE FOR ADVERTISING MATTER

These regulations are applicable to all magazines published by members of the Comics Magazine Association of America, Inc. Good taste shall be the guiding principle in the acceptance of advertising.

1. Liquor and tobacco advertising is not acceptable.

2. Advertisement of sex or sex instruction books are unacceptable.
3. The sale of picture postcards, pinups, "art studies," or any other reproduction of nude or seminude figures is prohibited.
4. Advertising for the sale of knives or realistic gun facsimiles is prohibited.
5. Advertising for the sale of fireworks is prohibited.
6. Advertising dealing with the sale of gambling equipment or printed matter dealing with gambling shall not be accepted.
7. Nudity with meretricious purpose and salacious postures shall not be permitted in the advertising of any product; clothed figures shall never be presented in such a way as to be offensive or contrary to good taste or morals.
8. To the best of his ability, each publisher shall ascertain that all statements made in advertisements conform to fact and avoid misrepresentation.
9. Advertisement of medical, health, or toiletry products of questionable nature are to be rejected. Advertisements for medical, health, or toiletry products endorsed by the American Medical Association, or the American Dental Association, shall be deemed acceptable if they conform with all other conditions of the Advertising Code.

Bibliography

[Alonso and Mizzaro, 2012] Alonso, O. and Mizzaro, S. (2012). Using crowdsourcing for TREC relevance assessment. *Information Processing & Management*, 48(6):1053 – 1066.

[Amer-Yahia and Senjuti Basu, 2016] Amer-Yahia, S. and Senjuti Basu, R. (2016). Human factors in crowdsourcing. *Proceedings of the VLDB Endowment*, 9(13):1615–1618.

[AnimeNewsNetwork, 2015] AnimeNewsNetwork (2015). Japanese Manga Book Market Rises to Record 282 Billion Yen. Retrieved from http://www.animenewsnetwork.com/news/2015-01-23/japanese-manga-book-market-rises-to-record-282-billion-yen/.83614.

[Arai and Tolle, 2010] Arai, K. and Tolle, H. (2010). Automatic E-Comic Content Adaptation. *International Journal of Ubiquitous Computing*, 1(1):1–11.

[Ariely et al., 2005] Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. (2005). Large stakes and big mistakes. *Review of Economic Studies*, 76(2):451 – 469.

[Aroyo and Welty, 2013] Aroyo, L. and Welty, C. (2013). Crowd Truth: Harnessing disagreement in crowdsourcing a relation extraction gold standard. In *ACM WebSci-13*, Paris, France. ACM.

[Artstein and Poesio, 2008] Artstein, R. and Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Journal of Computational Linguistics*, 34(4):555 – 596.

[Beaty, 2012] Beaty, B. (2012). *Comics versus Art*. University of Toronto Press, Toronto.

[Bechdel, 2006] Bechdel, A. (2006). *Fun Home – A Family Tragicomic*. Mariner Books, USA, june 5, 2007 – reprint edition.

[Blockheel et al., 2002] Blockheel, H., Bruynooghe, M., Dzeroski, S., Ramon, J., and Struyf, J. (2002). Hierarchical multi-classification. In *Proceedings of Workshop on Multi-Relational Data Mining*, pages 21–35, Ljubljana, Slovenia.

[Bongard et al., 2013] Bongard, J. C., Hines, P. D. H., Conger, D., Hurd, P., and Lu, Z. (2013). Crowdsourcing Predictors of Behavioural Outcomes. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(1):176–185.

[Brabham, 2008] Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into*

New Media Technologies, 14(1):75–90.

[Brabham, 2010] Brabham, D. C. (2010). Moving the crowd at threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, 13(8):1122 – 1145.

[Brabham, 2012] Brabham, D. C. (2012). Managing unexpected publics online: The challenge of targeting specific groups with the wide-reaching tool of the internet. *International Journal of Communication*, 6(0):20.

[Brabham, 2013] Brabham, D. C. (2013). Using crowdsourcing in government. Technical report, IBM Center for the Business of Government.

[Brodley and Friedl, 1996] Brodley, C. E. and Friedl, M. A. (1996). Identifying and eliminating mislabeled training instances. In *Proceedings of the 13th National Conference on Artificial Intelligence*, pages 799–805, Portland, Oregon, USA.

[Burwen and de Seve, 2012] Burwen, D. and de Seve, M. (2012). *CIA: Operation AJAX*. Cognito Comics, San Francisco, CA, USA, v.3.0.2 edition. Graphic Novel for iOS.

[Campbell, 2006] Campbell, T. (2006). *The History of Webcomics*. Antarctic Press.

[Cesa-Bianchi et al., 2006] Cesa-Bianchi, N., Gentile, C., and Zaniboni, L. (2006). Incremental algorithms for hierarchical classification. *The Journal of Machine Learning Research*, 7:31–54.

[Chittilappilly et al., 2016] Chittilappilly, A. I., Chen, L., and Amer-Yahia, S. (2016). A Survey of General-Purpose Crowdsourcing Techniques. *IEEE Transactions on Knowledge and Data Engineering*, 28(9).

[COLOURlovers, 2011] COLOURlovers (2011). Color + Design Blog / The Colors of Good vs. Evil: Comic Book Color Palettes [infographic]. Retrieved from http://www.colourlovers.com/blog/2011/09/15/the-colors-of-good-vs-evil-comic-book-color-palettes-infographic.

[Conboy et al., 2015a] Conboy, G., Duga, B., Gardeur, H., Kanai, T., Murata, M., and O'Connor, E. (2015a). EPUB Region-Based Navigation 1.0. Retrieved from http://www.idpf.org/epub/renditions/region-nav/.

[Conboy et al., 2015b] Conboy, G., Duga, B., Gardeur, H., Kroupa, B., Murata, M., and O'Connor, E. (2015b). EPUB Multiple-Rendition Publications 1.0. Retrieved from http://www.idpf.org/epub/301/spec/epub-publications.html.

[Das Gupta et al., 2016] Das Gupta, V., Rooney, N., and Schreibman, S. (2016). Notes from the Transcription Desk: Modes of engagement between the community and the resource of the Letters of 1916. In *Digital Humanities 2016: Conference Abstracts*, pages 476–477, Kraków, Poland.

[Diffallah et al., 2012] Diffallah, D. E., Demartini, G., and Cudre-Mauroux, P. (2012). Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms. In *CrowdSearch 2012 Workshop*, pages 26–30.

[Donmez et al., 2009] Donmez, P., Carbonell G., J., and Schneider, J. (2009). Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 259–268, Paris, France. ACM.

[Dooley, 2013a] Dooley, M. (2013a). The Comic Books that Corrupted our Kids. Print Magazine, Retrieved from http://www.printmag.com/uncategorized/comics-corrupted-our-kids/.

[Dooley, 2013b] Dooley, M. (2013b). Cover Your Eyes: the Graphic Horrors of 1950s Comics. Print Magazine, Retrieved from http://www.printmag.com/uncategorized/1950s-comics/.

[Dunn et al., 2012] Dunn, S., Hedges, M., and Waplington, A. (2012). Crowd-sourcing scoping study: Engaging the crowd with humanities research. Crowd-sourcing scoping study, Centre for e-Research, Department of Digital Humanities Kings College London, London, UK.

[Dunst et al., 2016] Dunst, A., Hartel, R., Hohenstein, S., and Laubrock, J. (2016). Corpus Analyses of Multimodal Narrative: The Example of Graphic Novels. In *Digital Humanities 2016: Conference Abstracts*, pages 178–180, Jagiellonian University & Pedagogical University, Kraków, Poland.

[Eagle, 2009] Eagle, N. (2009). Txteagle: Mobile crowdsourcing. In *Proceedings of the 3rd International Conference on Internationalization, Design and Global Development*, IDGD '09, pages 447 – 456, San Diego, CA, USA. Springer-Verlag.

[Eisner, 1978] Eisner, W. (1978). *A contract with God*. W. W. Norton & Company, New York, USA and London, UK.

[Eisner, 1985] Eisner, W. (1985). *Comics and Sequential Art: Principles and Practices from the Legendary Cartoonist*. W. W. Norton & Company, New York, USA and London, UK, reviewed edition.

[Eisner, 1986] Eisner, W. (1986). *The Dreamer*. W. W. Norton & Company, New York, NY, USA, november 17, 2008 – reprint ed. edition.

[Eisner, 2008a] Eisner, W. (2008a). *Expressive Anatomy for Comics and Narrative: Principles and Practices from the Legendary Cartoonist*. W. W. Norton & Company, New York, USA and London, UK.

[Eisner, 2008b] Eisner, W. (2008b). *Graphic Storytelling and Visual Narrative*. W. W. Norton & Company, New York, USA and London, UK.

[EPUB-31, 2016] EPUB-31 (2016). EPUB 3.1. Retrieved from http://www.idpf.org/epub/31/spec/epub-spec.html. [proposed specification].

[EPUB-AHL, 2012] EPUB-AHL (2012). Charter Proposal: Advanced/Hybrid Layouts. Retrieved from http://www.idpf.org/charters/2012/layout/.

[EPUB-CFI, 2016] EPUB-CFI (2016). EPUB Canonical Fragment Identifiers 1.1. Retrieved from http://www.idpf.org/epub/linking/cfi/epub-cfi.html.

[EPUB-ContentDocs, 2014] EPUB-ContentDocs (2014). EPUB Content Documents 3.0.1. Retrieved from http://www.idpf.org/epub/301/spec/epub-contentdocs.html.

[EPUB-MediaOverlays, 2014] EPUB-MediaOverlays (2014). EPUB Media Overlays 3.0.1. Retrieved from http://www.idpf.org/epub/301/spec/epub-mediaoverlays.html.

[EPUB-OA, 2014] EPUB-OA (2014). Open Annotation in EPUB - Editors Draft. Retrieved from https://docs.google.com/document/d/1imR8GM-UobvcjvxVpLTGgX7thxGpAaHe1QLzgpb7xvU/edit.

[EPUB-OCF, 2014] EPUB-OCF (2014). EPUB Open Container Format (OCF) 3.0.1. Retrieved from http://www.idpf.org/epub/301/spec/epub-ocf.html.

[EPUB-Previews, 2015] EPUB-Previews (2015). EPUB Previews 1.0. Retrieved from http://www.idpf.org/epub/previews/.

[EPUB-Publications, 2014] EPUB-Publications (2014). EPUB Publications 3.0.1. Retrieved from http://www.idpf.org/epub/301/spec/epub-publications.html.

[EPUB-SemVoc, 2016] EPUB-SemVoc (2016). EPUB 3 Structural Semantics Vocabulary. Retrieved from https://idpf.github.io/epub-vocabs/structure/#h_comics.

[Ewalt, 2011] Ewalt, D. M. (2011). The Coolest Costumes Of Comic-Con. Forbes, Retrieved from http://www.forbes.com/sites/davidewalt/2011/07/25/costumes-comic-con-cosplay/.

[Feng et al., 2009] Feng, D., Besana, S., and Zajac, R. (2009). Acquiring high quality non-expert knowledge from on-demand workforce. In *Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources*, People's Web '09, pages 51–56, Singapore. Association for Computational Linguistics.

[Forbes, 2010] Forbes, J. (2010). Guest editorial: Dear Manga, You Are Broken. Retrieved from http://mangablog.mangabookshelf.com/2010/03/26/guest-editorial-dear-manga-you-are-broken/.

[FRBR, 1998] FRBR (1998). Functional requirements for bibliographic records : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. final report viii, IFLA, München, Germany.

[Gale et al., 1992] Gale, W., Church, K. W., and Yankowsky, D. (1992). Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, pages 249 – 256, Newark, Delaware, USA. Association for Computational Linguistics.

[Gordon, 2016] Gordon, I. (2016). Superhero Entertainments. Coursera MOOC, Retrieved from https://www.coursera.org/learn/superhero.

[Grady and Lease, 2010] Grady, C. and Lease, M. (2010). Crowdsourcing document relevance assessment with mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pages 172 – 179, Los Angeles, California. Association for Computational Linguistics.

[Grossberg et al., 2005] Grossberg, L., Wartella, E. A., Whitney, D. C., and Wise, J. M. (2005). *MediaMaking: Mass Media in a Popular Culture*. First Serie. SAGE Publications, Inc, USA, 2nd edition.

[Hirth et al., 2013] Hirth, M., Hoßfeld, T., and Tran-Gia, P. (2013). Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modelling*, 57(11-12):2918 – 2932.

[Howe, 2006] Howe, J. (2006). The Rise of Crowdsourcing. *Wired*, 14(6):1–4.

[Hsueh et al., 2009] Hsueh, P. Y., Melville, P., and Sindhvani, V. (2009). Data quality from crowdsourcing: A study of annotation selection criteria. In *HLT '09 Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, HLT '09, pages 27–35, Boulder, Colorado, USA. Association for Computational Linguistics.

[Huiskes and Lew, 2008] Huiskes, M. J. and Lew, M. S. (2008). The MIR flickr retrieval evaluation. In *MIR '08 Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pages 39–43, Vancouver, Canada. Association for Computing Machinery, Inc.

[JBPA, 2014] JBPA (2014). An Introduction to Publishing in Japan 20142015. Technical report, Japan Book Publishers Association, Tokyo, Japan.

[Joe, 2014] Joe, R. (2014). Digital Comics, Digital Payments. Retrieved from http://www.publishersweekly.com/pw/by-topic/authors/pw-select/article/64160-digital-comics-digital-payments.html.

[Jones, 2005] Jones, G. (2005). *Men of Tomorrow: Geeks, Gangsters and the Birth of the Comic Book*. Basic Books, New York, NY, USA.

[Kanai et al., 2012] Kanai, T., Sorotkin, P., Webster, R., Lester, J., Kroupa, B., Conboy, G., Duga, B., Murata, M., O'Connor, E., Audrain, L., and Gardeur, H. (2012). EPUB 3 Fixed-Layout Documents [deprecated]. Retrieved from http://www.idpf.org/epub/fxl/. [deprecated].

[Kazai and Milic-Frayling, 2009] Kazai, G. and Milic-Frayling, N. (2009). On the evaluation of the quality of relevance assessments collected through crowdsourcing. In *Proceedings of SIGIR Workshop on Future of IR Evaluation*, Boston, USA. Association for Computing Machinery, Inc.

[Kendall, 1938] Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2):81–93.

[Kilgariff, 1998] Kilgariff, A. (1998). Gold standard datasets for evaluating word sense disambiguation programs. *Computer Speech & Language*, 12(4):453–472.

[Kolmogorov, 1933] Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4:83–91.

[Kosinski and Bachrach, 2012] Kosinski, M. and Bachrach, Y. (2012). Crowd IQ: measuring the intelligence of crowdsourcing platforms. In *WebSci '12 Proceedings of the 4th Annual ACM Web Science Conference*, pages 151–160, Evanston, Illinois, USA. ACM.

[MacDonald, 2005] MacDonald, H. (2005). Web Comics: Page Clickers to Page Turners. Retrieved from http://www.publishersweekly.com/pw/print/20051219/35322-web-comics-page-clickers-to-page-turners.html.

[Macias, 2006] Macias, P. (2006). Fans lift J-culture over language barrier. The Japan Times, Retrieved from http://www.japantimes.co.jp/culture/2006/09/07/arts/fans-lift-j-culture-over-language-barrier/.

[Marwaha, 2009] Marwaha, A. (2009). New service is all in a day's SMS. BBC News, Retrieved from http://news.bbc.co.uk/2/hi/technology/7881931.stm.

[McCarthy, 2006] McCarthy, H. (2006). *500 Manga Heroes & Villains*. Barron's Educational Series, London, UK.

[McCloud, 1993] McCloud, S. (1993). *Understanding comics - The invisible art*. William Morrow, An Imprint of HarperCollins Publishers, Inc.

[McCloud, 2000] McCloud, S. (2000). *Reinventing comics*. William Morrow Paperbacks, 1st perennial ed. edition.

[Miller, 1986] Miller, F. (1986). *Batman: The Dark Knight Returns*. DC Comics, New York, NY, USA, 2011 edition.

[Moore and Schneider, 1995] Moore, A. W. and Schneider, J. (1995). Memory-based stochastic optimization. In *Proceedings of Neural Information Processing Systems 8*, pages 1066–1072, Denver, CO, USA. MIT Press.

[Morrison and Cohen, 2005] Morrison, C. T. and Cohen, P. R. (2005). Noisy information value in utility-based decision making. In *Proceedings of the First International Workshop on Utility-based Data Mining*, pages 34–38, Chicago, IL, USA. Association for Computing

Machinery, Inc.

[M'Pouli Njanga Seh, 2016] M'Pouli Njanga Seh, S. (2016). *Annotation automatique des figures de comparaison dans les textes littraires*. Doctoral dissertation, Université Pierre et Marie Curie, Paris 6.

[Nallapati et al., 2012] Nallapati, R., Peerreeddy, S., and Singhal, P. (2012). Skierarchy: Extending the power of crowdsourcing using a hierarchy of domain experts, crowd and machine learning. In *Proceedings of the 21st Text REtrieval Conference (TREC 2012)*, volume TREC 2012 Crowdsourcing Track of *NIST Special Publication*, Gaithersburg, Maryland, USA.

[Nguyen et al., 2017] Nguyen, V., Rigaud, C., and Burie, J.-C. (2017). Comic characters detection using deep learning. In *(work in progress)*, Kyoto, Japan.

[nippon.com, 2015] nippon.com (2015). Japanese Publishing in Free Fall. Retrieved from http://www.nippon.com/en/features/h00092/.

[Nowak and Dunker, 2010] Nowak, S. and Dunker, P. (2010). Overview of the CLEF 2009 large-scale visual concept detection and annotation task. In *Multilingual Information Access Evaluation II. Multimedia Experiments Lecture Notes in Computer Science – 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009*, volume 6242 of *Lecture Notes in Computer Science*, pages 94–109. Springer Berlin Heidelberg.

[Nowak and Lukashevich, 2009] Nowak, S. and Lukashevich, H. (2009). Multilabel classification evaluation using ontology information. In *Proceedings of IRMLeS Workshop*, Heraklion, Greece.

[Nowak and Rüger, 2010] Nowak, S. and Rüger, S. (2010). How reliable are annotations via crowdsourcing? a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the 11th ACM ICMIR '10*, pages 557–566, Philadelphia, USA. ACM.

[Piersanti, 2016] Piersanti, S. (2016). Berrett Koehler Publishers. Retrieved from https://www.bkconnection.com/.

[Quarles, 2012] Quarles, P. (2012). Senate Subcommittee on Juvenile Delinquency: Wertham Versus Gaines On Decency Standards. Retrieved from http://www.wnyc.org/story/215975-senate-subcommittee-juvenile-delinquency-ii/?utm_source=sharedUrl&utm_medium=metatag&utm_campaign=sharedUrl. [Audio Recording].

[Ratier, 2015] Ratier, G. (2015). 2015 : lanne de la rationalisation. Technical report, Association des Critiques et Journalistes de Bande Dessines.

[Rigaud et al., 2013] Rigaud, C., Tsopze, N., Burie, J.-C., and Ogier, J.-M. (2013). Robust frame and text extraction from comic books. *Graphics Recognition. New Trends and*

Challenges Lecture Notes in Computer Science, 7423:129–138.

[Saffire, 2009] Saffire, W. (2009). On language. *New York Times Magazine*. Retrieved from http://www.nytimes.com/2009/02/08/magazine/08wwln-safire-t.html.

[Samson et al., 2016] Samson, C., Fiesler, C., and Kane, S. K. (2016). "Holy Starches Batman!! We are Getting Walloped!": Crowdsourcing Comic Book Transcriptions. In *ASSETS '16 Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 289–290, Reno, Nevada, USA. ACM.

[Saxton et al., 2013] Saxton, G. D., Oh, O., and Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management*, 30(1):2–20.

[Sharma, 2010] Sharma, A. (2010). Crowdsourcing critical success factor model. Retrieved from http://www.crowdsourcing.org/navigate-search?q=ANKIT%20SHARMA.

[Sheng et al., 2008] Sheng, V., Provost, F., and Ipeirotis, P. (2008). Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 614–622, Las Vegas, U.S. Association for Computing Machinery, Inc.

[SNE, 2016] SNE (2016). Syndicat national de l'édition Chiffres clés de lédition. Retrieved from http://www.sne.fr/enjeux/chiffres-cles/.

[Snow et al., 2008] Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). Cheap and fast - but is it good? evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 254 – 263, Honolulu, Hawaii, US.

[Soponis, 2006] Soponis, T. (2006). Publishers Look to Digital Comics. Retrieved from https://web.archive.org/web/20070121124515/http://www.publishersweekly.com/article/CA6393781.html. Archived: 2007-01-21.

[Spiegelman, 2011] Spiegelman, A. (2011). *Maus*. Pantheon Books, New York, 25th anniversary edition edition.

[Surowiecki, 2004] Surowiecki, J. (2004). *The Wisdom of Crowds*. Anchor, USA, reprint edition.

[Tankel and Murphy, 1998] Tankel, J. D. and Murphy, K. (1998). Collecting comic books: a study of the fan and curatorial consumption. In *Theorizing Fandom: Fans, Subculture and Identity*, page 257. Hampton Press.

[Thompson, 2012] Thompson, J. (2012). Why Manga Publishing Is Dying (And How It Could Get Better). Retrieved from http://io9.com/5874951/why-manga-publishing-is-dying-and-how-it-could-get-better. Accessed 2016-09-24.

[Tolle and Arai, 2013] Tolle, H. and Arai, K. (2013). Manga Content Extraction Method

for Automatic Mobile Comic Content Creation. In *Proceedings of the 2013 ICACSIS*, pages 321–328, Bali, Indonesia. Faculty of Computer Science, Universitas Indonesia.

[Vakili, 2012] Vakili, V. (2012). The Worlds Of Books, Connected. Retrieved from http://www.huffingtonpost.com/valla-vakili/small-demons-books_b_1757064.html.

[Verma, 2015] Verma, T. (2015). Laughing through our worries: The Indian web comics. Hindustan Times, Retrieved from http://www.hindustantimes.com/books/laughing-through-our-worries-the-indian-web-comics/story-e6RDl58hD3NGVTKiK5IF0K.html.

[von Ahn and Dabbish, 2004] von Ahn, L. and Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 319–326, Vienna, Austria. Association for Computing Machinery, Inc.

[Walsh, 2012] Walsh, J. A. (2012). Comic Book Markup Language: An Introduction and Rationale. *Digital Humanities Quarterly*, 6(1).

[Wieneke et al., 2016] Wieneke, L., Düring, M., and Guido, D. (2016). Introducing HistoGraph 2: Exploration of Cultural Heritage Documents Based on Co-Occurrence Graphs. In *Digital Humanities 2016: Conference Abstracts*, pages 400–402, Kraków, Poland.

[Wilsey, 2006] Wilsey, R. B. S. (2006). 'Fun Home: A Family Tragicomic,' by Alison Bechdel - The New York Times Book Review. *The New York Times*.

[Woo, 2012] Woo, B. (2012). Understanding understandings of comics: Reading and collecting as media oriented practices. *Participations: Journal of Audiences & Reception Studies*, 9(2):180–199.

List of figures

2.1. Building a definition for comics. [McCloud, 1993] starts from the general definition given in [Eisner, 1985] and gradually tries to refine it. 20

2.2. The lower part of "The Big Triangle" illustrates the evolution of graphic representation from photographic resemblance (left) to iconic abstraction (4th tile from left to right), all the way to full meaning (right), i.e., words. Source: Scott McCloud, *Understanding Comics* [McCloud, 1993, p. 52-53]. 21

2.3. Various stages in the production of grains. The Mortuary Chapel of Menna, superintendent of the estates of the king and of Amen, at the middle of the 18th dynasty, Thebes. Credit: Guity Novin's A History of Graphic Design – The Online TextBook. 22

2.4. Trajan's Column – Battle scene between the Roman and Dacian armies Attributed to Apollodorus of Damascus – Conrad Cichorius: "Die Reliefs der Traianssāule", Erster Tafelband: "Die Reliefs des Ersten Dakischen Krieges", Tafeln 1-57, Verlag von Georg Reimer, Berlin 1896. Credit: Wikipedia, Public Domain. 23

2.5. The messengers with Guy, with portrayal of medieval agriculture in the border. Credit: Wikipedia, CC BY-SA 3.0. 23

2.6. "Life of Raymond Lull" – 14th century manuscript featuring elements such as panels and speech areas similar to speech balloons. By Thomas le Myésier – Karlsruhe, Badische Landesbibliothek, Cod. St. Peter perg. 92, fol. 1v. Manuscript, reproduced in "An illustrated history of the Knights Templar", James Wasserman Credit: Wikipedia, Public Domain. 24

2.7. Mural painting at Voroneţ Monastery (Romania) – 15th century, featuring sequences of panels. Anonymous painter. Credit: Wikipedia, Public Domain. 25

2.8. Cover of Detective Comics #1 (Mar. 1937). Art by Vin Sullivan. Retrieved from `http://thebatman.bravepages.com/comics/dc001.jpg`, Fair use. . 27

2.9. Cover of *Action Comics #1* (Jun. 1938, DC Comics). Art by Joe Shuster (art), and Jack Adler (colour). Source: "Action Comics #1" at The Grand Comics Database. Retrieved October 31, 2006. (similar file if not the one originally uploaded.) Fair use. 28

2.10. The first appearance of Captain America in *Captain America Comics #1* (Mar. 1941). Cover art by Joe Simon (inks and pencils) and Jack Kirby (pencils). Retrieved: Grand Comics Database, Fair use. 29

2.11. Cover of *Atomic War! #1* (Nov. 1952), one of the many comics which reflected Cold War paranoia in 1950s America. Credit: Wikisource, Public Domain. 30

2.12. *Crime SuspenStories #22* (May 1954, EC Comics), art by Johnny Craig. The (in)famous issue used by Sen. Kefauver during the Senate hearing on comics. Retrieved: Grand Comics Database, CC BY 3.0. 31

2.13. Cover of *Showcase #4* (Sep.-Oct. 1956, DC Comics), featuring the first appearance of The Flash. Art by C. Infantino (pencils) and J. Kubert (inks) Source: "Showcase #4" at The Grand Comics Database. Retrieved Dec. 25, 2004, Fair use. 32

2.14. Teams of superheroes were instrumental in DC's and Marvel's efforts to relaunch the genre. Their success greately contributed to the Silver Age of superheroes. a *The Brave and the Bold #28* (Mar. 1960, DC Comics), the first appearance of the Justice League of America. Art by Mike Sekowsky and Murphy Anderson. Retrieved: Grand Comics Database, Fair use. b Cover of *Fantastic Four #1* (Nov. 8, 1961). By J. Kirby (art), S. Lee (writer). Source: Marvel, Fair use. 33

2.15. Covers of some of the most important underground "comix". Although short lived, most of them played an important role in the development of the movement and subsequent evolution of comics. a Mr. Natural visits the City – The famous cover of *Zap Comix #1* (Feb. 1968, Apex Novelties). The flagship magazine of the "comix" underground movement with its own "seal" reading *Fair warning: For Adult Intellectuals Only*, an irony to CCA's seal of approval. Art by R. Crumb. Source: "Zap Comix #1" at The Grand Comics Database. Retrieved Feb. 3, 2009., Fair use. b Cover of *Bijou Funnies #1* (summer 1968, Bijou Publishing Empire), one of the earliest underground "comix". Artwork by J. Lynch. Source: (WP:NFCC#4), Fair use. c Cover of the only issue of superhero parody "comix" *Feds'n'heads* (Feb. 1968). Just like *Zap*, it features a seal-resembling notice of its own *For Englightened Adults Only!*. Art by G. Shelton. Credit: `http://comixjoint.com/fedsnheads-1st.html`, Retrieved Nov. 25, 2016, Fair use. d Cover of satirical "comix" *Young Lust #1* (Oct. 1970, Company & Sons). The title itself is a parody of 1950s romance comics *Young Love*. Art by Bill Griffith and Jay Kinney. Source: (WP:NFCC#4), Fair use. . . 36

2.16. "Humans who look like animals" – the main allegory driving graphic storytelling in *Maus*. In these panels the author draws attention on the absurdity of categorising people and creating racial-based stereotypes. In the end, he acknowledges that "the metaphor self-destructs". Source: *Maus*, Art Spiegelman [Spiegelman, 2011, p. 210], Fair use. 38

2.17. The cover of Book Two of *The Dark Knight Returns* series – "Dark Knight Triumphant". Towards the end of Book One, Bruce Wayne finally embraces his identity as Batman, with repercussions in both the story and its artistic delivery. From this point on, Batman is committed to his mission of defending Gotham at all cost, even if this transforms him into a near-villain. The brutality of his actions is enhanced by the rugged "big-foot" drawing style adopted by the artist. Source: *The Dark Knight Returns*, Frank Miller [Miller, 1986, p. 78], Fair use. 39

2.18. The cover of *Ultimates #13*, art by Brian Hitch. Notice the modern, dynamic and vividly coloured drawings depicting the characters in cinematic postures. Credit: Wikipedia, Fair use. 41

2.19. An illustrative page of Allison Bechdel's graphic style and original use of photographic art in *Fun Home*. Source: *Fun Home*, Allison Bechdel [Bechdel, 2006, p. 102], Fair use. 42

2.20. The type of the panel border supports the action within. a The jagged, irregular border implies an emotionally intense action. b Border-less panels suggest an open, vast or limitless space. c The border can sometime be a structure (window, door, portal) and confines the characters – the opposite effect of border-less panels. d Wavy borders are used to depict thoughts, memories, imaginary or future actions. Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 48-49], Fair use. 45

2.21. In this sequence of panels from *The Spirit* story "Fluid X", we experience an intense emotional moment through the eyes of the character. Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 61], Fair use. . . . 46

2.22. One of the best illustrations of the concept of "ellipsis" and the engagement it demands from the reader, for it is the reader who "drops the axe", and decides who screams and why. Source: Scott McCloud, *Understanding Comics* [McCloud, 1993, p. 66], Fair use. 47

2.23. The six types of transitions between adjacent panels. Source: Scott McCloud, *Understanding Comics* [McCloud, 1993, p. 70-72], Fair use. 48

2.24. A wonderfully complex page from the graphic novella *The Dreamer*, an account of Will Eisner from the early years of American comic books. The page seeks to find the balance between its freedom as a meta-panel with its seemingly borderless panels and the dissimulated references (the window, the wall, even the panels drew by Eyron) to the structural and artistic limits of the panel. Source: Will Eisner, *The Dreamer* [Eisner, 1986, p. 23], Fair use. 50

2.25. The last tier from *Prisoner on the Hell Planet* – the comic story enclosed in *Maus*, Art Spiegelman. A wonderfully built allegory of both the complicated origins of *Maus* – paternity and maternity fuel Spiegelman's artistic endeavour – as well as a reminder of the limits of panels and the comics medium. Source: *Maus*, Art Spiegelman [Spiegelman, 2011, p. 105], Fair use. 51

2.26. The catalyst page in Frank Miller's Book One of *The Dark Knight Returns*. The bat breaking through the window symbolises the end of Bruce Wayne's struggle to accept his role as the Batman, protector of the law. At the same time, the bat is breaking the grid, a metaphor for Miller's reinvention of the Batman. Source: F. Miller, *The Dark Knight Returns* [Miller, 1986, p. 26], Fair use. 52

2.27. An early use of elements of the main comics medium (panels, balloons, motion lines) in F. Opper's *Happy Hooligan* (1905). Copyright 1905 American Journal-Examiner – marked at top right portion of page. Source: Frederick Burr Opper, *The Happy Hooligan* (Apr. 5, 1905) Retrieved: `http://chnm.gmu.edu/aq/comics/spectlex1.html#spect21`, Public domain 53

2.28. Two interesting instances of styling the letters in *Maus*. Notice word telegraphing and the special styling of the letter "S" resembling that used by the Nazi SS. Source: *Maus*, Art Spiegelman [Spiegelman, 2011, p. 176], Fair use. 54

2.29. A few of the most common types of balloons: speech, thought, caption, scream. 55

2.30. In this panel from *The Avengers #32* (2000), the android's speech balloon has a different styling: the straighter shape, the metallic colour and the particular type of border are all suggestive of its mechanical nature. Source: *The Avengers #32*, K. Busiek & G. Pérez (2000), Fair use. 56

2.31. The abstraction process of a drawing, from photographic realism to a simple, commonly understood representation of a human face. Source: Scott McCloud, *Understanding Comics*, [McCloud, 1993, p. 29], Fair use. 56

2.32. A micro-dictionary of common gestures and postures. Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 105], Fair use. 57

2.33. Depending on the message the artist wants to convey, a single representative posture is selected out of a sequence of continuous movements. Had the focus been on the finality of the action, the artist would have chosen the last posture. Source: *Comics and Sequential Art*, Will Eisner [Eisner, 1985, p. 109] 58

2.34. A panel from Will Eisner's story "P'gell from Paris". It describes two of the characters overlooking a street in Istanbul. The city is referenced both by name (in the dialogue), but also through a series of cultural cues (the streets, the landscape etc.) Source: "P'gell from Paris" in *Spirit #21*, Retrieved: Digital Comics Museum, Fair Use 58

2.35. Onomatopoeia describing the sound of an alarm ringing. Source: "The Payoff – Prison!!", in *Startling Terror Tales #12* (1952). Retrieved: Digital Comics Museum, Fair use. 59

2.36. Different uses of motion lines. a Fixed camera perspective – Simple motion lines. Source: "The Grumly Murder" in *The Spirit #21*, Will Eisner (Jun. 1950), Retrieved: Digital Comics Museum, Fair use. b Moving camera perspective. Source: *Understanding comics*, Scott McCloud [McCloud, 1993, p. 113], Fair use. c "Driver's seat" perspective. Source: *Understanding comics*, Scott McCloud [McCloud, 1993, p. 114], Fair use. 60

3.1. Digital comics sales in North America (2009-2015) Source: ICv2 76

3.2. Book and magazine sales in Japan, by value (1981-2013) Credit: nippon.com 76

3.3. Two examples of famous comics parodies: a "Superduperman", in *Mad #4* (May 1953). By W. Wood, H. Kurtzman. A satirical story of Superman and Captain America. In these panels, the author ridicules the famous scene where Clark Kent (here, Clark Bent) changes into Super(duper)man inside a phone booth. Credit: Wally Wood, Fair use. b The first appearance of Howard the Duck, a parody of "funny animals" (e.g., Donald Duck), later featuring in several other stories that satirise famous comics and important moments in comics history. Source: "The Enchanter's Apprentice", in *Adventure into Fear #19* (Dec. 1973). By S. Gerber(w), V. Mayerik (p), S. Trapani (i), S. Goldberg (col.), A. Simek (let.), R. Thomas (ed). New York: Marvel Comics Group., Fair use. 81

3.4. Sample page from the story "The Grumly Murder" in *The Spirit #21*, Will Eisner (Jun. 1950), Retrieved: Digital Comics Museum, Fair use. 84

3.5. a The *CBML* encoding of the comic book page in Figure 3.4. b Customisation of the `<teiHeader>` element, containing a definition of both the list of characters (`<castList>`) and the description of panel-to-panel transitions typology (`interpGrp`) used by the `@ana` attribute of the `<cbml:panel>` element. 85

3.6. Facsimile of the title page of the 2011 hard cover edition of *Maus*. Notice the bibliographic particularities of this publication, i.e. serialisation, progressive publishing and composition (blue), multiple copyright (green). In the bottom half, notice the multiple data and classifications made by the US Library of Congress Cataloguing (blue). Source: *Maus*, Art Spiegelman [Spiegelman, 2011] 87

3.7. "Superman" stories moved from newspaper strips (a), to series of stories in comic books (b), to the creation of the *Superman* comic book (c). More so, the magazine has been renamed and renumbered on several occasions (d) and special issues have been inserted inside the original series. a The first tier of the first Superman Sunday strip, Jerry Siegel (story) and Joe Shuster (art) (Nov. 5, 1939). The Superman strips ran from 1939-1966 and at their peak were published in almost 400 newspapers, with a readership of approximately 20 millions. Source: The Speeding Bullet, Fair use. b Cover of *Action Comics #1* (June 1938, DC Comics). Art by Joe Shuster (art), and Jack Adler (colour). *Action Comics* is the first comic book to host a series of "Superman" stories. Source: "Action Comics #1" at The Grand Comics Database. Retrieved October 31, 2006. (similar file if not the one originally uploaded.), Fair use. c Cover of *Superman #1* (Summer, 1939). Art by Joe Shuster. Following their success in *Action Comics*, "Superman" stories started featuring in their own title comic book starting with 1939. Source: Grand Comics Database, Fair use. d Cover of the last issue of *The Adventures of Superman #649* (Apr., 2006). Art by Ivan Reis. Between 1986-2006, *Superman* magazine was renamed *The Adventures of Superman*, with its "first" issue continuing from #424, while a brand new series used the *Superman* title. In May 2006, after 228 monthly issues, it was returned to its initial title and numbering. Source: Wikipedia, Fair use. 88

3.8. Group 1 Entities in the FRBR model. Source: IFLA. 93

3.9. A description of Art Spiegelman's "Prisoner on Planet Hell" using FRBR. 98

3.10. An extended bibliographic description of Art Spiegelman's *Maus* using FRBR. 99

4.1. The high-level structure of a typical EPUB package (a) and of the `META-INf/container.xml` document (b). Within, notice the reference to `OEBPS/content.opf`, the Open Package document. 104

4.2. The placement of the Content Documents within the EPUB Package (a). It contains all *HTML* files, images, stylesheets, fonts and auxiliary files necessary for the rendering of the *EPUB* file. All these elements come together in the `OEBPS/package.opf` document (b), in which we distinguish three parts: the metadata (orange), the manifest (green – a list of all the content documents in the package) and the spine (blue – the rendering order of the final content documents). 104

4.3. The following example shows two Renditions of *The Sandman* bundled in the same container, one optimised for screens 1920 pixels or wider. The Default Rendition will be used for screen sizes smaller than 1920 pixels by default. Credit: [Conboy et al., 2015b] 108

4.4. The following example shows two Renditions of a magazine bundled in the same container. Note that it is not necessary to state that the Default Rendition is reflowable, since Renditions are reflowable by default. Credit: [Conboy et al., 2015b] 108

4.5. The following example shows a multilingual EPUB Publication, with English, French and Spanish Renditions of the content. Credit: [Conboy et al., 2015b] 109

4.6. The following example shows an EPUB Publication with an image-based Rendition and a text-based serialisation available. Credit: [Conboy et al., 2015b] 109

4.7. The following example shows the `rendition:label` attribute being used to provide a human-readable name for a Rendition. Credit: [Conboy et al., 2015b] 110

4.8. The following example shows a multilingual EPUB Publication with each language in a separate Rendition. [Conboy et al., 2015b, Section 4] 111

4.9. The desired segmentation of a page into composing panels. Source: *Clues Tome 1 – Sur les traces du passé*, Mara (2008), Copyright ©Akileos, Retrieved from [Conboy et al., 2015b, Section 4] 112

4.10. Rendering the page in Figure 4.9 on a cellphone screen, using a semi-transparent letterboxing effect to obscure the surrounding areas. Source: [Conboy et al., 2015b, Section 4]. 114

4.11. The highest level of the `region-based nav` element inside the Data Navigation Document corresponding to the previous example. Notice the use of the `<meta property>` element to describe the characters inside a panel. Six such elements can be found in the EPUB Structural Semantics Vocabulary for comics. 115

4.12. A typical double spread (a) and the corresponding EPUB snippet (b). Notice how both regions (left and right) are starting from the top-left corner of their pages (0, 0) and extend 100% of the width and 50% of the height of each page. Credit: Stan Shaw (Nov. 17, 2014), for *Non Depression Magazine*, Retrieved: Draw Stanley 116

4.13. Illustration of the basic principle behind EPUB Sequences: the navigation between two adjacent navigation units is triggered by one or more user-performed actions and results in at least one (i.e. the page turn) navigation effect. 118

4.14. Preview Publication – The following example shows a minimally complete metadata section, where the Publication is identified as a preview, metadata links back to the source publication and a link to acquire it has also been included. [EPUB-Previews, 2015] 121

4.15. Embedded Preview – The following example shows how a publication can provide a collection that indicates which EPUB Content Documents are included in a preview (here, only part of chapter 3), as well as a manifest of the resources necessary for this preview. [EPUB-Previews, 2015] 122

4.16. A comprehensive example of using Open Annotations in EPUB. Notice the three distinct sections of an annotation: the metadata (red – publisher, creator, description etc.), the annotation target (green – identifies the type and the annotated publication) and the annotation body (blue – holds the actual content of the annotation). [EPUB-OA, 2014] 124

4.17. Two *CBML* extensions for annotating objects on a comic book page. The first solution we propose is based on the introduction of a new `<cbml:object>` element, whereas the second one uses a typed version of the existing *TEI* `<note>` element. 128

4.18. Two methods of describing visual cues in *CBML*. The first method we propose is based on the introduction of a new `<cbml:visualCue>` element, whereas the second one makes use of a double-typed *TEI*-stock `<note>` element. Notice how both of them rely on two typologies (senses described by the visual cues and types of motion perspectives) defined with the help of `<interpGrp>` elements and included in the `<teiHeader>`. 130

4.19. *CBML* encoding of the border type using on the `@ana` attribute of the current `<cbml:panel>` element. The values of this attribute rely on the typology defined in the `<interpGrp>` elements specified in the `<teiHeader>`. Notice how the `@ana` attribute can hold multiple such interpretation values separated by spaces. 131

4.20. *CBML* encoding for scenes spreading two or more pages. The method makes use of the `@ana/interpretation` mechanism applied to a `<div type=''panelGrp''>` element. The values of the `@ana` attribute rely on the typology defined in the `<interpGrp>` elements specified in the `<teiHeader>`. 131

6.1. The elements composing Small Demons' "story-verse". Participants on our platform will create a "comic-verse" by finding at least a subset of these elements: people, places, events and as many relevant objects as possible. . 183

6.2. Animated data flow diagram, giving an overview of the crowdsourcing annotation system. It highlights its main components: the data processing engines (in red), the data storage units (in blue), and the users and stakeholders (in green). See Figure 6.4 for a more detailed view of the *Comics++* crowdsourcing platform. 185

6.3. UML diagram of the *scribe* database, illustrating the collections used in a typical *scribe* crowdsourcing platform. 190

6.4. Data flow diagram of the *Comics++* crowdsourcing platform, illustrating the interplay between workflows, tasks, subjects and generated classifications (annotations). 192

6.5. The file structure of the *Comics++* crowdsourcing platform. 194

6.6. A typical configuration of the `project.json` file, in which the most important fields are `title`, `forum` and `terms_map`. 195

6.7. The outline of the `mark.json` file, composed of a header (green) and the tasks section (blue). 196

6.8. The task requiring a participant to mark the elements of a comic book page (a); the corresponding code (b). 198

6.9. The task requiring a participant to identify panels with special borders (a) and the corresponding code (b). 198

6.10. The task requiring a participant to identify special types of pages (a) and the corresponding code (b). 199

6.11. The task requiring a participant to transcribe the characters (a) and the corresponding code (b). 200

6.12. The task requiring a participant to transcribe onomatopoeia (a) and the corresponding code (b). 201

6.13. The task requiring a participant to transcribe places (a) and the corresponding code (b). 202

6.14. A snippet of a typical subjects *CSV* file, listing the individual subjects (i.e. pages) of a subjects group (i.e. comic book). Observe the last column (`set_key`), which identifies a set of subjects as members of the same group (i.e. `mws-02`). 203

6.15. The home screen of the *Comics++* crowdsourcing platform, featuring a simple design based on concepts drawn from Material and Bootstrap. . . . 205

6.16. A typical screen of the **mark workflow** in *Comics++*. The main areas are: the main menu (blue), the content pane (purple), the toggle bar (red), the question pane (green), the extras (yellow). 206

6.17. A typical screen of the **mark workflow** in *Comics++*. The main areas are: the main menu (blue), the central pane (purple), the annotation pane (green), the extras (yellow). 207

6.18. The colour palettes of our crowdsourcing platforms *Comics++* (a) and *(Dis)Similitudes* (b), based on Material Design principles. 208

7.1. Page structure - Distribution of task accomplishment times, for correctly and misclassified splash-pages. 216

7.2. Total counts for each type of marked element. 217

7.3. The page distribution of marked elements, for each of the stories: a) Dash Dillon, b) Nightro, c) Truant Toy, d) Shrinking Horror. 219

7.4. Daily cumulative number of contributed marks. 220

7.5. Distribution of transcribed elements. 221

7.6. The page distribution of transcribed elements, for each of the stories: a) Dash Dillon, b) Nightro, c) Truant Toy, d) Shrinking Horror. 222

7.7. Daily numbers of contributed transcriptions. 223

7.8. Distribution of marked elements. 229

7.9. Distribution of transcribed elements. 230

7.10. Distribution of transcribed elements. 230

A.1. Configuration of the "production" and "test" environments of the `mongoid.yml` file for the *Comics++ scribeAPI* project. 243

A.2. Two stages of the Actialuna mobile prototype of *Comics++*: a) page structure annotation step and b narrative elements annotation. 244

List of tables

1.1. A list of other worldwide GLAM institutions dedicated to comics. 12

3.1. A selection of comic book-specific attributes corresponding to Group 1 (upper) and 2 (lower) Entities in the *FRBR* model [FRBR, 1998, Chapter 4]. . . . 97

7.1. Special panels by story – annotated vs. actual vs. correct. (*) indicates special panels whose positions were correctly identified, but were assigned to the wrong type. 213

7.2. Special panels by type – annotated vs. actual vs. correct. (*) indicates special panels whose positions were correctly identified, but were assigned to the wrong type. 213

7.3. Page structure markings for pages in each comic. 215

7.4. Distribution of marked elements for each story. 218

7.5. Distribution of transcriptions and per-mark ratios for each type of element. 221

7.6. Distribution of transcribed elements for each story. 222

7.7. Transcribed elements: completed vs. bad flags 224

7.8. Examples of annotated characters. 225

7.9. Examples of objects transcriptions in each story. 227

7.10. Participants' choice of authentication modalities. 229