



HAL
open science

Métamodèle d'analytique des apprentissages avec le numérique

Camila Canellas

► **To cite this version:**

Camila Canellas. Métamodèle d'analytique des apprentissages avec le numérique. Environnements Informatiques pour l'Apprentissage Humain. Sorbonne Université, 2021. Français. NNT : 2021SORUS538 . tel-03824802

HAL Id: tel-03824802

<https://theses.hal.science/tel-03824802v1>

Submitted on 21 Oct 2022

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.



Sorbonne Université

École Doctorale Informatique, Télécommunications et Électronique de Paris

UMR 7606 Laboratoire d'Informatique de Paris 6 | Équipe Modèle et Outils en Ingénierie des
Connaissances pour l'Apprentissage Humain (MOCAH)

MÉTAMODÈLE D'ANALYTIQUE DES APPRENTISSAGES AVEC LE NUMÉRIQUE

par Camila Morais Canellas

Thèse de doctorat d'informatique

Dirigée par Vanda Luengo
Encadrée par François Bouchet et Thibaut Arribe

Soutenue le 08 novembre 2021, devant un jury composé de :
Nathalie GUIN. Maître de conférences HDR. Université Lyon 1. Rapporteur.
Serge GARLATTI. Professeur des universités. IMT Atlantique. Rapporteur.
Tewfik ZIADI. Maître de conférences HDR. Sorbonne Université. Examineur.
Sébastien IKSAL. Professeur des universités. Université du Mans. Examineur.
Vanda LUENGO. Professeur des universités. Sorbonne Université. Directrice de thèse.
François BOUCHET. Maître de conférences. Sorbonne Université. Encadrant.

Botas...

as botas apertadas são uma das maiores venturas da terra, porque,
fazendo doer os pés, dão azo ao prazer de as descalçar.

— Machado de Assis

χαλεπὰ τὰ καλά

RÉSUMÉ

Ce travail s'inscrit dans une démarche d'implémentation d'un processus d'analytique des apprentissages avec le numérique dans un contexte où la production documentaire est réalisée via une approche d'ingénierie dirigée par les modèles. Nous nous intéressons principalement aux possibilités qui pourraient émerger si une même approche est utilisée afin de réaliser une telle implémentation. Notre problématique porte sur l'identification de ces possibilités, notamment en s'assurant de permettre, via le métamodèle proposé, l'enrichissement d'indicateurs d'apprentissage avec la sémantique et la structure des documents pédagogiques consultés par les apprenants, ainsi qu'une définition en amont des indicateurs pertinents.

Afin de concevoir le métamodèle en question, nous avons d'un côté procédé à une étude exploratoire auprès des apprenants afin de connaître leurs besoins et la réception d'indicateurs enrichis. D'un autre côté, nous avons réalisé une revue systématique de la littérature des indicateurs d'interaction existants dans le domaine de l'analytique des apprentissages avec le numérique afin de connaître les éléments potentiellement à abstraire pour la construction d'un métamodèle qui les représente. L'enjeu a été celui de concevoir un métamodèle où les éléments nécessaires à l'abstraction de ce domaine soient présents sans être inutilement complexes, permettant de modéliser à la fois des indicateurs d'apprentissage se basant sur une analyse descriptive et ceux faisant une prévision ou un diagnostic. Nous avons ensuite procédé à une preuve de concept et à une évaluation de ce métamodèle auprès des modélisateurs.

ABSTRACT

This work is part of a process of implementing a learning analytics process in a context where documentary production is carried out via an engineering approach driven by models. We are mainly interested in the possibilities that could emerge if the same approach is used in order to achieve such an implementation. Our problematic concerns the identification of these possibilities, in particular by ensuring to allow, via the proposed metamodel, the enrichment of learning indicators with the semantics and the structure of the educational documents consulted by the learners, as well as an a priori definition of relevant indicators.

In order to design the metamodel in question, we first carried out an exploratory study with learners, aimed at knowing their needs and the reception of enriched indicators. On the other hand, we carried out a systematic review of the literature of existing interaction indicators in the field of learning analytics in order to gather the potential elements to be abstracted for the construction of the corresponding metamodel. The challenge was to design a metamodel where the elements necessary for the abstraction of this domain are present without being unnecessarily complex, making it possible to model both learning indicators based on a descriptive analysis and those used for a prediction or a diagnosis. We then proceeded to a proof of concept and an evaluation of this metamodel.

Abbiamo tutti dentro un mondo di cose : ciascuno un suo mondo di cose! E come possiamo intenderci, signore, se nelle parole ch'io dico metto il senso e il valore delle cose come sono dentro di me; mentre chi le ascolta, inevitabilmente le assume col senso e col valore che hanno per sé, del mondo com'egli l'ha dentro? Crediamo di intenderci; non ci intendiamo mai!

— Luigi Pirandello

REMERCIEMENTS

Je remercie ma famille, qui m'a toujours soutenue, me montrant le chemin de la force et m'inspirant de courage pour continuer. Sem vocês nada faria sentido.

Je remercie celui qui est tous les jours à côté de moi, me montrant qu'aucune difficulté n'est insurmontable et le vrai sens de tous les « co » (collaborer, coopérer, contribuer...).

Je remercie ma directrice et mes encadrants, pour leur générosité, leur patience. Merci de partager vos nombreuses connaissances, théoriques et pratiques, avec moi. Je vous remercie également pour tous les conseils et contributions qui m'ont permis d'avancer.

Merci également à toute l'équipe de Kelis, qui m'a accueillie et m'a fait me sentir chez moi, me faisant confiance, m'inspirant par son professionnalisme et m'encourageant dans ce travail. Je remercie aussi les étudiants, tuteurs et le responsable pédagogique de Libreours, pour leur implication et collaboration.

Enfin, je remercie tous mes amis, qui sont près ou bien loin, que j'ai connus il y a 8 ans ou 8 jours, avec qui j'ai passé peu de temps et que j'admire déjà, ou avec qui j'ai eu le plaisir de vivre longtemps, mais qui sont maintenant loin : mes plus sincères remerciements pour la coexistence, pour les échanges, pour l'apprentissage.

TABLE DES MATIÈRES

I INTRODUCTION

1	INTRODUCTION	3
1.1	Problématique et questions de recherche	5
1.1.1	Problématique	5
1.1.2	Questions de recherche	5
1.2	Plan	6

II CONTEXTE

2	MÉTHODE DE GÉNÉRATION DE SYSTÈMES	11
2.1	Introduction	11
2.2	L'ingénierie dirigée par les modèles	11
2.2.1	La définition d'un modèle	12
2.2.2	La définition d'un métamodèle	13
2.2.3	La définition de transformation	15
2.3	Discussion	16
3	CONTEXTE DE PRODUCTION DOCUMENTAIRE	17
3.1	Introduction	17
3.2	Ingénierie documentaire	17
3.2.1	Chaînes éditoriales numériques	19
3.2.2	Plateformes documentaires — plus que des espaces virtuels pour partager des documents	22
3.2.3	Terrain pratique : Scenari et Libre cours	23
3.3	Discussion	26
4	ILLUSTRATION DE PRODUCTION DOCUMENTAIRE PÉDAGOGIQUE	27
4.1	Introduction	27
4.2	Illustration de production d'un cours	27
4.3	Dans la pratique	28
4.3.1	Étape 1 — Le développeur	28
4.3.2	Étape 2 — Le modélisateur	28
4.3.3	Étape 3 — L'auteur	31
4.3.4	Étape 4 — L'apprenant	32
4.4	Discussion	33

III ÉTAT DE L'ART

5	CADRE D'ÉTUDE	37
5.1	Introduction	37
5.2	Analyser des traces d'apprentissage	37

5.3	Objectifs d'une démarche d'analyse de traces	38
5.3.1	Basés sur la prise de décision	38
5.3.2	Basés sur l'action	39
5.3.3	Basés sur les bénéfiques	40
5.3.4	Discussion	40
5.4	Parties prenantes et niveaux d'analyse	42
5.4.1	Discussion	43
5.5	Les traces d'apprentissage	45
5.5.1	Les types de traces	45
5.5.2	La collecte des traces	47
5.5.3	Discussion	52
5.6	Les méthodes pour analyser les traces	54
5.6.1	Discussion	58
5.7	La visualisation des résultats des analyses	59
5.7.1	Discussion	61
6	VERS UNE MODÉLISATION DE L'ANALYTIQUE DES AP- PRENTISSAGES	63
6.1	Introduction	63
6.2	Cadres et méthodes	63
6.3	Des cadres pour l'analytique des apprentissages	64
6.4	Vers des modèles pour l'analytique des apprentissages	67
IV CONTRIBUTION : LA PRÉPARATION DU MÉTAMODÈLE		
7	LES ÉTAPES ET PARTIES PRENANTES : LES CONTRIBU- TIONS DE CHACUN	73
7.1	Introduction	73
7.2	Une analyse issue de la littérature	73
7.2.1	Identification du besoin	74
7.2.2	Sélection des traces	75
7.2.3	Prise en main des traces	75
7.2.4	Analyse des traces	75
7.2.5	Rendre les résultats visibles	76
7.2.6	Utilisation des résultats	76
7.3	Une analyse suivant l'approche IDM	77
7.3.1	Identification du besoin	77
7.3.2	Identification de l'indicateur	78
7.3.3	Modélisation de l'indicateur	78
7.3.4	Rendre les résultats visibles	79
7.3.5	Utilisation des résultats	79
8	ÉVALUATION EXPLORATOIRE DES BESOINS	81
8.1	Introduction	81
8.2	Énoncé du problème	81

8.3	But de l'étude	81
8.3.1	Questions de recherche	82
8.4	Délimitations et limitations	82
8.5	Méthodologie	83
8.5.1	Procédures de collecte de données	83
8.5.2	Procédures d'analyse des données	84
8.5.3	Structure narrative	84
8.6	Dimensions	86
8.6.1	Dimension 1 — L'utilité perçue	86
8.6.2	Dimension 2 — Gérer différents supports de cours... et les traces	87
8.6.3	Dimension 3 — Des indications « manuelles »	88
8.6.4	Dimension 4 — Naviguer parmi les cours	89
8.6.5	Dimension 5 — Sémantique et structure	90
8.6.6	Dimension 5 — Au menu	92
8.6.7	Dimension 6 — Comparer ou ne pas comparer ?	93
8.6.8	Dimension 7 — Rétroaction	96
8.6.9	Dimension 8 — Accès aux données	96
8.7	Résultats : synthèse des retours	97
8.7.1	(Q1) Quels sont les besoins des utilisateurs en termes d'analytique des apprentissages avec le numérique ? Sont-ils suffisamment homogènes pour un même modèle documentaire ?	97
8.7.2	(Q2) Ces besoins (latents) incluent-ils des informations basées sur la sémantique des documents ?	98
8.7.3	(Q3) Les interprétations des utilisateurs pour chaque graphique affiché sont-elles correctes ?	98
8.7.4	(Q4) Y a-t-il une différence d'interprétation par les utilisateurs entre les visualisations basées sur la sémantique et celles qui ne le sont pas ?	98
8.8	Limitations	99
9	CONNAÎTRE LES INDICATEURS D'APPRENTISSAGE POUR LES MODÉLISER	101
9.1	Introduction	101
9.2	Questions de recherche	101
9.2.1	Dimension 1 — Indicateurs proposés	102
9.2.2	Dimension 2 — Parties prenantes	102
9.2.3	Dimension 3 — Données, analyses et visualisations	102
9.2.4	Dimension 4 — Contexte	103

- 9.3 Méthode 103
 - 9.3.1 Requêtes et stratégie de recherche 103
 - 9.3.2 Sélection des documents 106
 - 9.3.3 Processus de codage 106
- 9.4 Résultats : Dimension 1 — Indicateurs proposés 112
 - 9.4.1 (Q1.1) Quel est le type d'analyse effectuée pour construire un indicateur? 112
 - 9.4.2 Discussion 114
- 9.5 Résultats : Dimension 2 — Parties prenantes 115
 - 9.5.1 (Q2.1) Qui a proposé ou décidé d'ajouter l'indicateur? 115
 - 9.5.2 (Q2.2) Dans des cas où plus d'un acteur est impliqué dans la phase de conception, y a-t-il de la conception participative? 117
 - 9.5.3 (Q2.3) Qui est la cible de l'indicateur proposé? 118
 - 9.5.4 (Q2.4) Avec quelle partie prenante l'indicateur proposé est-il testé? 118
 - 9.5.5 Discussion 120
- 9.6 Résultats : Dimension 3 — Données, analyses et visualisations 121
 - 9.6.1 (Q3.1) Quel artefact numérique est à l'origine des interactions qui ont été tracées? 122
 - 9.6.2 (Q3.2) Le cas échéant, quels algorithmes sont utilisés pour calculer l'indicateur? 123
 - 9.6.3 (Q3.3) Comment les résultats des analyses sont-ils visualisés? 124
 - 9.6.4 (Q3.4) Quelle est la représentation graphique choisie pour l'indicateur? 124
 - 9.6.5 (Q3.5) Où ces visualisations sont-elles proposées? 125
 - 9.6.6 Discussion 126
- 9.7 Résultats : Dimension 4 — Contexte 127
 - 9.7.1 (Q4.1) Quel est le statut (conçu, testé ou déployé) de cet indicateur? 128
 - 9.7.2 (Q4.2) Comment a-t-il été évalué? 129
 - 9.7.3 (Q4.3) Quels sont les domaines d'apprentissage dans lesquels l'indicateur a été proposé? 130
 - 9.7.4 (Q4.4) Quel est le type de dispositif pédagogique en place pour lequel cet indicateur est proposé? 132
 - 9.7.5 (Q4.5) Quelle est la théorie de l'apprentissage sur laquelle repose l'indicateur proposé? 132

- 9.7.6 Discussion 134
- 9.8 Limitations 135

V CONTRIBUTION : LE MÉTAMODÈLE

- 10 UN MÉTAMODÈLE POUR L'ANALYTIQUE DES APPRENTISSAGES 139
 - 10.1 Introduction 139
 - 10.2 Description du métamodèle 139
 - 10.2.1 Modèle d'interaction documentaire 140
 - 10.2.2 Indicateur 141
 - 10.2.3 Visualisation 144
 - 10.3 Modélisation d'indicateurs 144
 - 10.3.1 Indicateur : Temps passé 144
 - 10.3.2 Indicateur : Regroupement 148
 - 10.3.3 Indicateur : Prédiction de réussite 151
 - 10.3.4 Indicateur : Auto-explication 155
 - 10.4 Discussion 159
- 11 PREUVE DE CONCEPT 161
 - 11.1 Introduction 161
 - 11.2 Métriques et fonctionnalités 161
 - 11.2.1 Analyse détaillée d'un document 161
 - 11.2.2 Séparation du contenu et de la forme de visualisation 162
 - 11.2.3 Interopérabilité 162
 - 11.2.4 Conformité à la confidentialité des données 163
 - 11.2.5 Maintenabilité 164
 - 11.2.6 Performance 164
 - 11.2.7 Multiformat 165
 - 11.2.8 Personnalisation 165
 - 11.2.9 Reproductibilité 166
 - 11.3 Discussion 167
- 12 MISE À L'ESSAI ET ÉVALUATION 169
 - 12.1 Introduction 169
 - 12.2 Évaluation du métamodèle par des modélisateurs 169
 - 12.2.1 Énoncé du problème 170
 - 12.2.2 Procédures 171
 - 12.3 Résultats : questionnaire aux modélisateurs 174
 - 12.4 Résultats : entretiens avec les modélisateurs 176
 - 12.4.1 Résultats : synthèse des retours 180
 - 12.4.2 Discussion 181

VI CONCLUSIONS

- 13 CONCLUSIONS 185

13.1	Synthèse	185
13.1.1	RQ1 Peut-on proposer un métamodèle de collecte, d'analyse et de visualisation des traces d'interaction prenant en compte l'ensemble de la chaîne de production de documents?	185
13.1.2	RQ2 Quels sont les composants nécessaires à ce métamodèle permettant de modéliser les indicateurs couramment utilisés dans le domaine, mais également permettant de les enrichir avec la connaissance fine de la sémantique des documents?	186
13.1.3	RQ3 Un modèle documentaire correspond-il à un besoin homogène en termes d'analytique des apprentissages avec le numérique et donc à une série d'indicateurs d'apprentissage bien identifiés?	186
13.1.4	RQ4 Y a-t-il une différence dans la réception des visualisations basées sur des indicateurs enrichis?	187
13.1.5	RQ5 Quels sont les indicateurs utilisés dans les systèmes d'analytique des apprentissages avec le numérique et ses différents éléments potentiellement à abstraire pour la construction d'un métamodèle qui les représente?	187
13.2	Limites et perspectives	187

VII ANNEXES

A	FORMULAIRE DE CONSENTEMENT DE L'ÉTUDE AUPRÈS DES UTILISATEURS	193
B	MAQUETTES DE L'ÉTUDE AUPRÈS DES UTILISATEURS	195
C	FORMULAIRE DE CONSENTEMENT DE L'ÉTUDE AUPRÈS DES MODÉLISATEURS	211
D	QUESTIONNAIRE DE L'ÉTUDE AUPRÈS DES MODÉLISATEURS	213
	BIBLIOGRAPHIE	217

TABLE DES FIGURES

FIGURE 2.1	Relations entre système, modèle, métamodèle et langage, repris et adapté de [52].	13
FIGURE 2.2	Notions basiques de l'IDM, issu de [17].	14
FIGURE 2.3	Pyramide de modélisation de l'OMG, par Jean Bézivin, cité dans [33].	15
FIGURE 3.1	Étapes de fonctionnement d'une chaîne éditoriale, issu de [69].	22
FIGURE 3.2	La suite logicielle Scenari, issu de [42].	24
FIGURE 3.3	L'architecture des composantes du poste client et des serveurs de Scenari.	25
FIGURE 3.4	Le portail de diffusion de Scenari.	25
FIGURE 4.1	Première étape Primitives documentaires.	28
FIGURE 4.2	Première étape Exemple pratique d'une primitive documentaire pour l'inclusion des textes dans les métamodèles.	29
FIGURE 4.3	Deuxième étape Métamodèle utilisant les primitives documentaires.	30
FIGURE 4.4	Deuxième étape Exemple pratique du métamodèle du modèle documentaire Opale.	31
FIGURE 4.5	Troisième étape Conception d'un cours utilisant le modèle documentaire.	31
FIGURE 4.6	Troisième étape Exemple pratique de la conception d'un cours utilisant le modèle documentaire Opale.	32
FIGURE 4.7	Quatrième étape Document généré prêt à être utilisé par les apprenants - cours issu de la plateforme Libre cours de [43].	32
FIGURE 5.1	Les quatre types d'analyses, issu de [59].	39
FIGURE 5.2	Les parties prenantes.	44
FIGURE 5.3	Les éléments du schéma CAM, issu de [148].	48
FIGURE 5.4	Exemple d'activité simple (ajout d'un lien sur un site) avec quelques détails supplémentaires, issu de [131].	49
FIGURE 5.5	Version 4 du <i>Learning Context Data Model</i> , issu de [12].	50

FIGURE 5.6	Deux exemples de déclarations sur l'utilisation de la vidéo, produites par deux outils différents : <i>BestPractice</i> , un outil fictif qui a adopté le profil xAPI pour la vidéo, et <i>Kaltura</i> , qui a implémenté sa propre structure de déclaration xAPI, issu de [84].	52
FIGURE 5.7	Types de méthodes d'analyse, issu de [146] (licence BY-NC-ND 4.0).	55
FIGURE 5.8	Les types de graphiques pour chaque partie prenante, issu de [125].	60
FIGURE 6.1	Dimensions critiques de l'analytique des apprentissages avec le numérique, issu de [67].	64
FIGURE 6.2	Modèle de référence pour l'analytique des apprentissages avec le numérique, issu de [25].	65
FIGURE 6.3	Représentation graphique du framework AL4LD, issu de [73].	67
FIGURE 6.4	Modèle conceptuel du métalangage UTL 2, issu de [30].	68
FIGURE 6.5	Exemple d'un M-Trace, issu de [39].	69
FIGURE 7.1	Les étapes d'implémentation : une analyse issue de la littérature.	74
FIGURE 7.2	Les étapes d'implémentation : une analyse suivant l'approche IDM.	77
FIGURE 9.1	Méthode utilisée dans cette revue systématique de la littérature.	107
FIGURE 9.2	Distribution du nombre d'indicateurs par article analysé.	111
FIGURE 9.3	Nombre d'articles contenant au moins un indicateur selon l'année de publication.	112
FIGURE 9.4	Les types d'analyse de chaque indicateur.	113
FIGURE 9.5	Les types d'analyse de chaque indicateur selon l'année de publication.	114
FIGURE 9.6	Les parties prenantes participant à la sélection d'un indicateur (avec codage multiple possible).	116
FIGURE 9.7	Les parties prenantes participant à la sélection d'un indicateur versus l'année de la publication.	117
FIGURE 9.8	Les parties prenantes cibles d'un indicateur.	118
FIGURE 9.9	Les parties prenantes cibles d'un indicateur versus l'année de la publication.	119

FIGURE 9.10	Les parties prenantes ayant participé à un test d'un indicateur. 119
FIGURE 9.11	Les parties prenantes ayant participé à un test d'un indicateur versus l'année de la publication. 120
FIGURE 9.12	Les parties prenantes impliquées dans les phases du cycle de vie d'un indicateur. 121
FIGURE 9.13	Les artefacts numériques sources des données utilisées lors du calcul des indicateurs. 122
FIGURE 9.14	Les environnements du type LMS sources des données utilisées lors du calcul des indicateurs. 123
FIGURE 9.15	Les modes de visualisation des indicateurs. 124
FIGURE 9.16	Les types de visualisation des indicateurs versus le type d'analyse. 125
FIGURE 9.17	Les environnements où les résultats des calculs (indicateurs) sont visualisés. 126
FIGURE 9.18	Les statuts des indicateurs versus le type d'analyse de l'indicateur. 128
FIGURE 9.19	Les méthodes d'évaluation des indicateurs citées. 129
FIGURE 9.20	Les domaines où l'indicateur a été proposé. 130
FIGURE 9.21	Les domaines et disciplines où l'indicateur a été proposé. 131
FIGURE 9.22	Les types de dispositifs pédagogiques où l'indicateur a été proposé. 133
FIGURE 9.23	Les types de dispositifs pédagogiques où l'indicateur a été proposé versus l'année de publication. 133
FIGURE 10.1	Métamodèle : la création des traces d'interaction. 140
FIGURE 10.2	Métamodèle : la création des indicateurs. 142
FIGURE 10.3	Métamodèle : la création des visualisations. 145
FIGURE 10.4	Modélisation de l'indicateur : temps passé. 146
FIGURE 10.5	Indicateur : le temps passé par un apprenant sur une partie du contenu. 147
FIGURE 10.6	Modélisation de l'indicateur : regroupement. 150
FIGURE 10.7	Indicateur : le regroupement d'apprenants selon leurs interactions avec le contenu. 151
FIGURE 10.8	Modélisation de l'indicateur : prédiction de réussite. 153
FIGURE 10.9	Indicateur : prédiction de réussite. 154

FIGURE 10.10	Modélisation de l'indicateur : auto-explication.	156
FIGURE 10.11	Indicateur : auto-explication.	158
FIGURE 12.1	Métamodèle : implémentation dans l'outil dédié.	169
FIGURE 12.2	Utilisabilité du métamodèle selon les participants.	174
FIGURE A.1	Étude auprès des utilisateurs — Formulaire de consentement	193
FIGURE B.1	Étude auprès des utilisateurs — Maquette d'accès au tableau de bord.	196
FIGURE B.2	Étude auprès des utilisateurs — Maquette 01.	197
FIGURE B.3	Étude auprès des utilisateurs — Maquette 02.	198
FIGURE B.4	Étude auprès des utilisateurs — Maquette 03.	199
FIGURE B.5	Étude auprès des utilisateurs — Maquette 04.	200
FIGURE B.6	Étude auprès des utilisateurs — Maquette 05.	201
FIGURE B.7	Étude auprès des utilisateurs — Maquette 06.	202
FIGURE B.8	Étude auprès des utilisateurs — Maquette 07.	203
FIGURE B.9	Étude auprès des utilisateurs — Maquette 08.	204
FIGURE B.10	Étude auprès des utilisateurs — Maquette 09.	205
FIGURE B.11	Étude auprès des utilisateurs — Maquette 10.	206
FIGURE B.12	Étude auprès des utilisateurs — Maquette 11.	207
FIGURE B.13	Étude auprès des utilisateurs — Maquette 12.	208
FIGURE B.14	Étude auprès des utilisateurs — Maquette 13.	209
FIGURE C.1	Étude auprès des modélisateurs — Formulaire de consentement.	211
FIGURE D.1	Questionnaire de l'étude auprès des modélisateurs (1/3).	213
FIGURE D.2	Questionnaire de l'étude auprès des modélisateurs (2/3).	214
FIGURE D.3	Questionnaire de l'étude auprès des modélisateurs (3/3).	215

LISTE DES TABLEAUX

TABLEAU 5.1	Tableau de comparaison des objectifs.	41
TABLEAU 9.1	Récapitulatif et détail des questions de recherche à l'étude.	104
TABLEAU 9.2	Articles retenus où au moins un indicateur a été trouvé.	109
TABLEAU 9.3	Les approches ou contextes pédagogiques.	134

ACRONYMES

ACM	Association for Computing Machinery
AL4LD	Analytics Layers for Learning Design
APC	Approche par les compétences
APP	Pédagogie par projet
CAM	Contextualised Attention Metada
CLOM	Cours en ligne ouverts et massifs
DSL	Domain-Specific Language
EFLA-4	Evaluation Framework for Learning Analytics
ENA	Epistemic Network Analysis
ERIC	Education Resources Information Center
HTML	Hypertext Markup Language
IDM	Ingénierie Dirigée par les Modèles
IEEE	Institute of Electrical and Electronics Engineers
IHM	Interfaces homme-machine
ITS	Intelligent Tutoring System
JEDM	Journal of Educational Data Mining
JLA	Journal of Learning Analytics
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
K-NN	K-Nearest Neighbors
LA	Learning Analytics
LAK	Conference on Learning Analytics and Knowledge
LCDM	Learning Context Data Model
LMS	Learning Management Systems
LRS	Learning Record Store
MIME	Multipurpose Internet Mail Extensions
MOOC	Massive Open Online Course
PLE	Personal Learning Environments
PDF	Portable Document Format
RGPD	Règlement Général sur la Protection des Données

SCENARI	Système de conception de Chaînes Éditoriales pour des contenus Numériques, Adaptables, Réutilisables et Interactifs
SCORM	Sharable Content Object Reference Model
STEM	Science, Technology, Engineering, and Mathematics
SUS	System Usability Scale
SVM	Support Vector Machine
TBR	Trace-Based Reasoning
TEL	Technology-Enhanced Learning
UEQ	User Experience Questionnaire
URL	Uniform Resource Locator
UTC	Université de technologie de Compiègne
UTL	Usage Tracking Language
VLE	Virtual Learning Environments
WYSIWYG	What you see is what you get
WYSIWYM	What you see is what you mean
XAPI	Experience Application Programming Interface
XIP	Xerox Incremental Parser

Première partie

INTRODUCTION

Cette première partie est dédiée à la présentation de notre problématique de recherche sur la manière d'inclure, dans le cycle d'analyses de traces, une connaissance fine du document structuré. Nous y présentons également les questions de recherche auxquelles nous envisageons de répondre dans la suite de ce manuscrit et présentons le plan de ce dernier.

INTRODUCTION

Avec la démocratisation des ordinateurs et appareils mobiles, les outils permettant l'apprentissage par le biais du numérique ont été largement adoptés avec le développement, par exemple, de plateformes de type LMS (*Learning Management Systems* — LMS, en anglais) ou encore des cours en ligne ouverts et massifs (CLOM — MOOC en anglais). Par conséquent, un besoin de comprendre et d'optimiser cet usage a été mis en évidence [60]. Ceci peut se faire notamment par le biais de traces, d'où la mise en place de systèmes de traçage dans les plateformes susnommées.

Ces traces, produites par les apprenants lors de leurs interactions avec ces systèmes, sont devenues l'objet d'études des chercheurs de plusieurs disciplines — telles que la science des données, les statistiques, l'informatique, la psychologie, entre autres¹ — regroupés autour de ce nouveau champ d'étude, appelé analytique des apprentissages avec le numérique [28], ou *Learning Analytics* (LA) en anglais. Ce champ d'étude relativement nouveau a eu sa première conférence, l'*International Conference on Learning Analytics and Knowledge* (LAK)², en 2011 et fait partie du domaine des technologies de l'information et de la communication appliquées à l'éducation et à la formation (*Technology-Enhanced Learning* — TEL) [25].

L'acte d'introduire l'analytique des apprentissages avec le numérique dans un environnement éducatif est appelé implémentation d'analytique des apprentissages avec le numérique [147]. L'idée est qu'en analysant des traces dérivées de tâches à médiation technologique, des parties prenantes puissent comprendre ce qui se passe lors d'un processus d'apprentissage donné, ce qui représenterait un levier permettant, par exemple, d'apporter des améliorations à un environnement d'apprentissage ou à un cours [92].

Depuis longtemps, des études ont été réalisées afin d'obtenir des informations sur les approches pédagogiques les plus pertinentes en matière de réussite. Cependant, les comportements lors des apprentissages des apprenants eux-mêmes n'étaient pas le centre d'attention [84]. Avec l'adoption de ces supports technologiques, il devient possible d'intégrer davantage les interactions des apprenants (avec les ressources à disposition, leurs collègues ou encore l'instructeur) aux

1. <https://solaresearch.org/about>

2. <https://tekri.athabascau.ca/analytics>

analyses. Mais, les données générées doivent correspondre aux besoins des analyses envisagées, ce qui d’après la littérature n’est pas toujours le cas.

L’objectif principal de l’analytique des apprentissages avec le numérique est alors d’améliorer processus et résultats des apprentissages en s’appuyant sur des traces d’interactions des apprenants avec ces systèmes numériques. Néanmoins, il n’y a pas, actuellement, de consensus sur les interactions effectivement pertinentes pour un apprentissage efficace [1] et naturellement, de consensus sur les traces à enregistrer et à analyser.

Notre travail se situe dans un contexte où la connaissance fine de la structure des documents avec lesquels les apprenants interagissent est connue a priori. Nous soutenons que cette connaissance représente un atout à explorer dans un contexte d’analytique des apprentissages avec le numérique, point sur lequel nous reviendrons et que nous chercherons à démontrer par la suite.

Dans notre contexte, cette connaissance fine des documents est possible, car ces documents structurés sont conçus et publiés avec SCENARI³ (Système de conception de Chaînes Éditoriales pour des contenus Numériques, Adaptables, Réutilisables et Interactifs). Scenari est une suite composée de plusieurs logiciels, unique par sa démarche conceptuelle et le spectre fonctionnel couvert, en particulier grâce à son outil de paramétrage et de modélisation de haut niveau SCENARIBuilder. Ce travail s’appuie sur l’ingénierie dirigée par les modèles (IDM). Dans ce contexte, un outil de modélisation (SCENARIBuilder) permet la modélisation de chaînes éditoriales (SCENARIchain), plateformes de diffusion et applications documentaires à travers un métamodèle. L’enjeu de ce métamodèle est d’abstraire au maximum la technicité des modalités de l’outil conçu. Par exemple, à partir des primitives documentaires — les « briques » qui servent de base pour la création des modèles documentaires —, il est possible de générer un éditeur de documents structurés. En parallèle de ce métamodèle, l’outil de conception inclut des algorithmes de génération qui transforment la modélisation en document (PDF, Web, etc.) matérialisé prêt à l’emploi.

Cette thèse s’inscrit ainsi dans la continuité des travaux de recherche et de développement menés depuis 1999 sur Scenari [8, 49]. En effet, Scenari est un système de conception de chaînes éditoriales numériques issu des laboratoires Costech et Heudiasyc de l’Université de technologie de Compiègne (UTC), prolongé par les projets d’innovation puis par la recherche et développement de Kelis⁴, créée en 2004,

3. <https://scenari.software/fr/>

4. <https://scenari.kelis.fr/index.html>

éditeur de la suite logicielle Scenari. Cette thèse vise à contribuer à la définition d'une génération de données enrichies qui seront utiles à des analyses envisagées préalablement, ainsi qu'à permettre aux utilisateurs finaux de mieux comprendre comment les données seront utilisées et avec quels objectifs.

1.1 PROBLÉMATIQUE ET QUESTIONS DE RECHERCHE

1.1.1 *Problématique*

Notre problématique porte sur la manière d'inclure, dans le cycle d'analyses de traces, la connaissance fine du document structuré. L'idée est de faire valoir cette connaissance, qui est connue en amont, dans des ressources pédagogiques produites par des outils auteurs générés par une approche d'ingénierie dirigée par les modèles.

1.1.2 *Questions de recherche*

Les thématiques qui découlent de notre problématique de recherche se répartissent autour de trois grandes parties : 1/ la proposition d'un métamodèle permettant l'implémentation de l'analytique des traces avec le numérique (**RQ1**) mais également d'indicateurs d'apprentissage basés sur la connaissance fine des documents (**RQ2**). Pour ce faire, nous interrogeons d'abord les futurs utilisateurs ainsi que la littérature afin de comprendre globalement 2/ quels sont les besoins parmi les utilisateurs d'un même modèle documentaire (**RQ3**) et quelle est la réception⁵ de ces indicateurs « enrichis » par cette connaissance des documents (**RQ4**); et enfin 3/ quels sont les indicateurs utilisés dans les systèmes d'analytique des apprentissages avec le numérique et ses différents éléments potentiellement à abstraire pour la construction d'un métamodèle qui les représente (**RQ5**)?

Ainsi, nous avons posé les questions de recherche suivantes :

- **RQ1** Peut-on proposer un métamodèle de collecte, d'analyse et de visualisation des traces d'interaction prenant en compte l'ensemble de la chaîne de production de documents ?
- **RQ2** Quels sont les composants nécessaires à ce métamodèle permettant de modéliser les indicateurs couramment utilisés dans le domaine, mais également permettant de les enrichir avec la connaissance fine de la sémantique des documents ?

5. Nous utilisons le terme en rapport avec l'action de recevoir un message et les processus liés à l'encodage/décodage de celui-ci. Cf. https://en.wikipedia.org/wiki/Reception_theory

- **RQ3** Un modèle documentaire donné correspond-il à un besoin homogène en termes d’analytique des apprentissages avec le numérique et donc à une série d’indicateurs d’apprentissage bien identifiés ?
- **RQ4** Y a-t-il une différence dans la réception des visualisations basées sur des indicateurs enrichis ?
- **RQ5** Quels sont les indicateurs utilisés dans les systèmes d’analytique des apprentissages avec le numérique et ses différents éléments potentiellement à abstraire pour la construction d’un métamodèle qui les représente ?

Notre objectif est alors d’apporter une proposition d’analytique des apprentissages avec le numérique considérant les informations fines relatives aux documents comme importantes et potentiellement capables d’enrichir ces analyses. Ainsi, l’ambition à terme, est de disposer des artefacts permettant la création d’indicateurs d’apprentissage qui permettent non seulement de définir la collecte, l’analyse et la visualisation des résultats, mais également que ces indicateurs soient enrichis des caractéristiques des documents à l’origine des interactions ayant produit les traces analysées.

Pour cela, nous proposons d’abstraire les différents éléments liés à l’analytique des apprentissages avec le numérique (primitives d’analytique) via un métamodèle, afin de concevoir des indicateurs d’apprentissage. À la différence des approches décrites dans l’état de l’art, nous proposons que la définition des modèles d’indicateurs se fasse en parallèle avec la définition des modèles documentaires (et non a posteriori), ce qui rend possible la prise en compte des informations fines des documents structurés en question. Cette approche a l’avantage de permettre la définition des traces nécessaires (et de leurs contenus), inversant ainsi la démarche habituelle [84] de partir des traces pour arriver à des indicateurs.

1.2 PLAN

Afin de répondre aux questions qui découlent de notre problématique, nous avons structuré notre travail et les prochaines parties et chapitres du manuscrit selon le plan qui suit.

Tout d’abord, nous proposons de nous tourner vers le contexte de notre étude dans la mesure où ce contexte, très particulier, a des implications fortes pour la suite des propositions. Ce contexte s’inscrit dans l’approche de l’ingénierie dirigée par les modèles (IDM). Il s’agit à la fois de la méthode utilisée pour la production des documents — dont l’utilisation en pédagogie sera l’objet d’analytique des appren-

tissages avec le numérique — mais également de la méthode qui nous permettra de métamodéliser les composants nécessaires pour mettre en œuvre la création des traces, leur analyse et la visualisation de ces résultats. Cette approche est présentée dans le Chapitre 2. Ensuite, comme nous le verrons dans le Chapitre 3, les chaînes éditoriales numériques représentent une solution au sein de l'ingénierie documentaire à la nécessité de la production de documents en masse et à la particularité d'avoir une tendance à produire des documents structurés physiquement, logiquement et sémantiquement. Enfin, dans le Chapitre 4, nous proposons d'illustrer de manière simplifiée la création d'un cours via des chaînes éditoriales numériques — cela permettra de comprendre les différentes étapes et acteurs impliqués.

Dans la troisième partie, nous interrogeons la littérature autour du déploiement de systèmes d'analytique d'apprentissages avec le numérique afin de comprendre, d'une part, les verrous scientifiques encore présents et, d'autre part, d'éclaircir les dimensions à prendre en compte lors du déploiement de ces solutions d'analytique d'apprentissages avec le numérique (objet de notre métamodèle). D'abord, dans le Chapitre 5, nous nous intéressons notamment aux trois moments identifiés — collecte, analyse, visualisation — permettant d'aboutir à des indicateurs d'apprentissage pertinents. Nous nous intéressons également aux parties prenantes qui jouent un rôle dans l'implémentation et l'usage de tels systèmes, aux objectifs de celles-ci ainsi qu'aux niveaux d'analyse. Enfin, dans le Chapitre 6, nous analysons les travaux allant vers la modélisation de tels systèmes.

Dans la quatrième partie, Chapitre 7, nous proposons la formalisation du cycle d'analytique des apprentissages avec le numérique, ce qui nous permettra de comparer avec les phases d'une même implémentation via une approche par l'IDM. Dans le Chapitre 8, nous présentons une étude préliminaire qui nous a permis de répondre aux questions de recherche **RQ3** et **RQ4**, où l'objectif est de comprendre les besoins de futurs utilisateurs des résultats des analyses utilisant notre métamodèle. Nous nous intéressons notamment à leur homogénéité, ou non, en termes de besoins en ayant comme prémisse leur usage d'un même modèle documentaire. Enfin, dans le Chapitre 9, nous présentons les résultats d'une revue systématique de la littérature s'intéressant aux différents indicateurs utilisés dans des systèmes proposant l'analytique des apprentissages avec le numérique. Cette étude nous permet de répondre à **RQ4** et ces résultats sont importants afin de connaître de manière précise les nombreuses caractéristiques de ces indicateurs et leurs contextes d'usage, connaissance indispensable pour la proposition de notre métamodèle.

Dans la cinquième partie, le Chapitre 10 est dédié à la présentation du métamodèle que nous proposons et qui répond à **RQ1** et **RQ2**. Nous présentons d'abord une description détaillée de ce métamodèle et les raisonnements derrière certaines décisions que nous avons prises. Ensuite, nous illustrons l'utilisation de ce métamodèle via la modélisation de quatre indicateurs. Dans le Chapitre 11 nous décrivons une série de contraintes que nous avons pour la création du métamodèle et comment elles peuvent influencer la mise en place d'un tel système en suivant l'approche proposée. Le Chapitre 12, est dédié à la communication des résultats d'une étude avec des modélisateurs ayant testé ce métamodèle suite à son déploiement dans un outil dédié.

Enfin, dans le Chapitre 13 nous présentons les conclusions que nous avons pu retirer de ces études et propositions. Nous présentons également les perspectives d'utilisation de ces résultats et de nouvelles questions de recherche qui pourraient être explorées.

Deuxième partie

CONTEXTE

Tout d’abord, dans cette deuxième partie nous proposons de nous tourner vers le contexte de notre étude dans la mesure où ce contexte, très particulier, a des implications fortes pour la suite des propositions. Ce contexte s’inscrit dans une approche de l’ingénierie dirigée par les modèles. Il s’agit à la fois de la méthode utilisée pour la production des documents — dont l’utilisation en pédagogie sera l’objet d’une analytique des apprentissages avec le numérique — mais également de la méthode qui nous permettra de métamodéliser les composants nécessaires pour mettre en œuvre la création des traces, leur analyse et la visualisation de ces résultats. Cette approche est présentée dans le Chapitre 2. Ensuite, comme nous le verrons dans le Chapitre 3, les chaînes éditoriales numériques représentent une solution au sein de l’ingénierie documentaire à la nécessité de la production de documents en masse et à la particularité d’avoir une tendance à produire des documents structurés physiquement, logiquement et sémantiquement. Enfin, dans le Chapitre 4, nous proposons d’illustrer de manière simplifiée la création d’un cours via des chaînes éditoriales numériques — cela permettra de comprendre les différentes étapes et acteurs impliqués.

2.1 INTRODUCTION

Plusieurs approches et méthodes ont été développées afin de définir des étapes, démarches ou manières de faire raisonnées, suivies pour parvenir à un but. La complexité croissante des systèmes informatiques a été un facteur clé menant vers l'apparition de l'Ingénierie Dirigée par les Modèles (IDM). Cette complexité pourrait être mieux maîtrisée via l'abstraction des différentes parties d'un système. Dans ce processus d'abstraction, l'idée est d'en garder uniquement les aspects de ce que l'on croit effectivement nécessaires au fonctionnement du système. Dans ce chapitre, nous verrons les principes de fonctionnement de cette approche afin d'analyser, par la suite (cf. Chapitre 3), comment elle a été utilisée par l'ingénierie documentaire avec l'apparition des chaînes éditoriales numériques, avec lesquelles nous travaillerons dans le cadre de cette thèse.

2.2 L'INGÉNIERIE DIRIGÉE PAR LES MODÈLES

Dans un contexte de développement de systèmes complexes, l'ingénierie dirigée par les modèles est une pratique qui permet de se concentrer sur un niveau plus abstrait que la programmation classique [33]. Cette pratique permet de décrire à la fois le problème posé et sa solution.

En informatique, la modélisation est utilisée pour tenter de contrôler une telle complexité [77], à la fois pour créer un logiciel ou pour le valider. Dans l'ingénierie axée sur les modèles, l'objectif est de créer et d'exploiter des modèles de domaine — dans notre cas, le domaine éducatif et plus particulièrement l'analytique des apprentissages avec le numérique — en tant que modèle conceptuel prenant en compte tous les sujets liés à cette tâche/domaine spécifique.

On s'est alors appuyé sur la modélisation, inhérente à l'IDM, pour essayer de maîtriser cette complexité, tant pour produire le logiciel (conception) que pour le valider (test). L'IDM est une forme d'ingénierie générative [33]. Elle a la particularité de suivre une démarche par laquelle tout ou partie d'une application informatique est généré à partir de modèles.

Dans le but d'atteindre un objectif donné, certains aspects de la réalité (ou une solution à un problème) sont simplifiés ; en d'autres mots, modéliser requiert forcément une simplification par rapport à la réalité. Cette modélisation peut également être utilisée pour séparer les différents besoins fonctionnels et les préoccupations extra-fonctionnelles (comme la fiabilité, les performances, etc.).

2.2.1 *La définition d'un modèle*

Un modèle est donc un concept ou un objet, perçu comme représentatif d'un autre. Cela peut être plus ou moins abstrait, plus ou moins précis. En informatique, un modèle est souvent utilisé pour représenter la conception ou le fonctionnement d'un ensemble de programmes informatiques.

Jézéquel et al. [77] proposent la définition suivante dans un contexte informatique, que nous adoptons pour la suite de notre travail :

DÉFINITION DE MODÈLE. « Un modèle est un ensemble de faits caractérisant un aspect d'un système dans un objectif donné. Un modèle représente donc un système selon un certain point de vue, à un niveau d'abstraction facilitant par exemple la conception et la validation de cet aspect particulier du système » [77, p. 4].

Un modèle suit le principe de substituabilité, il doit permettre de répondre à certaines questions en lieu et place du système qu'il est censé représenter, de la même façon que le système aurait répondu lui-même.

Il est difficile de répondre à la question de ce qui constitue un bon modèle, mais il est possible d'affirmer qu'il doit avoir le bon niveau d'abstraction afin de représenter le système, facilitant sa conception et sa validation [77].

Un modèle est une ressource opérationnelle créée par le modélisateur qui conçoit le système en utilisant les différents composants d'un métamodèle. La méthode consiste à générer des artefacts (code si l'on se réfère au développement logiciel, mais d'autres domaines peuvent également être adressés) en utilisant un modèle validé par le métamodèle et un algorithme de transformation.

2.2.2 La définition d'un métamodèle

Le métamodèle est celui qui permet de capitaliser un domaine de connaissances [77]. Il peut être compris comme le langage permettant d'exprimer l'ensemble de modèles constituant le système.

DÉFINITION DE MÉTAMODÈLE. « Un métamodèle est un modèle qui définit le langage d'expression d'un modèle [109], c'est-à-dire le langage de modélisation » [77, p. 10].

Le métamodèle est un modèle qui va, à son tour, permettre la représentation des modèles, c'est-à-dire, un « modèle des modèles » où un modèle est une instance de son métamodèle. Créer un métamodèle implique alors l'analyse et la définition de cadres, règles et contraintes utiles et nécessaires à la définition d'une classe définie de problèmes d'un domaine ou métier spécifique. Il s'agit d'étudier les composants, acteurs et relations afin d'en arriver à un métamodèle qui permet de représenter chacun de ses éléments.

Un métamodèle peut être utilisé en tant que langage prenant en charge une méthode ou un processus particulier, comme un schéma pour exprimer une sémantique supplémentaire à des données préalablement enregistrées, en tant que mécanisme pour créer des outils qui fonctionnent avec une série de modèles, etc.



FIGURE 2.1 – Relations entre système, modèle, métamodèle et langage, repris et adapté de [52].

Sur la Figure 2.1, nous avons l'exemple repris de [52] visant à éclaircir ces concepts et les relations entre eux, en s'appuyant sur la cartographie. La carte est ici un modèle, une interprétation de la réalité. Mais cette interprétation n'est pas sans intention : elle rend abstrait un

aspect en particulier de cette réalité, qui est le découpage administratif du territoire en régions et départements. Dans l'exemple cartographique, c'est alors la légende — le code couleur qui représente une région ou un département, par exemple, ou encore l'échelle utilisée ou l'orientation — qui permet de comprendre la carte, mais qui indique également le langage que l'on a suivi afin d'élaborer la carte. « Le langage dans lequel ce modèle est exprimé doit donc être clairement défini. De manière naturelle, la définition d'un langage de modélisation a pris la forme d'un modèle, appelé métamodèle » [33, p. 2]. Le modèle doit ainsi être conforme au métamodèle, c'est la relation (*conformsTo*) représentée par χ dans la Figure 2.1. La légende, indispensable en cartographie pour que l'on puisse savoir comment interpréter les cartes en question, a aussi une relation avec l'ensemble des cartes construites en utilisant ce langage. La légende est ainsi définie comme un modèle représentant (*representedBy*, μ) cet ensemble de cartes [33]. Enfin, chaque carte ainsi conçue est un élément appartenant (ϵ) à cet ensemble.

En plus de la notion qu'en IDM « *Everything is a model* », dans [17] Bé-zivin proposait deux relations primaires pour cette approche : *conformsTo* et *representedBy*. Dans l'exemple cartographique, ces deux relations « permettent ainsi de bien distinguer le langage qui joue le rôle de système, du (ou des) métamodèle(s) qui jouent le rôle de modèle(s) de ce langage » [33, p. 2]. Ces notions sont résumées par [17] sur la Figure 2.2 :



FIGURE 2.2 – Notions basiques de l'IDM, issu de [17].

Comme nous venons de le voir, les modèles ayant pour base un métamodèle donné sont dits conformes à ce métamodèle. Un méta-

modèle étant également un modèle, il peut lui-même être conforme à un métamodèle : le méta-métamodèle. Cette hiérarchie de modèles est généralement décrite (cf. Figure 2.3) sous une forme pyramidale dont le nombre de niveaux peut être variable.



FIGURE 2.3 – Pyramide de modélisation de l'OMG, par Jean Bézivin, cité dans [33].

2.2.3 La définition de transformation

Ce sont les transformations qui vont permettre aux modèles d'être représentés via les métamodèles, qui rendent les modèles opérationnels. Ce sont également les différentes transformations que modèles et métamodèles subissent qui augmentent considérablement la productivité des applications suivant cette approche [45]. « En effet, les transformations assurent les opérations de passage d'un ou plusieurs modèles d'un niveau d'abstraction donné vers un ou plusieurs autres modèles du même niveau (transformation horizontale) ou d'un niveau différent (transformation verticale) » [45, p. 12]. À noter que le modèle transformé est alors appelé « modèle source » et le modèle résultant de cette transformation « modèle cible ». Ces éclaircissements nous permettent d'arriver à une définition des transformations dans une approche par IDM :

DÉFINITION DE TRANSFORMATION. « Une transformation est la génération automatique d'un modèle cible à partir d'un modèle source. » [85, p. 44, notre traduction].

Keppel et al. [85] définissent également le concept de « définition de transformation », utilisée par la transformation. Il s’agit d’un ensemble de règles de transformation, qui sont des spécifications sans ambiguïté de la façon dont (une partie d’un) modèle peut être utilisée pour créer (une partie) d’un autre modèle.

Le processus de développement des systèmes via cette approche s’apparente à une séquence de transformations de modèles partiellement ordonnée, « chaque transformation prenant un ou des modèles en entrée et produisant un ou des modèles en sortie, jusqu’à l’obtention d’artefacts exécutables » [45, p. 2].

2.3 DISCUSSION

L’ingénierie dirigée par les modèles consiste donc à utiliser les modèles comme les éléments de base du développement. Comme nous le constaterons par la suite, cette approche constitue à la fois le cœur de notre contexte d’étude — la production documentaire via l’IDM —, mais également la méthode choisie pour l’implémentation d’un système d’analytique des apprentissages avec le numérique — via la proposition d’un métamodèle pour ce faire. Dans le chapitre qui suit, nous présentons les chaînes éditoriales qui instrumentalisent la production de contenus numériques.

À noter que dans la communauté des analytiques des apprentissages avec le numérique, en lien avec les systèmes de tuteurs intelligents ITS (*Intelligent Tutoring Systems*) [108], on retrouve souvent le terme de modèle pour faire référence par exemple au modèle de l’apprenant. S’il s’agit bien d’un modèle au sens qui est dit en 2.2.1, ce n’est en revanche pas exactement un modèle au sens de l’IDM. Nous utiliserons donc « modèle » dans la suite du manuscrit avec le sens donné ici et préciserons qu’il s’agit d’une référence au deuxième sens quand c’est le cas.

CONTEXTE DE PRODUCTION DOCUMENTAIRE

3.1 INTRODUCTION

Dans ce Chapitre, nous verrons comment l'émergence du besoin d'une production documentaire de masse a rencontré dans l'ingénierie dirigée par les modèles une méthode qui a permis à l'ingénierie documentaire de répondre à ce défi. Les chaînes éditoriales numériques peuvent ainsi être vues comme la matérialisation d'une réponse à ce besoin. Dans ce chapitre, après avoir défini des termes déterminants pour la suite de nos travaux, nous mettons en évidence les particularités non seulement de la manière de produire et réceptionner des documents, mais également des spécificités uniques inhérentes aux documents produits de telle façon.

3.2 INGÉNIERIE DOCUMENTAIRE

Crozat [41] cite deux moments où un contexte technique et d'évolution sociale a mené à des changements conséquents de la production documentaire : 1/ le premier de ces moments est lié à la révolution industrielle qui a notamment comme conséquence la « documentarisation de la société » [41, p. 184]; 2/ le second moment est celui de l'avènement du numérique, menant à une « redocumentarisation de la société, c'est-à-dire une réinvention de la nature documentaire » [41, p. 184].

Afin de comprendre ce que cette redocumentarisation signifie en termes de changements des modes de production et réception de documents, nous proposons de commencer par définir ce qu'est un document, via la définition fournie par Bruno Bachimont et reprise par Crozat.

DÉFINITION DE DOCUMENT. « [Un document] est une inscription de contenus sur un support pérenne, établie dans un contexte éditorial » [41, p. 184].

Crozat [41, p. 185] détaille quelques aspects de cette définition : « Un contenu est une forme d'expression pourvue d'une valeur culturelle associée à un véhicule matériel, il exprime une signification et suscite une interprétation ; une inscription est un contenu fixé sur un support matériel, tel qu'il lui apporte une permanence dans le temps ; un

contexte éditorial est l'association d'un contexte de production et d'un contexte de réception ».

Toutefois, outre cette définition de document générique, nous jugeons opportun de définir également ce qu'est un document numérique. Ceci nous conduira à évaluer les changements que son apparition entraîne aux contextes de production et réception documentaire. Nous avons retenu ici la définition proposée par Pédauque et citée par Crozat :

DÉFINITION DE DOCUMENT NUMÉRIQUE. « Un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs » [41, p. 186].

Guyot élargit la définition d'un document (y compris le document numérique) en relevant ses éléments constitutifs potentiels [71] et en considérant que le numérique permet à la fois la consultation, mais également un autre moyen de produire des documents :

1. Contenu — un énoncé porteur de sens
2. Code d'écriture — un langage, qui peut être plus ou moins formalisé
3. Règles techniques — le logiciel sur lequel il s'appuie, le formatage informatique
4. Support — papier, numérique, objet, etc.
5. Règles éditoriales conventionnelles — texte et paratexte, organisation interne
6. Marquages documentaires — identifiants, liens, classements ou encore intentions présumées
7. Règles socio-organisationnelles — cadre d'usage, règles de circulation
8. Contexte — conditions de production, de communication

Les règles de mise en forme feraient partie des règles d'organisation interne. En effet, ces règles s'appliquent à la présentation du texte, des images et de tout autre élément multimédia présent dans le document. De plus, l'organisation interne du document textuel, dans l'exemple fourni par [71], est souvent composée de chapitres, paragraphes, titres, etc., et a un déroulé séquentiel. En outre, le paratexte permet d'enrichir le document, par exemple via des glossaires, des liens internes ou externes, facilitant la navigation dans le contenu ou bien à trouver l'information de manière plus précise.

Selon Crozat [41], c'est le processus de documentarisation qui va permettre de définir le traitement subi par un contenu afin qu'il prenne

la forme d'un document. Ce processus ayant été bouleversé, mais également facilité par les avancées du numérique, c'est ainsi la stabilité du document qui ressort comme le changement le plus évident dû au contexte technique et d'évolution sociale relativement récent qu'est la numérisation. Le même chercheur souligne également d'autres aspects liés aux propriétés techniques du numérique : la fragmentation des contenus, la recombinaison de ces documents/fragments de documents, leur désémantisation/resémantisation.

Ces propriétés sont intimement liées à la rééditorialisation [41], tendance de l'écriture numérique, où l'on donnerait un nouvel objectif à un document ou à des fragments de celui-ci. Selon le même auteur [41], ceci se ferait via des processus comme la sélection des archives, leur déconstruction en fragments, la réécriture à partir de ces fragments avec éventuellement l'ajout d'une production originale et, enfin, une reconstruction de l'ensemble de ces fragments visant ce nouvel objectif. Un dernier processus, la publication, homogénéise les fragments avec une mise en forme adaptée au nouvel objectif et au nouveau contexte. Par exemple, un fragment du type « concept » dans le module I d'un cours donné, une fois passé par une rééditorialisation, pourrait être resémantisé en tant que « rappel » dans un document appartenant au module II de ce même cours.

L'ingénierie documentaire s'intéresse alors à la gestion numérique des documents, mais également à de nouvelles formes d'écriture, avec le numérique [41]. Elle a en outre permis de répondre à la problématique de production de masse de documents [8] et contribué à l'homogénéité des documents produits de cette façon, via la structuration et sémantisation de ces derniers.

3.2.1 Chaînes éditoriales numériques

Un exemple technique issu de l'ingénierie documentaire est celui des chaînes éditoriales numériques (dont nous utiliserons la forme plus courte « chaînes éditoriales » pour la suite du manuscrit). Les chaînes éditoriales émergent dans les années 80 dans un cadre de production documentaire scientifique — principalement avec \LaTeX ¹ — et technique — avec Standard Generalized Markup Language (SGML)² [62].

Il s'agit d'artefacts visant l'instrumentalisation de la production de contenus et la définition que nous retenons est la suivante :

1. <https://www.latex-project.org>
 2. <https://www.w3.org/Markup/SGML>

DÉFINITION DE CHAÎNE ÉDITORIALE NUMÉRIQUE. « Une chaîne éditoriale numérique est un système de production documentaire cherchant à instrumenter des fonctions d'écriture originales (dans ce que le numérique lui apporte et lui impose de spécifique) en prise avec la tendance du numérique, en particulier afin d'automatiser la rééditorialisation » [41, p. 190].

Cette instrumentalisation technologique et méthodologique s'inscrit dans l'approche de l'ingénierie dirigée par les modèles, que nous avons pu présenter dans le chapitre 2. La méthode consiste à créer « (...) un modèle de document, à assister dans les tâches de création du contenu et à automatiser la mise en forme » [40], cité dans [69].

C'est à travers l'utilisation de modèles documentaires que la production documentaire en masse a pu être assurée tout en gardant une homogénéité entre les documents. C'est également la structuration du contenu via des modèles qui rend possible la mise en forme et la publication automatiques [62]. Les chercheurs soulignent que c'est « cette mise en évidence de la structure [qui] permet d'instrumenter la séparation entre le fond et la forme » [9, p. 2], [13].

Revenons sur ce concept de document structuré : il s'agit d'un document qui contient des informations sur sa structure — physique, logique, sémantique [71]. Guyot nous propose l'exemple d'un curriculum vitae, ou encore d'une fiche technique. Il continue en précisant que des balises permettraient ainsi de s'interroger sur un « champ » spécifique. Par exemple si on considère le nom d'une personne associé à un champ « auteur », celui-ci peut permettre d'apprendre ce qu'il ou elle a écrit en tant qu'auteur, mais le même nom associé à un champ « jury » peut être utilisé pour savoir les textes — une thèse — pour lesquels il ou elle a participé au jury, etc. Par opposition, faire la même recherche sur un document non structuré, c'est faire la recherche sur le contenu entier, avec comme résultat savoir uniquement si le terme s'y trouve ou pas, mais sans information sémantique attachée permettant de faire une requête précise.

Une autre conséquence de cette structuration est la séparation fond/-forme. La mise en forme de chacune des parties — préalablement connues grâce au modèle documentaire — peut être définie et appliquée à toutes les parties correspondantes des documents suivant le modèle documentaire en question. L'auteur d'un document n'a donc plus à s'occuper de la mise en forme ou de l'apparence du document, mais seulement du contenu. C'est le concept *WYSIWYM* : *what you see is what you MEAN*, opposé notamment à *WYSIWYG* : *what you see is what you get*. Cela représente non seulement un changement de support, mais une reconfiguration du système technique de production et

de manipulation du document [41]. Cette séparation fond/forme est aussi intimement liée aux possibilités du polymorphisme [41] où un même contenu peut être publié automatiquement via des algorithmes de transformation en plusieurs formats, tels qu'un PDF, un document HTML ou une présentation en diaporama, selon les différents besoins et situations d'usage.

Définissons alors ce qu'est un modèle documentaire en retenant la définition proposée par Cohen :

DÉFINITION DE MODÈLE DOCUMENTAIRE. « Il s'agit d'un ensemble de balises qui va permettre, d'une part, de définir la structure logique d'un document indépendamment du support physique (telle partie est incluse dans telle autre partie) et, d'autre part, de définir la sémantique des différents blocs de contenus (bloc « définition », bloc « remarque », par exemple). » [32, p. 81].

Cohen souligne également que, du fait de l'utilisation d'un modèle, l'auteur doit se poser la question du statut de ce qu'il va écrire (de quel bloc) et de sa place hiérarchique dans le document. Ceci nécessite alors d'adopter « une écriture structurée dès le départ, ce qui implique une réflexion en amont » [32, p. 81].

Cette implication de l'auteur nous permet de nous tourner vers un autre aspect fondamental des modèles documentaires qui est leur niveau de spécificité. Par opposition à un modèle universel, un modèle dédié est un « modèle documentaire spécifique à un contexte d'usage métier en particulier » [9, p. 3].

Par conséquent, la conception des modèles est fortement liée au métier dans lequel il s'insère. C'est la description du métier qui mène à la définition du modèle documentaire, qui à son tour véhicule une sémantique liée au domaine en question. Le besoin documentaire est analysé puis formalisé dans un modèle, comprenant des schémas structurels : un modèle pédagogique d'un cours générique séquentiel, par exemple, aura besoin de blocs du type « définition », « exercice », « rappel », etc. ; tandis qu'un modèle conçu pour le domaine juridique aura des blocs dédiés nommés « numéro du procès », « délibération », etc. Naturellement, notre intérêt se concentre uniquement sur les modèles dont le contexte d'usage métier est pédagogique.

Les chaînes éditoriales sont un exemple de systèmes basés sur l'ingénierie dirigée par les modèles. Ces systèmes utilisent cette approche afin de créer des modèles qui seront ensuite utilisés dans la production de documents. La Figure 3.1 issue de Guillaume et al. [69] permet de synthétiser le fonctionnement d'une chaîne éditoriale, passant de la réalisation d'un modèle documentaire (schéma documentaire et algorithmes de publication), à la production des documents structurés

suivant ce modèle et enfin à des publications en plusieurs supports selon les algorithmes de transformation disponibles.



FIGURE 3.1 – Étapes de fonctionnement d’une chaîne éditoriale, issu de [69].

Plusieurs exemples pratiques d’outils de chaînes éditoriales sont mis en avant par [69], comme Arbortex³, ChainEdit⁴, FrameMaker⁵ et Scenari⁶. C’est sur ce dernier que nous nous concentrons maintenant, dans la mesure où il constitue notre contexte pratique d’étude.

3.2.2 Plateformes documentaires — plus que des espaces virtuels pour partager des documents

Une fois le document généré, une dernière étape nous intéresse dans le cadre de nos travaux : la mise en ligne de ces documents afin qu’ils puissent être consultés par les utilisateurs finaux.

Les documents numériques générés peuvent être rendus disponibles, par exemple, en tant qu’un site Web indépendant. Cependant, c’est la mise en ligne dans le cadre d’une plateforme pédagogique qui nous intéressera, car l’enjeu de travailler sur le couplage document/donnée y apparaît.

Il rend possible non seulement de centraliser l’accès à des documents à des fins pédagogiques, mais également d’être capable de proposer des parcours de consultation de ces documents aux utilisateurs. Ces parcours peuvent être séquentialisés ou non, mais représentent un ensemble de documents liés entre eux (par sujet ou matière, intérêt personnel, séquence didactique avec un objectif défini, etc.).

Notamment, nous retenons que les propriétés suivantes coopèrent à définir une plateforme documentaire à vocation pédagogique :

- Gestion d’utilisateurs : permet d’insérer des utilisateurs et de leur attacher des rôles définis au préalable, comme instructeur et apprenant.
- Gestion de sessions : au travers de l’identification des utilisateurs, permet par exemple l’inscription dans des parcours/séquences

3. <https://www.ptc.com/products/arbortext>

4. <https://chainedit.univ-rennes1.fr>

5. <https://www.adobe.com/products/framemaker.html>

6. <https://scenari.software/fr/>

- déterminés (voir ci-dessous), mais également de connaître l'interaction d'un utilisateur avec les différentes pages de la plateforme.
- Définir une séquence pédagogique : permet de déclarer une séquence pédagogique liée aux documents disponibles.
 - Définir une instantiation d'une séquence pédagogique : indiquer un calendrier défini à une séquence pédagogique, vraisemblablement avec une cohorte d'apprenants identifiée.
 - Effectuer une recherche transversale : pouvoir fouiller tous les documents déployés dans la plateforme de manière centralisée.
 - Recueillir les activités d'apprentissage réalisées par les apprenants : offrir la possibilité d'envoi d'un devoir, par exemple, ou d'enregistrer le résultat d'une évaluation, comme un questionnaire.
 - Instaurer une chaîne de communication entre les acteurs : par exemple, la création d'une page dédiée aux nouvelles relatives à une instantiation de séquence pédagogique — un rappel d'une date importante, un changement d'agenda du cours, etc.

3.2.3 Terrain pratique : Scenari et Libre cours

Comme nous l'avons vu précédemment (cf. Section 3.2.1), Scenari⁷ est un exemple de chaîne éditoriale et c'est sur cette suite logicielle que nous nous sommes appuyés afin de réaliser à la fois des investigations plus approfondies comme l'analyse des besoins des utilisateurs, les implémentations du métamodèle et les tests avec des modélisateurs expérimentés.

3.2.3.1 Suite logicielle Scenari

Cette suite logicielle consiste en un logiciel de modélisation (SCENARIBuilder⁸) qui utilise un framework (primitives documentaires) et un ensemble de transformations qui génèrent du code source ; et en un logiciel d'exploitation de ces modèles (SCENARIchain⁹), qui exécute le code généré pour permettre aux auteurs d'instancier le modèle documentaire via la création et la publication de documents. « [...] Scenari est le système technique générique qui permet aux développeurs de mettre à disposition des modélisateurs les composants paramétrables pour concevoir une chaîne éditoriale dans SCENARIBuilder » [42].

7. <https://scenari.software/fr/>

8. <https://doc.scenari.software/SCENARIBuilder/fr/>

9. <https://doc.scenari.software/SCENARIchain/fr/>

3.2.3.2 *Librecours*

Librecours¹⁰, de son côté, est une application de la suite logicielle Scenari, développée depuis la fin de l'année 2017. Il s'agit d'une plateforme de formation utilisée principalement à l'Université de Technologie de Compiègne (UTC) avec des contenus notamment dans le domaine de l'informatique (bases de données, systèmes et réseau, littérature numérique, etc.). Librecours est également un terrain de recherche technologique, une instance de développement expérimental dans laquelle cette thèse s'inscrit. L'enjeu pour nous est de s'intéresser au couplage document/traces d'interaction dans le cadre des principes de modélisation propres aux chaînes éditoriales.

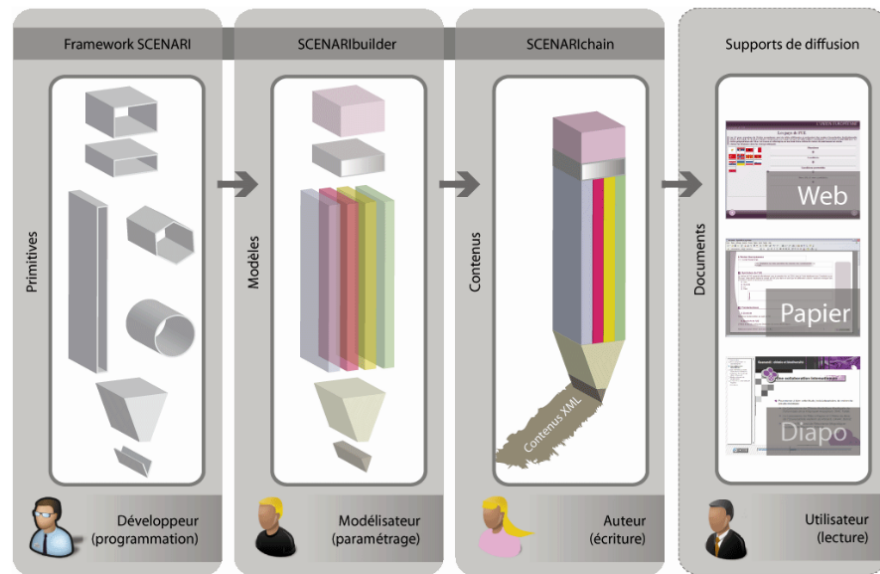


FIGURE 3.2 – La suite logicielle Scenari, issu de [42].

3.2.3.3 *Architecture des composantes du poste client et des serveurs*

Scenari se déploie en mode client-serveur, pour écrire et diffuser des documents (cf. Figure 3.3). L'environnement d'écriture Scenari peut également être déployé sous forme d'une application de bureau autonome, pour un usage personnel ou déconnecté.

3.2.3.4 *Le portail de diffusion*

Le dépôt Scenari correspond au portail de diffusion des documents. C'est une webapp Java couplée à un serveur de ressources NGINX ou Apache (cf. Figure 3.4).

10. <https://librecours.net>

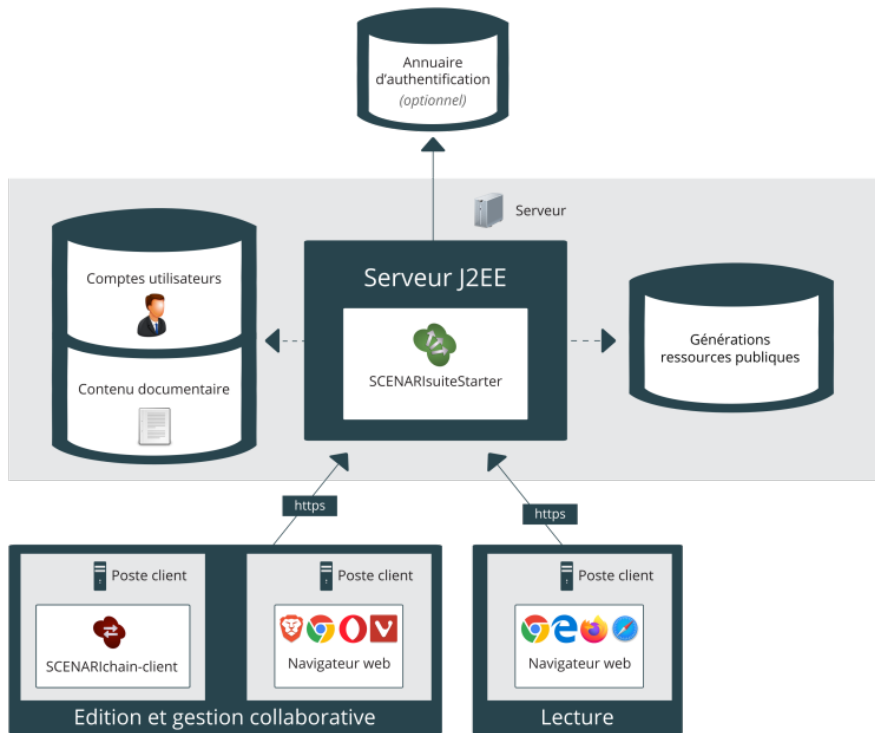


FIGURE 3.3 – L'architecture des composantes du poste client et des serveurs de Scenari.

Le moteur de recherche Elasticsearch peut également être installé en complément pour des fonctions de recherche avancées.

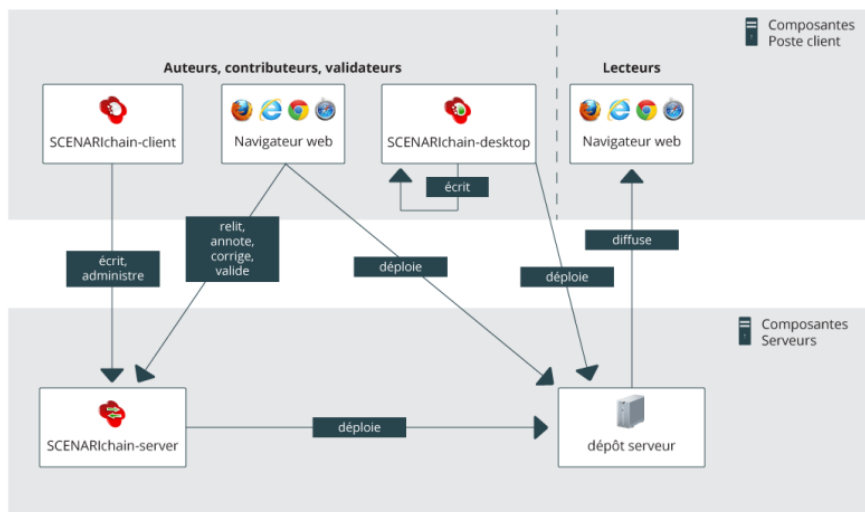


FIGURE 3.4 – Le portail de diffusion de Scenari.

3.3 DISCUSSION

L'ingénierie documentaire, basée sur l'approche IDM et appliquée via des chaînes éditoriales implique non seulement de nouvelles formes de production/réception de documents, mais également des documents particuliers sur plusieurs aspects.

L'aspect structuré de ces documents est la caractéristique essentielle dans le contexte de nos travaux. Comme nous avons pu le mettre en avant, connaître au préalable la structure — physique, logique, sémantique — d'un document permet à la fois 1/ de repérer les intentions des auteurs en leur proposant des éléments de contenu pertinents à leur contexte; 2/ d'harmoniser ces documents; 3/ de séparer le fond de la forme et de définir cette dernière en fonction de la sémantique de chaque partie. Naturellement, nous sommes menés à nous demander si cette connaissance — et cette manière de faire — pourraient représenter également un avantage lors de la mise en place d'un système d'analytique des apprentissages avec le numérique. Le fait d'ajouter des traces d'interaction aux chaînes éditoriales pourrait en effet avoir un intérêt à la fois pour enrichir les traces avec les caractéristiques de la chaîne en question; mais également, à partir de l'analyse de ces traces, d'améliorer les chaînes en question, soit en perfectionnant les documents créés via la chaîne (réingénierie) ou encore en rendant la chaîne éditoriale dynamique.

Comme nous le verrons dans la suite (cf. Chapitre 5), la communauté de l'analytique des apprentissages avec le numérique s'est d'abord concentrée sur les fonctionnalités de base, telles que l'authentification entre les systèmes et la création de listes d'étudiants inscrits, etc. [84]. Néanmoins, afin de produire des analyses qui soient effectivement utiles, un processus judicieux considérant, par exemple, « la sémantique, la sélection de caractéristiques et les problèmes de niveau supérieur associés aux algorithmes et à la création de sens » [84, p. 22, notre traduction] est essentiel.

ILLUSTRATION DE PRODUCTION DOCUMENTAIRE PÉDAGOGIQUE

4.1 INTRODUCTION

Ce chapitre permet d'illustrer le processus de création d'un cours avec l'utilisation des chaînes éditoriales numériques. De cette manière, et à partir d'un exemple simplifié, nous dégagerons les étapes qui sont nécessaires— de la création des primitives documentaires à l'utilisation du cours par les apprenants —, mais aussi les acteurs qui vont interagir avec les différentes parties du système (et que nous devons considérer pour la proposition de métamodèle d'analytique des apprentissages avec le numérique qui suivra).

4.2 ILLUSTRATION DE PRODUCTION D'UN COURS

Dans notre contexte, des développeurs ont défini des « briques », dites primitives documentaires, qui servent de base pour la création des modèles documentaires. Ceci correspond à avoir un *métamodèle* dans une démarche d'IDM.

Ensuite, un outil de modélisation (SCENARibuilder, l'outil de développement des systèmes Scenari d'écriture, de publication et de diffusion de documents) permet d'utiliser ces briques pour définir des modèles documentaires. Dans la pratique, celui qui utilise cet outil, le modélisateur, définit un *modèle de document* en utilisant les primitives documentaires disponibles. L'enjeu est d'abstraire au maximum la technicité des modalités de l'outil conçu.

Plus tard, un auteur (un enseignant, par exemple) utilise ce modèle pour créer un module de son cours dans SCENARichain (outil d'écriture structurée collaborative). L'outil d'écriture inclut des algorithmes de génération qui transforment la modélisation en document, permettant à l'enseignant de publier son cours selon plusieurs supports et formats tels que PDF, Open Document, Web, présentation synthétique, etc. L'intérêt ici est la séparation du fond et de la forme, ou de l'étape de l'écriture de celle de la publication. Le passage (génération) d'un format à l'autre se fait automatiquement à travers des algorithmes de transformation internes à l'outil [13]. Naturellement, nous nous intéressons exclusivement au *document* sous son format Web et à la

manière dont il va être utilisé ensuite par les apprenants. En effet, ce format est le seul qui permette d'envisager raisonnablement un mécanisme de génération de traces sans devoir y associer un outil dédié.

4.3 DANS LA PRATIQUE

Afin d'appréhender les étapes et rôles des parties prenantes dans la démarche de conception de documents à des fins pédagogiques via l'ingénierie dirigée par les modèles, nous proposons ici un exemple pratique très simplifié. Cet exemple reprend le vocabulaire du modèle Opale¹ de Scenari, qui est adapté à la création de cours pour l'enseignement supérieur ou secondaire.

4.3.1 Étape 1 — Le développeur



FIGURE 4.1 – Première étape | Primitives documentaires.

Dans un premier temps, un développeur définit des primitives documentaires telles que « texte », « multimédia », « quiz », « organisation », etc. Chacune de ces « briques » sert à construire le modèle documentaire (cf. Figure 4.1). Un exemple d'une partie d'une de ces briques est fourni dans la Figure 4.2.

4.3.2 Étape 2 — Le modélisateur

C'est à partir de ces briques (qui peuvent ressembler à un bloc multimédia ou quiz dans un LMS comme Moodle, par exemple) que le modélisateur va travailler, en définissant un modèle de document qui répond aux besoins d'un groupe de métiers. Le langage utilisé est donc adapté, il en va de même pour les algorithmes de traitement des modèles disponibles.

Un modélisateur utilise ainsi les primitives disponibles dans SCENARIBuilder pour définir des modèles documentaires. Par exemple, un modèle très simplifié est constitué d'« activités d'apprentissage ».

1. <https://doc.scenari.software/Opale/fr/>

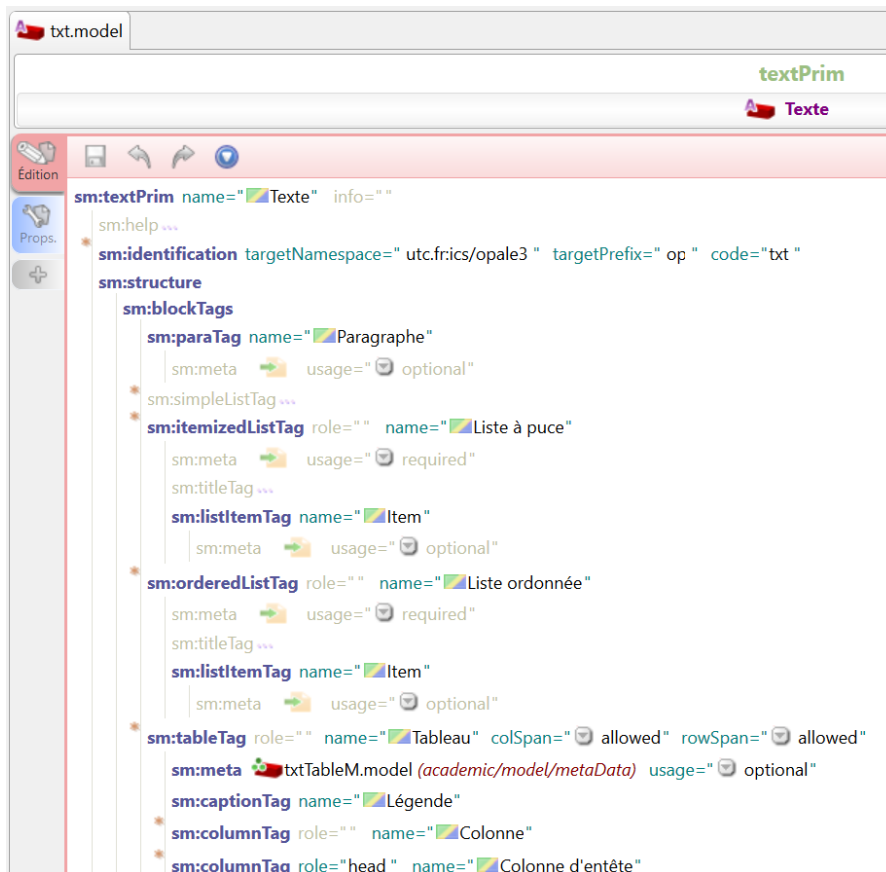


FIGURE 4.2 – Première étape | Exemple pratique d'une primitive documentaire pour l'inclusion des textes dans les métamodèles.

Celles-ci ont la possibilité (primitive d'organisation) d'avoir un(e) ou plusieurs :

- « Introduction » (primitive texte)
- « Concept » (primitive texte ou image)
- « Grain de contenu » : constitué à son tour de parties (primitives texte et images) intitulées « information » et « exemple ».
- « Conclusion » (primitive texte)
- « Pratique » : constitué à son tour de quiz (primitives quiz).

Un module peut avoir une ou plusieurs « activités d'apprentissage ». Une « activité d'apprentissage » doit obligatoirement avoir exactement une « introduction », une « conclusion » et une partie « pratique » à la fin, et inclure entre « introduction » et « conclusion » un ou plusieurs « concept » ou bien « grain de contenu ». Une partie « pratique » doit avoir un ou plusieurs quiz.

Le modélisateur définit aussi, lors de la publication en format Web, une page est créée pour chaque « activité d'apprentissage », que celle-ci contienne un ou plusieurs « concepts » ou bien « grains de contenu », un défilement de la page permet d'accéder aux contenus

s'ils sont longs. Un menu est également créé et affiché sur la gauche, permettant aux apprenants de parcourir le module, soit par les pages, soit par des sauts vers les différentes parties internes à la page.

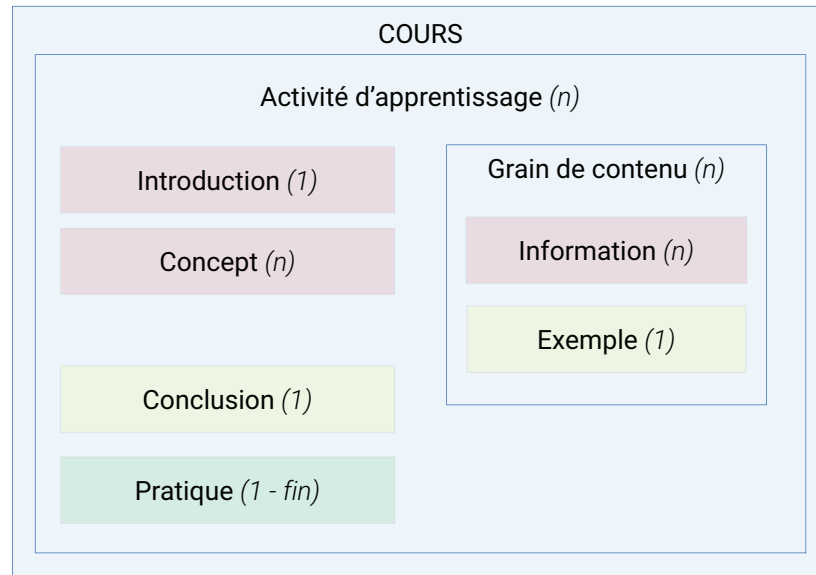


FIGURE 4.3 – Deuxième étape | Métamodèle utilisant les primitives documentaires.

Le modélisateur détermine de ce fait la structure (chaque partie et ses constituants possibles/obligatoires) ainsi que la sémantique de base (ce que chaque partie devra contenir) pour chaque document qui se fondera sur ce modèle documentaire (cf. Figure 4.3). En d'autres termes, les différentes « parties » possibles du document (les blocs) sont préétablies au moment de la modélisation et les types de contenu de chaque partie sont définis par leurs noms même, donc la sémantique sera connue en amont. Ainsi, l'auteur doit choisir le modèle correspondant aux besoins de son métier et cela aura des conséquences également dans la phase d'analyse des traces, notamment en ce qui concerne la sémantique. En d'autres mots, il ne pourra pas, par exemple, définir une autre sémantique pour son cours.

Enfin, le modélisateur utilise les algorithmes de génération de SCENARIBuilder afin d'obtenir un modèle (comme Opale) qui sera ensuite utilisé dans SCENARIchain par l'auteur. Un exemple d'une primitive du métamodèle à l'origine d'Opale est fourni dans la Figure 4.4.

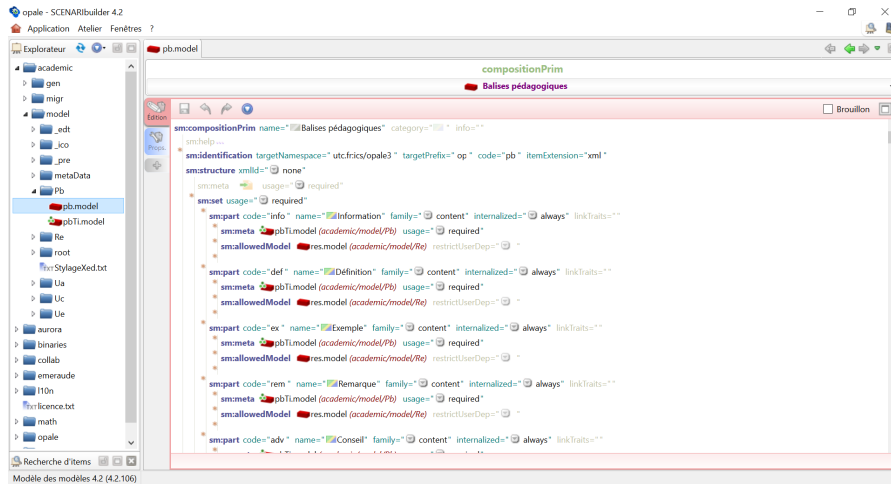


FIGURE 4.4 – Deuxième étape | Exemple pratique du métamodèle du modèle documentaire Opale.

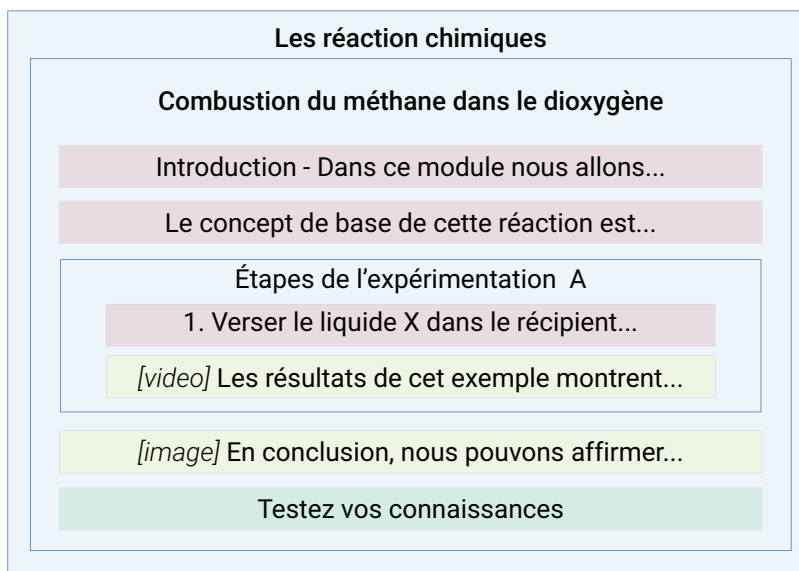


FIGURE 4.5 – Troisième étape | Conception d'un cours utilisant le modèle documentaire.

4.3.3 Étape 3 — L'auteur

Un auteur, par exemple un enseignant, utilise ce modèle pour créer un cours (cf. Figure 4.5). Il doit avoir le modèle choisi dans SCENARichain et a ainsi à sa disposition les éléments permettant sa création. Son cours est constitué de quatre « activités d'apprentissage », chacune avec une « introduction », deux « concepts » et quatre « grains de contenu », une « conclusion », suivie d'une activité « pratique » afin de

vérifier la compréhension du contenu théorique. Une fois le cours créé, l'enseignant peut le publier en format Web, en le plaçant sur la plateforme. La Figure 4.6 permet de visualiser une utilisation du modèle documentaire Opale pour la création d'un document pédagogique.

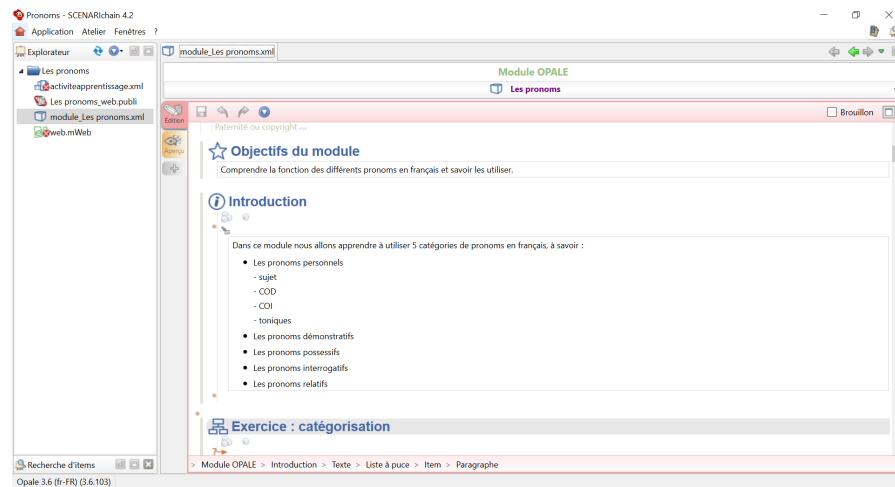


FIGURE 4.6 – Troisième étape | Exemple pratique de la conception d'un cours utilisant le modèle documentaire Opale.

4.3.4 Étape 4 — L'apprenant

Les apprenants accèdent aux contenus publiés, ouvrent les pages dans l'ordre qui leur convient, les font défiler, répondent aux quiz, etc. (cf. Figure 4.7).

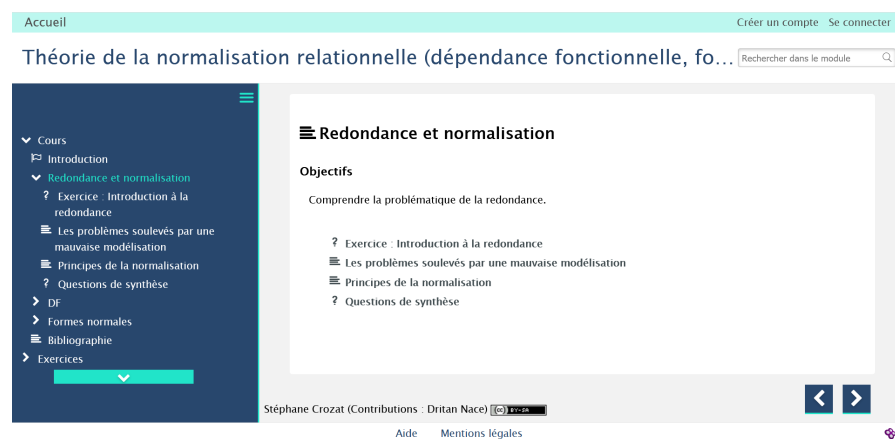


FIGURE 4.7 – Quatrième étape | Document généré prêt à être utilisé par les apprenants - cours issu de la plateforme Librecours de [43].

4.4 DISCUSSION

Avec cette illustration, outre les caractéristiques liées à l'ingénierie documentaire avec une approche par l'IDM, nous notons que l'usage de chaînes éditoriales dans un contexte pédagogique permet également de [69] :

- rendre explicites des pratiques pédagogiques ;
- mutualiser les pratiques ;
- optimiser la gestion de la production ;
- faire des économies dans la production documentaire.

Ces caractéristiques nous intéressent notamment en ce qui concerne les choix pédagogiques, qui seraient explicites et mutualisés. De ce fait, nous pourrions supposer qu'un même modèle documentaire représente un même choix globalement en termes de pédagogie et de conception d'apprentissage (*learning design*). De ce fait, il serait ainsi possible de supposer également des besoins similaires en termes d'analytique des apprentissages avec le numérique. Cette question est analysée dans le Chapitre 8.

Troisième partie

ÉTAT DE L'ART

Dans cette troisième partie, nous interrogeons la littérature autour du déploiement de systèmes d'analytique d'apprentissages avec le numérique afin de comprendre, d'une part, les verrous scientifiques encore présents et, d'autre part, d'éclaircir les dimensions à prendre en compte lors du déploiement de ces solutions d'analytique d'apprentissages avec le numérique dans le cadre de notre contexte (objet de notre métamodèle). D'abord, dans le Chapitre 5, nous nous intéressons notamment aux trois moments identifiés — collecte, analyse, visualisation — permettant d'aboutir à des indicateurs d'apprentissage pertinents. Nous nous intéressons également aux parties prenantes qui jouent un rôle dans l'implémentation et l'usage de tels systèmes, aux objectifs de celles-ci ainsi qu'aux niveaux d'analyse. Enfin, dans le Chapitre 6, nous analysons les travaux allant vers la modélisation de tels systèmes.

CADRE D'ÉTUDE

5.1 INTRODUCTION

Le domaine de l'analytique des apprentissages avec le numérique repose sur trois moments clés : la collecte, l'analyse et la visualisation des traces d'apprentissage. Analyser ces traces consiste à extraire des connaissances qui pourront être utiles à un ou plusieurs acteurs dans le but d'améliorer l'efficacité des processus liés à l'apprentissage.

Dans ce chapitre, nous présentons les travaux liés à l'analytique des apprentissages avec le numérique concernant ces trois moments ainsi que les objectifs sous-jacents d'une telle démarche, les acteurs et niveaux d'analyse tout en essayant de dégager à chaque fois les verrous scientifiques que notre approche pourrait contribuer à faire avancer. Ceci nous permettra également de dégager les composants à prévoir en vue de la modélisation de cette démarche qui n'est pas spontanée, mais itérative et complexe.

5.2 ANALYSER DES TRACES D'APPRENTISSAGE

La littérature nous indique que la définition, probablement la plus répandue, de l'analytique des apprentissages avec le numérique est la suivante : « Collecte, analyse, évaluation et communication des données relatives aux apprenants, leur contexte d'apprentissage, dans la perspective d'une compréhension et d'une optimisation de l'apprentissage et de son environnement » [129, notre traduction]. Cette définition nous permet de dégager trois aspects importants à examiner dans la littérature :

- La collecte des données : de quelles données parle-t-on et comment sont-elles collectées ?
- L'analyse des données : comment ces données sont-elles analysées ?
- La communication du résultat des analyses : une fois l'analyse faite et évaluée, comment les résultats sont-ils communiqués aux différentes parties prenantes ?

Cependant, répondre aux questions ci-dessus n'est pas suffisant dans notre objectif de modélisation d'une implémentation d'analytique des apprentissages avec le numérique : des chercheurs [1] nous

rappellent que l'analyse de ces données d'apprentissage via le numérique va permettre aux différents acteurs d'adapter leurs actions suite à l'interprétation de ces données. Par exemple, les enseignants, les concepteurs de cours ou encore les administrateurs d'environnements d'apprentissages virtuels pourront rechercher des motifs non observés et des informations sous-jacentes dans les processus d'apprentissages avec le numérique. Il nous paraît alors important d'examiner la littérature également afin de comprendre :

- Quels sont les acteurs — les parties prenantes — qui peuvent bénéficier (ou jouer un rôle lors) des analyses ?
- De quels niveaux d'analyse parle-t-on ?
- Quelles sont les différentes intentions ou objectifs de ces analyses ?

Ce sont ces six questions qui nous ont guidés dans la revue de la littérature présentée dans le présent chapitre. Ces informations seront nécessaires afin de pouvoir abstraire et modéliser une démarche d'implémentation d'analytique des apprentissages avec le numérique.

5.3 OBJECTIFS D'UNE DÉMARCHE D'ANALYSE DE TRACES

5.3.1 Basés sur la prise de décision

La création de valeur à partir des données dans le domaine de l'analytique des apprentissages avec le numérique consiste à favoriser une meilleure prise de décision et de meilleurs résultats liés aux différentes situations d'apprentissage. Différents types d'analyses permettent d'envisager différentes actions de la part des parties prenantes. C'est en essayant de répondre aux questions suivantes que [59] sont arrivés à une classification comportant quatre grands types d'analyses fondées sur des données :

Descriptive Qu'est-il arrivé ?

Diagnostique Pourquoi est-ce arrivé ?

Prédictive Que va-t-il arriver ?

Prescriptive Que devrait-on faire ?

Un aspect important à noter dans cette classification est la place donnée à la prise de décision et aux éventuelles actions de la part des parties prenantes à partir de l'analyse — en contraste avec une décision automatisée. Cet aspect est un des points qui distinguent deux champs assez rapprochés [130], celui de la fouille de données éducatives (*Educational Data Mining* (EDM), en anglais) et celui de l'analytique des apprentissages avec le numérique. Dans le premier, une automatisation de la prise de décision serait choisie davantage

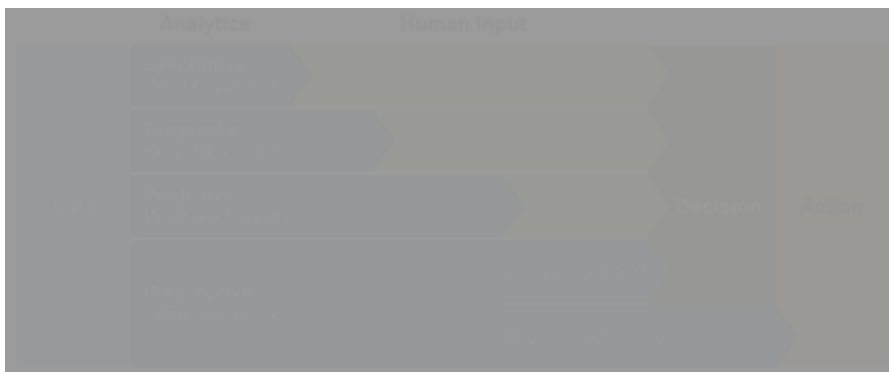


FIGURE 5.1 – Les quatre types d'analyses, issu de [59].

par rapport au second, même si parfois les algorithmes utilisés sont les mêmes. Comme le remarque [122], un aspect commun aux deux domaines est la nécessité d'une grande quantité d'observations sur le processus d'apprentissage. Ces observations peuvent provenir de différentes sources, dans différents formats, à des moments différents et doivent être traitées de manière à pouvoir obtenir l'un des résultats utiles souhaités par chaque champ.

5.3.2 Basés sur l'action

Khalil et Ebner [81] ont pour leur part déterminé six objectifs principaux liés aux démarches d'analytique des apprentissages avec le numérique, basés sur les actions à prendre :

- Prévisions : explorer une valeur inconnue telle que la performance, les connaissances, ou la note afin de prévoir les activités de l'apprenant et les performances futures.
- Interventions : déterminer les apprenants à risque et prévenir le décrochage, conseiller les apprenants.
- Recommandations : faire des recommandations aux étudiants en fonction de leurs activités, comme suggérer un cours, recommander une lecture, etc.
- Personnalisations : liées à la recommandation, permet par exemple de changer l'environnement d'apprentissage personnel et soutenir l'apprentissage de tous les apprenants, indépendamment de leurs profils.
- Réflexions et itérations : consiste à évaluer le travail passé pour améliorer l'expérience future et le transformer en apprentissage selon la personnalisation et l'adaptation.
- Analyses comparatives : vise l'identification des meilleures pratiques qui produisent des résultats supérieurs. Lié à la perfor-

mance, permet par exemple, d'identifier les points faibles de l'environnement d'apprentissage.

5.3.3 Basés sur les bénéfiques

Enfin, Ifenthaler et al., repris dans [124], classifient les bénéfiques potentiels d'une démarche d'analytique des apprentissages avec le numérique en trois catégories :

- Sommatif : fournit des informations détaillées suite à la fin d'une phase d'apprentissage, souvent comparées à des points de référence ou des repères préalablement définis. Par exemple, viser la compréhension des habitudes et des résultats d'apprentissage, établir un comparatif des parcours d'apprentissage, effectuer un suivi du progrès vers les objectifs établis, etc.
- Temps réel : utilise des informations en cours pour améliorer les processus par des interventions. Par exemple, soutenir la collaboration, effectuer des évaluations, recevoir des interventions (automatisées ou pas), avoir un feedback automatique, etc.
- Prédicatif : prévoir la probabilité de résultats afin de planifier les stratégies futures. Par exemple, optimiser les parcours d'apprentissage, s'adapter aux recommandations, augmenter l'engagement, augmenter le succès, etc.

5.3.4 Discussion

Le Tableau 5.1 résume l'ensemble des catégories des objectifs de l'analytique des apprentissages avec le numérique apportés par différents acteurs et présentés dans la section 5.3. Nous remarquons que deux critères en commun semblent être à l'origine : 1/ le temps, par exemple, quand des données passées sont utilisées pour faire de la prédiction sur le futur, ou quand des données en temps-réel sont utilisées afin de décrire ce qui est en train de se passer (une sensibilisation, une analyse qui décrit l'implication d'un groupe d'apprenants, ou encore des résultats d'une phase d'apprentissage); 2/ la décision ou action attendue suite à l'analyse, qui peut être plus au moins soutenue via les résultats de l'analyse eux-mêmes, passant d'une analyse descriptive de ce qui se passe laissant la possibilité aux acteurs de faire un diagnostic et décider d'une éventuelle action corrective, et pouvant aller jusqu'à une automatisation de l'action à prendre, comme l'adaptation ou personnalisation automatique du système numérique en question. Ceci nous permet d'arriver à une classification plutôt équivalente à celle basée sur la prise de décision, où l'aspect temporel

CATÉGORIE	ÉTUDE [59]	ÉTUDE [81]	ÉTUDE [124]
Descriptive	descriptive	sensibilisation personnalisations	sommatif
Diagnostique	diagnostique	réflexions et itération	sommatif
Prédictive	prédictive	prévisions	prédictive
Prescriptive	prescriptive	recommandations interventions personnalisations	sommative temps réel
Comparaison		analyses comparatives	

TABLEAU 5.1 – Tableau de comparaison des objectifs.

reste implicite — quelles données analyser, moment de l'analyse et moment de prise de décision. Cependant, un critère proposé par [81] échappe à cette classification : les analyses comparatives. En effet, des analyses comparatives pourraient en principe être effectuées afin de favoriser les décisions, que ce soit en décrivant une situation, en faisant un diagnostic ou encore une prédiction, et que ce soit sur des données passées ou sur une prescription sur des actions futures, fonctionnant ainsi comme un aspect qui peut être utilisé en conjonction avec les autres.

Nous retenons les différents objectifs des analyses à modéliser et notamment les aspects temporels et décisionnel liés à celles-ci : le contexte de nos travaux nous situe parmi ceux visant à analyser les données avec une intention claire de fournir le résultat de ces analyses à une des parties prenantes afin qu'il ou elle puisse prendre une décision et, éventuellement, accomplir une action. En d'autres mots, les résultats des analyses de données que nous modélisons ont vocation à être distillés et restitués à un usager et non à changer le comportement d'un système de manière automatique. Ce choix, lié à notre contexte, pourrait représenter un verrou du fait de ne pas permettre l'adaptation automatique des visualisations, par exemple. Néanmoins, cela devrait être atténué par le fait que la métamodélisation devrait permettre des analyses prescriptives, tout en soutenant les processus décisionnels par les parties prenantes, qui pourraient ainsi décider d'une adaptation selon un certain profil, par exemple.

5.4 PARTIES PRENANTES ET NIVEAUX D'ANALYSE

Dans la mesure où, comme nous venons de le voir, les résultats des analyses seront restitués à un usager, nous souhaitons comprendre qui sont ces potentielles parties prenantes qui auront un rôle assigné dans la modélisation, que ce soit des personnes impliquées pour parvenir à produire les analyses ou encore les destinataires de celle-ci. Selon Wise et al. [147], un des aspects spécifiques de l'implémentation d'un système de ce type est celui du guidage de son utilisation dans une prise de décision productive par les différentes parties prenantes. Dans les prochains paragraphes, nous nous intéressons à la littérature qui a pu identifier ces différentes parties prenantes et les possibles atouts qu'une telle démarche pourrait apporter à chacune d'entre elles.

Certains travaux, comme [1], s'intéressent à la catégorie que nous proposons de regrouper sous l'étiquette « instructeurs », c'est-à-dire, des enseignants, professeurs, concepteurs de cours, tuteurs, mentors, etc. La revue de la littérature nous permet d'affirmer que cette catégorie est la cible la plus fréquente pour la communication des résultats d'analyses, via des tableaux de bord [125]. Parmi les décisions que ces acteurs pourraient prendre suite à la consultation du résultat des analyses, nous pouvons citer l'adaptation ou la refonte du contenu d'un cours, une intervention ou une régulation auprès des apprenants à mi-parcours, etc.

Une autre catégorie issue de la littérature [25, 67] est celle des « apprenants », étiquette sur laquelle nous proposons d'inclure des étudiants, apprenants, élèves, etc. Bien que les apprenants soient à la fin les bénéficiaires principaux de l'intention d'améliorer les apprentissages portée par l'analytique des apprentissages avec le numérique, ils sont une autre cible directe de la visualisation des résultats via des tableaux de bord [125]. L'un des cadres d'étude le plus suivi menant à la proposition de ces tableaux de bord à l'intention des apprenants est celui de l'apprentissage autorégulé [114, 127, 140, 145].

Les résultats d'une étude qualitative récente avec des apprenants nous indiquent leurs attentes par rapport à un système d'analyse des traces numériques : « *Findings show that students expect learning analytics features to support their planning and organization of learning processes, provide self-assessments, deliver adaptive recommendations, and produce personalized analyses of their learning activities* » [124, p. 67].

La troisième catégorie, identifiée par les mêmes auteurs [25, 67], est celle des « administrateurs » (traduction que nous avons choisie par rapport à *institution* en anglais). Cette catégorie est moins représentée

dans des études du domaine, notamment si l'on considère que ces acteurs font l'objet d'études du domaine voisin appelé parfois *academic analytics* et qui est lui-même très proche des analyses du type *business intelligence* [122]. Un exemple du type d'analyse qui peut intéresser cette catégorie est la comparaison des taux d'inscription à différentes unités d'enseignement d'une année sur l'autre, ou encore les taux de réussite de différents groupes en fonction des instructeurs.

D'autres parties prenantes sont également citées rapidement par ces auteurs [25, 67], comme les chercheurs proposant des analyses ou encore des concepteurs de systèmes.

Cette identification des parties prenantes est importante, car elle permet de cibler les différents besoins d'utilisateurs d'un système et, naturellement, mène également à une distinction en termes de niveaux d'analyses identifiés par les chercheurs. Ifenthaler et al. [75] proposent quatre niveaux :

- mega : gouvernance, analyses interinstitutionnelles ;
- macro : institutionnelle, analyses à l'échelle de l'établissement, études des cohortes d'apprenants, réduction du décrochage et augmentation de la rétention ainsi que des taux de réussite ;
- meso : curriculum, s'intéresse à la conception du programme et au matériel d'apprentissage, aux processus d'apprentissage ;
- micro : cours, analyses visant le soutien de l'apprenant par des recommandations et des fonctions d'aide implémentées dans l'environnement d'apprentissage numérique.

Un point important à souligner lié aux parties prenantes est relatif aux choix des analyses à proposer. Les parties prenantes — notamment instructeurs et apprenants — ont des besoins distincts selon leurs contextes et profils et il peut s'avérer complexe d'entrevoir, parmi de nombreuses options d'analyses et visualisations (discutées plus loin dans ce Chapitre) lesquelles proposer. Ceci a conduit à des efforts de production de tableaux de bord adaptatifs [44], par exemple. D'autres, comme le souligne [127], optent pour offrir aux utilisateurs l'option de réaliser directement un filtrage des données disponibles.

5.4.1 Discussion

La Figure 5.2 permet de visualiser les parties prenantes. Comme vu précédemment, les caractéristiques de notre contexte d'étude, notamment la connaissance fine des documents pédagogiques — versus des données d'un niveau macro, comme la liste des cours suivis par un apprenant —, nous mène à privilégier les niveaux micro et meso et par conséquent les parties prenantes instructeurs et apprenants.

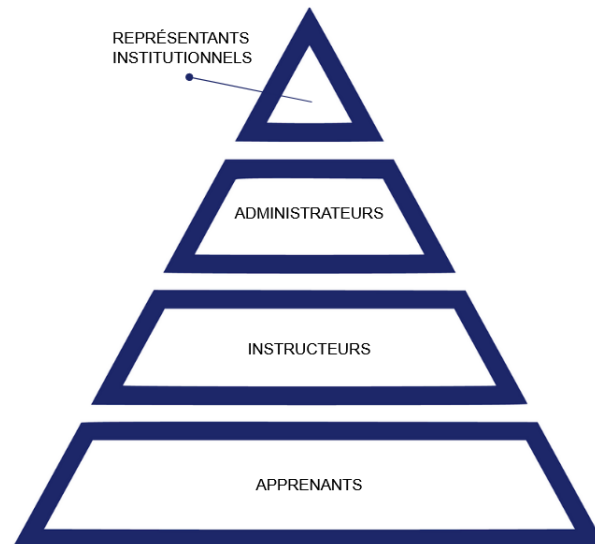


FIGURE 5.2 – Les parties prenantes.

Cependant, cela n'exclut pas que d'autres parties prenantes/niveaux d'analyse soient pris en compte dans des futurs efforts de modélisation, comme dans le cas où la plateforme proposée par Scenari deviendrait le choix d'outil LMS (*Learning Management Systems*) d'une institution.

La question de l'équilibre entre proposer un nombre fermé d'analyses et un accès aux données afin de permettre de les interroger à son gré reste ouverte. Le défi est de trouver un équilibre afin de permettre aux utilisateurs d'explorer les données selon leurs besoins (et donc avec un nombre important de choix à faire), et de fournir des tableaux de bord ou analyses prédéterminées (pas de choix du tout). D'un côté, la première option pourrait mener à une non-utilisation des analyses due au nombre de choix à faire vs le temps restreint pour ce faire ou encore parce que des interprétations des résultats deviendraient progressivement plus difficiles [127]. D'un autre côté, préconiser un nombre fermé d'analyses — souvent par des chercheurs (cf. Chapitre 9, Section 9.5.1) — risque de ne pas répondre aux réels besoins du terrain. La modélisation d'indicateurs (cf. Chapitre 10) qui peuvent être pertinents à un modèle documentaire spécifique en offrant l'option de les activer ou pas pourrait contribuer à répondre à cette question.

5.5 LES TRACES D'APPRENTISSAGE

5.5.1 *Les types de traces*

L'utilisation d'une variété de systèmes informatiques lors des situations d'apprentissage peut potentiellement produire des traces numériques d'apprentissage. Usuellement les traces analysées sont celles générées par des environnements d'apprentissage tels que les *Learning Management Systems* (LMS), les *Virtual Learning Environments* (VLE) ou les *Personal Learning Environments* (PLE), mais également des *Intelligent Tutoring Systems* (ITS), des simulateurs ou jeux sérieux, etc.

Ces systèmes peuvent capturer les actions des apprenants avec des degrés de granularité différents — d'un clic de la souris à l'estimation du niveau d'acquisition d'une compétence X chez l'apprenant Y — mais aussi avec une grande variété de formats et de contenus. « Ainsi, la collecte des données concerne l'observation de l'apprenant et le traçage de ses interactions médiées par les outils, le stockage des traces récoltées et l'import de traces dans les outils d'analyse » [28, p. 19].

L'analytique des apprentissages avec le numérique s'intéresse à des données qui peuvent être classifiées selon les questions posées. Schwendimann et al. [125], dans leur revue systématique de tableaux de bord, ont classifié des indicateurs en six catégories, selon qu'ils considèrent des informations sur les apprenants (compétences, âge, cours suivis, etc.), les actions (nombre de pages visitées, temps passé sur une tâche, etc.), le contenu (sentiment lié aux messages, nombre de concepts et lien entre eux, etc.), le résultat (note moyenne, distribution des notes au sein d'un groupe, etc.), le contexte (placement des apprenants autour d'une table, localisation géographique), ou les aspects sociaux (réseau montrant la communication avec les autres dans un forum, etc.). À noter que ces catégories ne sont pas mutuellement exclusives, par exemple, un indicateur pourrait donner la note d'un apprenant en fonction de son âge, auquel cas on croise « apprenants » et « résultats ».

Ainsi, la littérature nous montre que, parfois, des données autres que celles des interactions des apprenants avec les différents systèmes informatiques — les traces d'interactions — sont utilisées. C'est le cas des données du type démographique (âge, sexe, localisation géographique, niveau d'études des parents, etc.) qui sont utilisées [18, 34, 103, 133] notamment comme des entrées (*features*) pour faire du regroupement ou pour le calcul de prédictions. En revanche, il faut noter que même si ce type de données permet de réaliser ces calculs, les résultats se concentrent sur les apprenants eux-mêmes et sur leurs

caractéristiques plutôt que sur les activités des apprenants et les actions qui sont effectivement sous leur contrôle [4], par exemple si on compare une donnée d'entrée portant sur la scolarisation des parents versus les interactions de l'apprenant avec les documents d'un cours.

Avec l'usage de certains dispositifs techniques en particulier, d'autres types de données peuvent être cités. Une analyse multimodale pourrait inclure, en plus des traces traditionnelles, des informations provenant de casques d'immersion ou encore liées à des gestes, au regard, à la parole ou à l'écriture [110]. Enfin, des auteurs [26] citent également des données subjectives (*self-report*), de performance, physiologiques ou liées aux comportements qui viennent enrichir les analyses liées à des études sur la charge cognitive des apprenants. Ces données — multimodales et provenant de questionnaires — ont un rôle de complémentarité à d'autres sources de traces — typiquement celles des interactions avec les LMS/VLE/PLE — visant à obtenir une compréhension plus moins incertaine du processus d'apprentissage [110].

Ifenthaler et al. [75] proposent un cadre comprenant cinq types de données : 1/ le profil des parties prenantes, 2/ les interactions de celles-ci avec le Web social, 3/ les systèmes numériques, et 4/ les caractéristiques des formations. Enfin, à ces données, des 5/ informations provenant de l'extérieur du système éducatif peuvent être intégrées. Nous remarquons que la catégorie relative aux apprenants de Schwendimann et al. [125] est en quelque sorte élargie pour contenir des caractéristiques des apprenants, mais également d'autres parties prenantes. Les catégories des actions, des contenus et des résultats seraient regroupées dans celle nommée « systèmes numériques » par Ifenthaler et al. [75], tandis que celle des aspects sociaux serait semblable à « interactions avec le Web social » et le contexte aux « caractéristiques de la formation ».

Moore propose un raffinement au sein de la catégorie des interactions. Trois types d'interactions différentes associées à l'apprentissage à distance ont été distingués [104] :

1. Apprenant-apprenant : elles font référence aux échanges entre apprenants participant à un cours, que ce soit de manière synchrone ou asynchrone, par exemple, par l'utilisation de forums ou de chats ou de groupes de travail.
2. Apprenant-instructeur : ces interactions peuvent parvenir, par exemple, lors d'un tutorat ou via les réponses aux questions posées par les apprenants dans le système de messagerie, et contribuent au sentiment de proximité de la présence en ligne.
3. Apprenant-contenu : ces interactions se produisent lorsque les apprenants utilisent de nombreuses ressources de contenu pé-

dagogique, telles que des sites Web, des manuels, des PDF, du matériel de recherche, des vidéos, etc.

En établissant un rapport entre ces catégories d'interaction [104] et celles proposées par Schwendimann et al. [125], nous constatons que les interactions entre apprenant-apprenant et apprenant-instructeur correspondraient aux aspects sociaux tandis que celles entre apprenant-contenu correspondraient aux actions. Certains systèmes focalisent leur attention sur ces interactions avec les différents contenus [56]. Student Activity Meter (SAM) [64], par exemple, a comme métriques le temps passé et l'usage des ressources pédagogiques. CourseVis [98] rassemble les informations d'accès au cours, la lecture des supports de cours, la performance aux tests et la progression.

5.5.2 La collecte des traces

Concernant l'enregistrement des traces de ces interactions, la littérature nous montre que des efforts ont été menés afin de déterminer des spécifications de traces. Nous présentons dans la suite cinq de ces efforts.

5.5.2.1 Contextualised Attention Metadata

Un premier exemple est celui proposé par [148], *Contextualised Attention Metadata* (CAM) visant à capturer des données sur la façon dont les utilisateurs prêtent attention aux informations disponibles via des navigateurs, donc à partir de sources de documents variées. Comme le montre la Figure 5.3, il permet de tracer les interactions des utilisateurs lors de l'utilisation de différents documents où chaque item représente un document numérique spécifique. Chaque item peut alors être l'objet d'une interaction aussi spécifique, définie dans l'élément action lié à l'événement. Néanmoins, les informations relatives au document ne sont pas très riches : titre (à l'intention des usagers), guid (*global unique identifier*, normalement l'URL du document) et type (type technique du document MIME, par exemple HTML).

5.5.2.2 Activity Streams

Un autre exemple de spécification de format ouvert pour les protocoles de flux d'activité est *Activity Streams* [132], qui consiste en une collection sérialisée d'activités d'utilisateurs au format JSON (*JavaScript Object Notation*). Une activité, dans sa forme la plus simple, est constituée d'un acteur, d'un verbe, d'un objet et d'une cible. Dans sa version 2.0, *Activity Streams* est utilisé pour regrouper les activi-



FIGURE 5.3 – Les éléments du schéma CAM, issu de [148].

tés prises dans les applications et services Web, suivant le standard JSON-LD. Ce format (cf. 5.4) accepte l'extensibilité via d'autres types et propriétés. Une « extension » est toute propriété, activité, acteur ou type d'objet non défini par le vocabulaire d'activité de base [131]. La spécification précise de la prise en charge des extensions peut varier selon les implémentations, pourvu qu'aucun modèle de traitement normatif pour les extensions ne soit défini. De cette façon, nous notons que seulement des informations de base sur les activités seraient « normalisées », reléguant des informations sur le contexte pourtant essentielles à la plupart des analyses à la partie des extensions.

5.5.2.3 *Learning Context Data Model*

Une troisième approche différente est une initiative issue du projet *Learning Context* appelée *Learning Context Data Model* (LCDM) [12] et propose une API permettant de collecter des informations de contexte à partir de différents types d'appareils, se concentrant donc sur l'aspect de la mobilité, comme visualisé dans le schéma de la Figure 5.5.

Cette spécification précise, entre autres, l'application à l'origine de l'événement et le type de connexion utilisée (Web, ordinateur de bureau, mobile), l'action effectuée et l'entité à l'origine de l'événement, un document pédagogique, par exemple. Cette spécification vise un modèle d'apprentissage tout au long de la vie et se concentre sur les caractéristiques génériques de l'environnement d'apprentissage et les

```

{
  "context": "https://www.adl.org/ns/activitystreams",
  "summary": "Martin added an article to his blog",
  "type": "Add",
  "published": "2015-02-10T15:00:55Z",
  "actor": {
    "type": "Person",
    "id": "http://www.test.example/martin",
    "name": "Martin Smith",
    "url": "http://example.org/martin",
    "image": {
      "type": "Image",
      "href": "http://example.org/martin/image.jpg",
      "mediaType": "image/jpeg"
    }
  },
  "object": {
    "id": "http://www.test.example/blog/article/xyz",
    "type": "Article",
    "url": "http://example.org/blog/2015/02/entry",
    "name": "Why I love Activity Streams"
  },
  "target": {
    "id": "http://example.org/blog/",
    "type": "OrderedCollection",
    "name": "Martin's Blog"
  }
}

```

FIGURE 5.4 – Exemple d'activité simple (ajout d'un lien sur un site) avec quelques détails supplémentaires, issu de [131].

activités des apprenants [138]. Cependant, peu d'informations sont disponibles concernant le document ou l'activité en question.

5.5.2.4 *xAPI et IMS Caliper*

Récemment, deux spécifications utilisées dans les outils de collecte de traces ont vu le jour : xAPI et IMS Caliper. *Experience API* (ou xAPI¹) [76] est un standard propre aux technologies d'apprentissage permettant de collecter des données sur un large éventail d'expériences d'une

1. <https://github.com/adlnet/xAPI-Spec>



FIGURE 5.5 – Version 4 du *Learning Context Data Model*, issu de [12].

personne [87], parfois vu comme une version modernisée de SCORM (*Sharable Content Object Reference Model*). L'avantage de son usage est l'interopérabilité, qui permet à différents systèmes l'utilisant de communiquer entre eux en capturant et en partageant le flux d'activités à l'aide du vocabulaire défini [87], qui est extensible. Les instructions (*statements*) xAPI ont le format JSON et se basent sur l'utilisation d'API RESTful web-service. Les activités des utilisateurs sont enregistrées dans des *Learning Record Store* (LRS) et les déclarations prennent la forme de triplets « Nom, verbe, objet ».

L'IMS Global Consortium a de son côté proposé *Caliper Analytics*² qui vise à recueillir des données d'apprentissage. Ce standard se base sur un certain nombre de profils (annotation, assessment, forum, gra-

2. <https://github.com/IMSGlobal/caliper-spec>

ding, media, etc.) de métriques liées au domaine d'apprentissage, où chacun fournit un ensemble de termes et de concepts spécifiques [36] permettant d'utiliser un vocabulaire commun qui décrit les interactions courantes des utilisateurs. Le format utilisé est JSON-LD (les aspects des données liées ne sont pas soumis à la conformité). Par souci de conformité, IMS n'aborde pas les aspects des données liées et traite la charge utile comme du JSON formel.

En 2016, des spécialistes se sont réunis afin de vérifier les différences entre ces deux spécifications [37]. Ils ont noté qu'alors que xAPI (Actor/Verb/Object) et Caliper (Actor/Action/Activity) utilisent un modèle de données basé sur des triplets, il existe des différences considérables dans la structure détaillée et l'utilisation des définitions d'objet et d'activité. Ils ont néanmoins affirmé qu'il devrait être possible pour chaque spécification d'utiliser le verbe/l'action de l'autre. Ils ont également signalé le besoin d'impliquer davantage des scientifiques ayant une connaissance de l'analyse des apprentissages avec le numérique dans tout travail d'actualisation des deux spécifications afin de s'assurer de « produire des informations d'analyse d'apprentissage utiles et pas seulement des données » [37, notre traduction]. Outre le fait d'avoir deux spécifications concurrentes, il n'est pas clair à l'heure actuelle si l'une ou l'autre permet d'avoir de meilleurs résultats [84].

Un des aspects liés à ces spécifications qui a reçu des critiques est le manque d'informations pertinentes si l'on suit ces spécifications en mettant en place seulement les aspects obligatoires (*required*) (cf. [84], notamment l'appendix F). La solution pour les fournisseurs de systèmes semble être de commencer à mettre en œuvre des fonctionnalités optionnelles, mais qui résulte en une large gamme de formats de données, comme l'exemplifient [84] dans le cas de données liées à la visualisation d'une vidéo (cf. Figure 5.6). Kitto et al. soulignent que ce manque d'informations pertinentes mène à des situations où le système est considéré comme conforme à l'une ou l'autre des spécifications via des tests de certification, mais dans la pratique produit des flux de données avec des événements largement redondants d'un seul type, ou encore qui ne décrivent aucun comportement de l'utilisateur qui pourrait être utile pour les analyses. Un exemple est donné par ces auteurs avec les données générées par l'activité enregistrée au sein d'un LMS pendant deux semaines, avec un total de 10 millions de registres produits. De ce total, 76,3 % sont relatifs à des « visualisations », sans informations contextuelles à propos du type d'objet affiché. Et pourtant, ils soulignent que la différence entre la visualisation d'un item d'un quiz et la visualisation d'une activité d'apprentissage est

énorme. Les types de ressources sont assez différents et doivent être séparés si nous voulons obtenir des informations significatives sur les interactions des apprenants avec les ressources à disposition.



FIGURE 5.6 – Deux exemples de déclarations sur l'utilisation de la vidéo, produites par deux outils différents : *BestPractice*, un outil fictif qui a adopté le profil xAPI pour la vidéo, et *Kaltura*, qui a implémenté sa propre structure de déclaration xAPI, issu de [84].

5.5.3 Discussion

Bien que, comme nous l'avons vu, souvent une grande variété de types et de sources de données [20] sont utilisés dans l'analytique des apprentissages avec le numérique, nous nous focalisons en particulier sur les traces collectées par l'interaction d'apprenants avec un dispositif d'apprentissage — les traces, souvent enregistrées dans des fichiers de « log », correspondent aux actions et au contenu dans la classification de Schwendimann et al. [125] ou encore au troisième type d'interaction de Moore [104].

Il existe de nombreuses contributions qui s'appuient sur les données collectées par des dispositifs de type LMS, mais il y a des limites par rapport à ce qui serait souhaitable pour détecter des situations spécifiques [122]. Une alternative est le développement de modules d'extension où l'outil hébergeant les interactions est étendu pour collecter les données requises et les analyser, mais dans la pratique le

niveau de granularité reste souvent inférieur à ce qui pourrait être obtenu avec la connaissance fine du document en question. Dans de nombreux cas, les données actuellement produites sont trop fines par rapport à des concepts non pertinents pour l'analytique des apprentissages avec le numérique, tout en n'étant pas suffisamment riches sur des concepts d'une importance capitale pour le domaine, comme l'affirment Kitto et al [84].

En ce qui concerne le format des données et donc les spécifications et les standards, la diversité des spécifications ou approches (xAPI, Caliper, Learning Context Data Model, Activity Streams) et des formats (CSV, JSON, JSON-LD, XML) semble indiquer un compromis entre la couverture et la généralité. Plus la technique est générique, moins les détails d'activités d'apprentissage sont couverts [122]. Ainsi, ce problème a tendance à se poser là où un grand nombre de champs essentiels pour l'analytique des apprentissages avec le numérique est considéré comme optionnel dans les spécifications, ce qui réduit la capacité à fournir des analyses utiles. On prétend fréquemment que la qualité de nos données éducatives s'améliorera si nous utilisons les spécifications, mais même lorsque les spécifications sont suivies, on se rend compte que les données résultantes échouent souvent à éclairer la prise de décision attendue à la fin [84], ce qui serait en partie dû au manque de richesse des données. La cause, soutenue par ces auteurs, est liée à « un accident historique », où la plupart des données enregistrées par ces systèmes ont été initialement créées à d'autres fins, telles que le débogage des systèmes eux-mêmes, donnant lieu à des situations appelées « *clicks to constructs* », où il est, selon ces auteurs, très difficile de mapper les données de flux de clics de bas niveau vers un ordre supérieur pertinent sur le plan pédagogique.

Comme le soutiennent Kitto et al. [84], si la priorité est donnée à la production de données qui correspondent en effet à ses utilisations prévues à la fin, ensuite il serait plus facile de définir les construits (« *constructs* », en anglais) liés aux cadres théoriques en sciences cognitives, sociales et éducatives bien fondées qui seront utilisés, ainsi que les méthodes d'analyse pour y arriver et, enfin, les données nécessaires pour y arriver.

Avec l'introduction de la métamodélisation et du concept d'indicateur, les données collectées auraient tendance à contenir les informations nécessaires à une analyse donnée et donc contenant toutes les informations jugées pertinentes et nécessaires pour cette analyse. Ceci consiste également à inverser la logique en se demandant d'abord de quelles données nous avons besoin pour répondre à une question par opposition à ce que l'on peut apprendre des données (souvent peu

riches) déjà à disposition. Cette logique « inversée » est également soutenue par [84], qui affirment : « Les meilleurs résultats sont obtenus si des analyses significatives sur le plan éducatif sont identifiées d'abord, puis utilisées pour informer à la fois ce que nous mesurons (c'est-à-dire les champs pertinents) et les protocoles que nous utilisons pour les mesurer » [84, p. 8, notre traduction]. D'autres chercheurs croient que de meilleures données sont générées à partir d'une approche « descendante », éclairée par la pratique de la recherche du domaine pour ensuite arriver aux traces nécessaires [86], [30]. D'autres vont encore plus loin et soutiennent que cette manière de faire permet également d'expliquer plus précisément les intentions et objectifs d'une collecte de données des usagers, relevant ainsi de la conformité aux réglementations, mais également des questions d'éthique sous-jacentes [83].

5.6 LES MÉTHODES POUR ANALYSER LES TRACES

Outre des mesures de fréquence simples, quelques méthodes d'analyse ont été employées afin d'essayer de répondre à des questions posées par des parties prenantes ou encore dans le but de découvrir des motifs dans des collections de données brutes qui puissent, d'une manière ou d'une autre, contribuer à l'amélioration de différentes situations d'apprentissage.

Baker [14] propose une taxonomie de méthodes, comprenant 1/ la prédiction (classification, régression, estimation de la densité), 2/ le regroupement (une revue des méthodes utilisées pour faire du regroupement est disponible dans [18]), 3/ l'exploration de relations (exploration de règles d'association, exploration de corrélations, exploration séquentielle de modèles, exploration de données causales), 4/ distillation des données pour le jugement humain et 5/ découverte avec des modèles.

Bien que ces méthodes soient utilisées tant pour des démarches de fouille de données que pour des démarches d'analytique des apprentissages avec le numérique, certaines différences dans la manière de les utiliser ont pu être identifiées [130] dans la littérature : comme nous l'avons vu, il est fréquent que les recherches en fouille de données utilisent le jugement humain afin de fournir des étiquettes pour la classification, tandis que la recherche en analytique des apprentissages utilise la découverte automatisée souvent pour informer une des parties prenantes afin que des décisions puissent être prises : qu'on envisage un changement de comportement, qu'on évalue une intervention, etc. Cette différenciation est importante, car elle permet

de définir l'usage qui sera fait de ces calculs : soit une entrée pour un changement automatique de comportement du système (le cas d'un *Intelligent Tutoring System* (ITS), par exemple, où les exercices proposés à un apprenant dépendent de son niveau estimé de compréhension), soit une « sortie », où les résultats sont donnés à voir à un des acteurs afin qu'il évalue la nécessité, ou non, d'une adaptation (par exemple, un étudiant qui découvre son retard en termes de tentatives d'un quiz par rapport à sa promotion et donc qui va le refaire).

Bien évidemment, le choix des méthodes à utiliser dépend de la question à laquelle on essaie de répondre et des traces ou autres données à disposition. Viberg et al. [146] ont analysé les types de méthodes les plus choisies par les chercheurs entre 2012 et 2018. Les méthodes prédictives ont été parmi les plus choisies dans les articles analysés, même si un déclin a été perçu en 2016 et 2017, avec une augmentation des méthodes d'association (comme association rule mining, correlation mining, sequential pattern mining and casual data mining). L'utilisation des méthodes de distillation des données pour le jugement humain est aussi importante et stable.

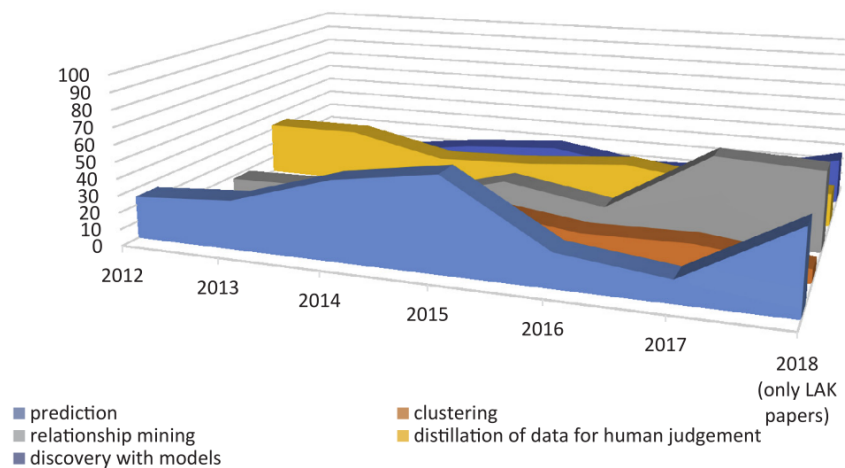


FIGURE 5.7 – Types de méthodes d'analyse, issu de [146] (licence BY-NC-ND 4.0).

Nous allons maintenant détailler certaines de ces méthodes les plus utilisées qui ont été identifiées et utilisées dans la littérature.

- A. Statistiques descriptives : consiste en la sélection, l'analyse et l'interprétation des données numériques grâce à l'élaboration de tables, graphiques et indicateurs numériques adéquats (Reis, Melo, Andrade et Calapez, 1999, cité dans [54]). Typiquement, ces analyses rentrent dans la classification « descriptive » de la typologie d'objectifs d'analyse vue dans la section 5.3 et comprend des analyses axées sur la description, la détection et la

création de rapports, y compris des mécanismes pour extraire des données de plusieurs sources, les filtrer et les combiner. Il s'agit surtout de la distillation des données pour le jugement humain, la catégorie 4 proposée par Baker [14].

- B. Prédiction : l'objectif est de prédire la classe ou l'étiquette d'un objet donné, dont le champ d'application le plus fréquent consiste à prédire les résultats académiques ou bien la performance des apprenants [10]. Asif et al. nous montrent que la recherche dans ce domaine a été menée à différents niveaux : au niveau du système de tutorat [53] pour prédire les résultats des tests des apprenants avec des informations de chronométrage et la quantité d'aide requise ou encore pour prédire si un apprenant est susceptible de réussir le prochain exercice de formation [116]; au niveau du cours, pour prédire la réussite à partir de données sociales [51], la situation académique et les notes [79, 80, 111, 133] ou la nature des questions posées par les étudiants [72]. La littérature montre également qu'il est possible d'envisager des analyses prédictives utilisant des indicateurs d'usage des documents pédagogiques. Ainsi, dans [139], les chercheurs ont calculé quatre indicateurs relatifs à l'utilisation du temps de la part des apprenants (soumission, procrastination, vitesse et régularité) qui ensuite ont été utilisés pour analyser leur effet de prédiction sur la performance de ces apprenants. La performance des différents algorithmes les plus utilisés pour effectuer ces analyses a été évaluée dans [133] par rapport à la prédiction de réussite/échec et de la note obtenue.

Merceron et al. [102] nous rappelle que les nombreuses études publiées dans ce domaine montrent qu'il est en effet possible de prédire l'abandon ou la performance avec une précision raisonnable, majoritairement supérieure à 70 %. Cependant, elle nous rappelle également qu'il n'existe pas un unique classifieur ou un unique ensemble de caractéristiques qui fonctionnent efficacement dans tous les contextes ; ainsi, il est nécessaire de rechercher les méthodes et fonctionnalités qui donnent la meilleure performance selon les données disponibles et le contexte d'application. Parmi les algorithmes les plus couramment utilisés pour ces tâches de prédiction, nous pouvons citer : arbre de décision, règles d'induction, K plus proches voisins (K-NN), classifieur bayésien naïf, réseau de neurones, machine à vecteurs de support (SVM), forêt d'arbres décisionnels.

- C. Exploration de relations : un des usages de ce type d'analyse est la découverte de motifs (*patterns*) dans la séquence des ac-

tions stockées dans les traces. Étant donné un ensemble de séquences, le problème est de découvrir des sous-séquences qui sont fréquentes (l'apparition de telles sous-séquences dépasse un support minimum, spécifié par le chercheur et en relation avec l'ensemble de sous-séquences trouvées) et qui peuvent alors avoir une relation à une certaine performance, par exemple. Parmi les principales approches de l'exploration séquentielle, nous trouvons celles basées « a priori » et celles basées sur la « croissance » [149], dont quelques exemples cités sont *AprioriAll*, *AprioriSome*, *DynamicSome*, *PrefixSpan*. Il est important de noter que l'application de ces techniques directement sur les traces brutes peut générer des instances excessives de motifs non pertinents et qu'un travail supplémentaire de prétraitement ou post-traitement est nécessaire afin de dégager des actions pertinentes.

- D. Regroupement : l'objectif de ce type d'analyse est de regrouper des objets dans des groupes (aussi appelés clusters) basés dans certaines similarités [10] entre ces objets. Par exemple, dans [50], les chercheurs utilisent les techniques de regroupement pour classifier et regrouper les énoncés selon les actes de parole, c'est-à-dire, l'action qu'une personne effectue en prononçant une phrase, comme poser une question, poser un problème, donner une réponse ou un indice, etc. De la même manière, [72] relie des clusters d'étudiants en fonction de la nature des questions qu'ils posent dans le contexte d'une classe inversée avec des caractéristiques connues de ces étudiants.

Bharara et al. ont présenté une revue des différents algorithmes de regroupement et de leurs applications respectives dans [18]. Entre les études citées, l'usage de ce type d'analyse est varié, allant de regrouper les apprenants selon l'usage qu'ils font du LMS à la découverte de sous-populations des apprenants d'un MOOC.

Parmi les algorithmes les plus couramment cités dans ces articles, nous trouvons : K-moyennes, C-moyennes, Fuzzy K-moyennes, K-prototypes, Fuzzy Clustering, Co-operative, PSO farthest first, Expectation Maximization, Hierarchical Agglomerative, Markov Clustering. Certains algorithmes prennent en paramètres un nombre prédéterminé de cluster, d'autres comme K-Medoids ou X-Means permettent de le déterminer automatiquement, et il est toujours possible d'utiliser le Bayesian Information Criterion pour retrouver le nombre optimal de clusters [50].

D'autres méthodes d'analyse pouvant être intéressantes pour l'analytique des apprentissages, mais qui sont relativement peu représentées, sont décrites dans [60]. La première, *l'epistemic network analysis* (ENA), permet l'analyse des réseaux construits sur la base de la co-occurrence de codes dans des artefacts tels que les processus d'apprentissage (par exemple, émergence dans le discours d'apprentissage), qui sont codés en utilisant différents schémas de codage. La deuxième est le cadre d'analyse de cohérence du discours dans la recherche sur la compréhension de texte, *text coherence analysis*, ou Coh-Metrix : l'une des premières utilisations de ce type d'analyse a été l'analyse des ressources et de matériels pédagogiques, ainsi que la recommandation, l'organisation et l'évaluation de ces ressources [65]. La dernière est la *microgenetic analysis* avec *sequence pattern mining*, qui peut être utilisée pour découvrir des modèles temporels d'apprentissage à partir de traces numériques.

Tout compte fait, il est important aussi de reconnaître que la puissance de certaines de ces analyses est liée à la qualité de données qui sont fournies en entrée pour les calculs. Shibani et al [128] nous rappellent l'exemple d'une étude [46] prédisant la réussite scolaire où il a été constaté que des modèles — dans un sens de l'analytique des apprentissages et non IDM — granulaires spécifiques au cours fournissent de meilleures informations aux instructeurs par rapport à l'utilisation d'un modèle générique. Selon les mêmes auteurs, cela a suscité un appel à des modèles prédictifs qui tiennent compte des conditions d'enseignement.

5.6.1 Discussion

Les méthodes d'analyse utilisées par des chercheurs du domaine de l'analytique des apprentissages avec le numérique varient largement. Si certaines analyses, comme celles descriptives, sont plus simples à implémenter, d'autres comme la prédiction et le regroupement nécessitent l'utilisation d'algorithmes plus complexes et qui ont chacun certaines particularités. Fréquemment, ces algorithmes nécessitent un ensemble de données qui seront utilisées afin d'être « calibrés », c'est-à-dire, 1/ trouver l'algorithme parmi ceux opérationnels qui aura la meilleure performance par rapport aux données disponibles (cf. [103], par exemple); 2/ trouver l'importance de chaque *feature*, soit parce que l'algorithme est intrinsèquement introspectable comme un arbre de décision, soit par l'utilisation de méthodes annexes comme via des explicateurs comme Shapley³; 3/ reprendre ces modèles de données

3. cf. <https://christophm.github.io/interpretable-ml-book/shapley.html>

découverts, par exemple, un arbre de décision permettant de classifier ou d'expliquer une classification afin de les appliquer à des nouvelles données.

Ainsi, la littérature nous permet de constater l'existence d'une grande variété de méthodes et qu'aspirer à les modéliser toutes à l'avance serait un travail impraticable. En revanche, il faut permettre de prévoir la possibilité de les intégrer et, peut-être, les modéliser au fur et à mesure des besoins exprimés.

À noter qu'en plus de l'utilisation d'outils statistiques généralistes (R, SAS, Stata, SPSS, etc.), des outils dédiés — comme RapidMiner, WEKA, Orange, Knime, etc. — ont la majorité de ces algorithmes intégrés, ce qui facilite la tâche, notamment lors du choix de celui le plus performant selon les données à disposition et de la construction du modèle de données.

À noter également que le fait d'avoir un nombre considérable de possibilités d'analyses renforce les questions soulevées à propos du nombre de choix qui restent aux utilisateurs, comme discuté auparavant (cf. 5.4.1), d'autant plus que, nous l'espérons, de nouvelles possibilités d'analyses seront rendues possibles — ou enrichies — grâce à l'enrichissement des traces que l'approche par l'IDM rendra possible.

5.7 LA VISUALISATION DES RÉSULTATS DES ANALYSES

Un sujet commun dans l'analytique des apprentissages avec le numérique — le deuxième le plus commun selon [113] —, est la visualisation et l'interprétation des résultats dans des tableaux de bord afin d'offrir un format plus simple pour l'interprétation et l'exploration de données sinon complexes et confuses [61]. Une préoccupation majeure concernant l'analytique des apprentissages avec le numérique, est que la visualisation pour tous les types de données doit permettre aux parties prenantes d'interpréter facilement les données et ainsi accroître leur prise de conscience et soutenir leur réflexion [31].

S'il existe une grande variété de manières de rendre les résultats des analyses, celle la plus utilisée est certainement la visualisation dans des tableaux de bord [4]. Il s'agit de montrer des résultats de(s) analyse(s) au sein d'une unique interface permettant ainsi de les observer en un coup d'œil. Les données sont fréquemment visualisées via des graphiques, listes ou tables. Mazza [97] affirme que, grâce à notre capacité de perception visuelle, choisir une représentation visuelle est souvent plus efficace que de fournir du texte brut ou des tables de données. [125, p. 37] définit les tableaux de bord de la façon suivante :

« *A learning dashboard is a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations* ».

Une revue systématique de la littérature a été conduite récemment sur le sujet des tableaux de bord. Schwendimann et al. [125] ont identifié quatre groupes pour lesquels des tableaux de bord ont été construits : instructeurs (75 % des papiers analysés), apprenants (51 %), administrateurs (11 %) et chercheurs (9 %). Cette étude a également évalué les objectifs déclarés des tableaux de bord, en trois groupes : autosuivi (*self-monitoring*) (51 %), suivi des autres (71 %) et suivi administratif (2 %). Trois articles (5 %) n'ont pas explicitement indiqué un objectif pour leur tableau de bord. Enfin, par rapport aux types de visualisations utilisées pour les indicateurs affichés, la revue a permis de découvrir que les histogrammes sont les plus choisis (60 %), suivis des graphiques linéaires (44 %), tables (38 %), diagrammes à secteurs (27 %) et graphiques en réseau (18 %). La Figure 5.8 permet de visualiser les types de graphiques choisis pour chaque partie prenante.



FIGURE 5.8 – Les types de graphiques pour chaque partie prenante, issu de [125].

Concernant la compréhension de différents types de visualisations utilisées dans des tableaux de bord, [22] suggère la nécessité d'enquêter davantage sur leur utilité relative. Cette étude a identifié en enquêtant sur 38 apprenants que, mis à part le diagramme radar, les nuages de mots et le cercle (réparti sur plusieurs critères), toutes les vues (skill meter, grille, table, réseau, carte des prérequis, arbre hiérarchique, carte conceptuelle) ont été comprises par au moins les deux tiers des participants. De plus, toute visualisation n'est pas adaptée à tout type d'information, certaines d'entre elles posent des problèmes déjà bien identifiés par la communauté IHM (Interfaces Humain-Machine).

D'autres études peuvent être citées, comme [127] et [27], où des efforts ont été entrepris afin de rendre les tableaux de bord interactifs, personnalisés [105] ou adaptables [44]. Des travaux [112] s'intéressent

également au besoin d'avoir des tableaux de bord différents au fil du temps.

5.7.1 *Discussion*

Bien que les tableaux de bord soient clairement le choix le plus fréquent permettant de rendre visibles les résultats d'une analyse à une — ou plusieurs — des parties prenantes, il ne s'agit pas de la seule façon de faire [31, 38, 57]. Nous pouvons citer, par exemple, une notification affichée à un apprenant quand il ou elle a atteint un seuil, disons un nombre de « concepts » visualisés lors d'une session d'étude ; un message envoyé à un instructeur toutes les semaines avec un récapitulatif de la situation des apprenants ; ou encore, des visualisations tout au long des pages d'un cours qui permettent de savoir, disons, le nombre de fois où chaque section a été accédée.

Concernant les informations affichées dans ces tableaux de bord, la littérature (cf. par exemple la revue systématique de la littérature dans le Chapitre 9) nous montre que souvent ces systèmes se concentrent sur des visualisations du type « descriptives » comme le temps passé en ligne ou des comparaisons avec les pairs. Néanmoins, les résultats d'une étude comparative de trois systèmes d'analyse des apprentissages ont montré la préférence pour des systèmes plus détaillés avec des analyses élaborées et des recommandations personnalisées [124]. Cela nous mène à nous assurer que des analyses plus poussées puissent être modélisées et rendues aux parties prenantes — mais sans oublier que, lors de l'implémentation dans des contextes pratiques, ces visualisations doivent être testées par rapport à leur compréhension effective. Un autre aspect est celui relatif à une implémentation où les traces, analyses et visualisations sont pensées (et modélisées) en amont : dans ce cas, les visualisations pourraient montrer les détails des documents, mais aussi s'appuyer sur la connaissance de leurs structures, afin de montrer les résultats des analyses sur chaque partie concernée, par exemple, plutôt que dans une interface dédiée séparée.

VERS UNE MODÉLISATION DE L'ANALYTIQUE DES APPRENTISSAGES

6.1 INTRODUCTION

Un système complexe, comme un système permettant d'effectuer l'analytique des apprentissages avec le numérique, pourrait être modélisé selon de nombreux points de vue (par exemple, structurel, dynamique ou encore fonctionnel) et ces points de vue pourraient également varier en termes d'abstraction (idées ou concepts, interfaces, composants abstraits, composants logiciels physiques) et de précision (ébauche, solution à améliorer, solution finale) [77].

Dans ce chapitre, nous présentons les définitions des termes « cadre » (*framework*) et « méthode », utilisés dans la suite du manuscrit. Nous verrons que la modélisation selon une approche IDM est l'utilisation d'une représentation simplifiée d'un ou plusieurs aspects du monde réel, avec un objectif donné [77]. Nous présentons les travaux qui se sont plus ou moins approchés de la modélisation d'une démarche d'analytique des apprentissages avec le numérique, qui manifestent un intérêt vers ce but d'abstraire les différents composants logiciel d'un système complexe.

6.2 CADRES ET MÉTHODES

Comme nous l'avons vu (cf. Chapitre 2), un modèle représente un système selon un point de vue choisi, à un niveau d'abstraction qui a pour but de faciliter la conception de cet aspect particulier du système [77, p. 4]. Un **modèle** dans ce domaine est usuellement représenté sous la forme d'un schéma, d'un état, d'objet ou d'une situation du monde réel, dans un format simplifié. Le point de vue (structurel, dynamique, fonctionnel, etc.) déterminera comment la situation est représentée, quelques modèles couramment représentés sont des modèles de cycle de vie, de processus, de *workflow*, etc.

Une **méthode**, de son côté, est une approche systématique adoptée afin d'atteindre un résultat ou un objectif spécifique. C'est une manière de penser et d'agir qui consiste à un ensemble de démarches. La définition retenue est la suivante : « Manière de mener, selon une

démarche raisonnée, une action, un travail, une activité ; technique » [89].

Un **cadre** (*framework*) se situe entre un « modèle » et une « méthode ». Un cadre définit la structure ou le système qui permettra d'arriver à un résultat envisagé. Parfois, un cadre contiendra un ou plusieurs modèles, mais les cadres tendent à être moins rigides et détaillés, donnant plutôt lieu à une vision d'ensemble à ses utilisateurs.

6.3 DES CADRES POUR L'ANALYTIQUE DES APPRENTISSAGES

Une fois ces termes précisés par rapport à leurs usages dans le présent texte, nous présentons deux des cadres les plus mentionnés dans le contexte de l'analytique des apprentissages avec le numérique, tous deux proposés en 2012. Le premier cadre que nous étudierons est celui proposé par Greller et al. [67]. Il s'agit d'un cadre de conception générique pouvant servir de guide pour la conception ou bien la configuration de systèmes envisageant l'analytique des apprentissages avec le numérique, basé sur des « dimensions ». La Figure 6.1 nous montre les dimensions retenues par ces auteurs : parties prenantes, objectifs, données, instruments, contraintes externes et internes.



FIGURE 6.1 – Dimensions critiques de l'analytique des apprentissages avec le numérique, issu de [67].

Selon ce cadre, les parties prenantes se divisent en sources de données (typiquement les apprenants) et les clients de l'analyse des données, qui vont prendre des décisions à partir de ces analyses, par exemple un enseignant qui utilise des données des apprenants. La

dimension « data » englobe à la fois les jeux de données (ouverts ou protégés), les indicateurs pertinents et l'échelle de temps. La théorie pédagogique, la technologie et la présentation sont regroupées dans la dimension appelée « instruments ». Les objectifs sont classés en prédiction et réflexion. Enfin, deux types de limitations sont à prévoir : les compétences (nécessaires à l'utilisation/interprétation des analyses) et l'acceptation du système par les parties prenantes (internes) et les limitations externes, comme les questions éthiques, la préservation de la vie privée et l'ajustement temporel (par exemple, est-ce que les résultats permettent une intervention à temps?).

Le deuxième cadre (selon la définition donnée ci-dessus, les auteurs, eux, parlent d'un modèle) que nous discuterons est celui proposé par Chatti et al. [25], basé sur quatre dimensions. La Figure 6.2 nous détaille les aspects concernés par chacune de ces dimensions.

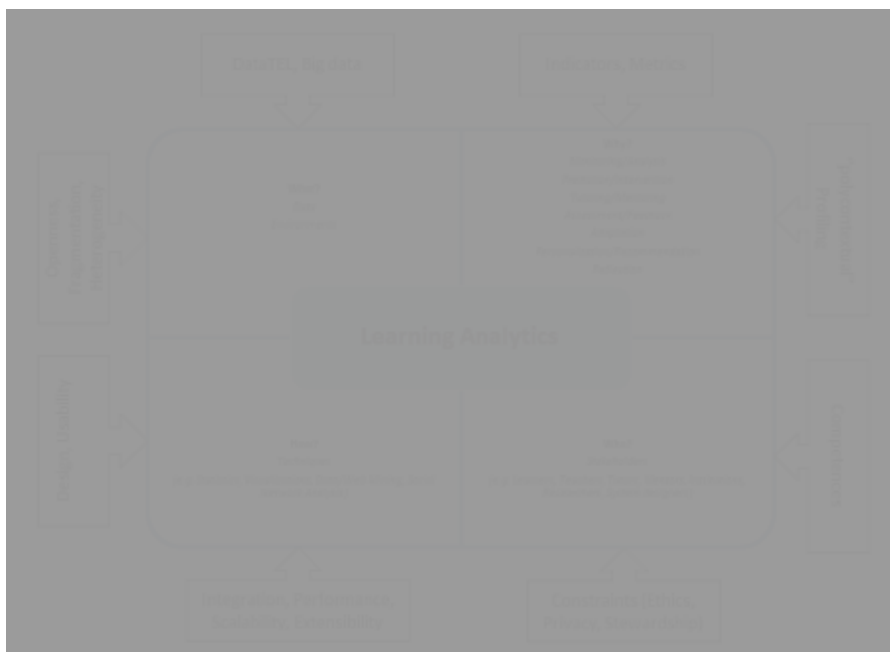


FIGURE 6.2 – Modèle de référence pour l'analytique des apprentissages avec le numérique, issu de [25].

Quoi Quel type de données le système collecte, gère et utilise-t-il pour l'analyse? Cette dimension concerne les données et les systèmes autour de la collecte de celles-ci et, par conséquent, mène à se poser les questions de leur ouverture, fragmentation et hétérogénéité, qui potentiellement influent sur les types d'analyses qui pourront être menées.

Qui Qui est ciblé par l'analyse? Il s'agit des parties prenantes, où celles identifiées auparavant (cf. Section 5.4) sont additionnées des concepteurs de systèmes. En outre, les auteurs soulignent l'import-

tance de la participation des différents acteurs comme la clé d'une acceptation et une utilisation plus large des outils visant à améliorer l'enseignement et l'apprentissage. Les notions de compétences et d'éthique complètent cette dimension.

Pourquoi Pourquoi le système analyse-t-il les données collectées ? Les différents objectifs (cf. 5.3) qui orientent les analyses (et donc leur typologie) se trouvent dans cette dimension, qui incorporent également les concepts d'indicateur et métrique (cf. Chapitre 9, Section 9.3.3).

Comment Comment le système effectue-t-il l'analyse des données collectées ? Cette dimension comporte le design du processus d'analyse et son utilisabilité, mais aussi plusieurs aspects autour des « techniques » employées. Cela inclut les méthodes d'analyse utilisées, les visualisations proposées, ainsi que la performance, l'intégration, la capacité à évoluer et l'extensibilité du système.

Un troisième cadre, moins généraliste par rapport à une implémentation d'analytique des apprentissages avec le numérique et plus intéressé par la scénarisation pédagogique (*learning design*) est AL4LD (*Analytics Layers for Learning Design*) [73]. Il capture les relations et les interactions entre les données éducatives et les modèles d'apprentissage en trois couches (cf. Figure 6.3) : l'expérience d'apprentissage, la scénarisation pédagogique elle-même et la communauté des éducateurs. Le fait de s'intéresser par le contexte (les intentions pédagogiques) aux interactions qui se produisent entre apprenants et contenus a l'avantage de soutenir une analyse avec des intentions plus poussées, comparant, par exemple, ce qui était attendu/préscrit à ce qui s'est effectivement passé. Cet aspect n'est pas présent — au moins explicitement — dans les deux cadres discutés précédemment.

6.3.0.1 Discussion

Les dimensions proposées par ces cadres (dont la plupart ont été discutées en détail dans le Chapitre 5) sont importantes pour toute implémentation d'une démarche d'analytique des apprentissages avec le numérique, mais l'organisation et le regroupement de ces différents aspects dans des dimensions ne se traduisent pas en une abstraction permettant de dégager les différentes parties d'un système qui permettra une effective implémentation dans la pratique. En résumé, ce sont des aspects qui doivent être considérés par une telle implémentation, mais qui ne se traduisent pas en parties d'un système informatique, il est donc intéressant d'explorer s'il est possible de proposer une modélisation se focalisant précisément sur ces aspects susceptibles de donner lieu à une implémentation.

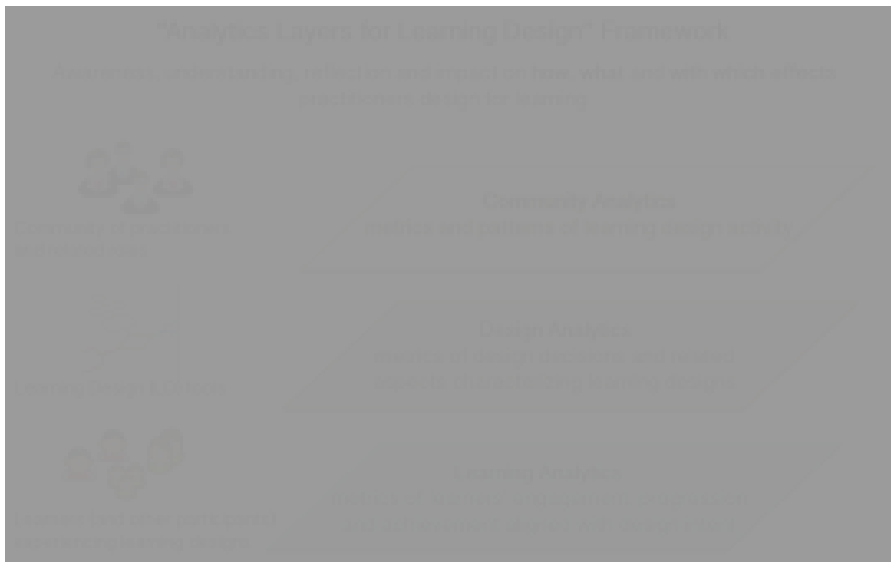


FIGURE 6.3 – Représentation graphique du framework AL4LD, issu de [73].

6.4 VERS DES MODÈLES POUR L'ANALYTIQUE DES APPRENTISSAGES

Parmi les travaux allant vers une modélisation d'analytique d'apprentissages avec le numérique, dans [29] les auteurs présentent un métalangage dédié à la définition de données et de besoins : l'*Usage Tracking Language* (UTL), qui permet de définir des informations qui vont enrichir les traces collectées et structurer l'information via la définition de données intermédiaires et des indicateurs. Ce modèle a l'avantage de s'adapter aux éléments qu'il décrit via l'ajout/modification des informations déjà disponibles et son focus est d'habiliter la prise en compte des scénarisations pédagogiques au processus de collecte et d'analyse des traces.

D'un côté, sa partie UTL/T (traces) permet de représenter la transformation des traces générées en les décrivant comme un ensemble de données comprenant des informations sur la source de celles-ci. Il est possible également de définir la relation de ces dernières avec les données primaires attendues afin de créer un indicateur et de définir la méthode d'extraction de ces informations qui seraient contenues dans les traces.

D'un autre côté, sa partie patron (UTL/P) se fonde sur le modèle « DGU » (*Defining, Getting, Using*) et propose de penser la trace d'usage en amont de la session d'apprentissage [30] et non une fois les traces collectées, qui semble être la démarche la plus choisie selon la littérature. Ceci vise à permettre une future comparaison de scénarios descriptifs avec des scénarios prédictifs de la situation pédagogique et

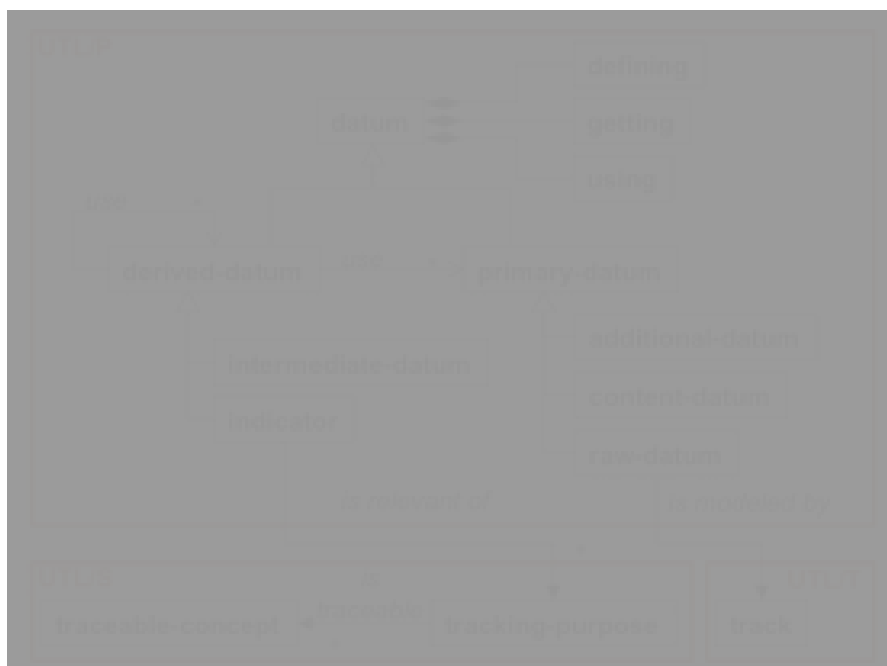


FIGURE 6.4 – Modèle conceptuel du métalangage UTL 2, issu de [30].

permet donc de décrire la structure d'un observable. Enfin, la partie UTL/S (scénario pédagogique) permet de lier de manière sémantique le scénario pédagogique à un indicateur donné.

UTL permet ainsi de définir les besoins en termes de données brutes et non la façon de les obtenir. Il permet aussi de décrire les concepts du scénario pédagogique à tracer en leur associant un usage observé.

Un deuxième effort lié aux traces et qui va vers une modélisation de ce traçage est Trace-Based Reasoning (TBR) [39]. Selon ses auteurs, TBR vise à permettre un processus de raisonnement plus dynamique et interactif, basé sur trois étapes : élaboration, récupération et réutilisation. La phase d'élaboration vise à répondre à la difficulté de prendre en compte le contexte dans la phase de raisonnement. En conséquence, chaque expérience (trace) est conservée avec son contexte, qui est alors accessible aux prochaines étapes.

Le modèle d'une trace en TBR (M-Trace) est la description formelle de la structure et du contenu d'une trace. Le modèle doit alors fournir des informations comme l'unité de temps utilisée, les « obsels » — « la contrepartie numérique d'un événement survenu dans le monde réel » [39, p. 364, notre traduction], voir le détail dans la Figure 6.5 — et sur les relations entre les éléments.

Les obsels sont ainsi formellement décrits dans le modèle de traces et chacun est caractérisé par un nom, un horodatage et un ensemble de propriétés, usuellement des paires attribut-valeur. Les auteurs men-



FIGURE 6.5 – Exemple d'un M-Trace, issu de [39].

tionnent rapidement qu'un métamodèle devrait définir ces propriétés afin d'assurer l'interopérabilité entre les différentes traces.

6.4.0.1 Discussion

Tant TBR que UTL s'intéressent à l'enrichissement des analyses des apprentissages avec le numérique via la prise en compte de la scénarisation pédagogique des activités qui produiront les traces à analyser.

Les chercheurs à l'origine d'UTL proposent un langage de modélisation à la fois des traces (données brutes), des transformations de ces traces (données intermédiaires ou indicateurs), mais également de la scénarisation pédagogique elle-même. Cette démarche permettrait d'enrichir ces données et de comparer, en fin de compte, ce qui aurait été scénarisé avec ce qui s'est réellement passé. Ces travaux nous montrent l'intérêt d'un enrichissement des analyses (et des traces pour le faire) avec des informations de contexte pouvant être modélisées, dans ce cas, du scénario pédagogique, dans notre cas, par des informations des différents documents consultés.

Un point intéressant est que les chercheurs à l'origine de TBR mentionnent que le nom obsel est choisi à la place d'événement afin de marquer le fait qu'un obsel est enregistré avec un propos (défini à l'avance). Cette démarche, de penser aux traces en amont — que nous avons examinée dans le Chapitre 5 —, ainsi que le fait que cette collecte soit faite en interne avec plus de collecte d'informations de contexte et selon les besoins, sont deux atouts. Cela conforte l'idée

d'une modélisation en amont dans notre cas. Néanmoins, même si l'enrichissement des traces dans TBR passe par une modélisation de ces derniers, les ingénieurs qui définissent ces modèles doivent le faire avec soin « afin qu'ils soient suffisamment riches pour prendre en charge des utilisations multiples » [39, p. 367, notre traduction], en d'autres mots, les traces sont enrichies/modélisées, mais sans avoir un but défini en termes d'analyses à prévoir. Il nous paraît ainsi intéressant d'explorer comment guider cette modélisation des traces non seulement en amont et via une modélisation, mais également en partant d'une question spécifique, traduite en un ou plusieurs indicateurs. En d'autres mots, c'est parce que nous essayons de répondre à une question (ou que nous avons un objectif donné, comme examiné dans 5.3) que nous enregistrons des interactions et que les traces sont potentiellement enrichies de certaines informations liées aux documents consultés.

Pour conclure, ici nous constatons un effort fait afin de modéliser les traces et indicateurs, mais aussi les scénarios pédagogiques, afin que l'un puisse enrichir l'autre lors des analyses. Notre approche est donc proche de celles citées, mais nous comptons explorer une démarche guidée par les documents pédagogiques afin de compléter et enrichir ces travaux. Comme nous l'avons vu, la production documentaire dans notre contexte s'appuie déjà sur la modélisation, ce qui nous permet d'explorer cette connaissance fine du document dans le cadre des analyses d'analytique des apprentissages avec le numérique.

Quatrième partie

CONTRIBUTION : LA PRÉPARATION DU MÉTAMODÈLE

Dans les chapitres précédents, nous avons mis en évidence que l'implémentation d'un système d'analytique des apprentissages avec le numérique est en réalité une succession d'étapes impliquant plusieurs parties prenantes. Dans le Chapitre 7, nous proposons la formalisation de ce cycle, ce qui nous permettra de comparer avec les phases d'une même implémentation via une approche par l'IDM. Dans le Chapitre 8, nous présentons une étude préliminaire qui nous a permis notamment de comprendre les besoins de futurs utilisateurs en termes de résultats des analyses utilisant notre métamodèle. Nous nous intéressons notamment à leur homogénéité, ou non, en termes de besoins en ayant comme prémisses leur usage d'un même modèle documentaire. Enfin, dans le Chapitre 9, nous présentons les résultats d'une revue systématique de la littérature s'intéressant aux différents indicateurs utilisés dans des systèmes proposant l'analytique des apprentissages avec le numérique. Ces résultats sont fondamentaux afin de bien connaître les nombreuses caractéristiques de ces indicateurs et leurs contextes d'usage, connaissance qui nous est indispensable pour la proposition de notre métamodèle.

LES ÉTAPES ET PARTIES PRENANTES : LES CONTRIBUTIONS DE CHACUN

7.1 INTRODUCTION

La littérature nous apprend que l'implémentation d'un système d'analytique des apprentissages avec le numérique est en réalité une succession d'étapes impliquant plusieurs parties prenantes [90]. Ces acteurs ont un objectif en commun de répondre à un besoin spécifique et lors des étapes chacun contribue avec sa spécialité. Afin de comprendre comment se déroule traditionnellement ce cycle de collecte, analyse et visualisation d'un ou plusieurs indicateurs dans un contexte d'analytique des apprentissages avec le numérique, nous proposons dans ce chapitre la formalisation de ce cycle. Cette formalisation — inspirée par [90] — nous sera utile afin de la comparer avec les phases d'une même implémentation via une approche par l'IDM.

7.2 UNE ANALYSE ISSUE DE LA LITTÉRATURE

Les LMS enregistrent des traces d'interaction avec la plateforme de la part des étudiants et instructeurs. La fonctionnalité de rapports de Moodle¹, par exemple, fournit des informations à trois niveaux : site, cours ou activité. Les différents rapports affichent des informations, entre autres, l'activité liée au cours (le plus actif, celui avec le plus grand nombre d'utilisateurs inscrits, les commentaires des utilisateurs, etc.), ou encore l'achèvement du cours et de l'activité, le temps passé pour terminer une activité et des informations sur les résultats obtenus, en forme de graphiques et statistiques avec une fonction de filtrage [35]. Comme le remarque Conde et al. [35], cet outil de rapport par défaut de Moodle offre une grande quantité d'informations, mais malgré des informations très détaillées (provenant de données « brutes »), il n'est pas capable de fournir des informations agrégées significatives, comme répondre à une question simple, par exemple : « Combien d'apprenants n'ont pas encore commencé un cours ? » Ou encore à des questions plus précises comme celles utilisant des algorithmes plus complexes, comme ceux mentionnés dans la Section 5.6.

1. <https://moodle.org>

Une autre manière d’extraire des connaissances est l’utilisation d’outils ad hoc, conçus pour suivre ou analyser des informations plus spécifiques et pour répondre à un besoin spécifique dans un contexte très concret — même si ces solutions sont généralement peu flexibles ou évolutives. C’est ce cas que nous tentons de décrire dans le texte qui suit, dont les étapes sont résumées dans la Figure 7.1.

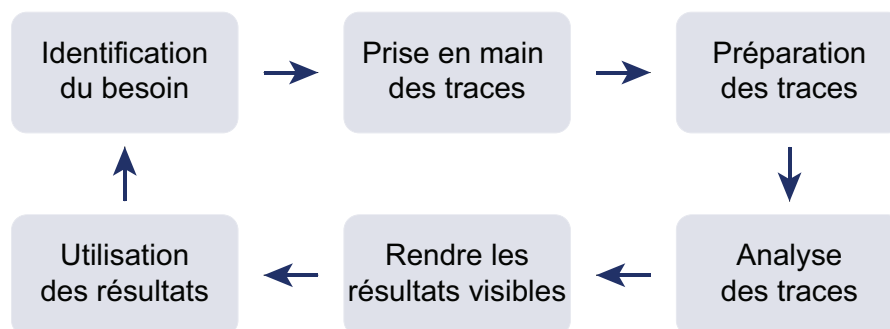


FIGURE 7.1 – Les étapes d’implémentation : une analyse issue de la littérature.

7.2.1 Identification du besoin

Une première étape correspond à l’identification du besoin. Dans cette étape, un besoin est exprimé par une des parties prenantes, que [90] propose d’appeler le « décideur ». L’analyse de la littérature nous permet d’identifier plusieurs acteurs ayant ce rôle, selon les différentes démarches adoptées. Un instructeur peut exprimer un besoin spécifique à son contexte et à la structure d’un cours qu’il souhaite améliorer, par exemple en découvrant quelles parties de celui-ci sont les plus vues par les apprenants ayant réussi le cours. Une institution peut, via son service pédagogique, mener un projet visant la création d’un tableau de bord générique disponible sur le LMS et à disposition de tout instructeur et apprenant de tous cursus confondus. Une entreprise proposant une solution (plateforme ou autre) pour l’apprentissage en ligne peut également proposer un ou plusieurs indicateurs par défaut liés à une telle solution. Enfin, un chercheur peut proposer un nouvel indicateur issu de ses recherches qu’il va tester lors d’un cours en partenariat avec une équipe enseignante. Les besoins et décideurs que nous venons de décrire ne représentent pas une liste exhaustive, mais devraient permettre d’appréhender les enjeux de cette première étape qui déclenche le cycle que nous étudions. Suite à l’identification du besoin, il doit être correctement décrit de manière à être compris par les différentes parties prenantes intervenant dans les étapes qui vont suivre. Ceci se fait de diverses manières selon les approches retenues, par exemple la conception centrée sur l’utilisa-

teur, la conception innovante, l'analyse fonctionnelle, la conception participative, etc.

7.2.2 *Sélection des traces*

La deuxième étape consiste en la sélection des traces pertinentes permettant de répondre au besoin qui vient d'être exprimé. Ce moment qui commence invariablement par faire intervenir l'expert de l'environnement d'apprentissage [90]. En effet, l'expert de l'environnement d'apprentissage possède une connaissance approfondie des traces collectées et de leur structure. Lebis [90] nous fait part de l'importance de ce spécialiste, avec des soucis prévisibles de communication si cette dernière se passe directement entre le décideur et l'analyste : « Ce rôle requiert de connaître l'environnement d'apprentissage dans son ensemble, afin de choisir méticuleusement quelles traces doivent être choisies » [90, p. 81].

7.2.3 *Prise en main des traces*

La troisième étape est la prise en main par l'analyste qui va préparer l'analyse qui sera effectuée lors de l'étape suivante. L'analyste peut être un chercheur ou un statisticien qui connaît bien les méthodes d'analyse, mais qui très souvent n'a pas de connaissances approfondies sur les traces à disposition ni sur le contexte d'apprentissage. Il va d'abord s'emparer des besoins exprimés par le décideur et des traces mises à disposition par l'expert de l'environnement d'apprentissage et également s'approprier le domaine d'apprentissage pour proposer, à un niveau conceptuel, l'analyse qui sera menée.

7.2.4 *Analyse des traces*

La quatrième étape consiste en l'implémentation de l'analyse définie auparavant. À partir des traces sélectionnées lors de la troisième étape par l'expert de l'environnement de l'apprentissage et de la définition conceptuelle de l'analyse qui va répondre aux besoins exprimés, l'analyste procède aux trois phases qui typiquement constituent l'analyse proprement dite : le prétraitement, l'analyse (ou application d'opérateurs [90]) et le post-traitement.

Comme indiqué précédemment, les méthodes de collecte de données sont souvent peu contrôlées, ce qui pourrait entraîner des valeurs hors limites (une note négative, par exemple, des combinaisons de données impossibles ou très improbables, des valeurs manquantes, etc.).

Ainsi, avant toute analyse, la représentation et la qualité des données doivent être vérifiées. Entre autres, le prétraitement des données comprend le nettoyage, la normalisation, la transformation, l'extraction et la sélection de caractéristiques. Certains de ces processus peuvent être effectués lors de l'étape de sélection des traces pertinentes, d'autres lors du prétraitement proprement dit.

L'analyse du jeu de données retenu et pré-traité peut alors se poursuivre. Finalement, un post-traitement peut être nécessaire afin de rendre le résultat plus compréhensible à ceux qui consulteront les résultats (via la visualisation dans un tableau de bord ou sous une autre forme).

7.2.5 *Rendre les résultats visibles*

La cinquième étape est le moment de rendre les résultats accessibles aux bénéficiaires pour qui l'analyse a été conçue. Comme nous l'avons vu (cf. Section 5.7), les instructeurs sont les premiers bénéficiaires d'une démarche d'analytique des apprentissages avec le numérique, suivis des apprenants et de différents administrateurs. Cette étape, menée par l'analyste avec l'aide du décideur ou bien d'un chercheur, joue un rôle essentiel dans l'utilisation effective des résultats obtenus et, en même temps, pourrait entraîner un mauvais usage/compréhension dans le cas où elle serait mal menée. Le risque consiste en une interprétation incorrecte ou imprécise de l'information fournie comme une mauvaise interprétation d'un graphe ou une démotivation de la part de l'apprenant quand ce que l'on cherchait à produire était exactement l'inverse (cf. par exemple [124]). La façon dont l'information qui a été trouvée sera rendue doit en effet être pensée dès les premières réflexions au sein d'un tel projet.

7.2.6 *Utilisation des résultats*

La sixième et dernière étape consiste en l'utilisation effective des informations produites, ce qui peut être fait avec ou sans intervention, comme mentionné auparavant. L'amélioration attendue liée au besoin exprimé au début du projet pourra alors être évaluée, formellement ou informellement. Une éventuelle action, selon le type d'analyse, pourra avoir lieu, par exemple une re-conception d'un cours, ou encore un changement du temps passé par les apprenants dans les différentes parties d'un cours. Naturellement, il est important, lors de l'évaluation, de s'assurer que les utilisateurs utilisent convenablement les

informations produites, en évitant qu'ils prennent des décisions non pertinentes dans le contexte en question.

7.3 UNE ANALYSE SUIVANT L'APPROCHE IDM

Dans la suite, nous précisons également les étapes envisagées et parties prenantes qui seraient susceptibles de prendre part à un cycle de collecte, d'analyse et de visualisation d'un ou plusieurs indicateurs dans un contexte d'analytique des apprentissages avec le numérique dans le cas où ces étapes passent par une métamodélisation, suivant donc une approche par IDM. Comme nous le constaterons, ce cycle diffère du cycle classique que nous venons de décrire par rapport à la succession de ces étapes, aux acteurs impliqués et, sans surprise, à sa proximité avec les étapes de conception des documents qui seront consultés et qui seront à l'origine de la production des traces via les interactions avec ces derniers. Ces étapes sont résumées dans la Figure 7.2.

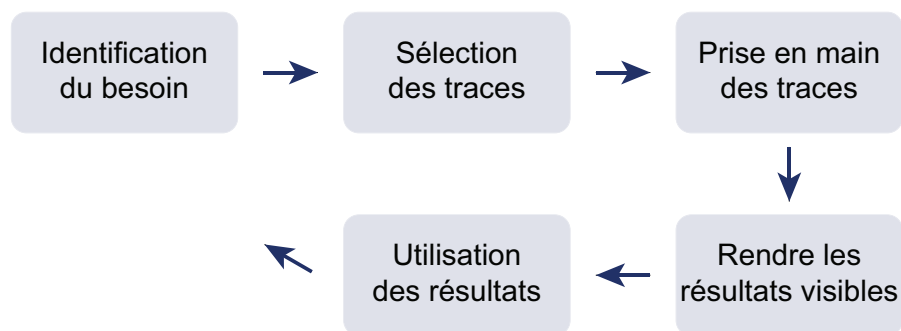


FIGURE 7.2 – Les étapes d'implémentation : une analyse suivant l'approche IDM.

7.3.1 Identification du besoin

La première étape consiste également en l'identification du besoin. Un besoin est exprimé par une des parties prenantes, que nous nommerons également le « décideur ». Ce décideur provient de la communauté utilisant le modèle documentaire en question : un instructeur, des apprenants, un administrateur pédagogique ou encore de la communauté rassemblée dans la « place des évolutions » du modèle documentaire². Il s'agit d'un service de l'association Scenari où les utilisateurs du modèle documentaire peuvent se rassembler, discuter

2. <https://scenari.org/evolutions/opale/co/home.html>

de l'avenir du modèle et regrouper leurs forces pour concrétiser les évolutions souhaitées.

7.3.2 *Identification de l'indicateur*

La deuxième étape consiste en la définition de l'indicateur (ou des indicateurs) permettant de répondre au besoin exprimé. L'idée est alors de se baser sur le concept d'indicateur, c'est à partir de la définition de celui-ci que débutera la modélisation. Lors de cette étape le modélisateur va appréhender le besoin du décideur et si besoin consulter un analyste concernant l'indicateur en question avec pour but de pouvoir le modéliser en ayant toutes les informations nécessaires : À qui est-il destiné ? A-t-il déjà été proposé dans la littérature ? Comment devrait-il être rendu aux utilisateurs qui vont le consulter (tableau de bord, type de graphique, message d'avertissement, etc.) ? L'idée est que la littérature, avec un vaste nombre d'études déjà réalisées puisse aider à guider la modélisation, évitant de se pencher à nouveau sur des questions auxquelles une réponse a déjà été apportée. Dans ce sens, la revue systématique de la littérature présentée dans le Chapitre 9 non seulement nous guide dans nos travaux de métamodélisation, mais aussi pourra servir aux modélisateurs lors de leurs modélisations de différents indicateurs.

7.3.3 *Modélisation de l'indicateur*

Une fois toutes ces informations définies, le modélisateur débute la troisième étape. Il procède alors à la modélisation de l'indicateur. Cette modélisation concerne trois grandes parties : les traces, l'indicateur et la visualisation.

Tout d'abord, le modélisateur définit les traces qui devront être générées pour que l'indicateur en question puisse être calculé. Il décide entre autres le type d'événement déclencheur d'un enregistrement d'une ligne dans le journal de traces, la structure qu'il prendra (c'est-à-dire les informations enregistrées) et les parties du document concernées, par exemple, toutes les pages, les sous-parties des pages qui sont un bloc « concept », toute tentative de réponse à un quiz, etc. Il définit également comment ces traces seront converties périodiquement dans des tables prêtes à l'emploi lors du calcul de l'indicateur.

Ensuite, le modélisateur modélise l'indicateur lui-même. Il définit les entrées et d'où/à quel moment viendra cette information (une date de début qui peut être fixe, calculée, définie par un des acteurs, etc.). Il insère l'analyse à être effectuée (application des opérateurs) et la sortie

attendue. Le prétraitement et le post-traitement se font au sein de ce code mais, d'une certaine manière, également par l'étape précédente et du fait d'avoir un contrôle plus important sur les données de départ. Le modélisateur définit également le planning des calculs — un agenda de la fréquence souhaitée des calculs — pour chaque indicateur selon le type d'analyse et objectifs exprimés.

Enfin, le modélisateur modélise comment le résultat sera rendu aux bénéficiaires.

7.3.4 *Rendre les résultats visibles*

La quatrième étape est le moment de rendre les résultats accessibles aux bénéficiaires pour qui l'analyse a été conçue. Les enjeux sur l'utilisation de ces résultats sont semblables à une démarche traditionnelle. Cependant, les décisions sur cette étape sont faites en amont (définition en étape 2 et implémentation en étape 3).

7.3.5 *Utilisation des résultats*

La cinquième et dernière étape consiste à utiliser les informations produites — par exemple un instructeur qui consulte les résultats d'analyse concernant les apprenants de son cours, ou des apprenants qui consultent leur progression par rapport à ses objectifs, etc. —, ce qui peut être fait avec ou sans intervention, comme mentionné auparavant. L'amélioration attendue liée au besoin exprimé au début du projet pourra alors être évaluée, formellement ou informellement. L'avantage ici, c'est que, le modèle documentaire étant utilisé dans différents contextes, une analyse des résultats peut être plus consistante du fait qu'une partie ou bloc, du même type, pourrait être analysée dans des contextes diversifiés.

ÉVALUATION EXPLORATOIRE DES BESOINS

8.1 INTRODUCTION

Nous avons mené une première étude exploratoire des besoins des utilisateurs des systèmes utilisant potentiellement notre métamodèle dans le futur. Dans ce Chapitre nous décrivons cette étude, nous focalisant notamment sur deux questions : 1/ les besoins et leur homogénéité — ou non — du fait de l'utilisation d'un même modèle documentaire et 2/ la réception (utilité, compréhension, etc.) des indicateurs qui nous servent d'exemple, principalement en ce qui concerne les indicateurs que nous appelons « enrichis » — de la structure ou bien de la sémantique des documents.

Ce Chapitre nous permet notamment de répondre aux questions de recherche **RQ3** Un modèle documentaire correspond-il à un besoin homogène en termes d'analytique des apprentissages avec le numérique et donc à une série d'indicateurs d'apprentissage bien identifiés ? et **RQ4** Y a-t-il une différence dans la réception des visualisations basées sur des indicateurs enrichis ? Ces questions, comme nous le verrons dans la suite (cf. Section 8.3.1), sont déclinées en des sous-questions suivant la structure de l'étude que nous avons menée.

8.2 ÉNONCÉ DU PROBLÈME

Nous avons mené une première étude qualitative exploratoire afin de comprendre les besoins des utilisateurs en termes d'analytique des apprentissages avec le numérique. Les résultats de cette étude devraient nous aider à définir le niveau d'abstraction nécessaire au métamodèle des indicateurs ainsi que les rôles/actions que chaque partie prenante devrait avoir. Ces résultats devraient ainsi guider les prochaines étapes vers la métamodélisation proposée dans le Chapitre 10.

8.3 BUT DE L'ÉTUDE

Le but de cette étude est de recueillir les retours des futurs utilisateurs (apprenants) sur les solutions qui seront potentiellement proposées sur la plateforme en ligne en termes d'analytique des ap-

prentissages avec le numérique. Nous nous sommes intéressés à deux thématiques, à savoir :

- La relation entre le modèle documentaire et les indicateurs à proposer : notre hypothèse est que chaque modèle documentaire correspondrait à des choix pédagogiques suffisamment similaires, qui correspondraient à leur tour à un nombre défini d'indicateurs utiles aux usagers de ce modèle documentaire.
- La réception des graphiques basés sur la sémantique du document : la réception des graphiques proposés par des démarches d'analytique des apprentissages avec le numérique a été étudiée, comme nous avons pu le constater dans l'analyse de la littérature (cf. chapitre 5). Cependant, nous nous demandons s'il y a une différence dans la réception des indicateurs enrichis, c'est-à-dire ceux basés sur la sémantique ou bien la structure des documents.

8.3.1 Questions de recherche

Comme nous le verrons dans la séquence, étant donné le caractère exploratoire de cette étude, nous tenons à laisser la liberté aux participants d'apporter à la discussion des sujets que nous n'avons pas anticipés (la structure narrative est détaillée dans 8.5.3). Néanmoins, une séquence narrative semi-structurée a été conçue avec la perspective de pouvoir répondre aux questions suivantes :

- (Q1) Quels sont les besoins des utilisateurs en termes d'analytique des apprentissages avec le numérique ?
- (Q2) Ces besoins (latents) incluent-ils des informations basées sur la sémantique des documents ?
- (Q3) Les interprétations des utilisateurs pour chaque graphique affiché sont-elles correctes ?
- (Q4) Y a-t-il une différence d'interprétation par les utilisateurs entre les visualisations basées sur la sémantique et celles qui ne le sont pas ?

8.4 DÉLIMITATIONS ET LIMITATIONS

Les utilisateurs participant à cette étude ne sont pas habitués à avoir des solutions d'analytique des apprentissages avec le numérique dans leurs environnements d'apprentissage et peuvent ne pas en savoir beaucoup sur le domaine, ce qui peut potentiellement représenter un frein à une participation effective. De ce fait, certaines techniques de la recherche qualitative, allant de la projection à l'exemplification,

pourraient s'avérer utiles si utilisées avec précaution afin de les aider à s'exprimer.

8.5 MÉTHODOLOGIE

Nous avons suivi une démarche à la fois inductive, via la présentation de maquettes, et déductive, laissant la parole aux participants ensuite, afin de pouvoir, par l'analyse de leurs discours, établir des dimensions ou thèmes. Le chercheur a le rôle de facilitateur. Le détail des procédures de collecte et d'analyse des données est décrit dans la suite du manuscrit.

8.5.1 *Procédures de collecte de données*

La collecte de données a été faite dans un cadre naturel familier aux personnes et aux lieux étudiés, une salle de classe dans leur université. Une méthode d'échantillonnage ciblé [117] est utilisée, c'est-à-dire l'échantillonnage par critère qui consiste à sélectionner les cas qui répondent à un critère pré-spécifié (leur participation au cours en question). La participation à l'étude a été volontaire et un formulaire de consentement (cf. annexe A) a été signé par chaque personne incluse. Tous les apprenants avaient utilisé la plateforme dans le cadre de leurs études. Un total de 28 apprenants a accepté de participer aux groupes de discussion, qui sont distribués comme suit :

1. Groupe 1 : Onze apprenants
2. Groupe 2 : Dix apprenants
3. Groupe 3 : Sept apprenants

Étant donné que les participants n'avaient pas d'expérience avec le domaine, nous avons créé une série de maquettes avec des indicateurs communs que nous avons instanciés en contexte dans l'environnement duquel sont familiers les apprenants, afin de guider les discussions. Ces maquettes sont disponibles dans l'annexe B. Certains de ces indicateurs ont été basés — ou enrichis — sur la structure ou bien la sémantique des documents provenant du modèle documentaire utilisé dans la plateforme en question. Nous avons également pris soin de proposer des visualisations variées pour un même indicateur, afin de vérifier s'il y a des préférences.

8.5.2 Procédures d'analyse des données

Les groupes de discussion ont été enregistrés et transcrits pour une analyse plus approfondie. Les transcriptions ont été analysées à l'aide d'un premier codage ouvert, puis d'un codage axial (détermination des dimensions autour de ce phénomène central) et enfin d'un codage de chaque ligne en utilisant les dimensions retrouvées.

8.5.3 Structure narrative

Les points ci-dessous sont un guide pour le facilitateur. Ils n'ont pas été tous utilisés, même si les phases ont suivi l'ordre indiqué. L'idée a été de suivre une séquence qui va du général au spécifique. Plus important encore, ont été les réactions et les réponses des participants qui ont guidé la conversation.

8.5.3.1 Première phase

Bienvenue (1 min)

- Merci de votre participation volontaire, elle est importante pour faire évoluer Libre Cours.
- Nous passerons une demi-heure ensemble à discuter de la plate-forme Libre Cours et de l'analytique des apprentissages avec le numérique (ou *Learning Analytics*).

Règles de base (1–2 min)

- Nous devons signer cet accord écrit aux fins d'administration de la recherche menée.
- L'étude est enregistrée afin que nous puissions l'analyser ensuite.
- Votre participation est totalement anonyme.
- Plus important encore, il n'y a pas de bonnes réponses, pas de jugement.

8.5.3.2 Deuxième phase

Questions d'engagement (3 min)

- Supposons que vous ayez accès aux données que vous avez produites lors de votre utilisation de la plate-forme Libre Cours. Par exemple, lorsque vous visualisez une section de cours, vous répondez à un questionnaire, etc.
- Comment utiliseriez-vous ces données ?
- Rappelez-vous du moment où vous suiviez le cours, quels types d'informations vous auraient été utiles ?
- Pourriez-vous dessiner ça ?

— Comment auriez-vous utilisé cette information ?

Projection (2 min)

- Qu'en est-il de vos collègues qui ont utilisé la même plate-forme, de quel type d'informations auraient-ils besoin ?
- À quels moments utiliseraient-ils / consulteraient-ils ces informations ?
- Comment cette information serait mieux visualisée (quels graphiques devraient être utilisés) ?

Exemplification (10 min)

- Voici quelques propositions qui pourraient être utiles aux utilisateurs de Libre Cours. Pour chaque proposition montrée :
- Qu'en pensez-vous ?
- Que montre-t-elle ? Qu'avez-vous compris ?
- Qu'est-ce que l'on apprend en la regardant ?
- Vous semble-t-elle plutôt claire ou ambiguë ?
- Est-elle utile (conscience / réflexion / motivation / changement de comportement) ? Pour qui ?
- Quelles sont vos favorites / les plus importantes (possible de faire une liste) ? Pourquoi ?

Vérification (10 min)

- Si nous séparons ces visualisations en deux groupes, comme tel. Pourriez-vous dire pourquoi elles sont dans un groupe et pas dans l'autre ? Quel est le critère utilisé ?
- Pensez-vous qu'il est utile de baser les graphiques sur la structure du document ? Pourquoi ?
- Cela aide-t-il à trouver les informations nécessaires ?
- Est-ce qu'il y a des graphiques qui semblent difficiles à envisager sans la structure du document ?
- Si vous pouviez améliorer ces graphiques, que changeriez-vous ?

8.5.3.3 Troisième phase

- Résumer avec confirmation.
- Examiner le but et demander si quelque chose a été oublié.
- Demander s'ils ont des questions sur l'étude et les méthodes utilisées.
- Remercier.

8.6 DIMENSIONS

8.6.1 Dimension 1 — L'utilité perçue

Une première dimension qui est ressortie de l'analyse est liée à l'utilité perçue de la démarche et des graphiques présentés. Globalement les participants semblent trouver de l'intérêt au fait de visualiser leurs données, surtout si ces informations leur permettent de se situer sur l'avancement de leur apprentissage par rapport (A) au contenu mis à disposition et par (B) comparaison avec l'avancement de leurs collègues :

[...] ça fait vraiment une vision globale de savoir comment on est en activité, si on est en retard par rapport aux autres, on voit aussi leurs résultats. Je trouve ça génial. On a toutes les informations.

Un intérêt de se « localiser » par rapport à ce qui a été fait et avec quels résultats est exprimé par des participants. Retrouver, par exemple, les résultats moins bons et savoir instantanément sur quelle partie du cours revenir afin d'améliorer ses résultats d'apprentissage est ainsi exprimé par un participant :

Je pense que ça pourrait être utile de, vu qu'on a plusieurs chapitres sur Libre Cours, qu'on voit les chapitres qu'on a les moins bonnes notes aux quiz. Comme ça je peux, par exemple, quand je devrai aller réviser plus tard, au moins on voit qu'on est plus faible sur tel chapitre que sur un autre.

Concernant l'utilisation d'un algorithme jugeant de la complexité du contenu, il n'y a pas eu de réfutation totale, mais la plupart ont préféré d'autres moyens de la juger. Certains des participants ont trouvé que pouvoir marquer soi-même des parties complexes pour y revenir plus tard était plus intéressant, d'autres ont proposé que cela soit jugé par l'instructeur lui-même. D'autres encore ont soutenu que l'utilisation d'un algorithme pourrait avoir un effet inverse à ce que l'on pourrait envisager :

Parce que là ça me semble un peu douteux, dans le sens où parfois il y a des exercices, des notions qui sont surtout très basiques, et on tombe forcément dans le piège, donc on a peut-être tous faux. Après avec l'explication d'après on voit que c'était pas vraiment compliqué. Et donc, si c'était un algorithme qui juge de la complexité, si tout le monde a faux... C'est pas forcément bien, ouais.

À qui un autre participant répond :

Je suis plus d'avis que le professeur, quand il crée son devoir, il met à côté, par exemple, un nombre d'étoiles, simple ou difficulté, simplement. Les professeurs ils ont vécu du coup plusieurs années, ils savent sur quels points en fait... C'est ça, oui. Où les élèves se trompent souvent.

En effet, le fait de chercher d'abord des situations où l'algorithme aurait « tort » pourrait être signe d'une certaine méfiance de la part des participants envers une automatisation de certaines fonctionnalités. Nous estimons qu'une attention particulière est importante dans ce sens par rapport à la mise en place d'indicateurs notamment avec une analyse prédictive.

Cette envie de pouvoir rajouter des informations aux documents — comme pour estimer de la complexité d'un contenu —, presque comme une annotation manuelle de ces derniers, apparaît dans tous les groupes et constitue une des dimensions qui est ressortie de l'analyse, cf. ci-après.

8.6.2 Dimension 2 — Gérer différents supports de cours... et les traces

Avoir plusieurs supports de cours, comme un PDF, un site, un jeu sérieux, une application dédiée pour apprendre à coder... peut représenter un avantage lorsqu'on exploite chacun de ces supports conformément à leurs buts, mais aussi un point d'attention, selon les propos des participants à notre étude :

[...] parfois ce qu'on fait en TD, on fait sur un outil et sur papier, enfin je préférerais avoir aussi la correction sur le site [inaudible] au même endroit, puis on travaille sur un cahier.

Ou encore :

Après, on peut passer le temps sur le cours, ça dépend, je ne trouve pas ça très intéressant si on passe du temps sur le PDF. Enfin, du coup, on passe moins de temps, en tout cas je sais que moi, personnellement, en général je passe plus de temps sur le PDF du cours que, sur le module en soi.

Ces différents usages ne sont pas tous centralisés et, surtout, ne sont pas tous tracés :

En fait, je suis mitigée, car si on imprime le cours... bah on va avoir tout à zéro ou très faible, parce qu'on a vu sur papier, du coup je ne sais pas trop. C'est bien, mais celui qui imprime tout, qui se connectera une fois...

À noter que dans le discours ci-dessus le fait de devoir gérer plusieurs supports de cours n'est pas, apparemment, un problème en soi.

Mais le devient si nous nous intéressons à des statistiques d'usage, par exemple. En effet, il n'était pas rare que des remarques homologues soient faites en se demandant après ce que l'instructeur va penser de ces résultats. Ceci peut laisser entrevoir une certaine méfiance par rapport à qui a accès aux résultats des analyses et ce qu'il va en faire. D'autant plus que dans ce contexte, l'instructeur fournit lui-même les différents supports et sait qu'un usage hors de la plateforme peut exister.

D'autres fois, cette presque nécessité d'avoir de bons résultats pourrait amener certains à un changement de comportement qui ne vise pas un meilleur résultat en termes d'apprentissage, mais pour créer un joli « graphique » :

Parce que je m'y verrais bien, me connecter juste pour avoir une jolie courbe, quoi.

8.6.3 Dimension 3 — Des indications « manuelles »

Comme souligné dans les dimensions précédentes, dans un contexte d'analyse automatique de traces d'apprentissage, les données analysées ne reflètent pas forcément la réalité des faits. C'est le cas, par exemple, quand le chercheur souhaite établir le temps passé à étudier une partie d'un cours à partir des traces laissées dans une plateforme en ligne. Si l'apprenant a ouvert la page en question et est parti ensuite prendre un café, le temps calculé serait plus long que la réalité. Ou encore, si un étudiant a choisi d'imprimer le support du cours et de l'utiliser pour étudier quand il n'est pas connecté, le temps calculé sera plus court que la réalité. De fait, ces écarts peuvent être plus ou moins importants selon la configuration présence-distance de la formation.

Une alternative à cette situation est de proposer aux utilisateurs la possibilité de marquer le contenu selon leurs besoins et de manière non automatique :

On a plein de cours sur le site, donc déjà pouvoir me noter sur chaque cours pour savoir où je suis. Ou plutôt que des notes, un code couleur. Comme ça on pourra rapidement voir les points qu'on doit relire et ce qu'on maîtrise déjà.

Mais ces indications faites manuellement sur les parties du contenu ne se limitent pas à cette fonction de savoir si l'on a déjà couvert une partie. La possibilité d'annoter, directement sur le contenu, mais de manière simple, est citée avec fréquence, avec des intentions similaires,

mais tout à fait personnalisées selon les intentions. Un participant indique l'utilisation qu'il ferait de ces annotations (une indication en cochant une case relative à une partie qui devient orange) :

[...] si c'est une chose que l'on a faite mais qu'on a encore des difficultés, alors l'avoir en orange, pour après aller voir le prof et lui dire : j'ai vu ce chapitre mais il y a cette partie-là que... que je ne sais pas.

Un autre participant élargit l'intérêt que cette fonctionnalité pourrait avoir, si de plus elle est partagée (visible par d'autres apprenants de la même promotion) :

Par exemple, en fait, pour chaque partie, on peut noter si on a bien compris ou pas et voir en fait combien de personnes n'ont pas compris. Si on est beaucoup, on peut peut-être poser la question en cours. Et si on est seul, on peut poser la question à un collègue.

Ces résultats sont concordants avec ceux trouvés par [124] dans leurs études sur les attentes des apprenants envers les fonctionnalités des systèmes d'analytique des apprentissages avec le numérique. Leurs résultats incluent le fait que les fonctionnalités prennent en charge leur planification et leur organisation des processus d'apprentissage. Bien que ce type de fonctionnalité ne soit pas strictement lié à notre sujet, ces résultats pourraient être explorés compte tenu du fait de son émergence spontanée et fréquente parmi les groupes. Un tel type de fonctionnalité pourrait servir à minimiser l'écart entre les traces recueillies par la plateforme en ligne et la réalité des faits. Mais aussi, grâce à la sémantique des annotations, cette forme d'instrumentation des utilisateurs serait potentiellement également source de traces qui pourraient être explorées par la suite afin de comprendre l'utilisation des contenus.

8.6.4 Dimension 4 — Naviguer parmi les cours

Une quatrième dimension qui a émergé est celle liée à la navigation au sein de la plateforme. Nous abordons cette dimension brièvement, car notre démarche s'intéresse surtout aux indicateurs au niveau des cours.

Les participants ont fait la remarque que la plateforme, à ce jour, n'incite pas à aller voir d'autres cours (ils sont pour la majorité librement accessibles) qui pourraient avoir un intérêt par rapport aux cours déjà suivis :

Je trouve que quand on va dessus, c'est... peut-être quelque chose de positif ou négatif, mais finalement on va juste sur le cours pour lequel on venait, on

va jamais vers les autres chapitres, on fait ce qu'on a à faire, parce qu'on sait, mais ensuite on ne va pas aller plus loin, ça pousse pas de l'approfondissement, ou aller voir même d'autres cours [...]

Également, l'accès à la liste des cours (à noter que le vocabulaire pour eux est souvent « séquence ») s'avère difficile pour certains :

Au début, sûrement, par exemple, quand on rentre dans une séquence — moi, ça m'est déjà arrivé — où je ferme la partie d'avant, justement celle qui donne accès à toutes les séquences et pour arriver à y accéder, moi j'ai mis un peu de temps [...]

Ce fait est expliqué par un autre participant :

Je crois, avec ce que tu viens de dire, il y a un problème, l'arborescence en gros. Ça veut dire que chaque fois qu'on clique sur une séquence, au lieu de pouvoir retourner directement sur la page avec les séquences et de revenir dans la séquence où on est, on peut uniquement revenir à la page de tous les modules.

8.6.5 Dimension 5 — Sémantique et structure

Le fait d'avoir certains indicateurs qui se basent sur la sémantique ou bien sur la structure des documents a été l'une des dimensions qui sont apparues lors des discussions. En ce qui concerne la sémantique, certaines des parties ainsi balisées semblent avoir plus d'intérêt que d'autres par rapport au besoin de consulter l'usage que l'on en a fait, mais aussi afin de pouvoir y revenir selon les différents moments d'un cours, comme pour des révisions. Quand, interrogé sur le fait d'avoir un graphique montrant les accès par type de contenu, un participant affirme :

[...] pour les définitions, avoir un endroit en fait qui regroupe les définitions au sein d'un même chapitre, comme ça on accède à tout. Et là on a toutes les définitions, les mots, qu'il fallait savoir...

Toujours dans cette envie d'avoir plus d'options pour la navigation des parties que nous avons remarquée dans la première dimension, il est suggéré qu'à travers ce graphique il soit possible d'accéder à une sorte de résumé des parties avec ce même type de contenu :

Apprenant — En fait, parce que sur nos cours on veut pas juste regarder juste les définitions, on est obligé de voir tout le cours... ou sinon on peut juste aller voir les différents encadrés. Après je pense que si on peut faire la

*distinction, c'est peut-être un peu plus utile, par exemple de pouvoir choisir, de voir juste les "définitions" et de voir justement les exemples...
Facilitateur — Si on peut, par exemple, revenir vers un graphique de ce type, cliquer sur la colonne "méthode" et avoir une page avec toutes les méthodes ?
Apprenant — C'est ça, avoir juste une page avec les "méthodes", parce que quand on a envie de réviser et qu'on veut juste revenir vers les définitions pour faire de fiches, alors là, ça serait génial.*

Ce fait est corroboré dans un autre groupe où la même idée/besoin revient de manière spontanée, avec un accent sur l'importance de certaines parties par rapport à d'autres aux différents moments de consultation du contenu :

Je pense pas que ce serait intéressant, surtout parce que "remarques", c'est souvent des trucs... enfin, c'est utile, mais ce n'est pas du tout indispensable pour l'apprentissage par rapport par exemple à "définition" [...] Il va falloir que je rouvre la définition. Par contre, si j'ai raté quelques "remarques" je ne suis pas sûre que j'y vais.

Ou encore par cet autre verbatim :

Ce qui me paraît très intéressant sur ces statistiques-là, c'est par exemple, la partie "notions fondamentales", parce que souvent, juste avec les titres par exemple sur une séquence on ne sait pas, par exemple, si on recherche vraiment quelque chose, on n'est pas sûr que c'est dans cette séquence-là, avoir la partie avec "notions fondamentales" on a juste besoin de regarder, on sait directement que ça c'est quelque chose...

Certains encore se demandent si certaines des données/graphiques ne seraient pas utiles surtout à l'instructeur, par exemple, à partir desquels il pourrait prévoir des révisions :

Par exemple, moi quand je me vois utiliser, quand il y a des exercices avec certaines parties où je reviens tout le temps, parce que je n'ai jamais retenu la "syntaxe" et tout, ça peut aussi voir, le prof peut aussi voir et se dire, bah, ils ont toujours besoin du cours sur la main et à l'examen, du coup... [rires] Ça peut être aussi l'idée, de se dire, s'il faut peut-être faire un cours spécialisé sur la "syntaxe", ça va être dur mais on va le faire, quoi.

Quant aux graphiques basés sur la structure, après avoir fait remarquer la différence et demandé explicitement si les participants trouveraient intéressant de les avoir lors de leurs cours, ils répondent de manière affirmative à l'unanimité.

Concernant la granularité de ce qui est tracé/présenté par contre il n'y a pas de consensus. Par exemple, concernant le fait d'avoir le

détail du nombre de visualisations d'une page ou encore le nombre de consultations des sous-parties de la même page, un participant affirme :

Mais si tu as envie de savoir... de vouloir tout refaire juste avant les examens et de pouvoir savoir où tu en es, au lieu de fouiller partout, parfois c'est [inaudible]

Nous notons également que le contexte de cette expérimentation a pu avoir une influence sur cette dimension et notre appréhension des besoins dans ce sens. Cela est dû au fait que la majorité du contenu à disposition des participants est conçu pour tenir sur la taille de l'écran, sans besoin de le faire défiler ou peu.

Et pourquoi pas, alors j'ai du mal à imaginer une page si grande qu'on met du temps à lire. Mais supposons qu'elles soient grandes, pourquoi pas un système qui indique où on en est resté la dernière fois ? Pour que la prochaine fois on y retourne.

Ce critère reste donc ouvert, mais il est fort probable que le niveau de granularité par rapport à la structure des documents soit en effet dépendant de chaque contexte et des modèles de documents exploités.

Somme toute, la compréhension des graphiques spécialement conçus afin de refléter cet enrichissement via la connaissance de la sémantique ou bien de la structure des documents ne semble pas poser de souci. Comme nous le remarquons dans l'exemple qui suit, même si certains n'estiment pas avoir compris un graphique, ce n'est pas le cas, car ils l'expliquent sans commettre d'erreur :

Apprenant — Je crois que je n'arrive pas à lire le graphique correctement. Du coup, est-ce que le point qui est le plus en haut, le plus à droite, ça veut dire que c'est une personne qui a fait énormément d'activités et globalement a des très bons résultats ?
Facilitateur — C'est exactement ça.

8.6.6 Dimension 5 — Au menu

Les indicateurs pouvant être explorés selon la connaissance de la structure du document ont été présentés selon deux possibilités : l'une dans un tableau de bord, avec une représentation visuelle qui reprenait cette structure (voir dans les annexes la Figure B.2) et l'autre qui montrait les mêmes résultats à côté du menu des pages du cours (voir dans les annexes la Figure B.14). Une préférence majoritaire a été exprimée dans le sens d'avoir les indicateurs selon cette seconde option, comme illustré par ces deux verbatims :

Je pense que si s'était une page à part comme ça, je n'y irais pas. Alors, si c'était à gauche, comme ça oui.

Dans les deux cas, il y a un clic à faire pour aller voir ça et ça sera moins intuitif que si c'était directement sur les titres. En fait, c'est, tous les autres graphiques qu'on a vus avant, on pourrait aussi les mettre d'une autre façon, mais directement sur le menu et ne pas les avoir sur une feuille en plus.

Cette option a été parfois exprimée même avant connaissance de l'option qui serait présentée, ce qui montre sa force et préférence de manière presque intuitive.

8.6.7 Dimension 6 — Comparer ou ne pas comparer ?

Le thème le plus récurrent qui est apparu dans les trois groupes et discuté parfois longuement, avec des points de vue distincts est celui relatif à la comparaison. Le thème n'a pas été abordé de manière explicite, mais plutôt implicitement via des graphiques qui affichaient, par exemple, le résultat d'un apprenant avec une comparaison à la moyenne obtenue par la classe (par exemple, cf. Annexe B, Figures B.7, B.8 et B.13).

Certains des participants ont montré un intérêt pour les comparaisons avec les pairs affirmant que ce type de données leur permettrait d'évaluer les besoins en termes de révision du contenu et aux performances de leurs collègues, pour ne pas se rendre compte trop tard d'un possible retard :

[...] ça permet de voir si, les autres, ils vont peut-être mieux que moi et donc peut-être qu'il faut revoir... travailler.

Il vaut mieux savoir, plutôt qu'après se planter aux mesures.

Pour certains, c'est exactement cette comparaison aux autres qui permettrait d'identifier des choses qui ne le seraient pas par une autre façon :

Moi, j'aime bien le fait que, s'il y a une activité qu'on a faite une fois ou zéro fois, alors que d'autres ont fait plusieurs fois on va se rendre compte qu'on a potentiellement [inaudible] dessus, alors que peut-être on n'aurait pas pensé, on aurait dit qu'on a autant révisé que les autres. Ça peut nous aider à identifier des points faibles, je pense que c'est utile.

Enfin, se comparer aux autres serait également une source de motivation, comme le mentionne ce participant :

Je pense que ça peut être presque motivant en plus, parce qu'on voit qu'il y a une corrélation entre le nombre d'activités faites et le résultat. C'est vrai que ça peut pousser à travailler.

Par contre, d'un autre côté, pour certains cette comparaison n'est pas utile :

Je trouve que c'est plus personnel, la promotion ce n'est pas important.

Certains soulignent même un côté assez négatif dans le fait de se comparer, que ce soit tout simplement sur des résultats obtenus, ou encore sur la relation entre ces résultats et l'effort entrepris pour l'avoir — une certaine sensation qu'un bon résultat n'est pas suffisant, qu'il fallait le faire plus vite, ou avec moins de temps dédié à la pratique :

[...] comparer toujours avec la promotion quand on fait des exercices, je pense que c'est déprimant.

Je trouve que c'est un graphe rigolo, mais après ça peut désespérer si on trouve que des gens qui ont fait très peu d'activités et qui ont de très bonnes notes, alors que nous on fait beaucoup d'activités et on va pas très bien, ça doit [inaudible] de voir ça.

D'autres semblent ne pas être sûrs, ou trouver que cela dépend du moment pédagogique sur lequel on se base pour effectuer une comparaison, comme s'il s'agissait d'une évaluation sommative ou formative :

Apprenant — Moi, je pense que c'est intéressant, parce qu'on cherche toujours à savoir si on est sur la moyenne ou non.

Facilitateur — D'accord, donc vous trouvez ça intéressant. D'autres ?

Apprenant — Moi, je trouve que c'est dommage d'obtenir tous les résultats parce que, justement, quand on apprend on se permet de faire des erreurs, on se teste, alors pour un examen on sait qu'on doit être prêt. Donc là on peut se comparer aux autres dans la période où l'on apprend et, enfin, je ne me sentirais pas à l'aise de devoir me comparer comment les autres ont appris, je n'aimerais pas en fait en avoir. Mais par contre [inaudible] j'aimerais avoir, mais je pense que c'est pas bénéfique au final, mais du coup...

D'autres trouvent qu'il s'agit tout simplement d'une préférence personnelle : le fait de se comparer à autrui pourrait avoir des conséquences positives ou négatives. C'est ce que témoignent les verbatims suivants :

Je pense que ça dépend vraiment des étudiants. Parce que je sais qu'il y en a que quand ils voient une courbe comme ça, ils vont se sentir plus motivés à travailler et d'un autre côté, je connais aussi d'autres personnes qu'une courbe comme ça, ça va vraiment énormément les stresser. Comme ça, ça dépend vraiment, vraiment de la personne.

En fait, c'est à double tranchant, ça peut motiver ou ça peut démotiver. Ça dépend des gens.

L'aspect de la comparaison qui semble en gêner d'autres semble plutôt lié au fait de se comparer à d'autres apprenants ; la comparaison avec des dates butoirs ou avec des objectifs que l'on a définis soi-même n'est pas négatif voire, au contraire, peut aider :

En revanche, sur l'idée des activités finalisées, je trouve pas intéressant de comparer par rapport aux autres, encore une fois, sachant que pour [titre du cours], chaque cours correspond à une date donnée, et donc, me dire que je suis en retard pour tel cours, pourquoi pas ? Parce que je suis en retard par rapport à des dates absolues et pas par rapport aux autres, pourquoi pas ? Sachant que celui qui ne fait pas les devoirs, parce que lui, par exemple, il a déjà acquis ces notions-là, ces notions par le passé, ça ne le culpabilisera pas.

Enfin, il a été exprimé plusieurs fois que les comparaisons n'étaient pas dans l'esprit de ce contexte spécifique. Les participants ont souligné que l'ambiance parmi eux — dont ils attribuent à une culture établie plus largement dans l'établissement — était surtout d'entraide et non de comparaison ou compétition, que ce soit une comparaison de l'achèvement d'activité ou des performances (notes obtenues) :

En fait, moi, tout ce qui est de comparaison, moi, j'ai un peu un problème, parce que ça me semble vraiment pas dans la modalité de l'[établissement]. Ça, c'est par exemple quelque chose que je verrais en prépa. C'est ça, les gens, tout le monde essaie d'être le meilleur, alors qu'à l'[établissement], on essaie vraiment tous de... je sais pas, mais il y a pas ce, vraiment, cette compétitivité...

Non seulement le contexte semble ne pas soutenir les comparaisons, mais pour certains, le fait que chacun ait un parcours différent (avant/durant) pourrait également biaiser ces comparaisons :

Pour rebondir sur l'idée que cela s'oppose à la mentalité de l'[établissement], il y a aussi, là ça aurait marché si tout le monde avait le même parcours. Mais le fait que tout le monde a un parcours différent, on va peut-être pas prioriser [titre d'un cours] en fonction de quelqu'un qui a des semestres très faciles, il va peut-être passer énormément de temps sur [titre d'un cours] et

va avoir une super bonne note. Quelqu'un qui suit [titre d'un cours] en même temps que [titre d'un cours], ou plutôt une [titre d'un cours], il va se contenter de faire juste le minimum, hein ?

En effet, les comparaisons — notamment envers ses collègues — s'avèrent une fonctionnalité polémique à proposer. Schumacher et al [124] soulignent avoir également rencontré un désaccord lors de leur étude : certains apprenants perçoivent de telles comparaisons comme motivantes, d'autres non. Il a été suggéré d'avoir la possibilité de n'afficher ce type d'informations que sur demande. C'est une option que nous considérons, en proposant par exemple deux indicateurs, avec ou sans comparaisons, avec le choix à faire entre ces options soit par l'instructeur, soit par les apprenants.

8.6.8 Dimension 7 — Rétroaction

Une autre dimension que nous avons pu observer est celle relative au besoin d'avoir de la rétroaction ou un espace de discussion intégré aux contenus des cours.

Une participante souligne que les explications montrées lorsque l'on répond correctement/incorrectement à un exercice sont importantes, notamment pour ceux qui ne suivent le cours qu'à distance :

Oui, les explications. Notamment moi, je ne suis que à distance, je ne vois le prof que [inaudible], il faudrait avoir les explications car sinon...

Pour ces mêmes apprenants, un espace de discussion — avec ou sans la participation de l'instructeur — intégré serait utile :

Ce qui est intéressant, en fait, pour ceux qui sont qu'à distance, qui fonctionnent avec LibreCours mais aussi avec Mattermost, je trouve ça intéressant qu'à la fin d'un module, on ait un espace pour qu'on puisse discuter...

Ce besoin de rétroaction semble plus important pour ces participants qui suivent un parcours ou une partie d'un parcours à distance et ces résultats se montrent cohérents avec ceux de [124, p. 70] *Interaction with fellow students but also direct contact to the lecturer via the system was essential for the students.*

8.6.9 Dimension 8 — Accès aux données

Le seul aspect lié à l'éthique, soulevé de manière spontanée par les participants, est celui de l'accès aux données. Pour une partie d'entre

eux, ces analyses ne seraient certainement pas partagées et disponibles seulement aux apprenants eux-mêmes. Surtout, il n'est pas question pour eux que l'instructeur puisse y accéder :

Tant que ça reste aussi strictement personnel, que le prof ne le voit pas.

Pour certains, imaginer que l'instructeur puisse savoir comment ils s'y prennent lors des sessions d'étude est particulièrement source de stress :

Je pense que c'est pas bon pour le moral. Non, je pense que ça stresse, à certains élèves qui peuvent se dire "si le prof, il regarde et qu'il voit que je me suis connecté très peu pendant la semaine".

Enfin, il semble important pour eux que ce soit clair de savoir qui peut avoir accès à chacune des visualisations disponibles et donc aux données :

Oui, non, non, c'est juste... Je suis d'accord, mais souvent quand on arrive au début, on a juste la partie avec toutes les statistiques, on ne sait pas forcément que c'est juste pour nous et, par exemple, le prof ne va pas forcément aller regarder dessus, et donc certains peuvent se sentir...

8.7 RÉSULTATS : SYNTHÈSE DES RETOURS

Nous proposons ensuite une synthèse des résultats de cette analyse, via les questions de recherche proposées en préambule.

8.7.1 (Q1) *Quels sont les besoins des utilisateurs en termes d'analytique des apprentissages avec le numérique ? Sont-ils suffisamment homogènes pour un même modèle documentaire ?*

L'analyse nous montre que les besoins concernant un seul contexte et un seul modèle documentaire peuvent varier considérablement, notamment en ce qui concerne les comparaisons, l'accès aux données et la rétroaction. Préalablement, dans le cadre de nos travaux, nous avons envisagé de mener l'étude et la proposition d'une série d'indicateurs « clés en main » spécifiques pour chaque modèle documentaire, par exemple, des indicateurs se concentrant sur les cheminements parcourus par les apprenants pour le modèle documentaire Topaze, qui n'est pas linéaire (parcours guidé par des activités conditionnelles), versus des indicateurs de complétude (temps passé, nombre de visualisations des parties, etc.) pour un modèle linéaire comme Opale. Cependant, ces résultats nous mènent à envisager de

permettre plutôt la création d'indicateurs selon un métamodèle, cela permettrait de créer et de les adapter facilement selon les contextes. Cela devrait permettre aux utilisateurs d'exprimer les besoins selon les contextes et préférences, puis aux modélisateurs de créer ou bien adapter facilement un indicateur — par exemple, ne le montrer qu'aux apprenants et non à l'instructeur, ou avoir un indicateur de résultat d'un test avec/sans comparaison — et, enfin, de concéder le choix aux utilisateurs de les activer ou pas.

8.7.2 (Q2) *Ces besoins (latents) incluent-ils des informations basées sur la sémantique des documents ?*

Les visualisations qui montraient des indicateurs se basant sur la sémantique ont été bien reçues et certaines des parties ainsi balisées semblent avoir plus d'intérêt que d'autres, notamment selon les différents moments de consultation de l'information (comme juste avant un examen). Quant aux graphiques basés sur la structure, les participants trouvent intéressant de les avoir lors de leurs cours à l'unanimité.

8.7.3 (Q3) *Les interprétations des utilisateurs pour chaque graphique affiché sont-elles correctes ?*

La compréhension des graphiques, y compris ceux spécialement conçus afin de refléter cet enrichissement via la connaissance de la sémantique ou bien de la structure des documents, semble ne pas poser de souci. Même si certains n'estiment pas avoir compris un graphique, ils l'expliquent sans commettre d'erreur par la suite.

8.7.4 (Q4) *Y a-t-il une différence d'interprétation par les utilisateurs entre les visualisations basées sur la sémantique et celles qui ne le sont pas ?*

Nous n'avons pas pu noter des différences qui pourraient être signe d'une difficulté particulière dans l'interprétation de ces graphiques. Cependant, une différence que nous pourrions citer par rapport à ces visualisations est qu'elles ont incité à des suggestions d'usages supplémentaires de la part des participants (par exemple, cliquer sur une partie du graphique pour avoir une page avec tous les contenus correspondants avec la même étiquette).

8.8 LIMITATIONS

Nous reconnaissons qu'il existe plusieurs limites qui ont une incidence sur les conclusions de cette étude.

D'abord, nous n'avons consulté des participants que d'un seul établissement dont les résultats ne sont pas destinés à être généralisés à d'autres établissements d'enseignement supérieur.

Deuxièmement, les résultats de cette étude représentent les expériences et les perspectives des apprenants à un moment donné. Nous reconnaissons que ces perceptions peuvent changer avec le temps et l'expérience avec ce type de système.

Malgré ces limites, des leçons peuvent être tirées de cette étude, qui sont mises en évidence dans les discussions et qui nous permettent de mieux prendre certaines décisions lors de la proposition du métamodèle.

CONNAÎTRE LES INDICATEURS D'APPRENTISSAGE POUR LES MODÉLISER

9.1 INTRODUCTION

Afin de connaître 1/ les indicateurs pouvant être modélisés en utilisant notre métamodèle et 2/ les différentes caractéristiques de ces indicateurs qui pourraient influencer les besoins en termes de modélisation, nous avons choisi de mener une revue systématique de la littérature. Une telle revue permet de résumer les informations existantes sur un phénomène de manière approfondie et impartiale; elle est en général menée dans le but de tirer des conclusions plus générales sur ce phénomène — lié à la question de recherche — mais également en prélude à d'autres activités de recherche [82], comme c'est notre cas, où nous nous appuyerons sur ces résultats afin de nous orienter lors de la métamodélisation.

Les auteurs de [82] résument une revue systématique de la littérature comme un processus de recherche permettant d'identifier, évaluer et interpréter les recherches disponibles pertinentes pour une question de recherche donnée. Il s'agit ainsi d'une forme d'étude dite secondaire, par opposition aux études individuelles — dites primaires —, sources d'une revue systématique.

Ce Chapitre nous permet notamment de répondre à la question de recherche **RQ5** Quels sont les indicateurs couramment utilisés dans les systèmes d'analytique des apprentissages avec le numérique et leurs différentes caractéristiques? Cette question sera ensuite déclinée d'abord en quatre dimensions et ensuite dans des sous-questions permettant de connaître ces indicateurs en détail et les classer.

9.2 QUESTIONS DE RECHERCHE

Bien que des revues systématiques aient été menées dans le champ de l'analytique des apprentissages avec le numérique, elles se sont concentrées sur les tableaux de bord proposés, c'est notamment le cas de [125] ou encore se sont concentrées sur une approche spécifique, comme l'autorégulation dans [95] et [145].

Cependant, contrairement aux études que nous venons de citer, nous nous interrogeons spécifiquement sur les indicateurs proposés

dans un contexte d'enseignement supérieur, mais aussi sur certaines de leurs caractéristiques et spécificités. Les auteurs de [91] affirment que le contexte de l'enseignement supérieur est particulièrement approprié pour réfléchir aux comportements des apprenants et apporter une aide adéquate de la part des enseignants ou des tuteurs ; mais également important et mis en pratique afin d'améliorer la coopération interinstitutionnelle. Ainsi, avec cette étude, nous avons examiné la littérature afin de répondre aux questions suivantes :

9.2.1 Dimension 1 — Indicateurs proposés

Quels sont les indicateurs listés par la littérature et leurs caractéristiques ? La description d'un indicateur passe par la réponse à la sous-question suivante :

- (Q1.1) Quel est le type d'analyse effectuée pour construire l'indicateur ?

9.2.2 Dimension 2 — Parties prenantes

Quelles sont les parties prenantes impliquées dans chacune des phases du cycle de vie de l'indicateur ? Le cycle de vie incluant une phase de conception, de test et d'utilisation, nous répondrons en particulier aux sous-questions suivantes pour chaque indicateur :

- (Q2.1) Qui a proposé ou décidé d'ajouter l'indicateur ?
- (Q2.2) Dans des cas où plus d'un acteur est impliqué dans la phase de conception, y a-t-il de la conception participative ?
- (Q2.3) Qui est la cible de l'indicateur proposé ?
- (Q2.4) Avec quelle partie prenante l'indicateur proposé est-il testé ?

9.2.3 Dimension 3 — Données, analyses et visualisations

Quel est le processus pour arriver au déploiement de l'indicateur ? D'où viennent les données et comment sont-elles présentées ? Voici les questions précises qui nous guident afin de le comprendre :

- (Q3.1) Quel artefact numérique est à l'origine des interactions qui ont été tracées ?
- (Q3.2) Le cas échéant, quels algorithmes sont utilisés pour calculer l'indicateur ?
- (Q3.3) Comment les résultats des analyses sont-ils visualisés ?
- (Q3.4) Quelle est la représentation graphique choisie pour l'indicateur ?

- (Q3.5) Quel est l’artefact numérique où ces visualisations sont proposées ?

9.2.4 Dimension 4 — Contexte

Enfin, nous nous intéressons à des caractéristiques des indicateurs qui pourraient représenter des particularités lors de leurs usages, comme le type de dispositif pédagogique en place ou encore son statut :

- (Q4.1) Quel est le statut (conçu, testé ou déployé) de cet indicateur ?
- (Q4.2) Comment a-t-il été évalué ?
- (Q4.3) Quels sont les domaines d’apprentissage dans lesquels l’indicateur a été proposé ?
- (Q4.4) Quel est le type de dispositif pédagogique en place pour lequel cet indicateur est proposé ?
- (Q4.5) Quelle est la théorie de l’apprentissage sur laquelle repose l’indicateur proposé ?

Nous fournissons un récapitulatif de ces questions de recherche dans la Table 9.1.

9.3 MÉTHODE

9.3.1 Requêtes et stratégie de recherche

Sur la base de notre objectif, nous avons créé une requête pour identifier les articles scientifiques qui ont proposé des indicateurs dans les domaines de l’analytique des apprentissages avec le numérique et le domaine voisin de la fouille de données éducatives, dans un contexte d’enseignement supérieur. Nous nous concentrons sur ce domaine afin de réduire le nombre de variables pouvant avoir un rôle dans les résultats, mais également étant donné que la plupart des usages des outils sur lesquels nous nous appuyons se font dans un contexte d’enseignement supérieur.

Notre requête de recherche a été conçue via la combinaison de mots-clés et l’utilisation des opérateurs booléens AND (ET) et OR (OU) sur un ensemble de mots-clés de requête pour la recherche d’articles. Les termes utilisés pour identifier les articles sur les domaines de l’analytique des apprentissages avec le numérique et de la fouille de données éducatives ont été : *learning analytics* OR *educational data mining* OR *educational datamining*. Les mots-clés visant à identifier des articles possédant un ou plusieurs indicateurs ont été : *indicator* OR *metric* OR *dashboard*.

TABLEAU 9.1 – Récapitulatif et détail des questions de recherche à l'étude.

Thème	Question	Type	Valeurs et exemples
Indicateurs proposés	(Q1.1) Quel est le type d'analyse effectuée pour construire l'indicateur ?	Catégorique	Descriptive, diagnostique, prédictive, prescriptive
Parties prenantes	(Q2.1) Qui a proposé ou décidé d'ajouter l'indicateur ?	Catégorique	Chercheurs, instructeurs, apprenants, administrateurs, NS
	(Q2.2) Dans des cas où plus d'un acteur est impliqué dans la phase de conception, y a-t-il de la conception participative ?	Catégorique	Oui, non, NS
	(Q2.3) Qui est la cible de l'indicateur proposé ?	Catégorique	Chercheurs, instructeurs, apprenants, administrateurs, NS
	(Q2.4) Avec quelle partie prenante l'indicateur proposé est-il testé ?	Catégorique	Chercheurs, instructeurs, apprenants, administrateurs, NS
Données, analyses et visualisations	(Q3.1) Quel artefact numérique est à l'origine des interactions qui ont été tracées ?	Exploratoire	LMS, outil spécialisé, Web, plateforme MOOC...
	(Q3.2) Le cas échéant, quels algorithmes sont utilisés pour calculer l'indicateur ?	Exploratoire	Oui, non, NS
	(Q3.3) Comment les résultats des analyses sont-ils visualisés ?	Exploratoire	Tableau de bord, diaporama, notification...
	(Q3.4) Quelle est la représentation graphique choisie pour l'indicateur ?	Exploratoire	Histogramme, table, tracé de ligne, carte thermique, camembert
	(Q3.5) Quel est l'artefact numérique où ces visualisations sont proposées ?	Exploratoire	LMS, outil spécialisé, Web, plateforme MOOC...
Contexte	(Q4.1) Quel est le statut de cet indicateur ?	Catégorique	Conçu, testé, déployé, NS
	(Q4.2) Comment a-t-il été évalué ?	Exploratoire	Questionnaire, entretien, méthodes mixtes...
	(Q4.3) Quels sont les domaines d'apprentissage dans lesquels l'indicateur a été proposé ?	Exploratoire	Sciences exactes, santé, sciences humaines...
	(Q4.4) Quel est le type de dispositif pédagogique en place pour lequel cet indicateur est proposé ?	Catégorique	Présentiel, hybride, distanciel, NS
	(Q4.5) Quelle est la théorie de l'apprentissage sur laquelle repose l'indicateur proposé ?	Exploratoire	Behaviorisme, cognitivisme, constructivisme, socio-constructivisme...

D'un côté, le terme métrique peut apparaître sans faire référence à la notion d'indicateur (par exemple si des auteurs mentionnent qu'ils utilisent une métrique comme le taux d'attrition pour évaluer l'impact d'une intervention, sans que l'indicateur associé ne soit rendu visible) mais a tout de même été inclus pour ne pas exclure trop vite des articles n'utilisant pas explicitement le mot « indicateur », qui n'est pas forcément standardisé.

De l'autre côté, le terme tableau de bord a également été inclus car, souvent, les auteurs n'identifient pas explicitement ce qu'ils affichent dans le tableau de bord comme étant des indicateurs, mais présentent néanmoins des tableaux de bord composés d'un ou plusieurs indicateurs.

Enfin, nous avons utilisé les termes suivants afin de restreindre des études portant sur le contexte de l'enseignement supérieur : *higher education* OR *academic* OR *academia* OR *college* OR *university*. Chacun de ces groupes de mots devait être présent au moins dans le champ titre, mots-clés ou résumé d'un article pour que celui-ci soit sélectionné ; l'expression logique 9.1 résume la stratégie de notre requête de recherche.

$$\begin{aligned}
 & \text{corpus} = \\
 & ("learning\ analytics" \vee "educational\ data\ mining" \vee "educational\ datamining") \\
 & \quad \wedge \\
 & ("indicator" \vee "metric" \vee "dashboard") \\
 & \quad \wedge \\
 & ("higher\ education" \vee "academic" \vee "academia" \vee "college" \vee "university")
 \end{aligned}
 \tag{9.1}$$

Les bases de données bibliographiques électroniques recherchées comprenaient celles qui sont indexées via : Association for Computing Machinery (ACM) Digital Library¹, Institute of Electrical and Electronics Engineers (IEEE)², SpringerLink³, ScienceDirect⁴, Wiley Online Library⁵, et Education Resources Information Center (ERIC)⁶. Nous avons également consulté les 100 premiers résultats de Google Scholar⁷ ainsi que les bases d'articles des revues scientifiques associées aux deux communautés de recherche les plus directement liées à notre

-
1. <https://dl.acm.org>
 2. <https://www.ieee.org/publications/index.html>
 3. <https://link.springer.com>
 4. <https://www.sciencedirect.com>
 5. <https://onlinelibrary.wiley.com>
 6. <https://eric.ed.gov>
 7. <https://scholar.google.com>

contexte d'étude : Journal of Learning Analytics (JLA)⁸, et Journal of Educational Data Mining (JEDM)⁹. Ces bases de données et revues ont été choisies en raison de leur large spectre de couverture de revues scientifiques évaluées par des pairs et leur pertinence au sujet de notre étude. À noter que certaines de ces bases de données permettent une requête de recherche limitant la recherche aux trois champs supracités (c'est-à-dire titre, mots-clés et résumé) tandis que d'autres ont nécessité un filtrage a posteriori (cf. 9.3.1, le nombre d'articles ici est ainsi à N = 556).

9.3.2 *Sélection des documents*

Une fois cette sélection initiale basée sur ces trois champs réalisée par un chercheur, nous avons procédé à une première exclusion d'articles remplissant au moins un des critères d'exclusion qui suivent via la lecture des respectifs titres, mots-clés et résumés :

1. Articles contenant 3 pages ou moins ;
2. Articles utilisant le mot « indicateur » avec un autre sens de celui de notre contexte (par exemple, « *The lack of information for stakeholders is an indicator of the difficulties for understanding the field* ») ;
3. Articles écrits dans une autre langue que l'anglais ;
4. Articles conceptuels ou reportant une revue systématique de la littérature.

Après cette étape, un total de 157 articles a été retenu. À noter que quelques articles ont été exclus à partir d'un des critères que nous venons de citer et qui n'ont pas été identifiés lors de la lecture de ces trois champs (9 cas, par exemple d'un article ayant un résumé en anglais, mais écrit dans une autre langue) et que pour 22 des articles nous n'avons pas pu accéder à l'intégralité du texte. Les N = 126 articles restants ont été lus dans leur intégralité et codés. La Figure 9.1 synthétise ces étapes.

9.3.3 *Processus de codage*

Nous avons commencé par définir ce qu'est un indicateur. Dyckhoff et al utilisent la définition suivante d'indicateur : « [...] *specific calculators with corresponding visualizations, tied to a specific question* » [48, p. 60]. Lors des premières lectures afin de définir le codage, nous avons rapidement

8. <https://learning-analytics.info/index.php/JLA>

9. <https://jedm.educationaldatamining.org/index.php/JEDM>

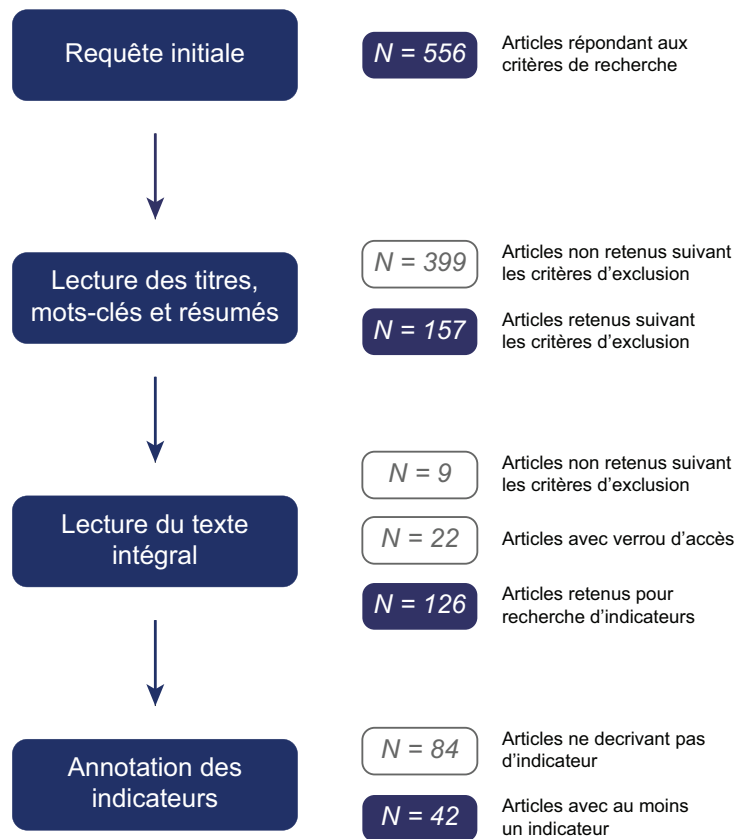


FIGURE 9.1 – Méthode utilisée dans cette revue systématique de la littérature.

remarqué le besoin d'une définition plus précise. La définition retenue nous permettant de réaliser le codage est la suivante :

DÉFINITION D'INDICATEUR. « Nous considérons qu'un indicateur d'apprentissage utilisé dans une démarche d'analytique d'apprentissage avec le numérique est une mesure (quantitative ou qualitative) calculée [propriété de calculabilité] liée à un comportement ou une activité instrumentée par le numérique [propriété de traçabilité] d'un ou plusieurs apprenants, donnée à voir à un usager [propriété de visibilité] et pouvant être utilisée dans le calcul d'autres indicateurs ».

Ensuite, chaque article a été codé afin de contrôler s'il contenait au moins un indicateur (ainsi que les caractéristiques de chaque indicateur trouvé, comme indiqué dans la Table 9.1). Avant et pendant les premières lectures, certains exemples ont été relevés et discutés entre les lecteurs, ce qui nous a permis de préciser cette définition avec quelques exemples concrets que nous indiquons ci-après.

Plus précisément, ne sont pas des indicateurs d'apprentissage à notre sens les éléments suivants :

- La page X a été vue ou non par l'apprenant Y [ne respectant pas la propriété de calculabilité].
- Le résultat global de l'apprenant (car non lié — directement — à une activité) [ne respectant pas la propriété de traçabilité].
- Le genre d'un apprenant [ne respectant pas la propriété de traçabilité ni de calculabilité].
- La motivation si elle est issue uniquement de données déclaratives, comme d'un questionnaire [ne respectant pas la propriété de traçabilité].
- Le nombre de pages vues par un apprenant dans un système utilisant cette donnée pour calculer un indicateur de risque de décrochage, si seul ce dernier est remonté à l'instructeur (car le nombre de pages vues n'est alors qu'une variable interne au système, non accessible par un usager) [ne respectant pas la propriété de visibilité].

Inversement, sont des indicateurs d'apprentissage à notre sens :

- La page X a été vue avant une date donnée [propriété de calculabilité].
- La page la plus consultée après la page X ou un graphe de pages visualisées [propriété de calculabilité].
- La motivation si elle est inférée de données comportementales (pages vues, temps passé), éventuellement renforcée par du déclaratif (questionnaire de motivation) [propriété de traçabilité].
- Un indicateur non rendu à un usager (par exemple, un instructeur ou un apprenant) dans le contexte de l'article en lecture, mais pour lequel les auteurs envisagent explicitement de le faire dans le futur, ce qui refléterait différentes publications selon les phases de la recherche en cours [propriété de visibilité].
- La note ou le résultat d'une activité d'apprentissage si elle est liée aux traces ou comparée à une autre donnée provenant des traces, par exemple le nombre de pages vues vs la note obtenue [propriété de traçabilité].

À noter que certains articles décrivaient des indicateurs issus d'une autre étude qui étaient seulement cités dans celui en lecture, par exemple, lors d'une revue de la littérature (ou *related work*). Ces indicateurs n'ont pas été inclus. D'un autre côté, un indicateur issu d'une autre étude et réutilisé (ou amélioré, retravaillé, etc.) dans l'article en lecture a été inclus.

La liste d'articles (cf. Table 9.2) retenus où au moins un indicateur — toujours selon la définition ci-dessus — a été identifié est la suivante :

TABLEAU 9.2 – Articles retenus où au moins un indicateur a été trouvé.

Début de la Table 9.2

Publication	Titre	Référence
2012	Learning Analytics for Collaborative Writing : A Prototype and Case Study	[100]
2012	Design and Implementation of a Learning Analytics Toolkit for Teachers	[48]
2013	Learning Analytics and Formative Assessment to Provide Immediate Detailed Feedback Using a Student Centered Mobile Dashboard	[4]
2013	A Software Suite for Efficient Use of the European Qualifications Framework in Online and Blended Courses	[55]
2013	Monitoring Learning Activities in PLE Using Semantic Modelling of Learner Behaviour	[135]
2014	Competency Map : Visualizing Student Learning to Promote Student Success	[66]
2015	Using Learning Analytics to Visualise Computer Science Teamwork	[136]
2016	Generating Actionable Predictive Models of Academic Performance	[115]
2016	Real-Time Indicators and Targeted Supports : Using Online Platform Data to Accelerate Student Learning	[93]
2017	Understanding Student Interactions in Capstone Courses to Improve Learning Experiences	[107]
2017	A Novel Web-Based Approach for Visualization and Inspection of Reading Difficulties on University Students	[101]
2017	A Demonstration of Evidence-Based Action Research Using Information Dashboard in Introductory Programming Education	[96]
2017	Using Sequential Pattern Mining to Explore Learners' Behaviors and Evaluate Their Correlation with Performance in Inquiry-Based Learning	[143]
2017	Short Answers to Deep Questions : Supporting Teachers in Large-Class Settings	[99]
2017	Learning Analytics as Assemblage : Criticality and Contingency in Online Education	[126]
2017	A Demonstration of Evidence-Based Action Research Using Information Dashboard in Introductory Programming Education	[96]
2017	Using Sequential Pattern Mining to Explore Learners' Behaviors and Evaluate Their Correlation with Performance in Inquiry-Based Learning	[143]

Publication	Titre	Référence
2017	Short Answers to Deep Questions : Supporting Teachers in Large-Class Settings	[99]
2017	Learning Analytics as Assemblage : Criticality and Contingency in Online Education	[126]
2018	Multi-Institutional Positioning Test Feedback Dashboard for Aspiring Students : Lessons Learnt from a Case Study in Flanders	[21]
2018	XIPIt : Updating the XIP Dashboard to Support Educators in Essay Marking at Higher Education	[16]
2018	Personalizing Computer Science Education by Leveraging Multimodal Learning Analytics	[11]
2018	Assessing Learning Analytics Systems Impact by Summative Measures	[70]
2018	Practicing the Scholarship of Teaching and Learning with Classroom Learning Analytics	[3]
2018	Raspberry Pi based learning center usage tracking system for optimal resource allocation	[106]
2018	Dimensional Morphing Interface for Dynamic Learning Evaluation	[57]
2018	Learning Analytics Dashboard Analysing First-Year Engineering Students	[142]
2018	Affordances and Limitations of Learning Analytics for Computer-Assisted Language Learning : a Case Study of the VITAL Project	[63]
2018	Monitoring Attendance and Its Impact on Engineering Students	[137]
2019	Student Centred Design of a Learning Analytics System	[119]
2019	It's My Data! Tensions Among Stakeholders of a Learning Analytics Dashboard	[134]
2019	Adaptive Learning Based on AI with Predictive Algorithms	[120]
2019	Investigating Learners' Perception of Learning Analytics Dashboard to Improve Learning Interaction in Online Learning System	[140]
2019	An Infographics-based Tool for Monitoring Dropout Risk on Distance Learning in Higher Education	[19]
2019	Utilizing Web Analytics in the Context of Learning Analytics for Large-Scale Online Learning	[121]
2019	IKARion : Enhancing a Learning Platform with Intelligent Feedback to Improve Team Collaboration and Interaction in Small Groups	[38]
2019	Smart Learning Analytics : Conceptual Modeling and Agile Engineering	[141]
2019	Gamification and Learning Analytics to Improve Engagement in University Courses	[24]

Publication	Titre	Référence
2019	Observing Learner Engagement on Mind Mapping Activities Using Learning Analytics	[23]
2019	An Integrated Framework for Course Adapted Student Learning Analytics Dashboard	[5]
2019	Empowering Online Teachers Through Predictive Learning Analytics	[74]
2019	A Handheld Classroom Dashboard : Teachers' Perspectives on the Use of Real-Time Collaborative Learning Analytics	[94]
2019	Teaching with Analytics : Towards a Situated Model of Instructional Decision-Making	[6]
2020	How Patterns of Students Dashboard Use Are Related to Their Achievement and Self-Regulatory Engagement	[123]
2020	A Customizable and Incremental Processing Approach for Learning Analytics	[118]
2020	Learning Analytics Dashboard for Teaching with Twitter	[68]

Fin de la Table 9.2

Parmi les 126 articles lus, $N = 42$ proposaient un ou plusieurs indicateurs au sens ci-dessus. Nous avons codé un total de $N = 208$ indicateurs, ce qui représente une moyenne de $\bar{m} = 4,95$ ($\sigma = 4,43$) indicateurs par article. La Figure 9.2 permet de visualiser la distribution du nombre d'indicateurs proposés par les articles et la Table 9.2 regroupe les publications où au moins un indicateur a été trouvé.

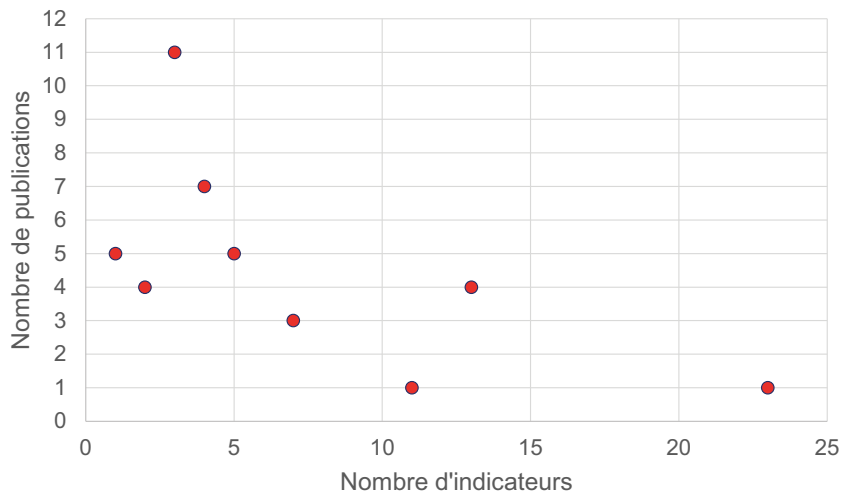


FIGURE 9.2 – Distribution du nombre d'indicateurs par article analysé.

En analysant le nombre d'articles où au moins un indicateur était présent, nous remarquons une progression depuis 2014, comme le

montre la Figure 9.3. À noter que les publications ont été collectées en mai 2020, donc l'année 2020 n'est pas une année complète.

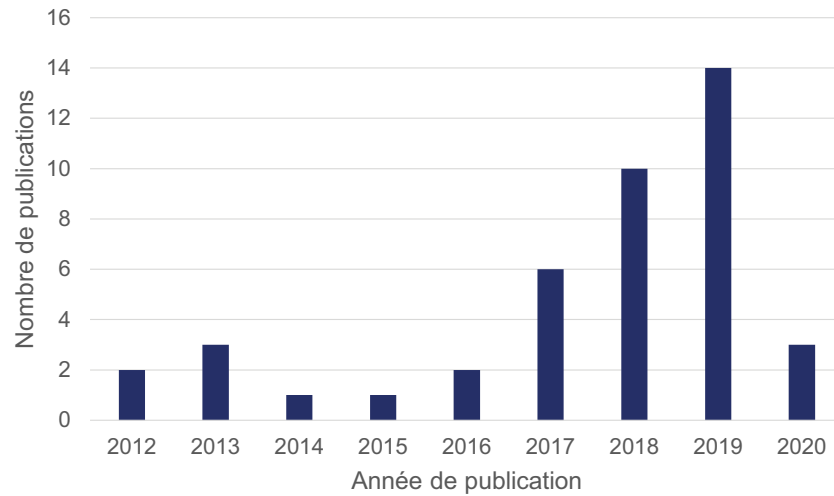


FIGURE 9.3 – Nombre d'articles contenant au moins un indicateur selon l'année de publication.

Dans la prochaine section, nous analysons ces résultats selon les dimensions et questions de recherche énoncées ci-dessus.

9.4 RÉSULTATS : DIMENSION 1 — INDICATEURS PROPOSÉS

La première dimension que nous avons étudiée se concentre sur les indicateurs et les types d'analyses conduites afin d'arriver aux résultats et sur les théories de l'apprentissage sur lesquelles repose chaque proposition.

9.4.1 (Q1.1) *Quel est le type d'analyse effectuée pour construire un indicateur ?*

Le premier aspect que nous avons étudié est le type d'analyse effectuée qui aboutit à la visualisation d'un indicateur. Nous nous sommes basés sur la classification de [59] pour le codage, basé sur la prise de décision et les actions mises en œuvre de la part des parties prenantes suite à l'analyse (en contraste avec une décision automatisée, voir la section 5.3). Ce schéma de codage distingue 4 niveaux croissants de types d'analyse :

Descriptive Qu'est-il arrivé ?

Diagnostique Pourquoi est-ce arrivé ?

Prédictive Que va-t-il arriver ?

Prescriptive Que devrait-on faire ?

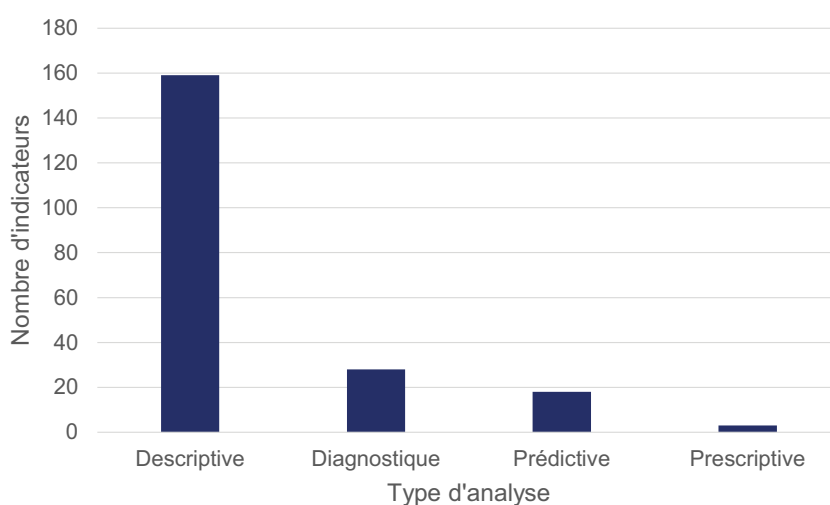


FIGURE 9.4 – Les types d'analyse de chaque indicateur.

Les résultats nous montrent une forte prédominance d'indicateurs portant sur une analyse descriptive de ce qui s'est passé, soit 159 indicateurs avec un total de $N = 208$ recensés (76,44 %). Ce type d'indicateur est suivi par ceux proposant une analyse diagnostique ($N = 28$; 13,46 %) et une analyse prédictive ($N = 18$; 8,65 %). Très peu d'indicateurs proposaient une analyse prescriptive ($N = 3$; 1,44 %). Les analyses diagnostiques ont été proposées dans 12 articles et celles prédictives dans six articles. Les analyses prescriptives n'ont été présentées que dans deux des articles analysés.

Si nous cherchons à repérer une éventuelle évolution temporelle des indicateurs proposés en nous concentrant sur l'année des publications contenant ces analyses (cf. Figure 9.5), nous remarquons une tendance croissante des indicateurs à vocation non descriptive, passant des analyses nécessitant une intervention humaine avant une décision ou action, aux analyses où une aide à la décision est fournie, voire une automatisation de la décision est proposée.

Cette tendance pourrait être liée à plusieurs facteurs. Une possibilité, à laquelle nous avons essayé de répondre sur la base des données recueillies, serait que des publications de mêmes auteurs progressent au fil du temps d'une proposition d'indicateurs uniquement de type descriptif vers des indicateurs de type prescriptif. Ainsi, ces derniers verraient le jour à la fin d'une série de publications provenant de mêmes auteurs. Afin de vérifier cette possibilité, nous avons analysé, pour les analyses classées comme prédictives ou prescriptives, si les auteurs correspondants avaient publié préalablement sur le même outil (proposant alors des analyses plus « simples »). Nous n'avons

cependant pas pu trouver de cas correspondant à cette situation sur la base d'articles retenus ayant au moins un indicateur.

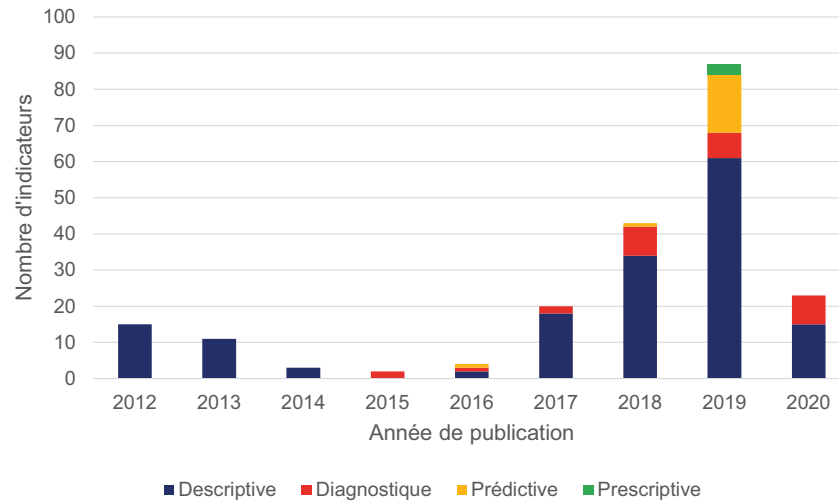


FIGURE 9.5 – Les types d'analyse de chaque indicateur selon l'année de publication.

9.4.2 Discussion

Nous remarquons une forte prédominance d'indicateurs se basant sur une analyse qui décrit ce qui s'est passé. Néanmoins, une tendance a été observée vers des analyses proposant automatiquement un diagnostic, réalisant une prédiction ou encore présentant une prescription. À partir de ces résultats nous remarquons 1/ une tendance croissante au non-descriptif, qui peut s'expliquer par (h1) une progression naturelle liée à l'évolution de la recherche, c'est-à-dire un reflet du moment de la recherche qui prendrait un certain temps avant d'aboutir à une implémentation sur le terrain et 2/ nous pouvons inversement nous étonner que la part du descriptif reste tout de même assez forte, en dépit d'encouragements à aller vers des analyses non descriptives. Ce phénomène pourrait s'expliquer par (h2) une certaine crainte d'une difficulté de compréhension du résultat de ces analyses plus automatisées ou par (h3) une réticence des parties prenantes présentes qui pourrait bloquer ces propositions. Cette étude ne permet pas de trancher parmi ces possibilités, mais soulève des questions de recherche intéressantes pouvant être poursuivies *a posteriori*. Cependant, c'est dans le sens de (h3) que les travaux récents d'Alasalmi semblent mener où les participants ont trouvé que l'analyse descriptive était la plus importante et même fondamentale. De plus, les caractéristiques

des analyses prescriptives et prédictives ont été « accueillies avec une attitude plus douteuse » [2, p. 140, notre traduction].

Nous retenons, pour des fins de métamodélisation, que les primitives proposées devraient permettre de modéliser ces quatre types d'indicateurs. Ceci implique de permettre l'inclusion d'algorithmes « plus puissants » (*machine learning*) ou encore juste des résultats d'un modèle (*data model* résultant d'une analyse faite, sous la forme d'un arbre de décision, par exemple) qui a été entraîné sur un jeu de données, typiquement pour faire des prévisions.

9.5 RÉSULTATS : DIMENSION 2 — PARTIES PRENANTES

La deuxième dimension que nous avons étudiée dans la revue est celle liée aux différentes parties prenantes impliquées et aux différents rôles de ces derniers dans chacune des phases du cycle de vie de l'indicateur.

9.5.1 (Q2.1) Qui a proposé ou décidé d'ajouter l'indicateur ?

Pour chaque indicateur recensé lors de la lecture du corpus retenu, nous avons tenté d'identifier qui était à l'origine du choix de ces indicateurs. Nous avons codé chaque indicateur parmi les choix suivants :

- Chercheurs : les chercheurs à l'origine de l'article.
- Instructeurs : les instructeurs (enseignants, enseignants-chercheurs, etc.) des cours ou formations pour lesquels les indicateurs ont été proposés.
- Apprenants : les apprenants participant aux cours ou formations, très souvent à l'origine de la production des données analysées (ce n'est pas le cas pour certaines des analyses prédictives, qui utilisent souvent des données des promotions passées).
- Administrateurs : tout personnel participant indirectement à ces cours, par exemple des ingénieurs ou conseillers pédagogiques, des conseillers étudiants *student advisors*, des administrateurs de plateformes pédagogiques, etc.

Parfois, un même indicateur a été proposé par plus d'une partie prenante, donc un codage multiple a été fait pour ces cas (par exemple, « chercheur » et « instructeur »).

Les résultats nous montrent que, dans la majorité des cas, ce sont les chercheurs qui sont à l'origine de la publication qui proposent eux-mêmes les indicateurs en question, c'est le cas pour 117 indicateurs. Parfois cela se fait en accord avec d'autres parties prenantes : avec des

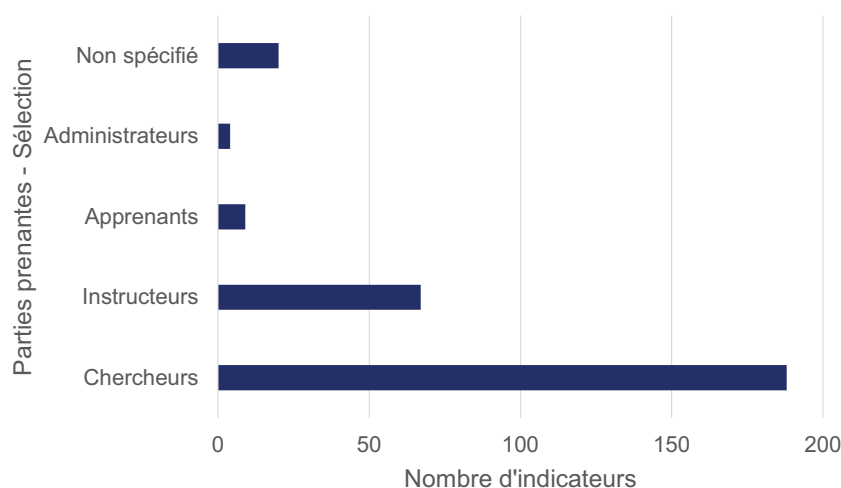


FIGURE 9.6 – Les parties prenantes participant à la sélection d'un indicateur (avec codage multiple possible).

administrateurs (N = 4 indicateurs), instructeurs (N = 58 indicateurs), instructeurs et apprenants (N = 9 indicateurs). Pour un total de 20 indicateurs, nous n'avons pas pu identifier la partie prenante à l'origine de leur proposition. La Figure 9.6 montre le total de fois où un des acteurs est impliqué dans la proposition d'un indicateur.

Une analyse des parties prenantes qui sont à l'origine de la proposition des indicateurs selon l'année de publication (voir Figure 9.7) nous montre une tendance, depuis 2018, à impliquer davantage d'autres acteurs plutôt que les chercheurs. C'est le cas notamment de l'implication d'instructeurs. En 2012, 13 indicateurs ont ainsi été définis par des chercheurs et des instructeurs, mais appartenant à un seul article [48]. En 2018, deux articles ([16] et [63]) ont proposé 13 indicateurs de cette manière et en 2019, c'est le cas de quatre articles ([119], [5], [94] et [6]). Enfin, la tendance semble se poursuivre en 2020 — avec des données partielles — avec deux articles ([118] et [68]) proposant un total de 20 indicateurs où les instructeurs s'associent aux chercheurs dans la phase de conception.

Les administrateurs n'ont été impliqués aux côtés des chercheurs dans la conception des indicateurs que dans un des travaux étudiés, en 2018 ([21], quatre indicateurs). Les apprenants, quant à eux, n'ont participé à la conception qu'une seule fois ([119], de 2019), avec des instructeurs et des chercheurs.

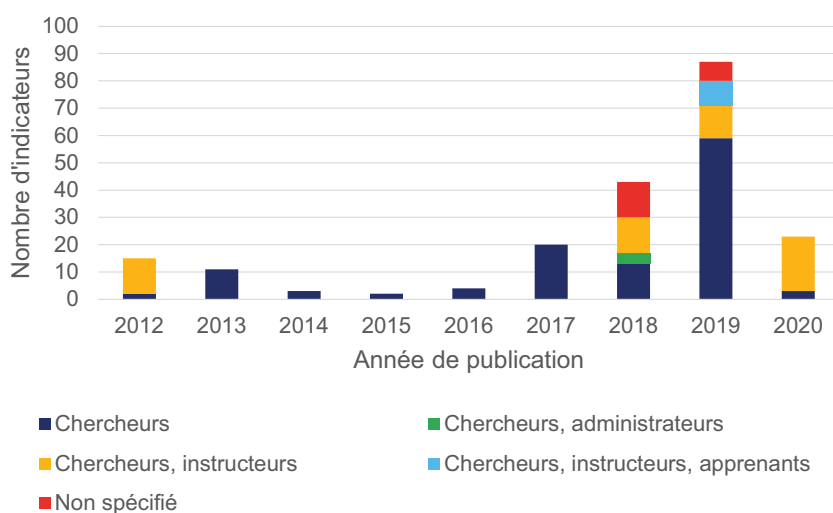


FIGURE 9.7 – Les parties prenantes participant à la sélection d’un indicateur versus l’année de la publication.

9.5.2 (Q2.2) Dans des cas où plus d’un acteur est impliqué dans la phase de conception, y a-t-il de la conception participative ?

La méthode de conception participative avec le(s) cible(s) d’un service ou d’un produit se base sur le principe d’un travail partagé, où l’utilisateur final est impliqué de l’idée au prototypage. Souvent utilisée dans la méthode du *Design Thinking*, cette méthode de conception est centrée sur l’utilisateur qui a un rôle actif lors des phases de conception, et non seulement au moment des tests (voir par exemple [112]).

Nous avons vérifié que seulement 8 (19,05 %) des articles de notre corpus indiquent avoir proposé des indicateurs via une démarche de conception participative. Les instructeurs sont les plus impliqués (N = 6 articles) suivis par des administrateurs (un seul article, [21]) et instructeurs et apprenants ([119]). Nonobstant, parfois les cibles de ces indicateurs (utilisateurs finaux, voir ci-dessous) ne correspondent pas à l’acteur participant à la conception. C’est le cas de [21], avec apprenants et administrateurs comme cibles, mais avec une conception participative qu’avec des administrateurs et de [118], avec apprenants et instructeurs comme cibles, mais avec une conception participative seulement avec des instructeurs. L’article [119], de son côté, cite une conception participative avec instructeurs et apprenants pour des indicateurs ayant comme cible les apprenants. L’article [21] cite une participation des instructeurs, mais ne spécifie pas clairement la cible de ces indicateurs.

9.5.3 (Q2.3) *Qui est la cible de l'indicateur proposé ?*

Nous avons pu observer également la ou les cibles des indicateurs proposés par les articles de notre corpus. La majorité de ces indicateurs, comme le montre la Figure 9.8 est à l'intention des instructeurs (N = 130 indicateurs) et des apprenants (N = 100 indicateurs). Les administrateurs sont la troisième cible (N = 22 indicateurs). Cette information est présente pour tous les articles analysés, ce qui n'est pas toujours le cas pour d'autres informations sur les parties prenantes (comme la conception ou les éventuels tests).

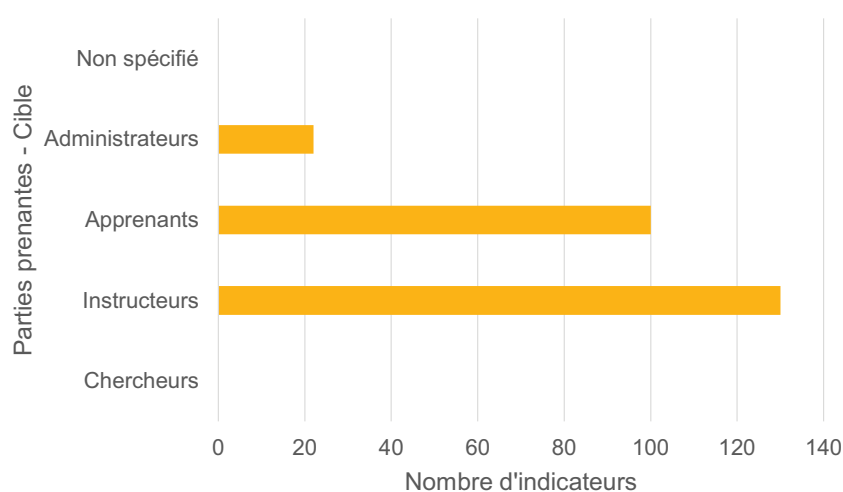


FIGURE 9.8 – Les parties prenantes cibles d'un indicateur.

La Figure 9.9 permet de visualiser les différentes cibles des indicateurs versus l'année de publication ; elle nous montre également que certains indicateurs semblent être conçus pour être utilisés par plus d'une partie prenante. Nous remarquons également que certains indicateurs sont également utilisés pour faire d'autres calculs, souvent en tant qu'entrée pour le calcul d'un autre indicateur.

9.5.4 (Q2.4) *Avec quelle partie prenante l'indicateur proposé est-il testé ?*

Nous avons pu observer en outre les utilisateurs avec qui les indicateurs proposés par les articles de notre corpus ont été testés. Comme précédemment, il faut noter que le codage n'a pas été exclusif pour certains indicateurs — s'il a pu être testé avec des apprenants et des instructeurs, par exemple.

Bien que la majorité des indicateurs recensés soit à l'intention des instructeurs, c'est avec les apprenants que la plupart des tests reportés

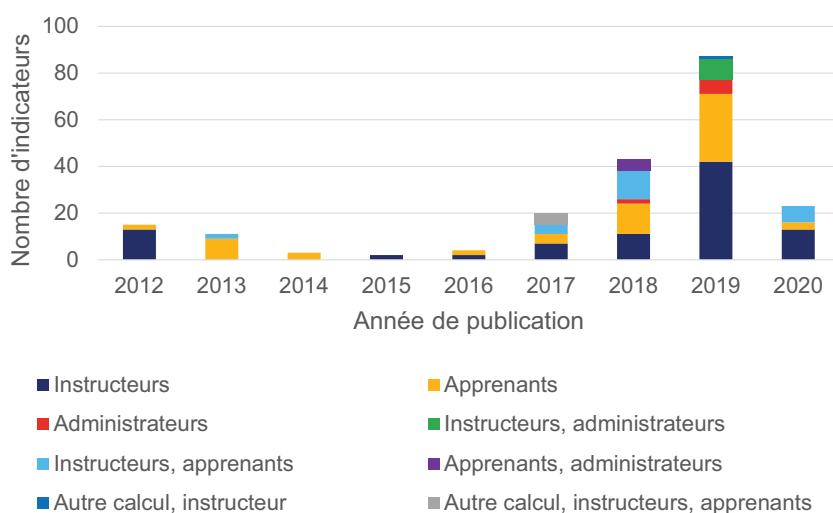


FIGURE 9.9 – Les parties prenantes cibles d'un indicateur versus l'année de la publication.

ont été réalisés (cf. Figure 9.10). Pour un total de 80 indicateurs, les apprenants ont été cités comme faisant partie des tests (contre 100 les ayant comme cible). Les instructeurs, à leur tour, ont participé aux tests de 70 des indicateurs recensés (en ayant été la cible de 130 indicateurs), dès lors seulement un peu plus de la moitié (53,84 %) de ces indicateurs ont été testés auprès des utilisateurs finaux. Les administrateurs ont participé à des tests dans le cas de l'implémentation de 13 indicateurs (et la cible de 22 indicateurs). Aucun test avec des utilisateurs n'a été mentionné pour 70 (33,65 %) des indicateurs retrouvés dans les articles retenus.

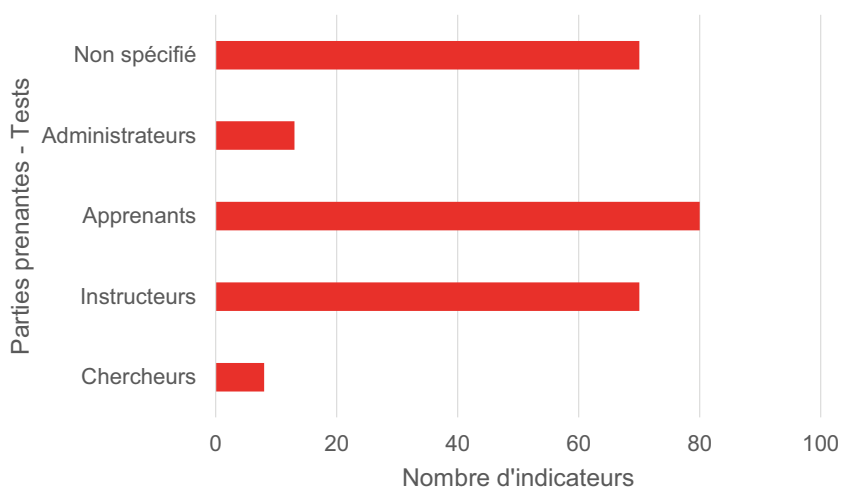


FIGURE 9.10 – Les parties prenantes ayant participé à un test d'un indicateur.

L'analyse des parties prenantes ayant participé à un test d'un indicateur versus l'année de la publication (cf. Figure 9.11) montre une certaine augmentation de travaux ne mentionnant pas de tests depuis 2016. À noter que 1/ les données de 2020 ne comprennent pas l'année entière et que 2/ il est possible d'avoir une augmentation de travaux de recherche qui n'ont pas encore atteint la phase de tests au moment de la publication, mais ce facteur reste à être mieux étudié.

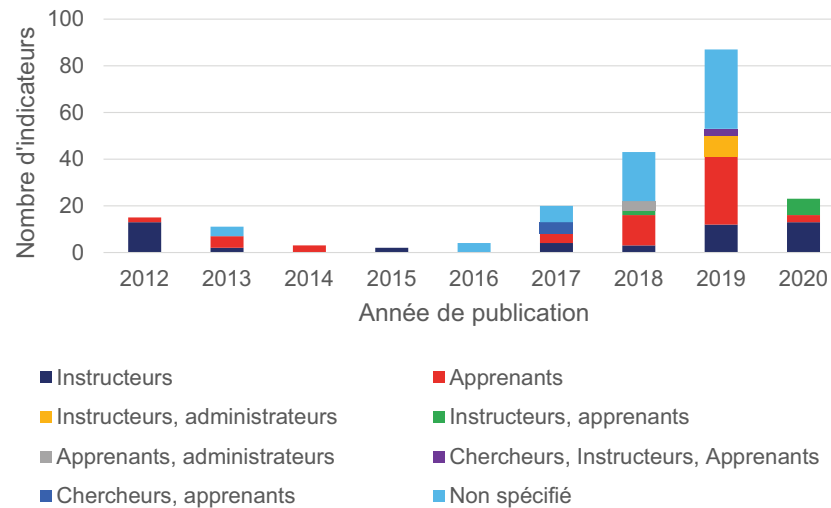


FIGURE 9.11 – Les parties prenantes ayant participé à un test d'un indicateur versus l'année de la publication.

Plus d'informations sur le format de ces tests sont disponibles dans la dimension « Contexte », section 9.7.2 ci-dessous.

Pour conclure sur l'implication des parties prenantes aux différentes phases du cycle de vie d'un indicateur, la Figure 9.12 résume le nombre de fois où un des acteurs a été impliqué sur chacune des phases (sélection, cible et test). Les instructeurs sont les plus impliqués et participent activement aux trois phases. Les chercheurs sont surtout présents lors de la sélection des indicateurs proposés, étant ceux à l'origine de la plupart des propositions. Les apprenants sont surtout la cible des indicateurs et participent aux tests ; ce sont également les phases où participent les administrateurs qui sont moins nombreux.

9.5.5 Discussion

Les parties prenantes autour de l'analytique des apprentissages sont nombreuses et peuvent intervenir potentiellement à plusieurs moments, de la proposition aux usages quotidiens, avec des besoins souvent distincts.

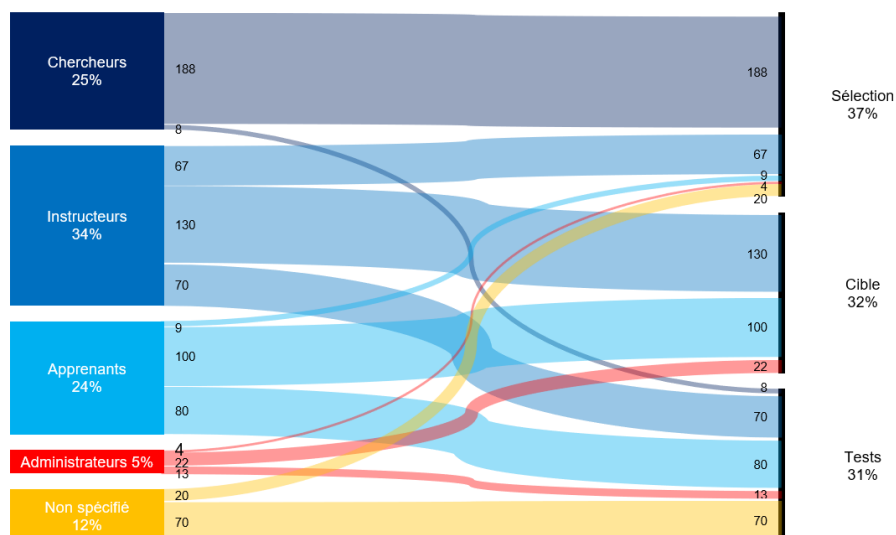


FIGURE 9.12 – Les parties prenantes impliquées dans les phases du cycle de vie d'un indicateur.

À travers cette revue systématique, nous avons pu observer que les indicateurs ont été proposés par des chercheurs la plupart du temps, mais une tendance à impliquer d'autres acteurs semble se dessiner, ce qui reste à être confirmé les prochaines années. De plus, nous avons pu remarquer que le choix d'une conception participative est peu présent.

Les instructeurs continuent à être la cible des indicateurs recensés, même si de plus en plus de propositions à l'égard des apprenants voient le jour. Les administrateurs sont peu nombreux comme cibles de ces analyses, pourtant ces résultats peuvent être biaisés du fait que nous avons restreint la définition des indicateurs à ceux utilisant des traces d'interactions.

Nous soulignons également que relativement peu de tests avec des utilisateurs ont été reportés, encore moins avec les instructeurs, principales cibles.

Dans cette troisième dimension, nous nous sommes questionnés sur l'origine des données, l'analyse de celles-ci et, enfin, la visualisation des résultats.

9.6.1 (Q3.1) Quel artefact numérique est à l'origine des interactions qui ont été tracées ?

Notre premier questionnaire porte sur l'origine des données, c'est-à-dire, quel est l'artefact numérique à l'origine des interactions qui ont été tracées. La Figure 9.13 nous montre que la majorité des articles (35,71 %) indique que les données proviennent d'un LMS (*Learning Management System*). Ce chiffre est suivi de près par des outils spécialisés, c'est-à-dire conçus spécifiquement pour un besoin, que ce soit dans le cadre d'un projet de recherche ou encore un outil développé et utilisé par une université spécifique, par exemple. Un total de 13 des 42 articles (30,95 %) indique avoir utilisé des données provenant de ce type d'outil. Un bon nombre d'articles ne mentionne pas la provenance des données utilisées (26,19 %), 2 articles (4,76 %) utilisent des données provenant d'outils basés sur le Web et un seul (2,38 %) déclare avoir utilisé des données provenant d'une plateforme de MOOC.

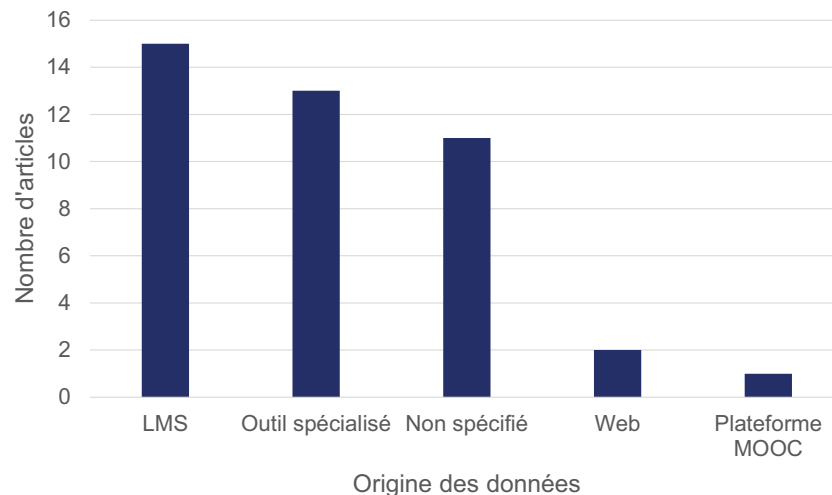


FIGURE 9.13 – Les artefacts numériques sources des données utilisées lors du calcul des indicateurs.

En ce qui concerne les LMS cités comme étant à l'origine des données, Moodle (26,67 %) a été le plus cité, suivi de Canvas (20,00 %). Cependant, la majorité n'a pas spécifié le LMS utilisé (33,33 %). La Figure 9.14 donne un aperçu des LMS mentionnés. Les outils spécialisés, à leur tour, n'ont été cités qu'une fois chacun, il s'agit de : BEDA, Educational Software Tool, Engineering Learning Center, Lab4CE, NetBeans, Piazza, Quiz My Class Understanding (QMCU), RUBRICS-360 et Statway, avec 4 outils non nommés. Les outils Web qui ont été cités sont Twitter et Google Docs ; la plateforme de MOOC est HPI. Nous remarquons des solutions liées à des sessions de laboratoire ainsi que des outils

liés à l'apprentissage de la programmation informatique, simulateurs, entre autres.

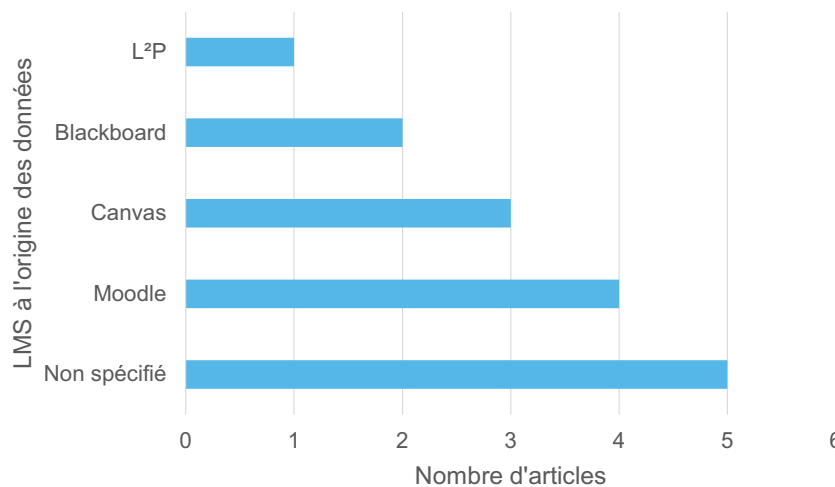


FIGURE 9.14 – Les environnements du type LMS sources des données utilisées lors du calcul des indicateurs.

9.6.2 (Q3.2) *Le cas échéant, quels algorithmes sont utilisés pour calculer l'indicateur ?*

Plusieurs algorithmes ont été proposés par des chercheurs permettant de faire de la classification, de la prédiction ou encore du regroupement, comme indiqué dans 5.6. Nous avons lu les publications sélectionnées afin de vérifier lesquels de ces algorithmes sont cités comme utilisés lors du calcul des indicateurs présents dans chaque article.

Uniquement 7 des 42 articles retenus citent l'usage d'un algorithme spécifique (16,67 %). Les auteurs de [115, 119] indiquent l'utilisation de l'arbre de décision. [134] cite faire de la classification, mais ne précise pas l'algorithme utilisé. [16] a utilisé Xerox Incremental Parser (XIP) pour une analyse syntaxique textuelle. [120] a appliqué les algorithmes K plus proches voisins et forêt d'arbres décisionnels pour faire de l'apprentissage automatique. [107] a utilisé Petri net dans l'algorithme de heuristic miner de ProM. [11] a utilisé la méthode k plus proches voisins, machine à vecteurs de support et l'arbre de décision. Enfin, [68] cite avoir utilisé le Natural Language Toolkit, une librairie de Python, sans plus de détail.

9.6.3 (Q3.3) Comment les résultats des analyses sont-ils visualisés ?

La grande majorité des indicateurs présents dans les articles sélectionnés sont visualisés dans des tableaux de bord : 35 des 42 articles disent utiliser ce mode de visualisation (83,33 %) — ce qui correspond à 162 des 208 indicateurs (77,88 %), avec en plus 9 indicateurs visualisés via des tableaux de bord et des mails envoyés aux dessinateurs de ceux-ci, ainsi qu'un indicateur visualisé dans un tableau de bord et via une notification.

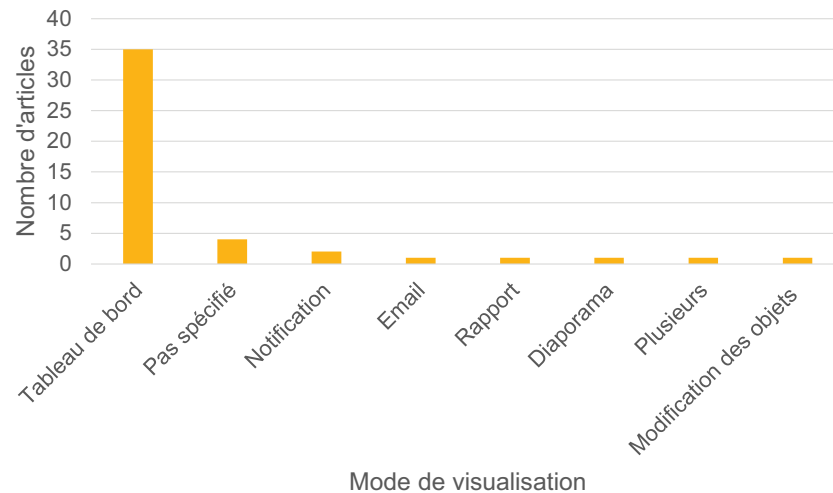


FIGURE 9.15 – Les modes de visualisation des indicateurs.

9.6.4 (Q3.4) Quelle est la représentation graphique choisie pour l'indicateur ?

Nous avons ensuite vérifié les types de visualisation — notamment des graphiques — choisis afin de rendre visibles les résultats obtenus. Vu la variété de formes graphiques qui pourraient être choisies, nous avons suivi au maximum la nomenclature disponible dans Data to Viz¹⁰ lors du codage ; c'est également pourquoi nous avons fait le choix de garder ces termes en anglais dans la Figure 9.16 qui synthétise ces résultats. Cette Figure nous montre les types de visualisation choisis par les articles sélectionnés, où la catégorie « Plusieurs » regroupe ceux où plus d'un mode de visualisation est cité (comme exemplifié ci-dessus).

Une grande variété de formes graphiques a été choisie afin de permettre la visualisation de ces résultats dans des tableaux de bord — ou

10. cf. <https://www.data-to-viz.com/>

autre mode. L'histogramme est le choix prioritaire, cité 42 fois (20,19 % des 208 indicateurs recensés). La catégorie regroupant des indicateurs — « *mixed* » — où plus d'un type de graphisme est utilisé est la deuxième option, suivie d'une variété d'autres formes de graphismes, ou encore des simples listes ou tables.

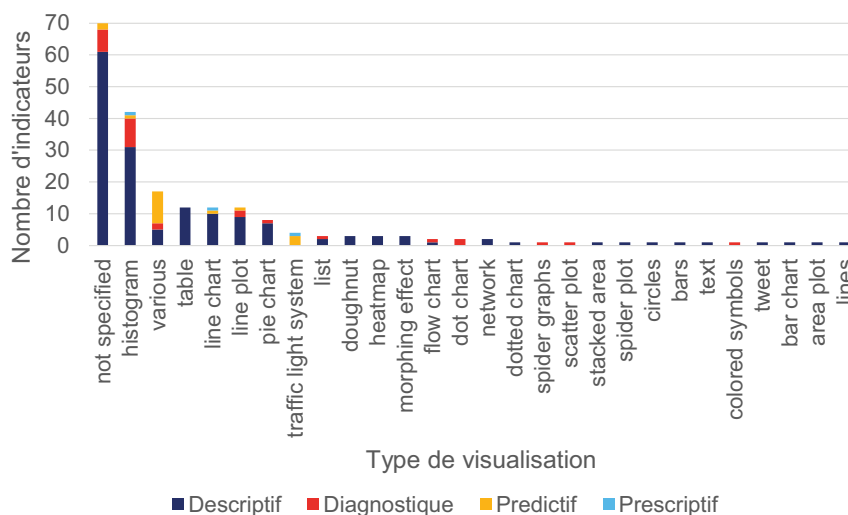


FIGURE 9.16 – Les types de visualisation des indicateurs versus le type d'analyse.

Cependant, pour beaucoup d'indicateurs ($N = 70$, soit 33,65 %), aucune information n'est fournie sur comment ces résultats sont affichés dans les tableaux de bord ou autre, ce qui pourrait nuire aux études comparatives sur la perception ou compréhension de chaque graphique, par exemple. Les types d'analyses conduites pour arriver à l'indicateur ne semblent pas être déterminants pour le type de visualisation choisi, mais une certaine tendance semble se dessiner pour des histogrammes lors des analyses diagnostiques et pour des systèmes de feu de circulation (rouge/jaune/vert) lors des analyses prédictives.

9.6.5 (Q3.5) Où ces visualisations sont-elles proposées ?

La dernière question que nous avons posée concernant la visualisation des indicateurs est celle de l'artefact numérique où ces visualisations sont proposées. La majorité des publications que nous avons étudiées n'affichaient pas cette information, soit un total de 15 articles (35,71 %). Parmi ceux qui citent l'environnement où sont affichés les indicateurs, les outils spécialisés sont les plus nombreux, cités dans 13 articles (30,95 %), suivis des solutions où un plugin est développé ou choisi afin de permettre l'affichage souhaité dans un LMS (23,81 %, 10 articles). Certains ont cité uniquement que les résultats étaient dis-

ponibles dans un LMS, sans citer un plugin ou sans donner d'autres informations à ce sujet (3 articles, 23,81 %). Un seul article a cité une plateforme de MOOC (HPI, *ad hoc*). Ces informations sont visibles dans la Figure 9.17.

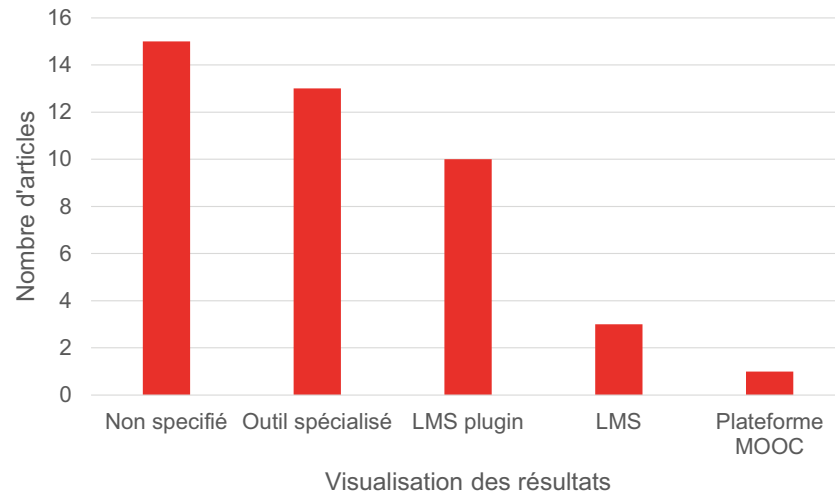


FIGURE 9.17 – Les environnements où les résultats des calculs (indicateurs) sont visualisés.

Les outils spécialisés cités où les indicateurs sont visualisés sont les suivants, cités une seule fois chacun : AMBA, HPI, iFrame, IT-LAT, Lab4CE, LAD, Learning Analytics Backend (LAB), MonitorView, MyLA, OU Analyse (OUA), PADA, PredictCS, Quiz My Class Understanding (QMCU), Raspberry-Pi Solution, SOLAR, Student Explorer, SuiteC, UATU, VeeU2.0 et XIPIt. Les outils Web qui ont été cités sont Twitter et Google Docs ; la plateforme de MOOC est HPI.

9.6.6 Discussion

Ces résultats nous interpellent par le nombre d'articles qui ne citent pas l'origine des données utilisées pour le calcul des indicateurs décrits. Ceci va à l'encontre des efforts dans le sens d'avoir des données ouvertes, accessibles et, enfin, reproductibles, puisque l'on ne connaît même pas l'origine des données. Quant aux données provenant des LMS, on peut supposer qu'il serait possible d'envisager une capitalisation de ces indicateurs, comme le définit [90], car il s'agirait d'un même format/typologie de données si l'on utilise le même LMS. Cette potentielle capitalisation serait également possible pour les outils basés sur le Web et les plateformes de MOOC. Ces données, par contre, sont produites indépendamment du processus d'analytique des apprentissages avec le numérique qui aura lieu dans un second temps, ce

qui est l'inverse de ce que nous envisageons avec la métamodélisation des indicateurs et des traces nécessaires pour les calculer.

Pour beaucoup des indicateurs analysés, peu d'informations sont fournies, que ce soit sur le graphisme choisi pour afficher les résultats, ou encore l'artefact numérique qui doit être accédé par les différents acteurs qui sont la cible de ces visualisations. Pour ceux qui détaillent ces informations, le choix semble être surtout des graphiques simples, comme l'histogramme, le graphique linéaire ou le diagramme circulaire.

Nous retenons, pour la métamodélisation, une envie de proposer un choix de formes différentes de visualisations de données, supplémentaires aux tableaux de bord, malgré une prédominance de ces derniers dans la littérature. Ce souhait provient du fait de 1/ prendre en compte les résultats de l'étude décrite dans le Chapitre 8 où les apprenants ont montré une préférence pour un affichage au long du document dès que possible et, 2/ de profiter de la connaissance de la structure des documents produits dans notre contexte, ce qui permettrait de développer cela plus aisément.

Du fait du manque d'informations préalables sur lesquelles l'on pourrait s'appuyer, deux interprétations seraient possibles : l'une où le fait de regarder des informations provenant d'indicateurs au long du contenu des documents pourrait nuire au *flow*¹¹ de la session d'étude et l'autre où le fait de ne pas avoir à s'arrêter et aller à un tableau de bord afin de consulter cette information faciliterait le *flow*.

Spécifiquement pour des fins de modélisation, nous retenons également le besoin de permettre d'indiquer, à chaque étape, qui est la cible ou les cibles de chacun des indicateurs modélisés, comment ils seront affichés pour chacune de ces parties prenantes, ainsi que, pour chaque besoin d'une entrée qui permettra de calculer l'indicateur le cas échéant.

9.7 RÉSULTATS : DIMENSION 4 — CONTEXTE

La dernière dimension de notre revue systématique autour des indicateurs de l'analytique des apprentissages avec le numérique porte sur les contextes où ils sont proposés. Nous nous intéressons sur la phase de conception où ils se trouvent, de la proposition théorique à un usage généralisé dans la pratique, le domaine et type de dispositif pédagogique où il a été proposé.

11. « État de motivation et d'attention focalisées qui permet de mobiliser au mieux ses facultés et d'atteindre une performance optimale » [150]

9.7.1 (Q4.1) Quel est le statut (conçu, testé ou déployé) de cet indicateur ?

Afin de connaître le statut des indicateurs, nous avons procédé au codage suivant, pour chaque indicateur :

- Conçu : l'indicateur en question a été proposé, par exemple, via un prototype, mais il n'a pas encore été testé.
- Testé : l'indicateur a été testé (plus d'information sur les tests ci-dessous), mais il n'est pas encore utilisé dans des contextes pratiques.
- Déployé : l'indicateur est déjà utilisé dans la pratique, c'est-à-dire, il est utilisé dans des situations d'apprentissage réelles et non seulement à des fins de tests.

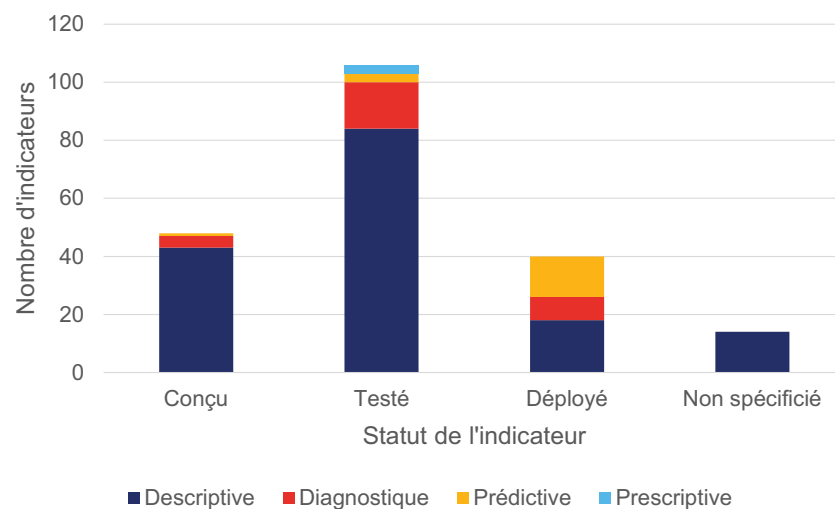


FIGURE 9.18 – Les statuts des indicateurs versus le type d'analyse de l'indicateur.

À peu près la moitié des indicateurs trouvés dans notre étude ont été testés, mais pas encore utilisés dans la pratique, c'est le cas de 106 des 208 indicateurs (50,96 %). Nous avons trouvé dans notre corpus 48 indicateurs (23,08 %) encore sous un statut jugé « conçu » et seulement 40 sont décrits comme étant déployés et utilisés dans la pratique (19,23 %). Peu d'articles ne permettaient pas d'appréhender sous quel statut les indicateurs décrits se trouvaient (un total de 14 indicateurs, provenant de deux publications : [70] et [126]). La Figure 9.18 résume ces résultats à propos du statut des indicateurs et rajoute l'information du type d'analyse effectuée.

Nous avons également vérifié s'il existe une relation entre les types d'analyse des indicateurs et leurs statuts lors de la publication (cf. Figure 9.18). Les indicateurs déployés ont plus tendance à avoir une analyse de type descriptif, en comparaison avec les autres statuts.

Même si des indicateurs portant sur une analyse diagnostique ont été identifiés pour tous les statuts, ils sont comparativement plus nombreux en phase de test, pour les indicateurs que nous avons pu recenser. Enfin, des indicateurs ayant une analyse prescriptive n'ont été rencontrés que dans des indicateurs en phase de test.

9.7.2 (Q4.2) Comment a-t-il été évalué ?

Les méthodes utilisées afin d'évaluer les indicateurs auprès des utilisateurs sont nombreuses. Cette information est présente dans 19 des 42 articles retenus (45,24 %) (cf. Figure 9.19). La grande majorité des évaluations suit des méthodes dites qualitatives (13 articles, soit 63,16 % des articles ayant mentionné une méthode d'évaluation auprès des utilisateurs). D'autres citent avoir utilisé des questionnaires (26,32 % des 19 articles) et seulement 2 articles (10,53 %) mentionnent avoir utilisé des méthodes mixtes pour ces évaluations.

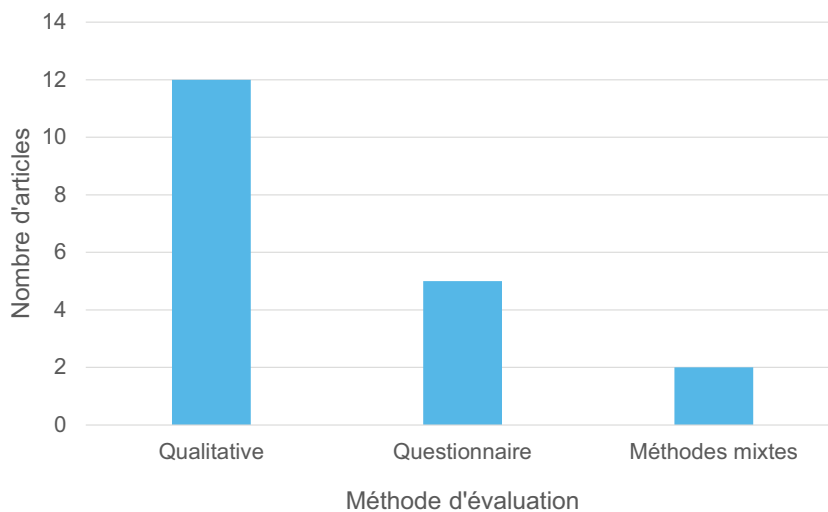


FIGURE 9.19 – Les méthodes d'évaluation des indicateurs citées.

Les méthodes et techniques qualitatives d'évaluation des indicateurs proposés ayant été mentionnées sont : *thinking aloud* [24, 48], *case study* [100], *evidence-based action research* [96], *retrospective interviews* [6], *heuristic evaluation*, *cognitive walkthrough*, *pluralistic walkthrough* [48]. D'autres n'ont pas cité une méthode spécifique, ayant mentionné juste « étude qualitative » ou « entretien semi-structuré ».

Seul un article mentionne une méthode ou technique spécifique concernant le questionnaire, c'est le « *User Experience Questionnaire (UEQ)* ». Les mêmes auteurs mentionnent avoir basé certaines des questions sur le « *Evaluation Framework for Learning Analytics (EFLA-4)* ».

Enfin, certaines études ont été évaluées via la comparaison des données d'usage (sans consulter directement un utilisateur) ou encore, certains ont testé uniquement les résultats de différents algorithmes, c'est les cas, par exemple, de [5, 74, 141]

9.7.3 (Q4.3) *Quels sont les domaines d'apprentissage dans lesquels l'indicateur a été proposé ?*

Concernant les domaines de compétence où un indicateur a pu être proposé (ou testé), nous avons codé les articles selon qu'ils ont été proposés à des apprenants provenant de cours ou formations appartenant :

- au domaine dit STEM (pour *sciences, technologie, ingénierie et mathématiques*, en anglais¹²). Il s'agit d'un terme utilisé pour une politique curriculaire où l'on donne la priorité à deux des trois branches des sciences, les sciences dites naturelles et les sciences formelles, souvent dans une approche interdisciplinaire et appliquée ;
- à toute autre discipline non considérée comme appartenant au domaine des STEM. Il s'agit notamment de la branche des sciences dite sciences sociales, ainsi que les sciences humaines et les arts.

Nous avons fait ce premier choix, très noué à des classifications anglophones, du fait que plusieurs articles mentionnait seulement ce terme et non la branche ou discipline subjacente qui permettrait de suivre une autre classification.

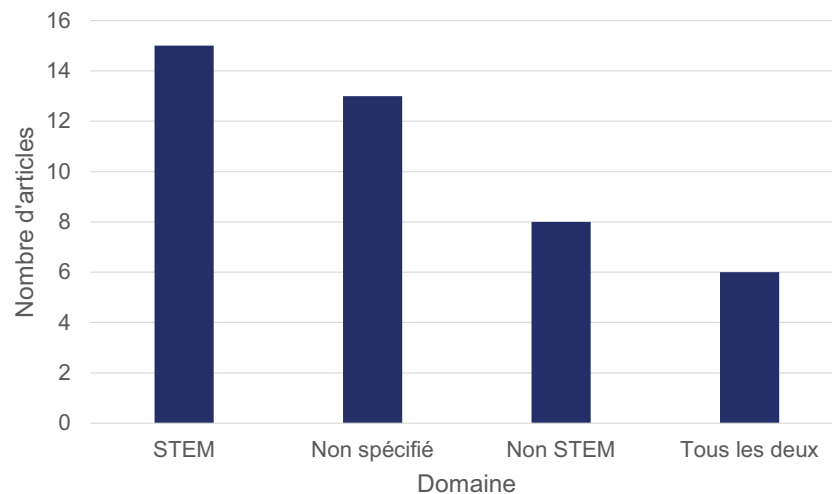


FIGURE 9.20 – Les domaines où l'indicateur a été proposé.

12. Voir : [Science, technology, engineering, and mathematics — Wiki](#)

Nous remarquons dans la Figure 9.20 que la plupart des publications mentionnaient avoir proposé des indicateurs — implémentés, testés, déployés — pour des cours ou formations faisant partie des STEM. C’est le cas de 15 des 42 articles recensés (35,71 %). Seulement 8 publications mentionnent avoir proposé des indicateurs pour des cours ou formations n’appartenant pas aux STEM (19,05 %), tandis que 6 publications ont proposé des indicateurs pour les deux (14,29 %). Plus d’un tiers des publications (13, soit 30,95 %) ne mentionne pas le domaine des cours ou formations pour lesquels les indicateurs ont été proposés.

Nous avons ensuite cherché à détailler ces données, quand cela est possible selon les informations présentes dans les publications, en codant les informations relatives à la classification des sciences¹³ axée par la nature (catégories) :

- les sciences formelles (ou sciences logico-formelles) ;
- les sciences physiques ;
- les sciences de la vie ;
- les sciences sociales.

Puis, nous avons pris note de la discipline spécifique mentionnée, si c’était le cas. Le résultat de ces codages est montré dans la Figure 9.21 où les données sont affichées sous forme hiérarchique, d’un ensemble de rectangles imbriqués, dont la zone est proportionnelle à sa valeur en nombre d’occurrences parmi les publications.

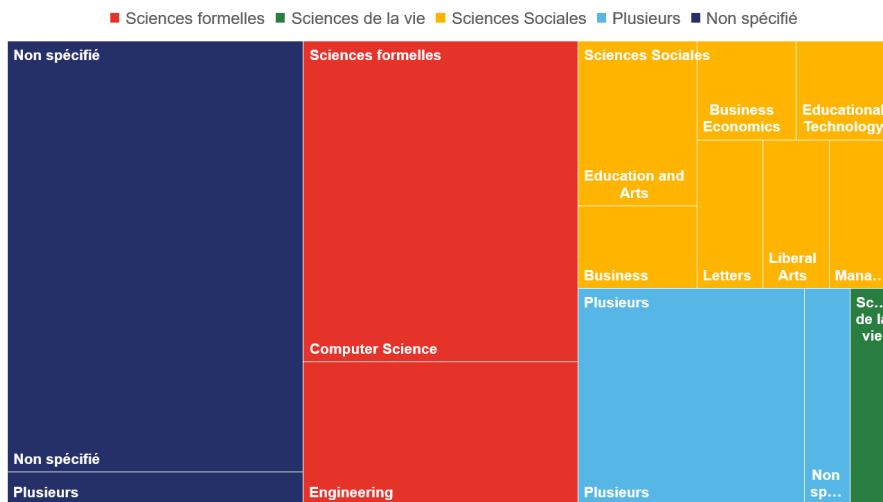


FIGURE 9.21 – Les domaines et disciplines où l’indicateur a été proposé.

13. Voir : <https://fr.wikipedia.org/wiki/Science>

9.7.4 (Q4.4) *Quel est le type de dispositif pédagogique en place pour lequel cet indicateur est proposé ?*

Notre dernier questionnaire à propos des indicateurs recensés porte sur le type de dispositif pédagogique pour lequel il est proposé. Nous avons codé les indicateurs dans trois catégories selon les informations présentes dans l'article :

- Présentiel : typiquement des cours ou formations classiques dispensés en face-à-face, avec des supports qui peuvent être disponibles ou non en ligne ;
- Distanciel : il s'agit de cours ou formations où il n'y a pas de face-à-face physique, la présence d'un environnement technologique — souvent une plateforme — est utilisée ;
- Hybride : cours ou formations où une partie se fait en distanciel et une partie en présentiel.

Seul l'article [6] mentionne deux types de dispositifs pédagogiques pour lesquels les indicateurs ont été proposés, nous avons alors décidé de le coder dans deux des catégories (hybride et présentiel). Cela porte le codage total à $N = 43$ pour cette Figure et nos considérations pour cette partie.

La Figure 9.22 nous permet de remarquer que plus de la moitié des articles retenus n'ont pas précisé cette information, c'est le cas de 22 des 42 articles (51,16 %, des 43 codages). L'hybride (11 articles, 25,58 %) et le distanciel (7 articles, 16,28 %) sont les deux types de dispositifs pédagogiques les plus choisis pour la proposition d'indicateurs parmi notre corpus. Les cours ou formations en présentiel n'ont été cités comme type de dispositif pédagogique que par 3 articles (6,98 %).

Une analyse des types de dispositifs pédagogiques cités versus l'année de publication (cf. Figure 9.23) nous montre que le distanciel n'est cité qu'à partir de 2017 et jusqu'en 2019 et le présentiel en 2018 et 2019. Une autre tendance semble être une augmentation du nombre d'indicateurs présents dans des articles où le type de dispositif pédagogique n'est pas mentionné. Nous rappelons que les données pour 2020 sont incomplètes.

9.7.5 (Q4.5) *Quelle est la théorie de l'apprentissage sur laquelle repose l'indicateur proposé ?*

La proposition d'un indicateur n'est pas exempte d'une certaine intention d'utilisation qui, finalement, conduirait à une amélioration des résultats d'apprentissage. Selon [78], il existe une notion commune dans la communauté de l'analytique des apprentissages avec le nu-

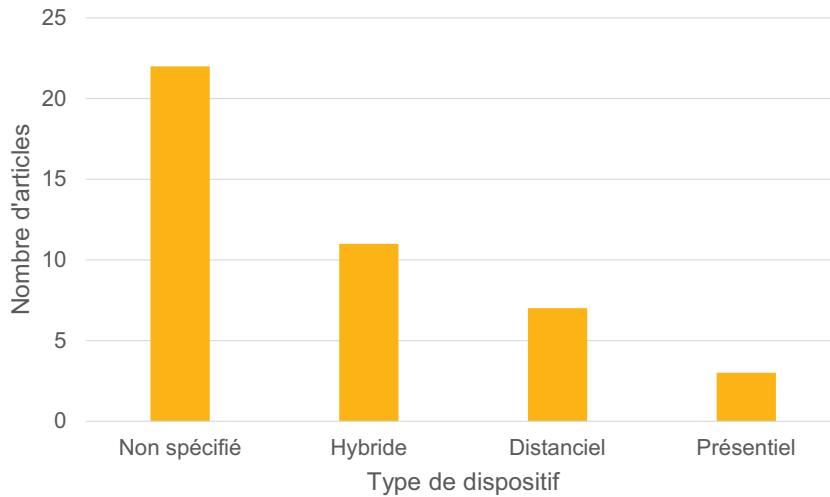


FIGURE 9.22 – Les types de dispositifs pédagogiques où l'indicateur a été proposé.

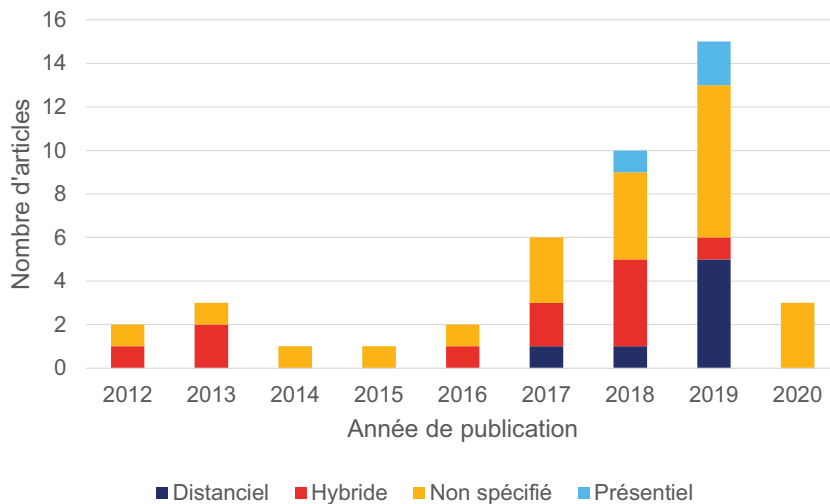


FIGURE 9.23 – Les types de dispositifs pédagogiques où l'indicateur a été proposé versus l'année de publication.

mérique selon laquelle la recherche doit être profondément ancrée dans les sciences de l'apprentissage. De toute évidence, l'indication — ou la décision — d'utiliser un indicateur spécifique gagne en force si elle est fondée sur l'une des différentes théories d'apprentissage liées à la pédagogie et à l'enseignement. Pour ces raisons, nous avons cherché à coder les différentes théories d'apprentissage sur lesquelles les indicateurs trouvés dans les articles lus se seraient basés.

Nous n'avons pu repérer qu'un seul article [99] citant une théorie de l'apprentissage servant de base pour la conception d'un indicateur, le constructivisme social. En élargissant le codage à des approches ou

contextes cités, nous avons pu construire la Table 9.3 avec les 14 publications ayant cité une approche ou un contexte (33,33 %). Ceci nous permet de constater que l'apprentissage autorégulé est l'approche la plus citée (trois fois). Cette approche bénéficie naturellement de tableaux de bord avec des indicateurs décrivant ce qui s'est passé, permettant ainsi aux apprenants d'évaluer leurs parcours et de les comparer avec les objectifs établis préalablement. Les deux autres approches, citées chacune deux fois, sont l'approche par les compétences (APC) et la pédagogie par projet (APP).

Nous n'avons pas pu repérer, dans notre corpus, d'études visant à croiser deux théories d'apprentissage ou approches pédagogiques dans une optique d'analytique des apprentissages avec le numérique. Par exemple, il pourrait être intéressant d'évaluer la pertinence ou l'usage d'un même indicateur dans deux contextes distincts par rapport aux approches pédagogiques en place, ou encore par rapport à deux conceptions pédagogiques distinctes d'un même cours.

TABLEAU 9.3 – Les approches ou contextes pédagogiques.

Approche	Total
Apprentissage autorégulé	3
Approche par compétences	2
Pédagogie de projet	2
Classe inversée	1
Conception centrée sur l'utilisateur	1
Constructivisme social	1
Enseignement entre pairs	1
<i>Inquiry-based learning</i>	1
Ludification	1
<i>Scholarship of Teaching and Learning</i>	1
<i>Teamwork Model (Dickinson and McIntyre's)</i>	1
Total	15

9.7.6 Discussion

Très semblablement, un même indicateur ne répond pas forcément aux mêmes besoins selon le type de dispositif pédagogique où il sera utilisé (cf. par exemple les remarques des participants dans 8.6.8).

Un autre point est celui relatif aux domaines d'études où l'on propose et teste les résultats des analyses effectuées. Les besoins peuvent varier et selon [128], le défi de développer des systèmes d'analytique des apprentissages avec le numérique spécifiques à une discipline et contextualisés demeure, car la plupart de ces systèmes

sont développés pour être généralisables et ouverts à des contextes plus larges. Les mêmes auteurs rappellent que « un modèle prédictif formé sur les données d'une discipline peut ne pas être généralisable à une autre discipline s'il existe des facteurs contextuels qui affectent les performances du modèle » [128, p. 210, notre traduction]. Pour rappel, la majorité des publications ne précise pas cette information et parmi celles qui la spécifient, presque la moitié cite les sciences formelles.

Peu d'indicateurs semblent être conçus sur la base d'une théorie de l'apprentissage établie qui permettrait d'envisager au préalable son utilisation et le(s) bénéfice(s) attendu(s). Ces résultats semblent ne pas différer de ceux de [78], où les auteurs suggèrent que la conception des tableaux de bord doit être mieux ancrée dans les sciences de l'apprentissage.

Il sera intéressant de vérifier la pertinence d'un même indicateur modélisé pour un modèle documentaire et utilisé dans des contextes ou bien types de dispositif pédagogique différents. Ou encore, l'utilisation d'un même indicateur dans des modèles documentaires différents. Ceci nous incite à métamodéliser tout en permettant des variations d'un même indicateur afin d'essayer de prendre en compte ces possibles variations qui seront nécessaires sur le terrain. Ceci devrait être possible via le nombre de briques et choix internes de ces dernières mis à disposition des modélisateurs où, d'un côté un changement simple permettrait d'adapter l'indicateur à un contexte spécifique et, de l'autre, une activation ou désactivation d'un indicateur modélisé permettrait à l'utilisateur de l'inclure ou pas dans son contexte.

9.8 LIMITATIONS

Les limites qui ont une incidence sur cette étude sont surtout liées au fait que le terme « indicateur » ne soit pas utilisé de manière homogène et systématique au sein de la communauté contribuant à l'analytique des apprentissages avec le numérique. De cette façon, nous reconnaissons que plusieurs articles proposant un indicateur puissent ne pas figurer dans nos analyses.

De plus, les informations concernant les différentes caractéristiques d'un indicateur que nous avons pu repérer dans les quatre dimensions ne sont souvent pas affichées dans les textes analysés. Ce fait ne permet pas, par exemple, qu'une autre équipe puisse reprendre les résultats obtenus et décider d'implémenter le même indicateur, car le contexte n'est pas informé. Avec cette situation, la capitalisation des résultats est anéantie et la répliquabilité de la recherche pourrait être compromise.

Malgré ces limitations, cette étude met en lumière les différents indicateurs et leurs caractéristiques. Ces informations nous permettent de mieux prendre les décisions lors de la proposition du métamodèle. Elles pourront être utiles également aux équipes lors de la définition et modélisation des indicateurs dans la pratique.

Cinquième partie

CONTRIBUTION : LE MÉTAMODÈLE

Le Chapitre 10 est dédié à la présentation du métamodèle que nous proposons et qui répond à **RQ1** et **RQ2**. Nous présentons d'abord une description détaillée de ce métamodèle et les raisonnements effectués derrière certaines décisions que nous avons prises. Ensuite, nous illustrons l'utilisation de ce métamodèle via la modélisation de quatre indicateurs. Dans le Chapitre 11 nous décrivons une série de contraintes que nous avons pour la création du métamodèle et comment elles peuvent influencer la mise en place d'un tel système en suivant l'approche proposée. Le Chapitre 12, est dédié à la communication des résultats d'une étude avec des modélisateurs ayant testé ce métamodèle suite à son déploiement dans un outil dédié.

10.1 INTRODUCTION

Dans ce chapitre, nous présentons le métamodèle conçu et ses primitives. C'est le modélisateur qui va l'utiliser pour créer le modèle documentaire en question. Tout d'abord, nous présentons une description générale du métamodèle, puis nous décrivons comment le processus de modélisation se fait via l'étude de cas de la modélisation de quelques indicateurs.

Ce Chapitre nous permet notamment de répondre aux questions de recherche **RQ₁** Peut-on proposer un métamodèle de collecte, d'analyse et de visualisation des traces d'interaction prenant en compte l'ensemble de la chaîne de production de documents? et **RQ₂** Quels sont les composants nécessaires à ce métamodèle permettant de modéliser les indicateurs couramment utilisés dans le domaine, mais également permettant de les enrichir avec la connaissance fine de la sémantique des documents? Cette proposition sera ensuite testée auprès de modélisateurs expérimentés dans le Chapitre 12.

10.2 DESCRIPTION DU MÉTAMODÈLE

La proposition du métamodèle doit permettre, une fois instancié, de fournir les spécifications indispensables aux différents processus de transformation et de génération. Le métamodèle conçu est composé de trois grandes dimensions :

- Interactions avec le document : il s'agit des potentielles sources de traces d'interactions pouvant être enregistrées.
- Indicateurs : le cœur du métamodèle, avec les primitives permettant de définir un indicateur.
- Modes de visualisation : permet de décider comment la visualisation des résultats obtenus sera proposée aux différents acteurs.

À noter que le métamodèle qui suit reprend seulement les éléments essentiels appartenant au système et qui auront un impact ou seront impactés directement par la présente proposition. En d'autres mots, certaines primitives documentaires, souvent déjà présentes dans le système, ne sont pas mentionnées, par souci de concision.

10.2.1 Modèle d'interaction documentaire

Les différentes interactions avec les documents produits qui pourraient potentiellement être capturés sont étroitement liées aux primitives documentaires existantes, qui permettent la création des documents. Ainsi, nous avons procédé à une investigation de ces interactions en nous basant sur les primitives documentaires que nous décrivons ensuite.

Une primitive documentaire dans notre contexte peut être typée selon six catégories : *organization*, *resource*, *meta-data*, *generics*, *evaluation* et *spatio-temporal*. Les primitives du type *meta-data* et *generics* ne sont pas concernées par les interactions que nous pourrions tracer et ne sont pas abordées ici.

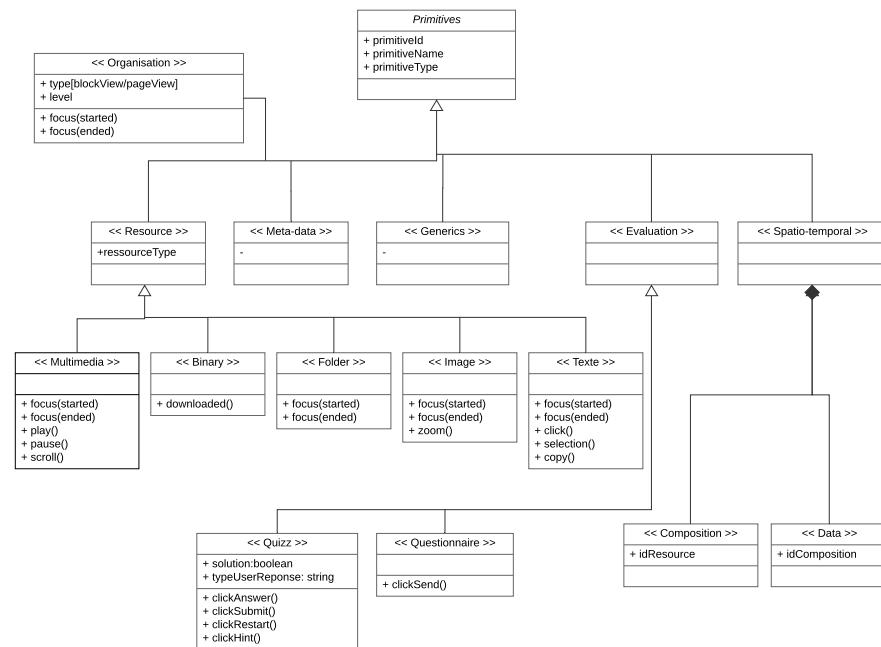


FIGURE 10.1 – Métamodèle : la création des traces d'interaction.

Une primitive documentaire du type *organization* sert essentiellement à définir la structure — physique, logique — du document. Elle définit « les liens méreologiques entre un ensemble et ses parties » [8, p. 35]. En ce qui concerne les documents numériques, il s'agit notamment de la définition de l'arbre structurant le document en pages et en « blocs », les différentes parties ou sections de la page. Le concept de niveau, ici *level*, fait référence à cette structure : un premier niveau, celui de la page, permettrait de définir un traçage de haut niveau comparé à un traçage plus fin indiquant qu'une section d'une page a été visualisée, par exemple. C'est également via les primitives documentaires du

type *organization* que la séquentialité logique du document est définie, qu'elle soit séquentielle ou non, c'est-à-dire dépendante de l'action d'un utilisateur.

Une primitive documentaire du type *resource* permet d'indiquer notamment la structure sémantique des documents. C'est l'ajout d'une primitive de ce type qui détermine les parties possibles d'un document créé à partir du modèle en question. Ces primitives sont intimement liées au métier en question ou bien à l'objectif pour lequel le modèle documentaire est créé. Le métamodèle des interactions avec le document nous montre que ce type de primitive documentaire peut être composé de plusieurs genres de ressources : *multimedia*, *binary*, *folder*, *image* et *text*, par conséquent les types d'interactions varient également.

Un autre type de primitive documentaire qui nous intéresse est celle liée aux évaluations (*evaluation*). Deux possibilités existent : 1/ une ou plusieurs questions traitées de manière isolée *quizz* ou 2/ un ensemble de questions (*questionnaire*) traitées comme un questionnaire, avec par exemple une note totale ou un temps déterminé.

Enfin, *spatio-temporal* est le type de primitive documentaire permettant de créer un lien spatio-temporel entre deux parties de contenu. Les deux possibilités au sein de ce type permettent de 1/ spécifier l'endroit en question (multimédia, image, etc.) et 2/ de créer le lien lui-même.

10.2.2 Indicateur

La primitive principale permettant de modéliser l'indicateur (*indicator*) a quelques attributs traditionnels permettant de l'identifier (*code*, *name* et *description*), ainsi que l'attribut permettant d'indiquer qui a accès à cette primitive (*user*). Ce dernier aspect est déjà présent et modélisé dans le système en question et donc un lien entre ces primitives suffira concernant la primitive proposée. Les trois autres attributs se réfèrent aux entrées (*inputs*), analyses (*analysis*) et sorties (*output*) nécessaires afin de définir l'indicateur qui sera modélisé.

Plusieurs types d'entrées peuvent être fournis :

- Un paramètre : il s'agit de définir une entrée nécessaire qui est assez flexible pour le modélisateur, par exemple une date de début et de fin qui sera utilisée pour réaliser l'analyse à partir des journaux des traces, ou encore un chiffre qui servira comme seuil pour déclencher un message à un utilisateur, etc.
- Une table : une entrée peut également être une table avec des données plus au moins prêtes à être utilisées par l'analyse en question.

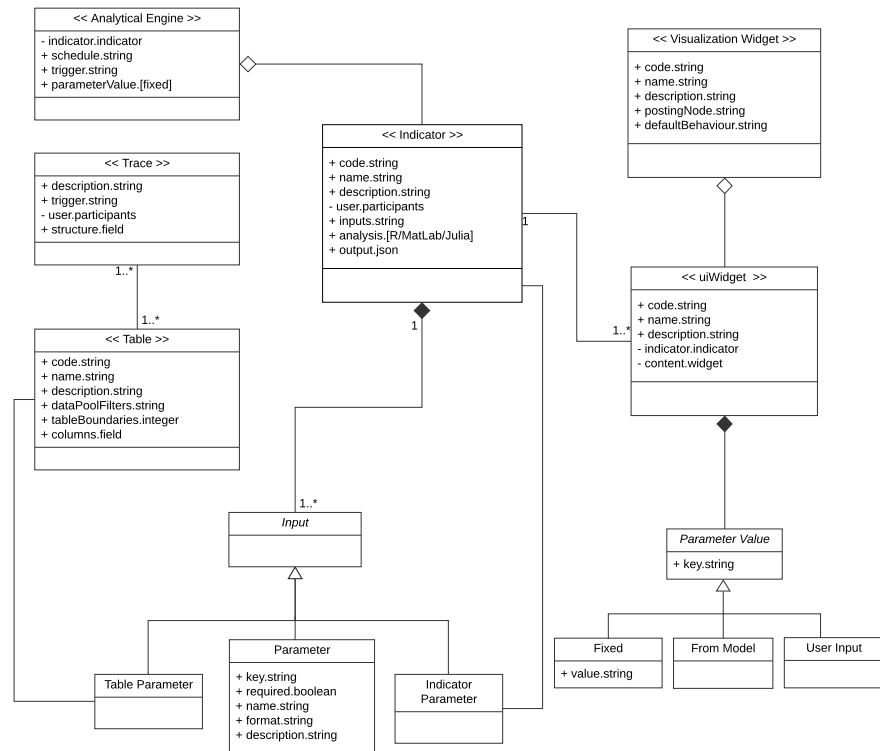


FIGURE 10.2 – Métamodèle : la création des indicateurs.

- Un indicateur : certains indicateurs peuvent avoir comme entrée un autre indicateur calculé préalablement, par exemple, un indicateur visant à afficher l'implication d'un apprenant pourrait avoir pour entrée le temps passé sur du contenu ou bien le nombre de connexions, etc.

Le calcul relatif à l'indicateur doit également être indiqué dans cette primitive. Nous aurions pu métamodéliser cette analyse plus en détail, par exemple en fournissant des opérations standard comme l'agrégation, la somme, etc. Toutefois, nous avons choisi de ne pas le faire vu que l'intention principale était de permettre la modélisation d'un nombre d'indicateurs très variés, incluant des méthodes mathématiques et algorithmes assez complexes comme celles décrites dans le Chapitre 5, Section 5.6.

Les traces qui seront produites à partir de l'interaction avec les documents sont définies dans la primitive correspondante (*trace*). Il s'agit surtout de déterminer le type d'interaction (ouverture d'une page, clic sur le bouton de lecture d'une vidéo, etc.) déclenchant le traçage, mais également les composants documentaires sur lesquels cette règle doit s'appliquer, comme toutes les ouvertures de pages ayant l'étiquette de "conclusion". Cela se fait au travers l'option « dans le modèle », c'est-à-dire, dans le modèle documentaire, via une clé

d'identification. Incontestablement, cette fonctionnalité a un lien étroit avec les interactions possibles décrites dans la Section 10.2.1 et une liste de ces possibilités pourrait être affichée pour faciliter le choix des modélisateurs dans une version définitive implémentée. La primitive relative au traçage permet également de définir ce qui sera écrit dans les lignes de traces produites, par exemple l'ajout de « typeBlock : Concept » aux traces, ce qui permettra de les filtrer ensuite afin de réaliser les calculs.

Il faut savoir que certaines des informations de chaque trace seront ajoutées par défaut et donc elles n'ont pas besoin d'être modélisées à chaque fois. Il s'agit surtout des éléments indépendants par rapport aux analyses envisagées, mais nécessaires technologiquement : c'est le cas du type d'événement (par exemple "evt" : "distrib :logMsg" pour une trace liée à une réponse à un questionnaire ou "evt" : "session :login" pour une connexion — ces éléments peuvent être stockés dans des services différents), de l'adresse de la page en question (par exemple "depotPath" : "/Partenaires/UL/UL-Tech-2017.zip"), de l'identification du projet dans la plateforme ("projectId" : "6g"), de l'identification du participant ou user ("participantId" : "8ux") ou encore de l'horodatage ("ts" : 1504681092278).

Une autre primitive (*table*) vise à permettre aux modélisateurs de définir des tables à partir des traces brutes, qui seraient prêtes à utiliser pour le calcul de certains indicateurs. Cette primitive doit permettre de déterminer le filtrage des traces qui seront prises en compte pour le remplissage de la table en question. Un système booléen permet de faire cette sélection, par exemple en choisissant « (typeBlock = Concept AND (focus = Started OR focus = Ended)) ». Il est également possible de définir des calculs à être effectués lors du remplissage, pour chaque colonne, si nécessaire. Ensuite, le modélisateur doit pouvoir définir les colonnes de cette table et les informations présentes dans les traces qui rempliront chacune de ces colonnes, ainsi que le type de donnée correspondant. Il est également possible de déterminer le temps de conservation de telles données, pour des fins techniques (libération d'espace en mémoire) mais également de conformité (adhérence notamment au règlement général sur la protection des données¹ (RGPD)).

La primitive *Analytical Engine* sert à la modélisation des déclencheurs des calculs des indicateurs. C'est avec cette primitive que le modélisateur détermine l'agenda des calculs effectifs, pour chaque indicateur. Des paramètres, comme des dates de début et fin, peuvent également être définis à ce niveau. Cela devrait permettre, par exemple, de gagner

1. <https://gdprinfo.eu/fr>

en performance dans le cas d'indicateurs où un calcul incrémental évite de le recalculer à partir de la totalité des données.

10.2.3 *Visualisation*

La visualisation des indicateurs peut se faire selon trois grands choix : dans un tableau de bord, à côté du menu ou des parties du document suivant sa structure, ou encore via des messages contextuels. Chacune des spécificités liées à ces trois moyens de visualisation est détaillée dans le métamodèle correspondant, affiché dans la Figure 10.3.

À noter que cet aspect de la visualisation des résultats n'a pas été déployé dans l'outil correspondant lors de l'évaluation auprès des modélisateurs décrite dans le Chapitre 12.

10.3 MODÉLISATION D'INDICATEURS

Afin de mieux appréhender comment le métamodèle sera utilisé dans la pratique, dans cette section, nous présentons des diagrammes de séquence de la modélisation de quatre indicateurs, chacun portant sur un type d'analyse. Cette présentation permet également de comprendre les différentes actions, de la part des utilisateurs et composants de la suite logicielle.

10.3.1 *Indicateur : Temps passé*

Le premier indicateur modélisé à l'aide du métamodèle a été nommé « Temps passé » et il reflète le temps total passé par un apprenant sur une certaine ressource du cours. Il s'agit d'un indicateur que nous considérons de bas niveau (un calcul simple et dépendant seulement des traces). L'analyse faite est classée comme descriptive selon la classification qui se base sur la prise de décision décrite dans le Chapitre 5, Section 5.3.

10.3.1.1 *Description de l'indicateur « Temps passé »*

L'indicateur repose sur les vues des ressources en se basant sur la fonction de *focus* sur une page ou section d'une page, il est possible de calculer chaque tranche de temps (*focus out* moins *focus in*) puis un total de ces délais pour chaque ressource².

2. N.B. : il est aujourd'hui possible de créer des entrées dans le journal des traces où ce calcul (*focus out* moins *focus in*) est déjà fait. Néanmoins, cette fonctionnalité

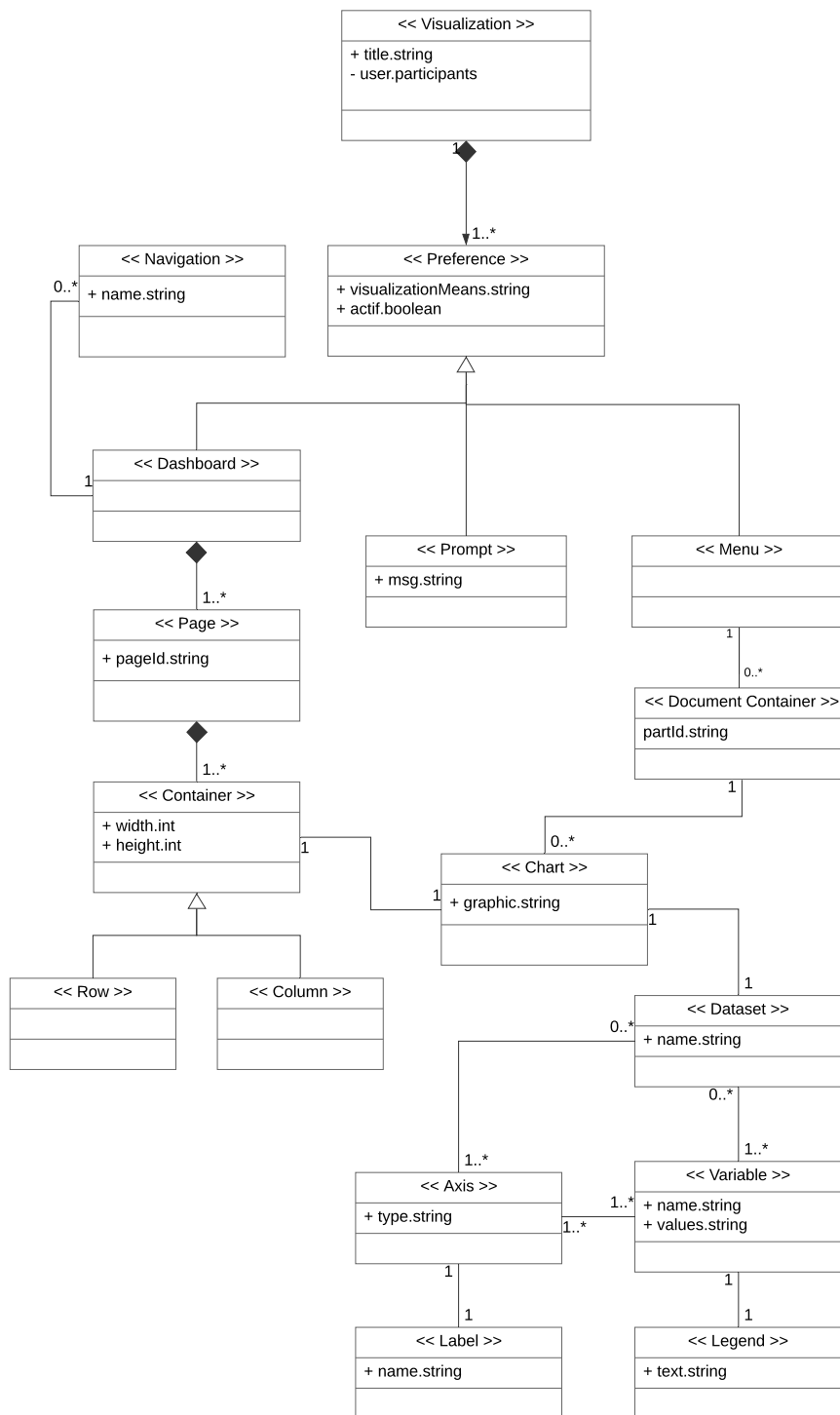


FIGURE 10.3 – Métamodèle : la création des visualisations.

Cet indicateur retourne une liste avec une valeur numérique par ressource concernée. Deux modes de visualisation sont envisagés dans étant dépendante des choix technologiques de chaque navigateur Web, nous optons pour garder cette forme simplifiée dans nos exemples.

ce cas, 1/ une visualisation graphique, avec un histogramme placé sur un tableau de bord et 2/ une visualisation sur la base de la structure du document, via une barre horizontale colorée (le plus foncé, le plus de temps passé) à côté de chaque entrée du menu. La modélisation de cet indicateur à partir du métamodèle est présentée dans la Figure 10.4.

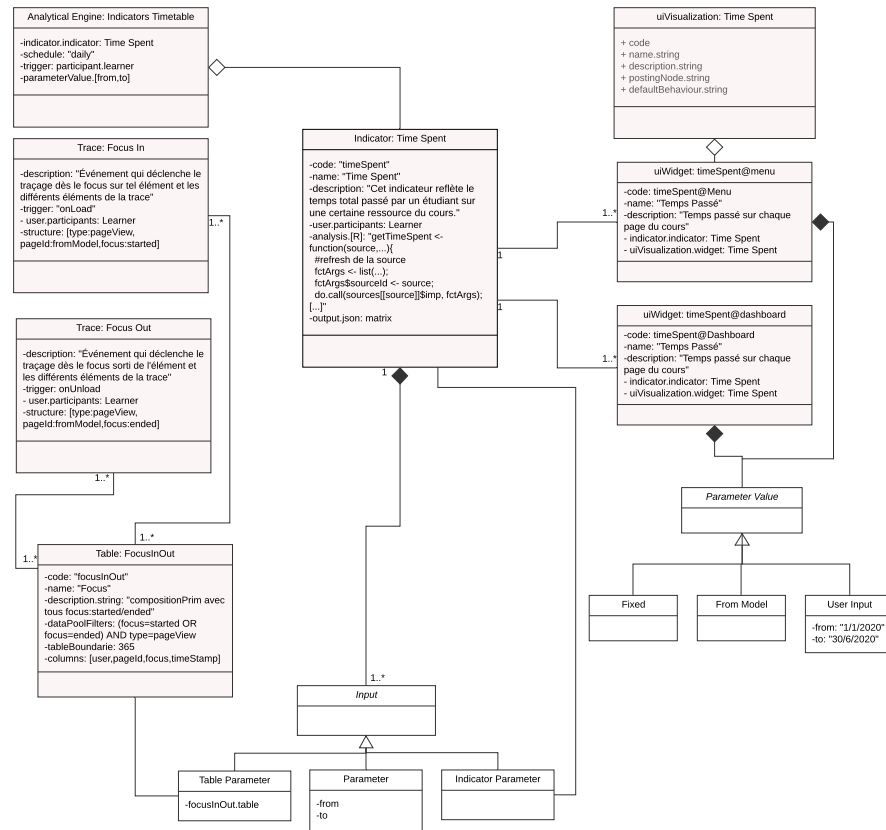


FIGURE 10.4 – Modélisation de l'indicateur : temps passé.

Concernant l'usage envisagé, l'objectif est de comprendre comment le temps passé sur chaque partie d'un cours peut affecter l'apprentissage. Ainsi, il peut être exploité comme un outil d'autorégulation de l'apprentissage concernant le temps passé versus le temps estimé. Il peut être modélisé à l'intention des apprenants comme des instructeurs. L'indicateur peut être utilisé, par exemple, au long du cours — comme mentionné, pour de l'autorégulation — ou encore à sa fin, par l'instructeur, à des fins de réflexion/analyse qui pourraient mener à une refonte du cours ou d'une partie de celui-ci.

En ce qui concerne l'ambiguïté d'interprétation de l'information, cet indicateur doit être utilisé avec précaution, car il ne représente que le temps passé en ligne, l'apprenant aurait pu télécharger le matériel afin de l'imprimer et ce temps ne serait pas pris en compte. Ainsi, le

type de dispositif pédagogique en place pour lequel cet indicateur est proposé est important et doit être pris en compte.

10.3.1.2 Séquence de l'implémentation de l'indicateur « Temps passé »

La Figure 10.5 nous montre le diagramme des phases et acteurs concernés par l'implémentation et l'utilisation de l'indicateur « Temps passé ».

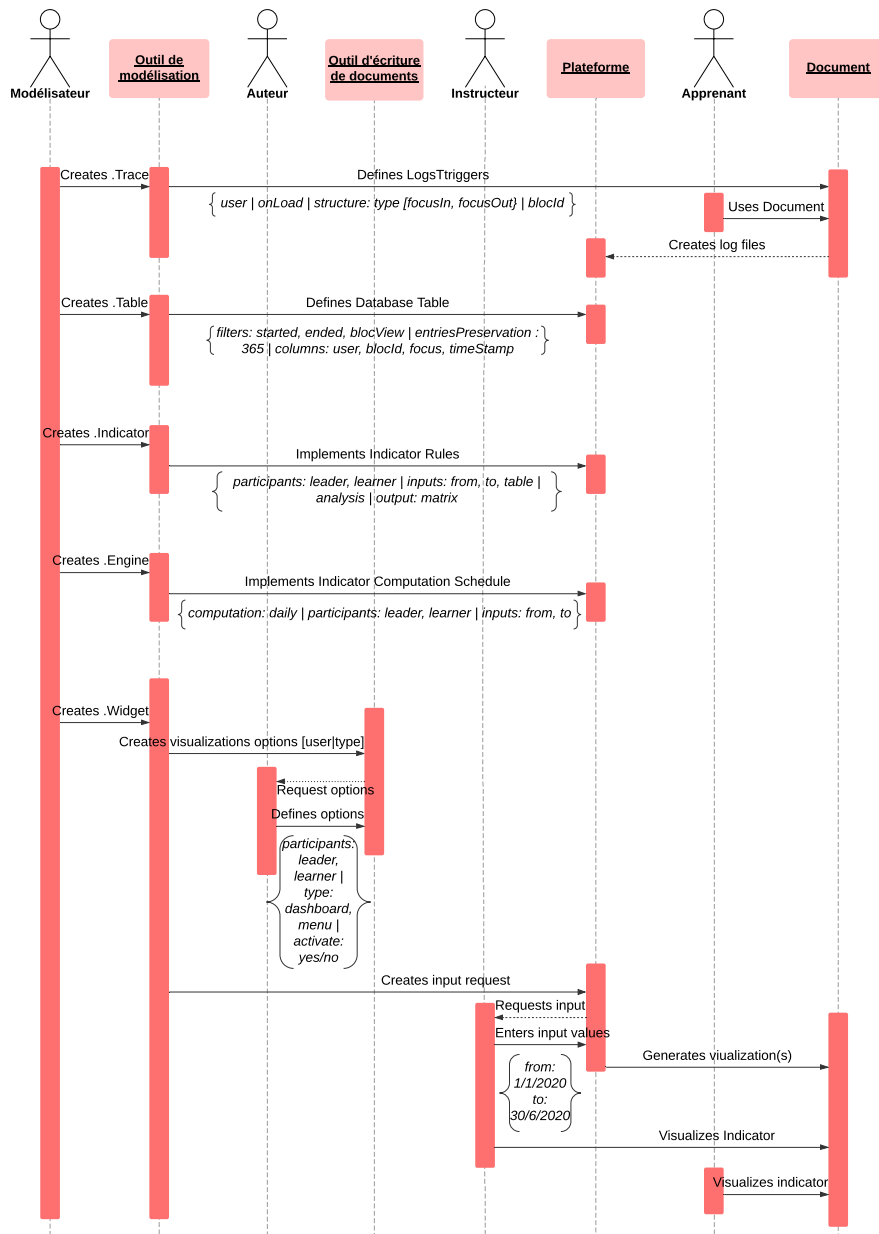


FIGURE 10.5 – Indicateur : le temps passé par un apprenant sur une partie du contenu.

Le modélisateur commence par s'assurer que les traces nécessaires au calcul de l'indicateur sont créées. Pour ce faire, il définit les évé-

nements (*log triggers*) qui entraîneront l'ajout d'une ligne au journal de traces et les informations qui seront inscrites sur cette ligne. Ici, il crée deux déclencheurs : lorsqu'un bloc de contenu est visible sur le navigateur [*focus in*] et lorsqu'il cesse de l'être [*focus out*].

Il crée ensuite la structure de table nécessaire pour calculer facilement l'indicateur (et ne pas lire les journaux de traces à chaque fois). Il définit les filtres qui permettront d'identifier les lignes sur les journaux de traces et comment ils sont utilisés pour remplir la table correspondante, ainsi que le planning de cette action.

Le modélisateur crée alors l'indicateur, définissant les participants qui l'utiliseront. Il définit également les entrées nécessaires, dont l'événement, l'analyse (code) et le format de sortie.

À l'aide du moteur d'indicateurs (commun), le modélisateur peut définir un planning de traitement de l'indicateur, ici il est traité quotidiennement, pour les apprenants et les instructeurs.

Les différentes options de visualisation sont créées à l'aide d'un *widget*. Chaque *widget* peut définir un type de visualisation parmi le tableau de bord, le menu ou un message (prompt). Ici le modélisateur définit une visualisation sur le tableau de bord et une sur le menu. Comme défini précédemment sur le *widget*, les entrées pour les dates sont insérées par l'instructeur. Après sa saisie, toutes les visualisations sont générées pour tous les utilisateurs concernés. Enfin, les instructeurs et apprenants peuvent consulter les résultats.

10.3.2 *Indicateur : Regroupement*

Le deuxième indicateur modélisé, « Regroupement », vise à trouver des groupes d'apprenants dans un contexte d'apprentissage en fonction des comportements liés à l'utilisation des ressources disponibles. Il s'agit d'un indicateur considéré comme de niveau moyen, car il utilise d'autres indicateurs comme entrées.

10.3.2.1 *Description de l'indicateur « Regroupement d'apprenants »*

L'indicateur regroupe les apprenants en fonction de l'utilisation des documents. Dans cet exemple, d'autres indicateurs servent comme entrée à un algorithme de regroupement. De cette façon, l'indicateur repose sur la combinaison des deux indicateurs déjà calculés à partir des journaux de données. Il fait un diagnostic à partir de l'usage des documents et regroupe les apprenants selon ces comportements.

Pour ce faire, un algorithme de regroupement sera utilisé. Il est possible qu'un travail préalable soit nécessaire afin de déterminer quel algorithme convient au mieux par rapport aux données qui seront

utilisées comme entrée (les *features*) et qui résultent dans des groupes plus homogènes et préférablement plus explicables. Ceci pourrait se faire à partir de données d'une promotion antérieure ou encore à partir de données partielles (avant la fin d'un cours, par exemple, le contexte et les objectifs détermineront les choix/étapes dans des contextes pratiques). Les parties prenantes et modélisateurs pourront s'appuyer sur des travaux comme celui de Bharara et al. [18], par exemple, qui liste les différentes possibilités en termes d'algorithmes qui pourraient être appliqués ainsi que les étapes nécessaires, que nous reprenons ici :

1. Collecter les données.
2. Analyser chaque fonctionnalité et son implication, afin de mieux comprendre les fonctionnalités à sélectionner.
3. Effectuer le prétraitement des données, ce qui comprend essentiellement la normalisation de certains attributs et également la réduction de la dimensionnalité.
4. Éventuel nettoyage des données après l'étape de prétraitement.
5. Sélectionner les fonctionnalités pour définir les données d'entrée, supprimer les données inappropriées et trouver une nouvelle catégorie de fonctionnalités.
6. Appliquer l'algorithme choisi sur les fonctionnalités sélectionnées afin de trouver des groupes d'apprenants avec des comportements similaires.

La visualisation prévue inclut un graphique du type nuage de points positionnant chaque participant dans les groupes, avec une éventuelle description de ces groupes (les *clusters* trouvés) et est à l'intention des instructeurs. La modélisation de cet indicateur à partir du métamodèle est présentée dans la Figure 10.6.

L'objectif est d'aider les instructeurs à comprendre comment la quantité d'interactions et temps passé sur les ressources du système peuvent collaborer à définir des actions ciblées.

10.3.2.2 Séquence de l'implémentation de l'indicateur « Regroupement d'apprenants »

La Figure 10.7 nous montre le diagramme des phases et acteurs concernés par l'implémentation et utilisation de l'indicateur « Regroupement d'apprenants ».

Le modélisateur commence par s'assurer que les entrées nécessaires au calcul de l'indicateur sont créées. Il n'a pas besoin de définir les événements (*log triggers*) qui entraîneront l'ajout d'une ligne au journal

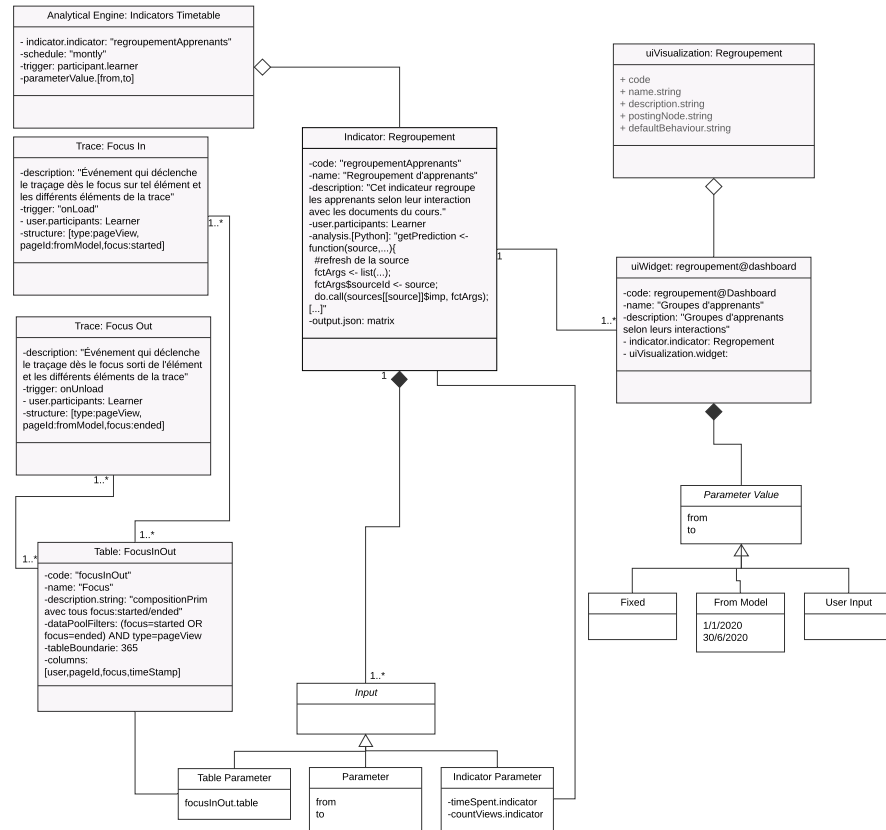


FIGURE 10.6 – Modélisation de l'indicateur : regroupement.

de traces, car ils sont déjà définis par les indicateurs qui servent comme entrée, après calcul, à l'indicateur en question. Il définit alors juste ces indicateurs comme entrées (à noter que dans la Figure 10.7 nous avons gardé les étapes nécessaires aux indicateurs qui servent comme entrée, au cas où ils n'étaient pas encore modélisés). Il définit également que les entrées nécessaires (ici dates de début et fin à tenir en compte pour les calculs) proviennent de l'étape de déploiement documentaire (ici la période d'activité du module) dans la plateforme.

Le modélisateur créera alors l'indicateur, définissant les participants pour qui cet indicateur sera proposé. Il définit également l'analyse (code) comprenant les étapes que nous venons de décrire et le format de sortie.

À l'aide du moteur d'indicateurs (commun), le modélisateur peut définir un planning de traitement de l'indicateur, ici il est traité une fois par mois, pour les instructeurs.

L'option de visualisation *widget* définie est celle du tableau de bord, avec un graphique d'un nuage de points. Les entrées pour les dates sont définies dans le modèle. Ensuite, toutes les visualisations sont générées pour tous les utilisateurs concernés, dans ce cas l'instructeur.

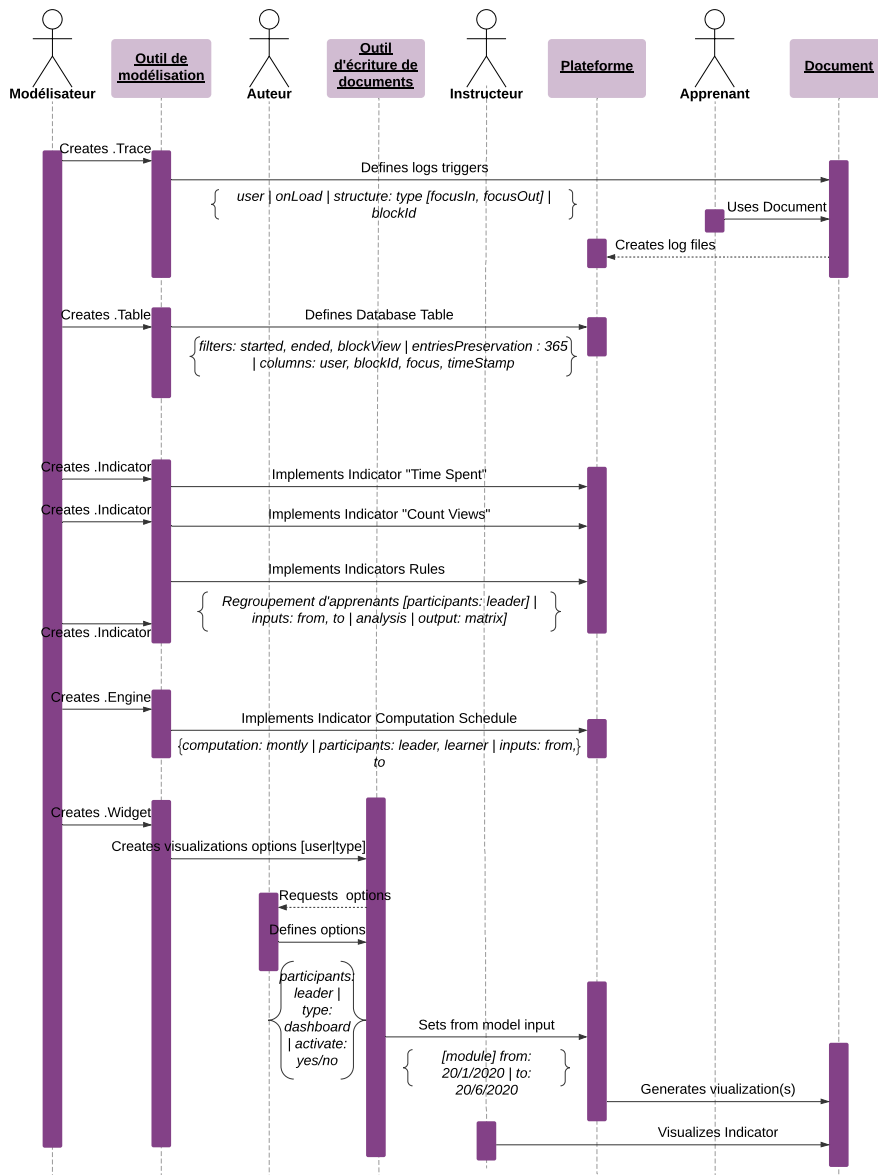


FIGURE 10.7 – Indicateur : le regroupement d’apprenants selon leurs interactions avec le contenu.

10.3.3 Indicateur : Prédiction de réussite

Cet indicateur vise à prédire les performances des apprenants en fonction de différents comportements. Il se concentre sur l’impact de l’interaction des apprenants avec le système d’apprentissage en ligne et vise à prédire les performances en fonction de ces caractéristiques comportementales de l’apprenant.

10.3.3.1 Description de l'indicateur « Prédiction de réussite »

L'indicateur repose sur la combinaison des journaux de données et des indicateurs déjà disponibles : « Nombre de visualisations » et « Temps passé » et également de la réussite. À noter que ces dernières données peuvent être extérieures au système.

Nous classons cet indicateur comme de haut niveau, parce qu'il utilise comme entrée d'autres indicateurs liés aux interactions avec le système. Son objectif est de faire la lumière sur le comportement d'interaction avec le contenu de l'apprenant lors des processus d'apprentissage et son impact sur la réussite des apprenants. Les résultats de l'analyse peuvent soutenir la prise de décision de l'instructeur, qui va regarder la prédiction de réussite et prendre la décision de comment encourager les apprenants à risque à s'appropriier du contenu.

Il s'agit d'un indicateur dont l'analyse est classée comme prédictive, à travers l'analyse de ces comportements d'interaction, le système cherche à dire la réussite ou non de l'apprenant à la fin du cours.

Cet indicateur représente ce que l'on pourrait appeler un indicateur dont les choix d'analyse dépendent des données d'entrée (cf. par exemple [103]). En d'autres mots, le choix de l'algorithme et comment il sera calibré est dépendant des données elle-mêmes, car les performances peuvent varier. Un exemple des étapes de calcul de ce type d'indicateur est donné dans [7], qui utilisent des réseaux de neurones artificiels, des arbres de décision et la classification naïve bayésienne. Ensuite, des méthodes d'ensemble (Bagging, Boosting et Random Forest (RF)) sont appliquées pour améliorer les performances de ces classificateurs. Les étapes suivies sont [7] :

1. Collecter et nettoyer les données.
2. Procéder à la normalisation des données.
3. Faire la discrétisation des valeurs.
4. Sélectionner les fonctionnalités de calibrage disponibles.
5. Réaliser la classification.
6. Effectuer les tests et la validation des résultats.

Suite à ces étapes, l'algorithme ayant le meilleur résultat permettant une prédiction plus précise (de la majorité des cas) est choisi. Le résultat peut, par exemple, être expliqué par un arbre de décision.

Spécifiquement pour ces cas, une option dans le métamodèle permet au modélisateur d'insérer l'arbre de décision résultant : il s'agit d'éviter, dans un premier temps au moins, de devoir avoir une multitude d'algorithmes — et leurs différentes configurations de calibrage — implémentés dans le système. L'idée dans ces cas est alors de permettre d'effectuer ces analyses et tests dans des outils dédiés (comme

ceux cités auparavant, cf. la discussion du Chapitre 5, Partie 5.6) et d'ajouter les résultats dans la partie dédiée à l'analyse du métamodèle en indiquant bien qu'il s'agit d'un "data model", qui est suffisant pour la classification des données en cours et création de la visualisation de l'indicateur. La modélisation de cet indicateur à partir du métamodèle est présentée dans la Figure 10.8.

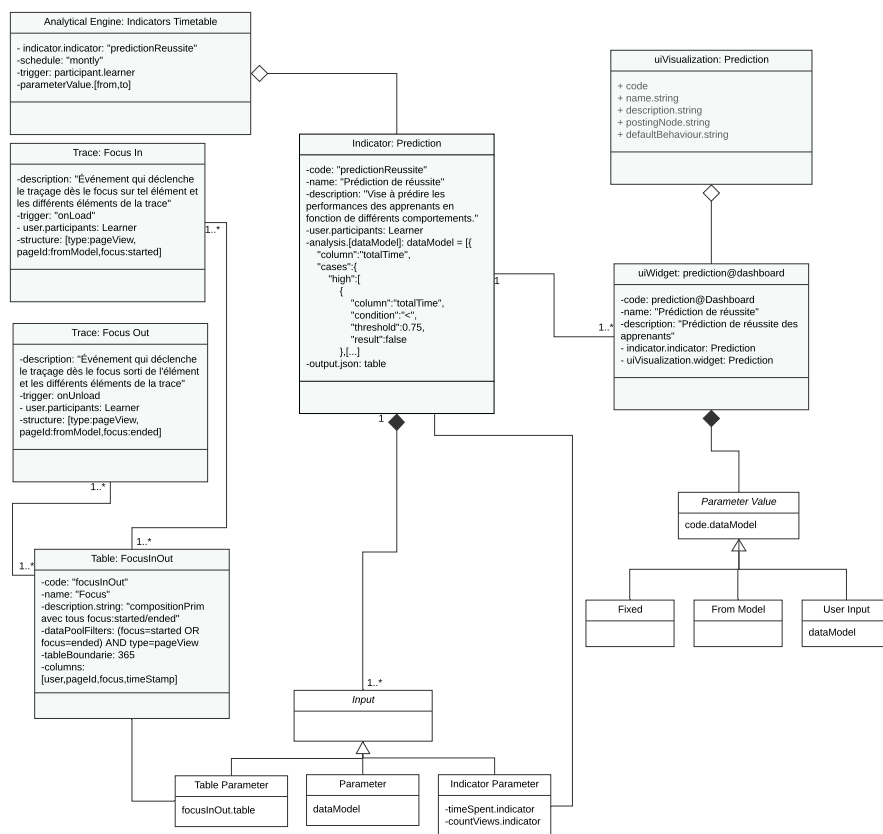


FIGURE 10.8 – Modélisation de l'indicateur : prédiction de réussite.

À partir des deux résultats de deux valeurs numériques (ID de l'apprenant et risque ou non de ne pas réussir), non ordonnées, une visualisation des résultats peut se faire, par exemple, via un système de couleurs (rouge/jaune/vert).

Dans cet exemple, l'indicateur est utilisé pendant le cours, mais en se basant sur une analyse faite avec les données des cohortes précédentes pour construire le modèle de données.

10.3.3.2 Séquence de l'implémentation de l'indicateur « Prédiction de réussite »

La modélisation de cet indicateur se déroule comme suit : le modélisateur s'assure d'avoir les tables (ainsi que les traces, analyse, etc.) correspondantes aux deux indicateurs qui seront utilisés comme en-

trées à l'indicateur en question. Ceci correspond aux six premières actions dans la Figure 10.9. Le modélisateur définit également le type de visualisation envisagée.

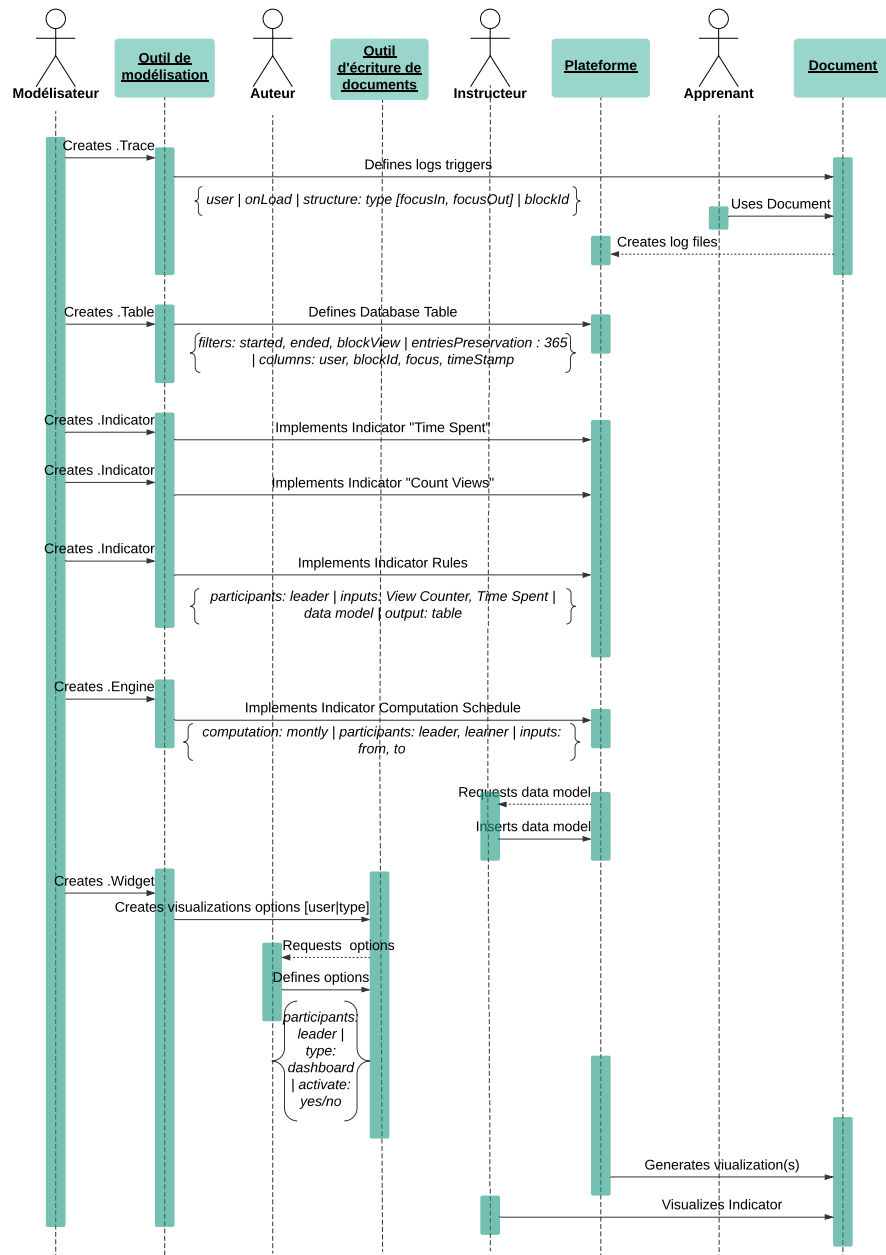


FIGURE 10.9 – Indicateur : prédiction de réussite.

Ensuite, à l'aide de ces tables, la définition du modèle de données "data model" est réalisée, dans un outil dédié, comme mentionné auparavant. Les données nécessaires (de la promotion précédente) sont alors téléchargées (ou un accès est donné à l'outil, selon les possibilités technologiques choisies) et complétées par les données de réussite, qui peuvent être disponibles dans le système ou dans un autre outil

d'administration, par exemple. Une fois les étapes décrites ci-dessous et le modèle de données défini, la plateforme demandera l'insertion de ce modèle de données à l'acteur défini dans la modélisation (ici l'instructeur). Par la suite, le calcul de l'indicateur est réalisé et la visualisation peut être faite et consultée par l'instructeur.

10.3.4 *Indicateur : Auto-explication*

Le dernier indicateur sur lequel nous nous intéressons est celui dénommé « Auto-explication ». Cet indicateur se concentre sur la reconnaissance de points potentiels dans le temps pour pratiquer la technique d'auto-explication qui consiste à expliquer comment de nouvelles informations sont liées à des informations connues, ou à expliquer les étapes prises lors de la résolution de problèmes. Sur la base du contenu consulté par les étudiants, des messages juste à temps vont inciter les apprenants à adopter une telle pratique.

10.3.4.1 *Description de l'indicateur « Auto-explication »*

Cet indicateur repose sur les études se concentrant sur les techniques d'apprentissage (humain) efficaces et montre comment le métamodèle et notre contexte pourraient amener à la création d'indicateurs qui encouragent l'adoption de telles techniques. Cet indicateur s'inspire également d'une attente de la part d'apprenants dans les travaux de [124], où un souhait que le système guide leurs pauses en analysant leur productivité et en suggérant quand il est temps de faire une pause a été exprimé.

Nous nous basons sur le travail de [47] où 10 techniques d'apprentissage ont été analysées et discutées, avec la proposition de recommandations sur l'utilité relative de chacune. Les techniques comprennent l'interrogation élaborée, l'auto-explication, la synthèse, la mise en évidence (ou le soulignement), le mot-clé mnémonique, l'utilisation de l'imagerie pour l'apprentissage de texte, la relecture, les tests pratiques (évaluation formative), la pratique distribuée et la pratique entrelacée. Afin de proposer des recommandations sur l'utilité relative de chacune de ces techniques, les auteurs ont évalué si leurs avantages se généralisaient à quatre catégories de variables : les conditions d'apprentissage, les caractéristiques des élèves, le matériel et les tâches-critères.

Parmi les recommandations de ces auteurs [47], deux techniques sont classées comme ayant une forte utilité (les tests pratiques et la pratique distribuée). Trois autres sont classées comme d'utilité moyenne : l'interrogation élaborée, l'auto-explication et la pratique entrelacée.

Nous nous concentrons sur la technique d'auto-explication pour la création de cet indicateur en particulier, mais d'autres, notamment l'interrogation élaborée, pourraient également être envisagés, ayant une modélisation et des objectifs très semblables.

La technique d'auto-explication est définie comme suit : « *Explaining how new information is related to known information, or explaining steps taken during problem solving.* » [47, p. 6]. Nous songeons à la création d'un indicateur qui, à partir des traces permet de savoir que l'apprenant a parcouru une partie théorique du contenu (une série de parties du contenu étiquetée comme des concepts, par exemple) et ainsi lui proposer de s'entraîner à faire une auto-explication (ou une explication à des pairs) du contenu en question ; au lieu de simplement relire le contenu, pratique courante, mais moins efficace pour l'apprentissage.

Cet indicateur repose fortement sur les fichiers journaux contenant l'information relative à la sémantique des documents consultés. Il peut être classé comme une analyse permettant de faire de la recommandation, même si son calcul est simple. Nous le classons comme ayant un niveau moyen, dû au fait qu'il repose sur des données enrichies par la sémantique et que son calcul doit se (ré)faire en temps réel. La modélisation de cet indicateur à partir du métamodèle est présentée dans la Figure 10.10.

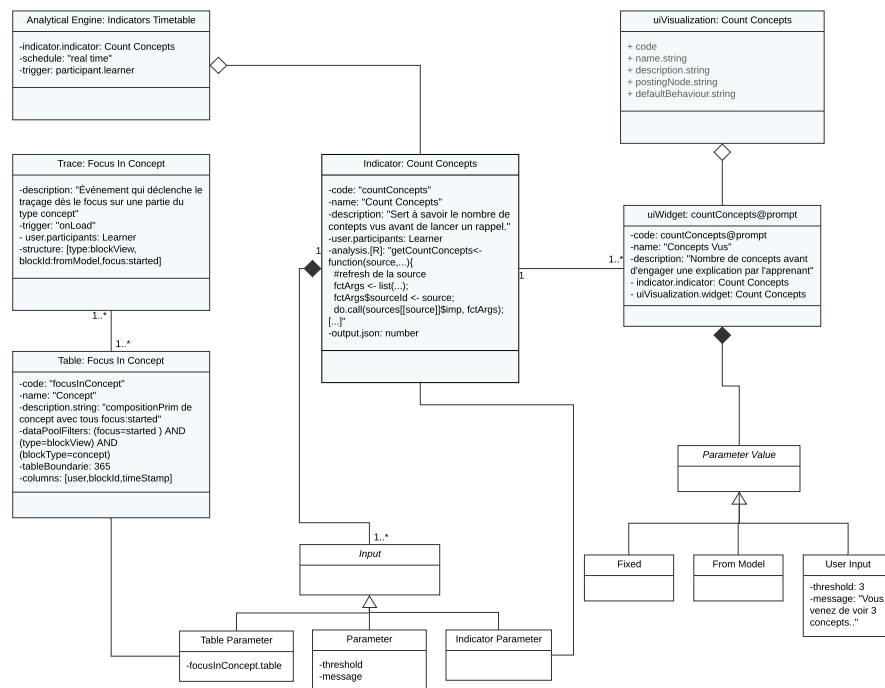


FIGURE 10.10 – Modélisation de l'indicateur : auto-explication.

Une fois qu'il est modélisé, les étapes liées à l'utilisation de l'indicateur sont les suivantes :

1. Identification en temps réel des séquences de contenu étiquetées comme « concept » vues par l'utilisateur.
2. À partir du calcul de la somme de ces types de visualisation, vérification que le seuil (prédéterminé) a été atteint.
3. Un message est déclenché encourageant l'utilisateur à s'arrêter et à pratiquer la technique en question (expliquer les X derniers concepts vus avec ses propres mots, en l'associant avec d'autres informations, etc.).
4. En option, la réponse de l'utilisateur (l'avez-vous fait ou non) peut être demandée (ce qui peut servir également pour une analyse postérieure de l'efficacité de l'indicateur).

Comme pour tous les autres indicateurs, les usagers peuvent activer ou désactiver son utilisation.

Cet indicateur nous semble un exemple intéressant, car il 1/ repose sur le bénéfice d'avoir l'information relative à la sémantique du document consulté, 2/ est basé sur une théorie/technique éducative étudiée auparavant, 3/ permet d'encourager l'adoption de cette technique au bon moment (indépendamment du cheminement choisi par l'apprenant, surtout lors de l'utilisation de modèles documentaires non-linéaires). L'adoption effective de changement de comportement aurait ainsi tendance à se faire plus facilement que seulement via une explication théorique des techniques plus efficaces à utiliser lors de ses sessions d'études.

Bien entendu, l'indicateur est à l'intention des apprenants et a vocation à être utilisé pendant le cours. Idéalement, les apprenants devraient être informés de cette technique d'apprentissage humain. Cet indicateur a le potentiel de faire l'objet d'études ultérieures sur le changement ou non de leur comportement. L'apprenant utilise les messages comme des rappels opportuns pour adopter une technique d'apprentissage humain censée être plus efficace, ces rappels sont adaptés au chemin que suit effectivement l'apprenant.

10.3.4.2 Séquence de l'implémentation de l'indicateur « Auto-explication »

Comme indiqué dans la Figure 10.11, le modélisateur définit la création des traces selon la visualisation des pages ou parties du contenu qui sont du type « concept ». La table correspondante est créée via le filtre des journaux de traces ne contenant que ce type de bloc.

La primitive définissant l'indicateur doit permettre de réaliser la somme du nombre de visualisations de ce type.

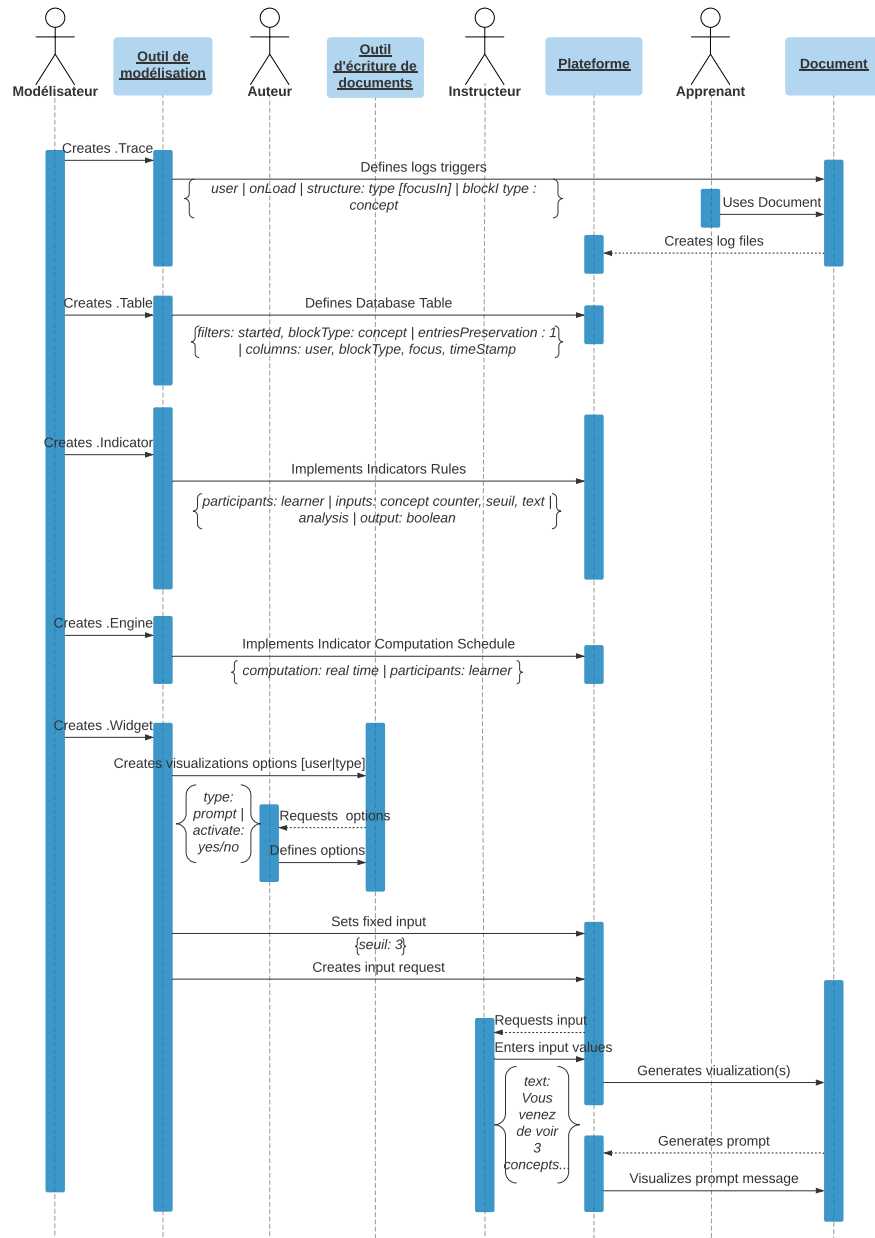


FIGURE 10.11 – Indicateur : auto-explication.

Le *widget* correspondant aux options de visualisation demande à l'acteur correspondant, ici l'instructeur, un seuil pour que le message soit déclenché. Dans notre exemple, l'instructeur définit qu'après une visualisation de trois concepts, le message incitant à une pause pour auto-explication de ces derniers soit affiché. Une fois le seuil atteint, la génération de ce message se fait.

10.4 DISCUSSION

Dans ce Chapitre, nous avons détaillé le métamodèle proposé, sa structure, ses composants et les relations entre eux. Nous avons présenté ensuite le détail de l'utilisation de ce métamodèle via la modélisation de quatre indicateurs, chacun basé sur un type d'analyse (descriptive, diagnostique, prédictive, prescriptive) relevée par l'état de l'art (cf. Chapitre 5). Ceci nous a permis de signaler, via des séquences de l'implémentation, les actions nécessaires de la part de chacune des parties prenantes, ainsi que les raisons de décisions prises concernant la proposition du métamodèle (niveau d'abstraction, composants, sources des paramètres, etc.). Dans le Chapitre qui suit, nous examinons quelques métriques et fonctionnalités qui découlent de l'utilisation du métamodèle proposé et dans le Chapitre 12 nous étudions la prise en main du métamodèle par des modélisateurs.

11.1 INTRODUCTION

Ce Chapitre est consacré à mettre en lumière les aspects fonctionnels particuliers qui découlent de l'utilisation du métamodèle proposé lors de son utilisation pour mettre en place une démarche d'analytique des apprentissages avec le numérique et certaines métriques qui pourraient représenter un avantage de l'utilisation de celui-ci.

11.2 MÉTRIQUES ET FONCTIONNALITÉS

11.2.1 *Analyse détaillée d'un document*

Un premier aspect à considérer est que, en utilisant une approche basée sur l'IDM pour produire à la fois le document lui-même et l'indicateur, nous pouvons créer des indicateurs enrichis (ou ajouter des détails à un existant) en tenant compte (dans les traces elles-mêmes) de la sémantique du document ou bien de sa structure. Ce fait permet, par exemple, de créer facilement les indicateurs donnés en exemple dans le Chapitre 10 où chaque type de parties de contenu peut être étudié séparément, car les informations (type de bloc utilisé) peuvent être ajoutées aux traces que l'on souhaite collecter en quelques clics, reliant la primitive documentaire à la primitive d'analytique des apprentissages avec le numérique. Alors que dans une approche classique, il faudrait baliser *a posteriori* chaque partie (en partant du principe que les documents sont structurés de manière consistante), ce qui devrait être fait presque manuellement, en liant chaque identifiant de partie à un type de contenu. De plus, toute modification du contenu du cours nécessiterait soit de refaire ce balisage, soit il serait de la responsabilité de l'instructeur de s'assurer que le balisage reste valide — une tâche supplémentaire sujette aux erreurs pour laquelle il n'est pas un expert. La facilité de mise en œuvre de ce niveau d'analyse détaillée, intégrée au système, peut certainement être considérée comme un avantage lorsque l'on considère les indicateurs destinés à la refonte des cours. De plus, l'utilisation de ces informations détaillées comme données d'entrée lors de l'ingénierie des fonctionnalités (*features*) pour des indicateurs plus complexes tels que la prédiction, l'analyse de

séquence, entre autres, est une voie prometteuse qui pourrait produire des résultats plus précis.

Ainsi, grâce à l'approche utilisée, les détails (sémantique et structure) des documents peuvent enrichir les indicateurs et ceci de manière automatique, contrairement à une approche traditionnelle, où ces informations auraient dû être ajoutées *a posteriori* et au cas par cas. Par conséquent, le métamodèle proposé permet de faire une seule fois par indicateur et en amont.

11.2.2 *Séparation du contenu et de la forme de visualisation*

Comme nous l'avons remarqué dans la revue systématique de la littérature présentée dans le 9, habituellement les indicateurs sont implémentés en utilisant un graphisme spécifique comme résultat souhaité de visualisation, souvent disponible pour consultation dans un tableau de bord.

Disposer des primitives spécifiques d'analytique des apprentissages avec le numérique pour le calcul et la visualisation d'un indicateur présente l'avantage de séparer le contenu et la forme, au même titre que les documents créés en suivant la même approche. De ce fait, nous pouvons imaginer des situations où le même résultat de calcul peut être utilisé dans un tableau de bord, destiné au formateur, et à côté du menu du document pour les apprenants, comme envisagé et préféré lors de l'étude exploratoire dont les détails sont analysés dans le Chapitre 8. Concernant l'exemple d'un indicateur relevant du temps passé dans chaque partie : le premier ira directement à un tableau de bord pour voir quelles parties prennent le plus de temps pour les apprenants, tandis que le second, qui se fera dans l'arborescence/menu du cours, pourra voir le temps passé sans avoir à arrêter de consulter le contenu et d'aller vérifier un tableau de bord, maintenant ainsi le flux d'une session d'étude.

11.2.3 *Interopérabilité*

L'interopérabilité d'un système d'analytique des apprentissages avec le numérique assure la compatibilité de tout type d'environnement d'apprentissage virtuel en permettant l'intégration de différentes sources de données [48]. Plus largement, selon l'ingénierie des systèmes, il s'agit d'une caractéristique d'un système lui permettant de fonctionner avec d'autres produits ou systèmes, actuels ou futurs, que ce soit dans sa mise en œuvre ou en accès.

Comme indiqué ci-dessus, les traces produites à l'aide de l'approche proposée ont le potentiel d'être enrichies de détails de documents. Cependant, si certaines situations l'exigent — cela peut être nécessaire lorsque les traces seront analysées par un outil externe —, il est également possible de créer de telles traces en utilisant n'importe quelle norme donnée (par exemple, `xAPI`). Cela peut augmenter l'interopérabilité du système, mais diminue l'avantage de travailler avec des traces enrichies, car les résultats ne pourront pas se baser sur cette information détaillée.

Un même indicateur peut être visualisé dans plusieurs déploiements de la plateforme. Cependant, pour que les traces produites puissent être traitées par un système extérieur, elles doivent être conçues en suivant le standard en question et sans les enrichissements cités.

11.2.4 *Conformité à la confidentialité des données*

Le respect des différentes lois sur la confidentialité des données est certainement un point de plus en plus important concernant les exigences innées d'un processus d'analyse de données personnelles. Les données collectées ainsi que leur accès et leur traitement doivent être soigneusement documentés.

Afin de justifier le traçage, pour chaque donnée, la documentation doit déterminer et justifier pourquoi ce traçage est effectué, qui y a accès, les utilisations et traitements qui en résultent, etc. De plus, le consentement explicite des utilisateurs est requis.

Ce point a été brièvement mentionné auparavant, car une différence clé dans l'approche discutée ici est le fait que les traces sont collectées, car elles vont être utilisées lors du calcul d'un certain indicateur qui a été modélisé comme requis par une partie prenante. Il s'agit d'un changement par rapport à l'approche selon laquelle toutes les actions effectuées dans l'environnement d'apprentissage sont collectées et seules les parties prenantes ultérieures décideront de la manière dont elles pourraient l'utiliser.

Comme les traces sont connectées à des indicateurs, il sera probablement plus facile de créer une documentation conforme aux principes du règlement général sur la protection des données (RGPD) tels que la communication de finalités spécifiques pour les données collectées et les acteurs impliqués dans le traitement des données. L'idée clé est de ne tracer que ce qui est nécessaire pour répondre à des questions prédéfinies, cela est une approche intrinsèquement conforme aux réglementations.

Ainsi, chaque indicateur est lié aux traces nécessaires afin qu'il soit calculé. Cette information pourrait être générée automatiquement afin de préciser aux utilisateurs quelles traces sont enregistrées et pourquoi.

11.2.5 *Maintenabilité*

Un autre avantage fondamental de cette approche est la maintenabilité de la solution proposée. La maintenabilité mesure l'effort nécessaire pour corriger ou transformer le logiciel, mais aussi son extensibilité, c'est-à-dire le peu d'effort nécessaire pour ajouter de nouvelles fonctions. C'est potentiellement là que nous pouvons trouver le plus grand avantage à utiliser l'ingénierie basée sur les modèles afin de créer une implémentation d'analytique des apprentissages avec le numérique.

Compte tenu de notre contexte où de nombreux modèles documentaires et contextes d'utilisation s'appuieront sur la même solution, pouvoir choisir et adapter facilement des indicateurs est essentiel. Ainsi, un modélisateur peut créer un indicateur pour un modèle documentaire spécifique et, le fait qu'il sera potentiellement utilisé par un certain nombre d'utilisateurs dans des contextes différents, c'est-à-dire tout document produit à l'aide du modèle documentaire, à partir d'une source unique (la modélisation) est une grande amélioration de la maintenabilité. De plus, sur le long terme, toutes les modifications ou corrections nécessaires doivent être apportées en un seul code/modélisation même si l'indicateur a été implémenté dans plusieurs instanciations de la plateforme ou dans plusieurs modèles documentaire ou encore dans plusieurs documents utilisant un même modèle documentaire. Cela permettrait de réduire le temps de développement, les ressources et les sources d'erreurs potentielles.

Par conséquent, la source de modélisation d'un indicateur est centralisée, ce qui permet d'avoir des gains en termes de maintenabilité. De plus, via le concept de fragmentation, un nouvel indicateur pourrait bénéficier de certains composants d'un autre indicateur déjà modélisé, avec un gain de temps et une réduction des sources potentielles d'erreurs.

11.2.6 *Performance*

La performance peut être mesurée en fonction du rapport entre la quantité de ressources nécessaires (moyens incluant le personnel, le temps, le matériel, etc.) et les résultats fournis.

Par rapport à l'approche proposée, nous pouvons parler de performance en ce qui concerne la création d'un indicateur utilisé dans de nombreux contextes (tous utilisant le même modèle documentaire), d'une série d'indicateurs — dans un tableau de bord par exemple — ou encore d'une sorte d'industrialisation des processus d'analyse, qui permettrait la création de différents tableaux de bord adaptés aux besoins spécifiques d'une série de modèles documentaires, avec un gain de temps potentiel pour effectuer cette tâche, car elle n'est effectuée qu'une seule fois et répliquée selon ces contextes multiples.

Ainsi, la modélisation d'un indicateur a le potentiel d'être utilisée dans une grande quantité de contextes et documents provenant d'un même modèle documentaire, cette sorte d'industrialisation d'analyse représente un gain potentiel de temps et ressources.

11.2.7 *Multiformat*

Comme nous l'avons vu, dans notre contexte une même ressource pédagogique peut être automatiquement transformée en plusieurs formats comme PDF, Open Document, Web, présentation synthétique, etc. Dans une situation où un apprenant va imprimer un ou deux chapitres d'un cours afin d'étudier pendant ses vacances, nous pouvons imaginer que, sur le PDF résultant, un indicateur — disons le temps qu'il a déjà passé sur chaque partie de chaque chapitre — sera imprimé à côté du contenu. Inutile de dire qu'il n'y aura aucune trace enregistrée de sa session d'étude à partir du PDF résultant, mais toutes les données déjà disponibles au moment de la transformation de la ressource en cours pourraient y être annexées.

11.2.8 *Personnalisation*

La possibilité de personnaliser un indicateur ou un tableau de bord a été étudiée par plusieurs auteurs [44]. Il peut s'agir de proposer des alternatives de filtrage ou des modes de visualisation différents, dont le choix est fait par les utilisateurs en fonction de leurs préférences et des besoins ressentis lors de leurs pratiques. Le fait que chaque modèle documentaire puisse avoir ses indicateurs spécifiques liés à ses usages réels pourrait également être compris comme une forme de personnalisation.

Offrir la possibilité de personnalisation est certes un atout, mais nécessite néanmoins de s'assurer que les utilisateurs comprennent les choix qui s'offrent à eux, ainsi qu'une attention particulière à ne pas submerger l'utilisateur avec un nombre de choix — et donc

de décisions — trop important qui pourrait par conséquent inhiber l'utilisation du système¹.

En utilisant l'approche proposée, il est possible de personnaliser un indicateur en proposant plus d'une option, comme une visualisation à l'intérieur d'un tableau de bord ou à côté du menu, permettant aux utilisateurs de choisir celle qu'ils préfèrent. Un même indicateur peut être visualisé dans plusieurs formats — tableau de bord, menu, etc. — avec un simple changement dans le *widget* correspondant du métamodèle, ce qui permettrait de le rendre personnalisable en laissant ces deux choix à l'utilisateur. De fait, en partant d'un indicateur modélisé, il est possible de le personnaliser en changeant, en quelques clics, les parties des documents à tracer ou encore les destinataires des visualisations, etc.

11.2.9 Reproductibilité

Un autre aspect important concernant les implémentations de l'analytique des apprentissages avec le numérique est lié à la nécessité de permettre la reproductibilité de tous les indicateurs déjà en place. Selon [90], la reproductibilité est constituée de trois propriétés, à savoir :

Répliquabilité : la capacité à reproduire une analyse identique, sans nécessairement prendre en compte les variables liées au contexte. Plus précisément, dans un processus d'analyse, la répliquabilité est considérée comme la succession ordonnée d'opérateurs d'un tel processus. En effet, les opérateurs impliqués dans ce processus sont clairement identifiés et leur ordre d'application est également bien défini [90]. Un indicateur qui a été modélisé dans un modèle documentaire peut facilement être utilisé (avec ou sans adaptations) dans un autre modèle documentaire, car les opérateurs nécessaires et leur ordre d'application sont connus, seules les modifications apportées aux traces d'entrée (les parties du modèle ne sont pas les mêmes) doivent être adaptées, ainsi que les modes de visualisation, si nécessaire/désiré dans le nouveau contexte.

Répétabilité : propriété d'un processus d'analyse à effectuer plusieurs fois, sur le même ensemble de données et avec les mêmes configurations, ayant les mêmes résultats de sortie. Autrement dit, cette qualité permet de retracer les résultats produits et leur cohérence [90]. La vérification de la qualité de la répétabilité des indicateurs mis en œuvre devrait être facilitée, car la même mise en œuvre serait disponible/testée dans divers contextes pratiques.

1. https://en.wikipedia.org/wiki/The_Paradox_of_Choice

Réutilisabilité : capacité d'un logiciel (ou d'un processus d'analyse) à être réutilisé dans une application autre que celle pour laquelle il a été conçu ou dans un autre contexte. Le même processus doit ensuite être facilement réutilisé sur un autre jeu de données plus ou moins similaire au jeu initial. Ainsi, d'éventuelles adaptations peuvent être apportées afin d'adapter l'analyse à un contexte similaire, par exemple, à un autre modèle documentaire, mais avec une attention particulière pour garantir que le fondement du processus d'analyse (un apprentissage, une théorie à partir de laquelle l'indicateur a été envisagé, par exemple) est respecté et reste cohérent.

Ainsi, un indicateur qui a été modélisé dans un modèle documentaire peut facilement être utilisé (avec ou sans adaptations) dans un autre modèle documentaire, car les opérateurs nécessaires et leur ordre d'application sont connus, seules les entrées seraient potentiellement à changer.

11.3 DISCUSSION

Dans ce Chapitre, nous avons analysé les différentes métriques et fonctionnalités liées à l'utilisation du métamodèle proposé et qui représenteraient un avantage dans certaines situations et contextes de son implémentation par rapport à un système plus traditionnel.

La majorité de ces aspects sont liés au fait d'avoir adopté une démarche inspirée par l'ingénierie dirigée par les modèles dans un contexte de production documentaire qui suit également cette approche. Le domaine de l'analytique des apprentissages avec le numérique y a été intégré via l'analyse et l'abstraction de ses éléments et par conséquent la proposition des primitives nécessaires à une modélisation efficace — sans pour autant avoir la prétention de créer un langage dédié (DSL, *domain-specific language*) pour répondre aux contraintes des analyses du domaine d'application du *big data*.

Cela a permis de concevoir un métamodèle permettant de définir les exigences liées à chaque indicateur dans le but de les traiter et générer des résultats qui prennent en charge les différentes tâches nécessaires pour atteindre les objectifs d'information des utilisateurs de manière objective.

Une étude qualitative a été menée avec les deux modélisateurs expérimentés qui s'occupent quotidiennement de la création de modèles documentaires. Les objectifs, étapes et résultats de cette étude sont décrits dans ce Chapitre.

12.1 INTRODUCTION

Une fois les primitives du métamodèle implémentées dans l'outil dédié aux modélisations (un exemple est donné dans la Figure 12.1), nous avons cherché à vérifier la bonne compréhension et prise en main de la part de ses futurs utilisateurs.

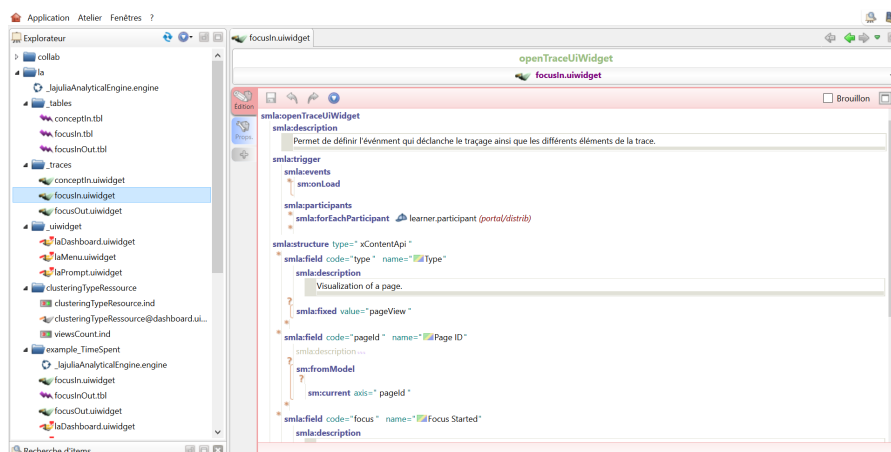


FIGURE 12.1 – Métamodèle : implémentation dans l'outil dédié.

12.2 ÉVALUATION DU MÉTAMODÈLE PAR DES MODÉLISATEURS

Afin d'évaluer le métamodèle nous avons choisi de le faire auprès de ses futurs utilisateurs, avec une prise en main via la modélisation d'un indicateur. À noter que les différentes transformations ne sont pas encore implémentées, donc l'étude se concentre sur la prise en main du métamodèle.

12.2.1 *Énoncé du problème*

Il s'agit de la volonté de comprendre la prise en main des modélisateurs du métamodèle proposé pour l'analytique des apprentissages avec le numérique, tout en se focalisant sur comment ils s'y prennent pour créer des indicateurs à l'aide des primitives proposées.

12.2.1.1 *But de l'étude*

Le but est celui de comprendre plusieurs aspects — utilité, avantages/inconvénients, compréhension, points de blocage, etc. — de l'utilisation du métamodèle par les modélisateurs expérimentés. Nous nous sommes intéressés à observer notamment :

- Comment se passe la prise en main : quelles actions sont prises, avec quel but et dans quel ordre (prévu vs action réelle).
- Les points de blocage ou d'amélioration : tout ce qui pourrait être un point de blocage ou des éventuelles améliorations lors de la modélisation d'un indicateur.
- Leurs perceptions sur plusieurs aspects : les avantages/inconvénients de notre approche.

12.2.1.2 *Questions de recherche*

Les questions auxquelles nous cherchons à répondre avec cette étude sont :

1. Comment se passe la prise en main du métamodèle par des modélisateurs ?
2. Les primitives sont-elles utilisées comme prévu ?
3. Quels sont les points de blocage lors de cette prise en main ?
4. Y a-t-il des améliorations à prévoir, lesquelles ?
5. Le vocabulaire utilisé est-il clair pour un modélisateur ?
6. Quelle est l'utilisabilité perçue du méta-modèle par des modélisateurs ?
7. Comment les différentes primitives proposées sont utilisées afin d'encourager la maintenabilité, la réutilisation, la personnalisation des indicateurs ?

12.2.1.3 *Délimitations et limitations*

Les modélisateurs ne sont pas forcément des experts en l'analytique des apprentissages avec le numérique, même s'ils ont une excellente

connaissance du métier de la modélisation documentaire. La première phase de l'étude doit ainsi être conduite avec soin afin de les accompagner à appréhender ces nouveaux concepts et vocabulaire.

12.2.2 *Procédures*

Nous avons défini plusieurs étapes afin de pouvoir présenter le métamodèle, guider la prise en main de celui-ci et évaluer les résultats. Nous nous appuyons particulièrement sur une prise en main exprimée par l'exemplification et l'explicitation. La procédure est décrite en détail ci-dessous.

12.2.2.1 *Approche de recherche qualitative utilisée*

Cette étude se base sur une triangulation comprenant trois phases :

1. Présentation du modèle avec exemplification par un indicateur.
2. Modélisation (par les modélisateurs expérimentés) d'un indicateur en autonomie et réponse à un questionnaire.
3. Entretien d'explicitation portant sur la modélisation d'un indicateur en autonomie.

12.2.2.2 *Rôle du chercheur*

Le rôle du chercheur lors de cette étude est celui d'un guide.

12.2.2.3 *Méthode d'échantillonnage*

Une méthode d'échantillonnage ciblé [117] est utilisée, c'est-à-dire l'échantillonnage par critère — qui consiste à sélectionner des cas qui satisfont à un critère prédéfini — dans notre cas, des modélisateurs expérimentés.

12.2.2.4 *Méthode et techniques : Triangulation en trois phases*

PHASE 1 Lors d'une séance de présentation avec les modélisateurs, le chercheur introduit rapidement l'analytique des apprentissages avec le numérique, les indicateurs (celui qui sert d'exemple et celui à modéliser en autonomie), et ensuite se concentre sur le nouveau métamodèle dédié à l'analytique des apprentissages avec le numérique, son but, ses principes.

Ensuite, les différentes primitives proposées sont démontrées et un indicateur est modélisé à titre d'exemple/prise en main par les chercheurs et participants ensemble.

À la fin de la séance, le chercheur fournit les caractéristiques d'un autre indicateur, qui sera ensuite modélisé en autonomie par les modélisateurs.

Les supports utilisés lors de cette phase sont :

- Formulaire de consentement (cf. Annexe C).
- Métamodèle installé dans l'outil dédié à la modélisation.
- Présentation de l'analytique des apprentissages avec le numérique, du métamodèle dédié et des diagrammes des acteurs et phases pour l'implémentation (cf. Figures 10.5, par exemple).
- Détails de l'indicateur modélisé par les participants et chercheurs ensemble.
- Instructions pour les prochaines étapes : modéliser un indicateur en autonomie, répondre au questionnaire et participer à l'entretien.
- Détails de l'indicateur à être modélisé par les participants en autonomie.

PHASE 2 Les modélisateurs vont modéliser l'indicateur dont les détails ont été fournis lors de la séance de la phase 1, en autonomie. À la fin de la modélisation, ils doivent remplir le questionnaire fourni également, portant sur l'utilisabilité du métamodèle.

Le chercheur reçoit le questionnaire et l'analyse avant l'étape suivante.

À noter que la plupart des supports nécessaires à cette phase ont été fournis au préalable. Le seul support utilisé lors de cette phase est :

- Questionnaire comprenant une première partie sur l'utilisabilité du métamodèle et une deuxième portant sur leur expérience avec la modélisation, via des questions ouvertes, utilisant la technique de complétion de phrases (cf. Annexe D).

La première partie de ce questionnaire a été construite avec l'échelle d'utilisation du système (*System Usability Scale, SUS*), un instrument largement utilisé qui mesure l'utilisabilité subjective de produits et de systèmes. Nous avons utilisé la version proposée par [58, Appendix B, p. 15], qui a été traduite et validée grâce à une évaluation systématique, permettant de mesurer l'utilisabilité d'un système ou d'un produit avec une grande confiance [58].

La deuxième partie du questionnaire porte sur l'expérience avec la modélisation d'un indicateur en autonomie et utilise la technique pragmatique de complétion de phrases. Selon [88, p. 481], « [...] la complétion de phrases est une méthode qualitative pour évaluer l'expérience de l'utilisateur ou bien explorer ses besoins. Les débuts de

phrases sont des amorces qui doivent aider l'utilisateur à penser aux aspects expérientiels de l'interaction et à exprimer son ressenti de manière semi-structurée. » Les résultats ont tendance à fournir des informations plus détaillées, notamment sur des éventuelles émotions négatives.

PHASE 3 Lors d'entretiens d'explicitation, le chercheur vise à comprendre comment s'est passée la prise en main du métamodèle (quelles actions, quelles intentions, quel ordre d'actions). Il insiste également sur les points de blocage lors de cette prise en main et sur les améliorations à prévoir.

Ensuite, une dernière tâche est proposée aux modélisateurs : ils reçoivent une nouvelle demande (hypothétique) d'un client, qui souhaite une modification d'un indicateur/un nouvel indicateur. Cette modification devrait permettre la réutilisation des primitives déjà modélisées. Le chercheur observe le modélisateur faire ces changements en notant si les réutilisations/changements sont faits comme prévu.

Concernant la technique utilisée, selon [144], l'entretien d'explicitation est utile quand nous avons besoin de nous informer de l'action de l'autre ou de l'aider à mieux s'approprier ce qu'il a fait, de savoir comment il s'y est pris. Le focus est sur la mise en mots a posteriori de l'action, se concentrant ainsi sur la dimension procédurale de l'action plutôt que sur les seuls aspects déclaratifs ou intentionnels.

À travers des techniques de relances (questions, reformulations, silences), le chercheur guide la mise en mots de l'expérience vécue. L'explicitation dans ce type d'entretien se réfère en premier lieu à ce qui est implicite dans la réalisation d'une action, qui peut être mentale ou matérielle [15], comme l'utilisation d'une des primitives du métamodèle. Par la suite, le chercheur mène le participant à expliquer, préciser cette référence à l'action. La technique propose de prendre différents repères sur ce qui est verbalisé, avec l'intention de diriger la mise en mots vers la dimension vécue et procédurale de l'action tout en évitant des jugements et des émotions. Quatre fonctions de relance de l'entretien sont à moduler par le chercheur selon les différentes situations de dialogue : initialiser, focaliser, élucider et réguler.

Le support utilisé lors de cette phase est celui de la demande de changement de l'indicateur qui a été modélisé par les modélisateurs en autonomie. Les deux demandes transmises à l'oral sont :

- en fait, le client a réfléchi et demande maintenant que les dates pour le calcul de l'indicateur/visualisation soient entrées par l'instructeur ;

- en fait, le client a réfléchi et demande maintenant que le nombre de visualisations soit par bloc et non par page. La visualisation ne se fait alors que sur le tableau de bord.

12.3 RÉSULTATS : QUESTIONNAIRE AUX MODÉLISATEURS

Suite à la modélisation de l'indicateur qui sert d'exemple, les modélisateurs ont répondu au questionnaire dont les différentes parties sont disponibles dans l'Annexe D. Nous reportons ces résultats dans cette Section.

PREMIÈRE PARTIE DU QUESTIONNAIRE : *system usability scale*

La première partie du questionnaire était constituée des questions composant le *System Usability Scale* (SUS). Les réponses apportées sont affichées dans la Figure 12.2 ci-après (pour rappel, les phrases complètes de ce questionnaire sont dans l'Annexe D).

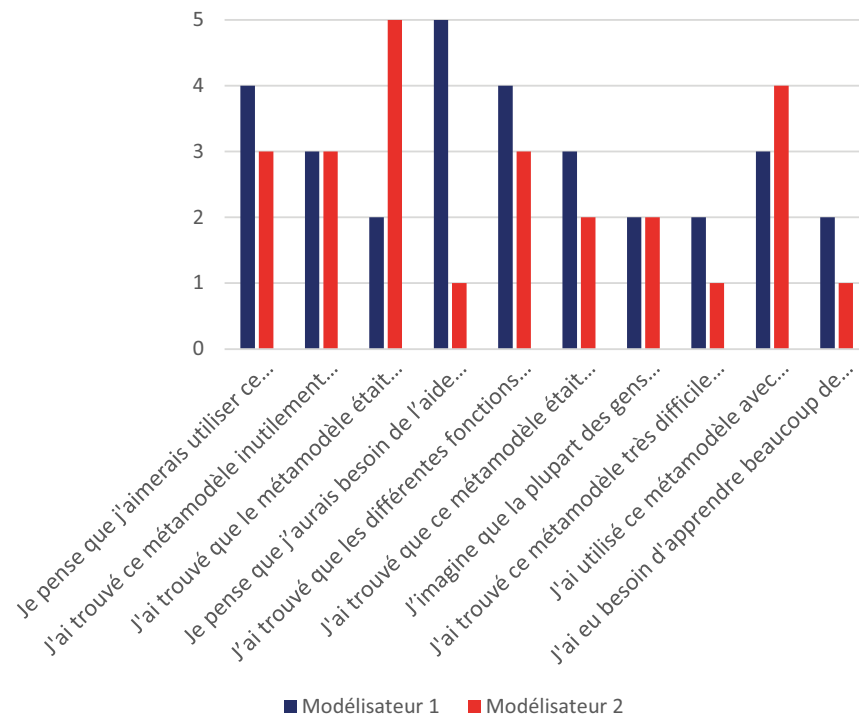


FIGURE 12.2 – Utilisabilité du métamodèle selon les participants.

Pour rappel, les réponses sont données selon une échelle de 1 à 5, où 1 correspond à « Pas du tout d'accord » et 5 à « Tout à fait d'accord ». Le score SUS pour les modélisateurs participants a été de 50 et 72,5, respectivement.

Nous remarquons que les modélisateurs, dans une première impression, trouvent le métamodèle complexe, mais pas inutilement complexe. Ce ressenti est corroboré par les propos qu'ils ont pu exprimer lors des entretiens : « Après on s'en sort aussi, quand on est en train de faire quelque chose, on essaie de ne pas se poser trop de questions, on utilise ce dont on a besoin et on voit qu'il y a d'autres choses en passage ». Cela pourrait indiquer une compréhensibilité du besoin de trouver un équilibre entre pouvoir modéliser la plupart des indicateurs versus avoir un métamodèle potentiellement plus simplifié, mais moins puissant.

DEUXIÈME PARTIE DU QUESTIONNAIRE : COMPLÉTION DES PHRASES

La deuxième partie du questionnaire visait à collecter le ressenti des modélisateurs par rapport à plusieurs aspects de l'utilisation du métamodèle.

Un des modélisateurs a complété la phrase « Quand j'utilise ce métamodèle je me sens. . . *face à un grand nombre de possibilités que je ne maîtrise pas encore* », montrant d'un côté que le métamodèle semble être complexe à première vue — mais permettant de modéliser un grand nombre d'indicateurs — avec plusieurs fonctionnalités que les modélisateurs n'ont pas encore pu utiliser lors de cette première prise en main. D'un autre côté, le même modélisateur affirme que « Les primitives du métamodèle sont. . . *peu nombreuses, mais concises* ». Ces réponses ont pu être renforcées/développées lors de l'entretien, où le modélisateur affirme que cette difficulté initiale est commune à tous les métamodèles, où un « temps d'entrée » est nécessaire. Il affirme encore que c'est commun d'avoir ce type de situation où l'on a « potentiellement plein de choses que l'on peut faire, c'est normal, cela prend un peu de temps et d'expérimentation pour tout comprendre ».

La primitive préférée du métamodèle est celle permettant de créer les traces, pour les deux modélisateurs. En effet, c'est la primitive la plus liée au modèle documentaire et qui permet d'enrichir les traces selon les besoins et avec les informations du document de manière automatique, ce qui semble plaire.

Ce qui est moins apprécié par l'un des modélisateurs est la modélisation des indicateurs, et par l'autre modélisateur, l'incohérence dans le nommage des extensions des primitives - par exemple, le « uiWidget » utilisé pour la visualisation devrait avoir le terme « visualisation » dans son nom afin de mieux comprendre son usage. Cette incohérence a été expliquée lors de l'entretien et faisait référence au nommage qui n'était pas clair pour l'une des primitives.

À la question relative aux suggestions d'améliorations : « Si je pouvais changer quelque chose à ce méta-modèle, je... » il a été répondu « détaillerais ou modifierais les infobulles d'aide (sm :help) » et « permettre de lier le opentrace dans la table ». En effet, le déploiement du métamodèle n'étant pas encore définitif, les textes d'aide sur chaque primitive/partie des primitives n'ont pas été complétés, à ce stade, mais ils le seront, ce qui permettra d'aider les modélisateurs dans leurs choix. Concernant la deuxième réponse, ce point a été abordé lors de l'entretien : il est question de la création d'un lien explicite entre deux primitives, ce qui est habituel dans l'outil de modélisation utilisé.

Selon les modélisateurs, l'utilisation de ce métamodèle serait surtout utile dans le contexte de projets avec des évolutions fréquentes sur les LA. Toujours selon les deux modélisateurs « Si je devais réutiliser un indicateur conçu dans le cadre d'un modèle documentaire dans le contexte d'un autre modèle documentaire, cela serait... *plus simple* ».

Ce gain en simplicité est renforcé par l'utilisation d'un métamodèle dans l'horizon de plusieurs projets/indicateurs et réaffirmé par les modélisateurs dans leurs réponses aux deux dernières questions : « Si un client me demande de changer quelque chose dans un indicateur, cela sera...*moins coûteux de passer par un métamodèle une fois que je le maîtriserai. / Simple* ». Également, « Si je devais modéliser la plupart des indicateurs de Learning Analytics auxquels je peux penser avec ce métamodèle cela serait... *un gain de temps et une meilleure maintenabilité du projet* ».

12.4 RÉSULTATS : ENTRETIENS AVEC LES MODÉLISATEURS

Chaque entretien avec les modélisateurs a eu une durée moyenne d'une heure. Ils ont fait l'exercice de raconter comment ils se sont pris, primitive par primitive, afin de créer la modélisation de l'indicateur proposé. Ils ont décrit cette prise en main selon l'ordre des primitives utilisées. Lors de l'entretien, la discussion a été faite à partir de cette modélisation, avec partage d'écran.

Le premier modélisateur a commencé par poser la primitive permettant de définir le traçage « Trace » et la table correspondante « Table ». Il dit comprendre comment définir ce qui va être injecté dans la publication du document et qui va ensuite alimenter les journaux de traces. Il se demandait seulement sur certaines informations, comme l'horodatage, qui n'étaient pas à modéliser, car faisant partie des informations insérées automatiquement dans la trace. Le chercheur a confirmé que certaines informations sont automatiques pour toutes les traces, information qui a été transmise lors de la séance de présentation du

modèle, mais qui n'a pas été retenue. Une liste avec ces informations devrait être fournie dans un guide aux modélisateurs afin de les aider à mieux comprendre ce point et les informations ne nécessitant pas d'être modélisées. À part cela, le modélisateur affirme être confiant dans ce qu'il faisait. Une suggestion apportée par le modélisateur a été d'ajouter la possibilité d'indiquer sur quel(s) élément(s) de l'interface les traces doivent être créées. Par exemple, si l'on veut tracer juste un bouton d'une page et pas tous, en indiquant l'identifiant de ce bouton provenant du modèle documentaire (à savoir que pour le moment il n'y a pas d'identifiant unique créé pour chaque bouton, donc des changements doivent être considérés dans le modèle documentaire).

Ensuite le premier modélisateur affirme avoir posé la primitive correspondant à la table qui va être créée à partir des traces : « Ça aussi, c'est une partie que j'ai bien comprise ». Concernant les filtres permettant de choisir les lignes et colonnes qui correspondent au besoin, la logique semble bien comprise et d'utilisation facile (à noter qu'un souci d'implémentation a fait que l'option « OU » a été oubliée, ce qui a été remarqué lors de la séance de présentation et expliqué aux participants). Ici, souvent, les informations insérées toujours de manière automatique aux traces doivent être utilisées, ce qui renforce la remarque précédente. Le modélisateur conseille que ces options puissent être proposées par le système sous forme de listes, afin que le modélisateur fasse seulement le choix parmi ce qui existe déjà — avec l'avantage que l'information de ce qui existe par défaut est également présent de cette façon.

La prochaine primitive qui a été utilisée par le modélisateur est celle qui permet de poser les informations relatives à l'indicateur lui-même « Indicator ». Le modélisateur a bien lié la table nécessaire en entrée au calcul effectué par la suite. La sortie est conditionnée par ce calcul, il faut remarquer que la typologie correspondante devrait aboutir à une liste définie, dont les valeurs n'ont pas encore été tranchées et que donc ce champ est ouvert pour le moment. L'objectif est d'avoir, dans ce champ, le type de sortie attendue, ce qui servira comme information pour la visualisation choisie dans la suite (une liste, une matrice, etc.). Dû au fait que la liste est à ce stade affichée dans un champ ouvert, le modélisateur demande des précisions, mais comprend la finalité du champ sans problème.

Le widget qui va permettre de préparer la visualisation de ces résultats « Visualization Widget » a été utilisé par la suite. Pour l'indicateur qui sert d'exemple, les paramètres qui servent comme entrée (date de début et date de fin) proviennent d'une entrée de la part de l'instructeur, mais cela pourrait également être fixé par le modélisateur au

moment de la modélisation, etc. Le modélisateur n'avait pas compris que c'est bien ici que l'origine de ces informations est définie. Avec l'explication lors de cet entretien, le modélisateur comprend comment définir ces entrées de plusieurs manières à l'aide de cette primitive.

Ensuite, c'est la primitive « uiWidget » qui est utilisée par le modélisateur, pour la modélisation en termes de moyen de visualisation choisie (tableau de bord, menu, etc.) pour chaque type d'utilisateur. Le modélisateur suggère de changer le nommage de la primitive afin d'avoir le terme "visualisation" présent, facilitant sa recherche/compréhension parmi toutes les primitives disponibles. Pour rappel, lors de cette étude, les détails concernant les visualisations type de graphique, etc. n'étaient pas encore implémentés.

Enfin, le modélisateur a utilisé la primitive « Analytical Engine » afin de définir la périodicité des calculs.

Le modélisateur précise somme toute que l'usage des primitives du métamodèle a été itératif pour certaines parties, où il se rendait compte d'un besoin et revenait d'abord à une autre primitive avant de continuer. Cette itérativité est naturelle et même attendue, elle témoigne de la compréhension par étape d'une part du métamodèle et d'autre part de l'indicateur lui-même.

Toujours selon le premier modélisateur, la partie du métamodèle correspondant à la visualisation semble un peu plus compliquée, notamment la définition des entrées comme les paramètres, mais a été néanmoins comprise. Cette difficulté pourrait être causée par le fait que le métamodèle correspondant à la partie n'a pas pu être implémenté dans sa totalité au moment des tests, hypothèse à confirmer.

Concernant le nommage, les autres primitives — à part « uiWidget » déjà mentionné — sont faciles à s'y retrouver. Pour finir, le modélisateur a affirmé :

[...] J'ai trouvé qu'il était concis quand même comme métamodèle, il n'y avait pas énormément d'item [primitive] et ça c'est pas mal.

À propos du changement demandé par le client (dernière tâche de la phase 3), le modélisateur a pu faire les changements nécessaires : « Je commencerais par changer les traces ». À la suite il affirme :

[...] Si je dois changer d'un événement à un autre, effectivement, le fait que c'est pris en charge par le métamodèle, c'est assez rapide, oui.

Le deuxième modélisateur a rapporté qu'il a commencé par la primitive relative à l'indicateur « Indicator ». En l'utilisant, il s'est rendu compte de quelles entrées il allait avoir besoin. Ainsi, il a alors pris en main la primitive relative aux traces « Trace », pour que ces

entrées soient tracées. Il a suggéré de changer le terme interne à cette primitive de « fromModel » à « callModel » afin de garder la nomenclature déjà présente dans l'outil concernant ces situations.

Ensuite, il a déterminé la table construite à partir des traces « Table ». Le modélisateur a fait le retour du fait que cette primitive n'est pas liée aux traces dont on aura besoin. Ce type de lien est commun dans les primitives de l'outil utilisé, permettant de raccrocher les différents composants et ainsi de ne pas les oublier (l'information est également présente dans les réseaux d'items, qui permet de vérifier à la fin de la modélisation ces liens).

Au moment où on définit les éléments nécessaires afin d'insérer les données aux tables, le modélisateur questionne le fait qu'une primitive puisse avoir une valeur ou plusieurs (*focus in* et *focus out*) ou que ces deux valeurs soient posées chacune dans un composant. Cela implique que le lien entre primitives juste mentionnées soit fait de deux manières, selon les besoins. Nous constatons qu'une réflexion sur ce sujet doit être effectuée afin de voir si la complexité apportée par cette flexibilité est justifiée — ce qui pourrait être le cas pour certains indicateurs — ou que ce soit défini qu'une seule valeur puisse être attribuée dans cette primitive. Cette réflexion donne lieu à deux possibilités qui viennent à l'esprit du modélisateur :

1. Avoir deux items à partir de cette primitive, chacun avec une valeur unique : on glisse chacun dans le composant pour créer la table (et créer ces liens) ;
2. Avoir un item de cette primitive avec plusieurs valeurs : on glisse l'item en question, puis on choisit la(les) valeur(s) concernée(s).

Le modélisateur affirme que la primitive dédiée à l'indicateur lui-même est claire, que la partie « Analysis » dépendra de la(les) technologie(s) choisie(s).

Concernant la primitive utilisée pour la modélisation de la visualisation de l'indicateur, le deuxième modélisateur questionne la nomenclature choisie, mais trouve la primitive claire :

L'item lui-même est hyper clair, c'est juste la nomenclature par rapport à l'existant...

Enfin, le modélisateur affirme que la primitive « Analytical Engine » est moins claire, car il est « un peu plus difficile de se représenter sans le contexte qui pourrait arriver ».

À propos de la dernière tâche de la phase 3, le changement demandé par le client, le modélisateur a pu la réaliser sans soucis.

Pour conclure, le deuxième modélisateur affirme qu'il aurait besoin d'une application pratique et d'un contexte réel pour des avis encore

plus précis, car c'est avec l'usage et le temps que l'on pourra trouver d'autres améliorations.

12.4.1 *Résultats : synthèse des retours*

Afin de synthétiser les résultats de cette analyse, nous reprenons les questions de recherche posées.

12.4.1.1 *Comment se passe la prise en main du métamodèle par des modélisateurs ?*

Les modélisateurs semblent prendre en main le métamodèle sans grandes difficultés. Certains points d'amélioration ont été proposés, certaines parties ont été moins claires, mais en général les primitives ont été comprises et utilisées correctement pour une toute première modélisation d'un indicateur.

12.4.1.2 *Les primitives sont-elles utilisées comme prévu ?*

Oui, elles sont utilisées comme prévu. L'ordre de l'utilisation des composants est logique et parfois itératif, ce qui est attendu, surtout pour une première utilisation.

12.4.1.3 *Quels sont les points de blocage lors de cette prise en main ?*

Une primitive liée à la visualisation a été moins claire pour l'un des modélisateurs ; pour l'autre, c'est la primitive liée à l'agenda des calculs. Dans le premier cas, cette difficulté pourrait être liée au fait que les visualisations ont été implémentées avec moins de détails au moment des tests, mais cela reste à vérifier. Dans le deuxième cas, le modélisateur affirme que la difficulté est liée au fait de devoir abstraire un projet/besoin et qu'une situation pratique permettrait probablement de se projeter sur l'usage de la primitive.

12.4.1.4 *Y a-t-il des améliorations à prévoir, lesquelles ?*

Oui, les deux améliorations principales sont liées à la nomenclature (voir ci-dessous) et au fait que certaines des primitives ne soient pas explicitement liées.

12.4.1.5 *Le vocabulaire utilisé est-il clair pour un modélisateur ?*

La plupart du vocabulaire semble compris sans problème, notamment celui relatif à l'analytique des apprentissages avec le numérique. Néanmoins, quelques suggestions de changements sont : 1/ ajouter

le terme « Visualization » au « uiWidget », ou encore ne pas considérer cette primitive comme un widget, car trop différent de l'existant; 2/ Dans « Trace », changer le terme de « fromModel » à « callModel » qui existe déjà.

12.4.1.6 *Quelle est l'utilisabilité perçue du méta-modèle par des modélisateurs ?*

L'utilisabilité perçue n'a pas de scores très élevés, mais lors des entretiens, il a été possible de comprendre que cela était dû au fait qu'il s'agit d'un métamodèle, dont une certaine complexité est presque « attendue ». Le fait que les modélisateurs ont pu modéliser un premier indicateur en autonomie et les entretiens nous permettent de relativiser cette complexité initiale perçue.

12.4.1.7 *Comment les différentes primitives proposées sont utilisées afin d'encourager la maintenabilité, la réutilisation, la personnalisation des indicateurs ?*

Cet aspect a été testé à travers la demande de changement d'un aspect de l'indicateur à la fin de la phase 3. Les modélisateurs ont pu comprendre et réaliser la tâche demandée. De plus, la maintenabilité, la réutilisation, la personnalisation de l'utilisation d'un indicateur via le métamodèle semblent bien perçues compte tenu des dernières réponses au questionnaire (et certainement aussi de l'expérience en modélisation).

12.4.2 *Discussion*

Les modélisateurs semblent prendre en main le métamodèle sans grandes difficultés, avec une utilisation des primitives selon un ordre logique et parfois avec des aller-retours, ce qui est attendu pour un métamodèle et surtout pour une première utilisation.

La plupart du vocabulaire semble compris et quelques suggestions de changements ont été faites, ce qui reste simple à remédier. Lors des entretiens, il a été possible de comprendre les scores relatifs à l'utilisabilité, qui n'est naturellement pas élevée, au vu de la complexité de l'utilisation inhérente. Il est à signaler que les modélisateurs ont pu modéliser un premier indicateur rapidement et que surtout certains aspects comme la maintenabilité, la réutilisation, la personnalisation (cf. Chapitre 11) d'un indicateur via le métamodèle semblent bien perçues. Ceci a pu être vérifié via la demande de changement d'un aspect d'un indicateur (phase 3), où les modélisateurs ont pu changer

seulement une partie d'une primitive — ou en poser une nouvelle — d'un indicateur pour le changer selon les nouveaux besoins ou pour en créer un nouveau à partir des primitives modélisées auparavant.

Sixième partie

CONCLUSIONS

Cette dernière partie est dédiée aux conclusions que nous avons pu retirer des études et propositions que nous avons pu présenter. Nous présentons également les perspectives d'utilisation de ces résultats et de nouvelles questions de recherche qui pourraient être explorées.

CONCLUSIONS

13.1 SYNTHÈSE

Cette thèse est à la croisée de l'analytique des apprentissages avec le numérique et de l'ingénierie dirigée par les modèles, nous avons exploré les bénéfices potentiels de l'utilisation de ce second afin d'implémenter ce premier.

Pour ce faire, notre recherche a consisté d'abord en un travail de fond sur différentes questions autour de la production documentaire dans un contexte pédagogique et des indicateurs d'apprentissage déjà utilisés par le domaine, mais aussi sur les potentiels enrichissements qui pourraient être apportés, notamment par la connaissance, fine et en amont, des documents.

Notre problématique s'est déclinée en cinq questions de recherche. Nous proposons de synthétiser les réponses à ces questions qui ont constitué le fil conducteur de ce travail.

13.1.1 **RQ1** *Peut-on proposer un métamodèle de collecte, d'analyse et de visualisation des traces d'interaction prenant en compte l'ensemble de la chaîne de production de documents ?*

Le métamodèle proposé a été conçu afin de répondre à plusieurs contraintes (cf. Chapitre 11). Son principal atout est celui de permettre la prise en compte, de manière automatique, de la sémantique et de la structure des documents sur lequel il repose. Ainsi, il habilite la modélisation des indicateurs habituellement utilisés dans des démarches d'analytique des apprentissages avec le numérique, mais permet également la modélisation d'indicateurs enrichis par cette connaissance fine des documents.

L'utilisation du métamodèle diffère également d'une démarche où l'on part des traces à disposition pour arriver aux indicateurs : il est proposé de partir de l'indicateur souhaité pour ensuite poser le traçage nécessaire afin d'effectuer les calculs et rendre la visualisation possible.

- 13.1.2 **RQ2** *Quels sont les composants nécessaires à ce métamodèle permettant de modéliser les indicateurs couramment utilisés dans le domaine, mais également permettant de les enrichir avec la connaissance fine de la sémantique des documents ?*

Les primitives retenues (cf. Chapitre 11) sont :

- Concernant les traces des documents : « Primitives, Organization, Resource (Multimedia, Binary, Folder, Image, Texte), Evaluation (Quizz, Questionnaire), Spatio-temporal (Composition, Data). »
- Concernant l'indicateur : « Indicator (Input/Parameter), Table, Trace, Analytical Engine, uiWidget (Parameter Value/Input), Visualization Widget. »
- Concernant la visualisation des indicateurs : « Visualization, Preference, Dashboard, Prompt, Menu, Navigation, Page, Container (Row/Column), Document Container, Chart, Dataset, Axis, Variable, Label, Legend. »

Le métamodèle et ses différents composants (les primitives d'analyse) ont été testés auprès de modélisateurs expérimentés (cf. Chapitre 12). Ces modélisateurs ont pu prendre en main le métamodèle sans problèmes et ont proposé quelques améliorations, notamment concernant le nommage et les liaisons explicites entre les composants modélisés.

Les particularités concernant la démarche dans laquelle le métamodèle est utilisé ont été commentées en comparant l'effort nécessaire à la création d'un versus plusieurs indicateurs, selon un nombre de métriques et fonctionnalités (cf. Chapitre 11).

- 13.1.3 **RQ3** *Un modèle documentaire correspond-il à un besoin homogène en termes d'analytique des apprentissages avec le numérique et donc à une série d'indicateurs d'apprentissage bien identifiés ?*

Les résultats de l'étude exploratoire (cf. Chapitre 8) auprès des futurs utilisateurs d'une implémentation suivant notre approche nous ont montré que les besoins concernant un seul contexte et un seul modèle documentaire peuvent varier considérablement, notamment en ce qui concerne les comparaisons, l'accès aux données et la rétroaction.

Ces résultats nous mènent à définir les primitives dans le cadre du métamodèle permettant de créer des indicateurs et de les adapter facilement. Ainsi, l'hypothèse qu'un même modèle documentaire représenterait suffisamment d'homogénéité en termes de besoins — et donc que la proposition d'une série d'indicateurs « clés en main » spécifiques pour chaque modèle documentaire serait la meilleure solution — a été réfutée. Les primitives du métamodèle ne correspondent

donc pas à des indicateurs, mais à des composants permettant de modéliser la plupart des indicateurs (cf. Chapitre 9) dont on aurait besoin selon les différents contextes.

13.1.4 **RQ4** *Y a-t-il une différence dans la réception des visualisations basées sur des indicateurs enrichis ?*

La compréhension des graphismes, y inclus ceux spécialement conçus afin de refléter cet enrichissement via la connaissance de la sémantique ou bien de la structure des documents, semble ne pas poser de souci (cf. Chapitre 8).

Également, nous n'avons pas noté de différences qui pourraient être le signe d'une difficulté particulière dans l'interprétation de ces graphiques particuliers. En effet, ces visualisations ont incité des suggestions d'usages supplémentaires de la part des participants, ce qui pourrait être le signe de leur utilité perçue, mais cela reste à confirmer.

13.1.5 **RQ5** *Quels sont les indicateurs utilisés dans les systèmes d'analytique des apprentissages avec le numérique et ses différents éléments potentiellement à abstraire pour la construction d'un métamodèle qui les représente ?*

Si l'étude exploratoire nous a permis de décider d'un métamodèle visant la création d'indicateurs via l'abstraction des différents aspects de ceux-ci, la revue systématique de la littérature (cf. Chapitre 9) nous a permis de connaître ces indicateurs en détail.

Nous avons ainsi pu appréhender les indicateurs qui seraient potentiellement modélisés en utilisant notre métamodèle et les différentes caractéristiques de ces indicateurs, comme le type d'analyse effectuée, le contexte, le statut de l'indicateur, le type de dispositif pédagogique mis en place, entre autres.

13.2 LIMITES ET PERSPECTIVES

Ce travail présente un métamodèle pour la conception et mise en place d'indicateurs d'analytique des apprentissages avec le numérique. Une des limites rencontrées a été la difficulté de conduire des comparaisons et tests pour ce métamodèle. Il est en effet compliqué de comparer les résultats similaires, mais provenant d'une utilisation d'approches distinctes. Par exemple, le fait de réutiliser un même indicateur modélisé pour un modèle documentaire dans un autre modèle

documentaire, avec peu ou sans modifications, comment mesurer ce fait par rapport à une démarche traditionnelle ? Nous avons contourné cela en mettant en lumière les fonctionnalités et les avantages de notre démarche dans le Chapitre 11, notamment en ce qui concerne les possibilités données par notre démarche — création automatique d'indicateurs enrichis de la sémantique et structure des documents et conception des indicateurs en amont — qui ne le sont pas par des approches traditionnelles que nous avons pu rencontrer lors de l'état de l'art. Nous avons également testé la prise en main du métamodèle avec des modélisateurs, même si en nombre réduit de participants (cf. Chapitre 12).

Les perspectives consistent en l'implémentation, dans l'outil dédié, des différentes transformations nécessaires à une utilisation effective du métamodèle dans des contextes pratiques. Étant donné que les métamodèles sont sujets à évolution, de nouvelles versions peuvent voir le jour avec de nouvelles primitives (ou des adaptations de celles présentées) qui capturent mieux l'activité d'analytique des apprentissages avec le numérique en conjonction avec la modélisation des documents. Certes, il sera nécessaire de prendre en compte l'impact de ces changements sur les éventuels artefacts et documents existants.

Cette utilisation dans la pratique devrait permettre d'analyser d'autres questions de recherche qui ont été soulevées au long de ce travail. Nous pensons notamment à :

- Évaluation auprès des utilisateurs d'une utilisation dans la pratique des indicateurs enrichis de la sémantique et structure des documents versus des indicateurs sans ces informations. Ceci pourrait se faire avec, par exemple, les indicateurs présentés dans le Chapitre 10 (sauf celui nommé « auto-explication », car il est à la base dépendant de la sémantique du document, voir ci-dessous).
- Étude de l'indicateur que nous avons présenté dans le Chapitre 10, nommé « auto-explication ». Cet indicateur en particulier nous semble intéressant à étudier étant donné qu'il est nouveau, basé sur la sémantique du document et n'est pas présent dans la revue de littérature. Il pourrait ainsi être étudié dans la pratique afin de vérifier si, via son usage, il est possible de remarquer un changement de comportement et une adoption de techniques d'apprentissage plus efficaces.
- Étude de l'utilisation des « indications » faites manuellement sur les parties du contenu, en d'autres mots, la possibilité d'annoter directement sur le contenu, selon ses intentions (cf. Chapitre 8, Section 8.6.3). Cela pourrait se faire à partir de traces de cette

utilisation, afin de comprendre l'usage qui est fait, les intentions, etc.

- Une dernière perspective de recherche serait celle de fournir en entrée (les « features ») les données enrichies de la sémantique du document afin de vérifier si la performance de certains algorithmes utilisés dans le domaine de l'analytique des apprentissages avec le numérique — des regroupements ou des prédictions, par exemple — a une meilleure performance.

Septième partie

ANNEXES



FORMULAIRE DE CONSENTEMENT DE L'ÉTUDE AUPRÈS DES UTILISATEURS

Formulaire de consentement libre, éclairé et exprès

Libre Cours

Laboratoire Informatique de Paris 6 (Sorbonne Université) / Kelis Conseil et Développement

Je certifie avoir donné mon accord pour participer à une étude qui désire rencontrer certains utilisateurs de *Libre Cours* afin d'observer et d'enregistrer leurs commentaires, réactions, ressentis dans le cadre de l'utilisation du service.

J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette expérience, j'accepte que soient recueillies des données sur mes réponses, notamment par l'enregistrement audio des verbalisations. Je comprends que les informations recueillies sont strictement confidentielles et à usage exclusif des investigateurs concernés.

J'ai été informé que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'autorise que des extraits de ce que j'ai dit puissent apparaître dans des publications scientifiques à condition qu'ils soient anonymisés. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par l'Unité de recherche. J'ai bien noté que le droit d'accès prévu par la loi « Informatique et libertés » s'exerce à tout moment auprès de l'unité de recherche.

Date :

Groupe :

Nom du volontaire :

Signature du volontaire (précédée de la mention « lu et approuvé ») :

Nom de l'expérimentateur :

Signature de l'expérimentateur :

FIGURE A.1 – Étude auprès des utilisateurs — Formulaire de consentement

B

MAQUETTES DE L'ÉTUDE AUPRÈS DES UTILISATEURS

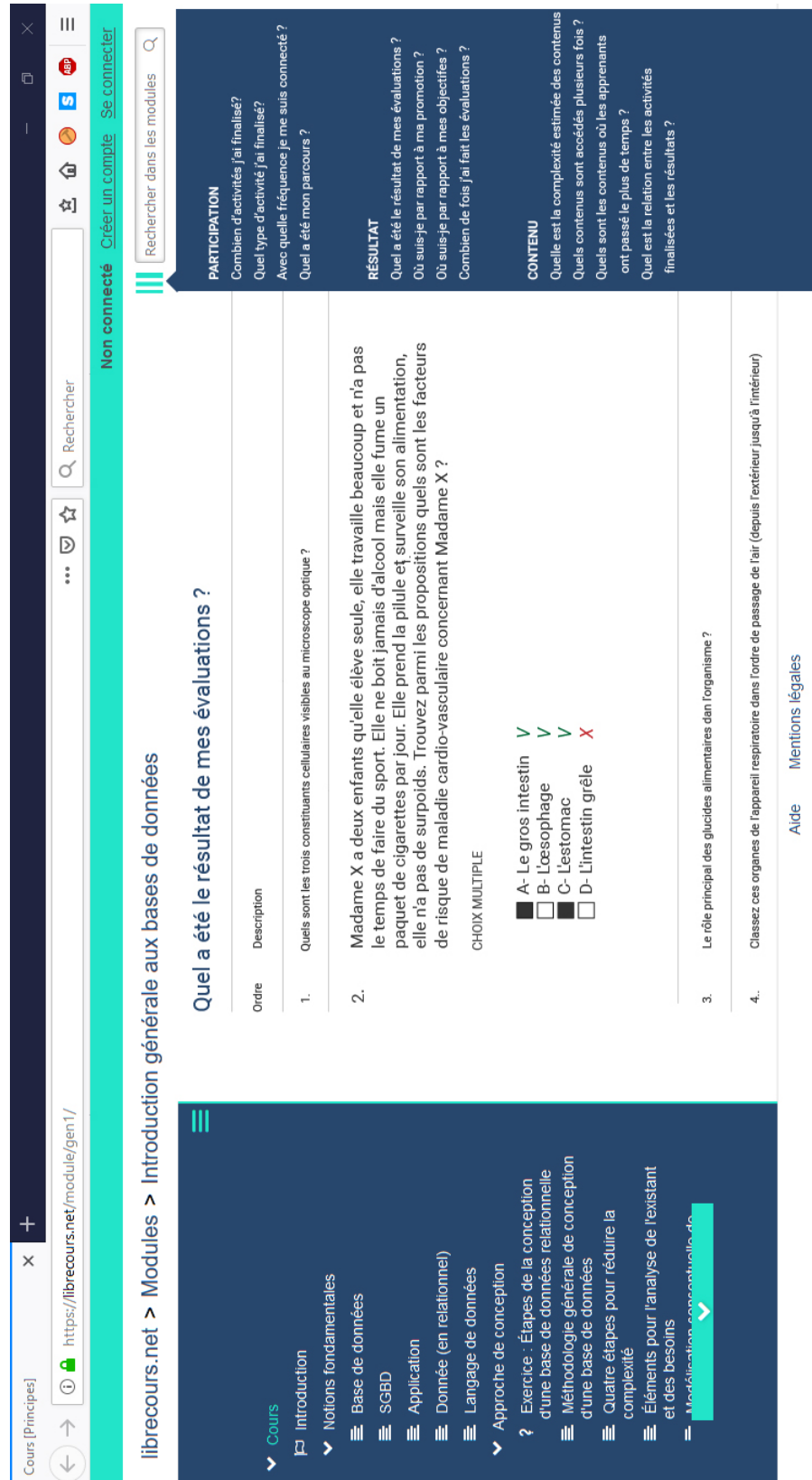


FIGURE B.1 – Étude auprès des utilisateurs — Maquette d'accès au tableau de bord.

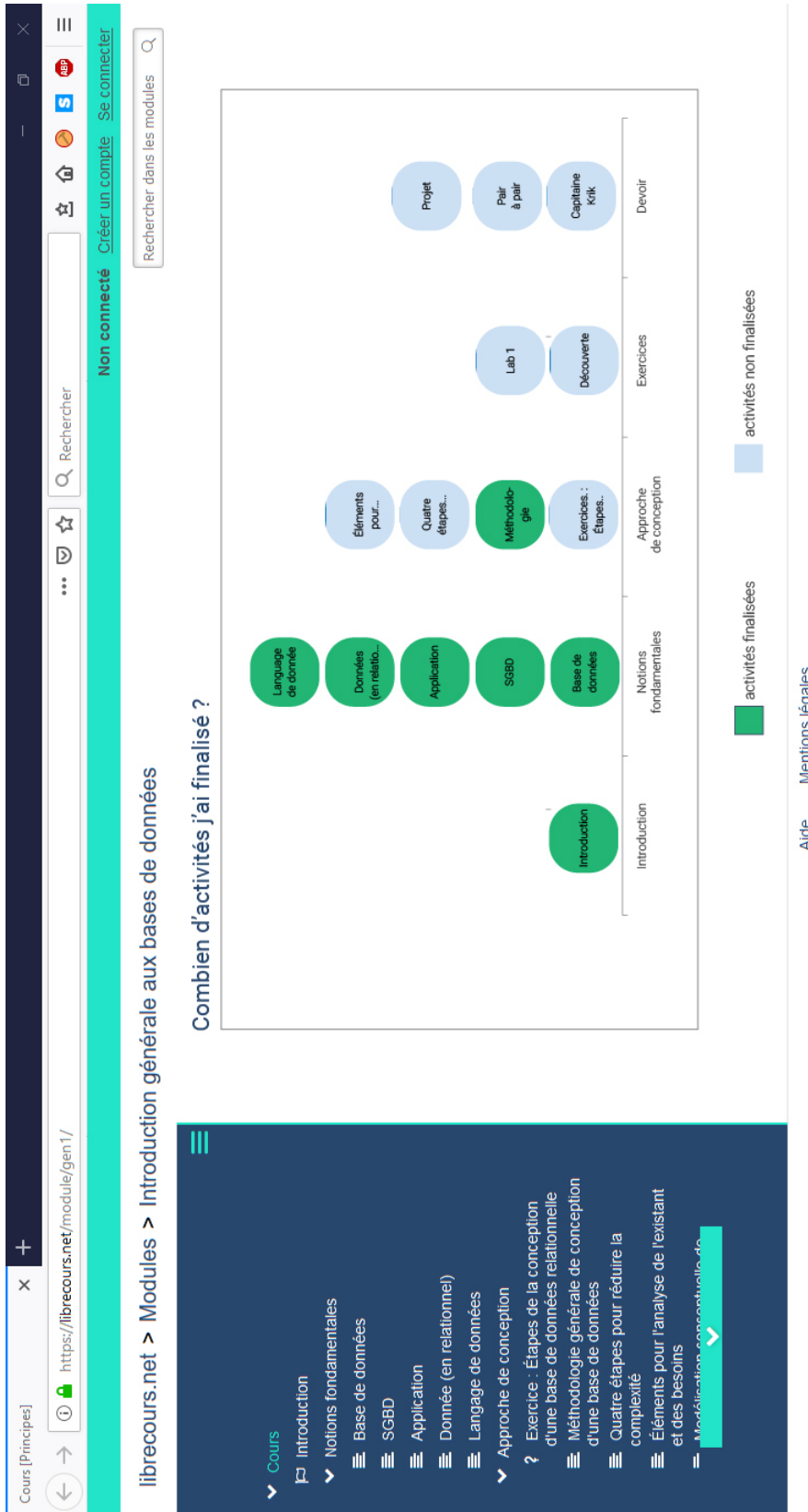


FIGURE B.2 – Étude auprès des utilisateurs — Maquette 01.



FIGURE B.3 – Étude auprès des utilisateurs — Maquette 02.

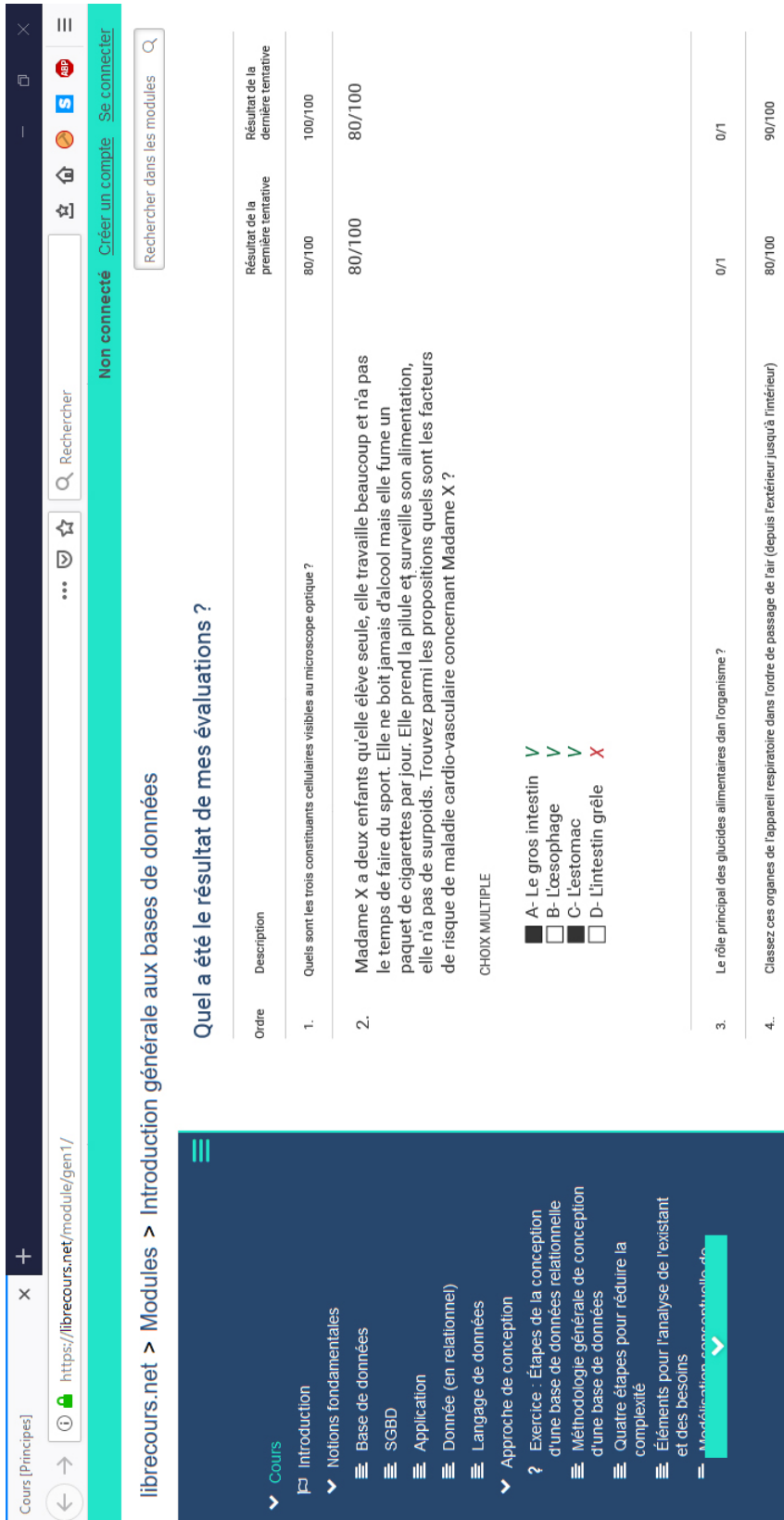


FIGURE B.4 – Étude auprès des utilisateurs — Maquette 03.

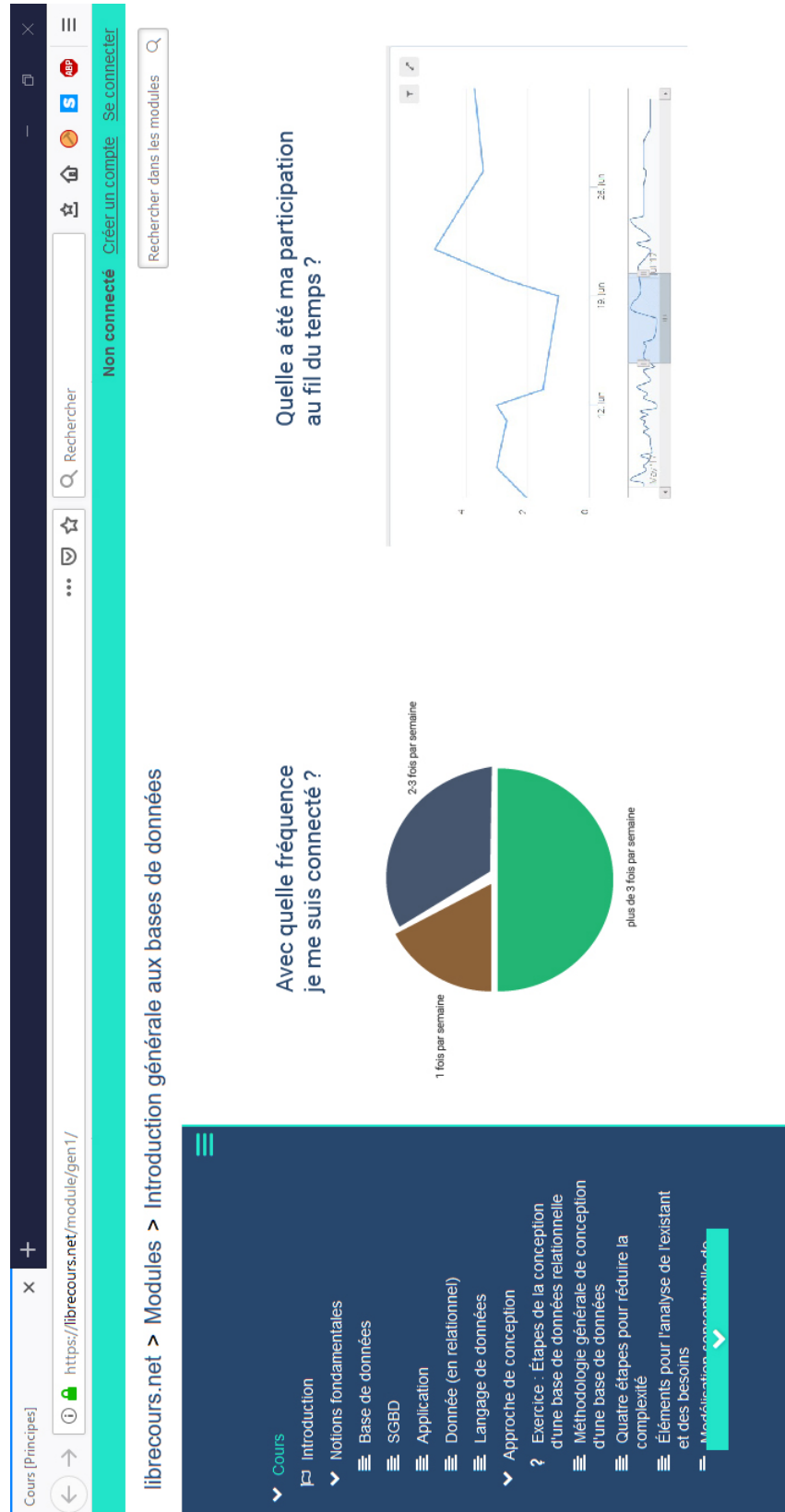


FIGURE B.5 – Étude auprès des utilisateurs — Maquette 04.

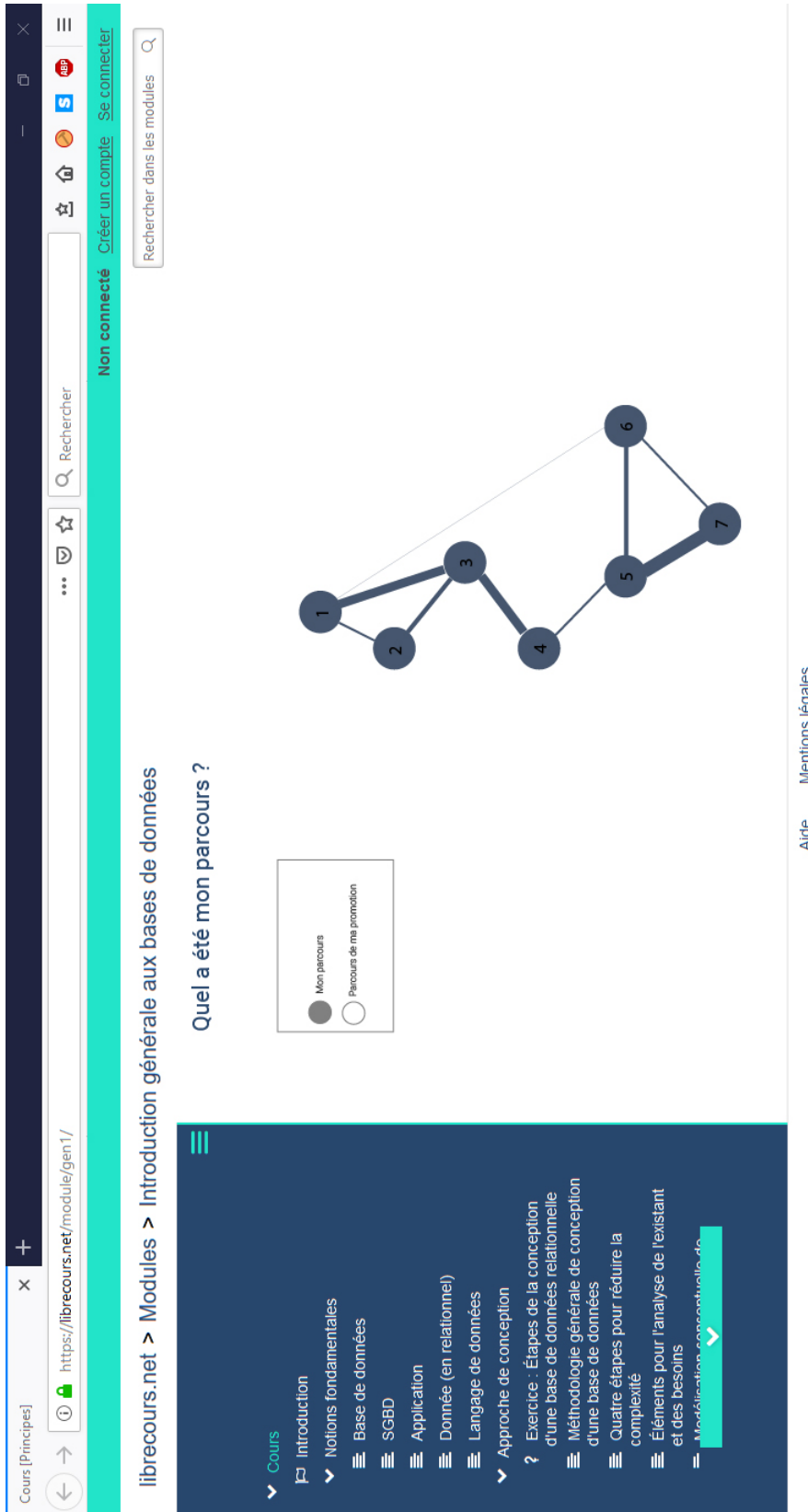


FIGURE B.6 – Étude auprès des utilisateurs — Maquette 05.

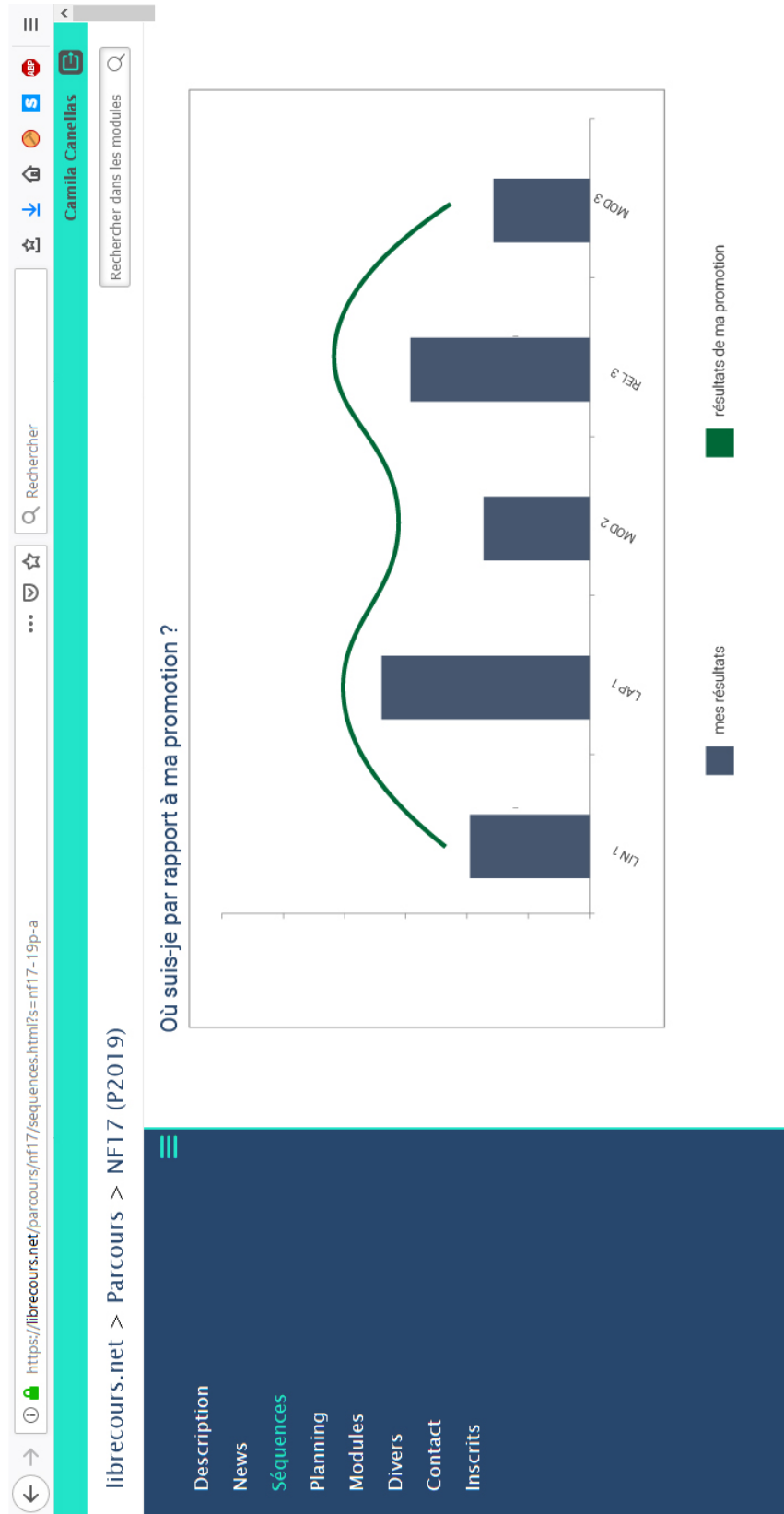


FIGURE B.7 – Étude auprès des utilisateurs — Maquette o6.

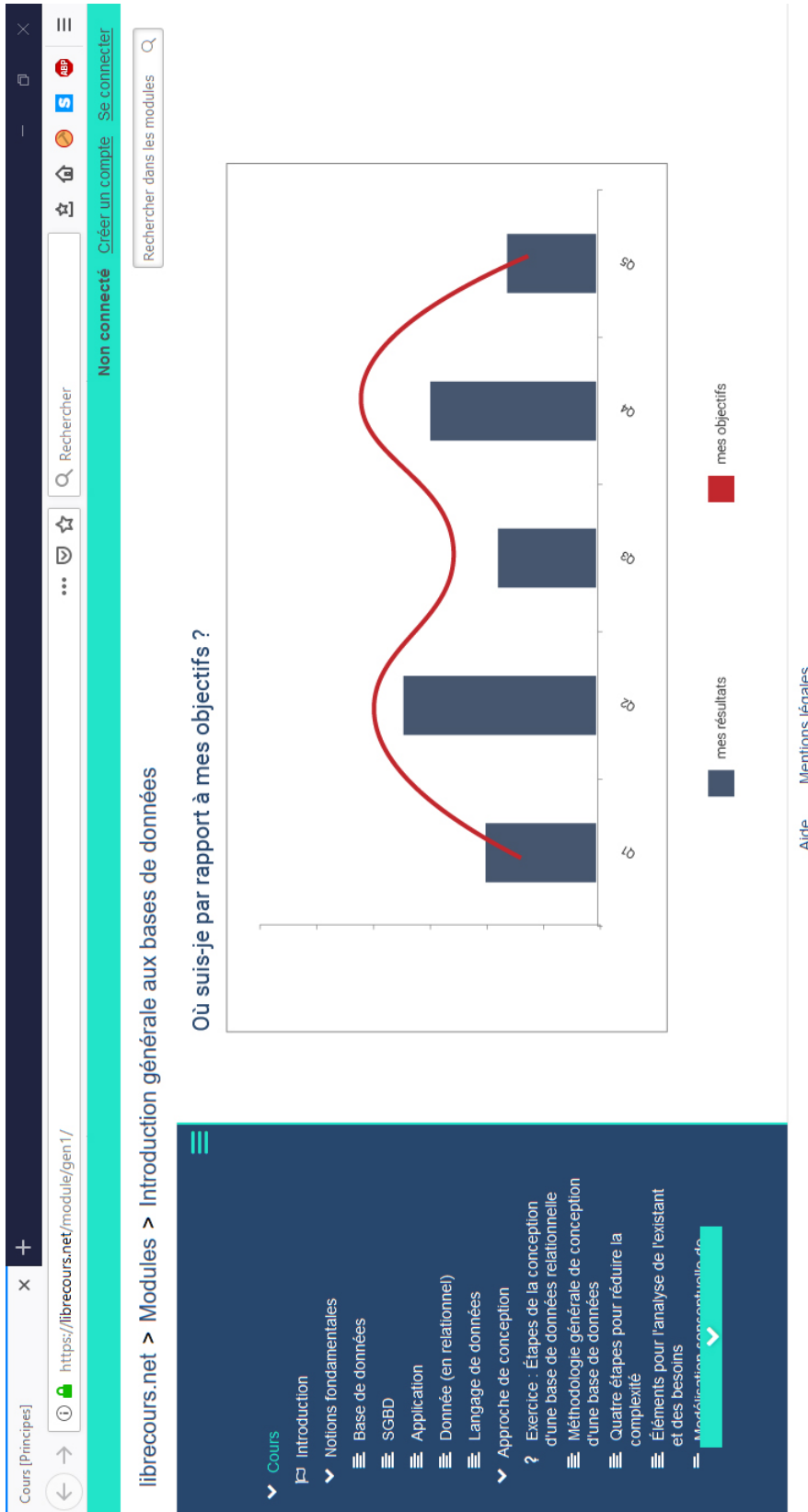


FIGURE B.8 – Étude auprès des utilisateurs — Maquette 07.

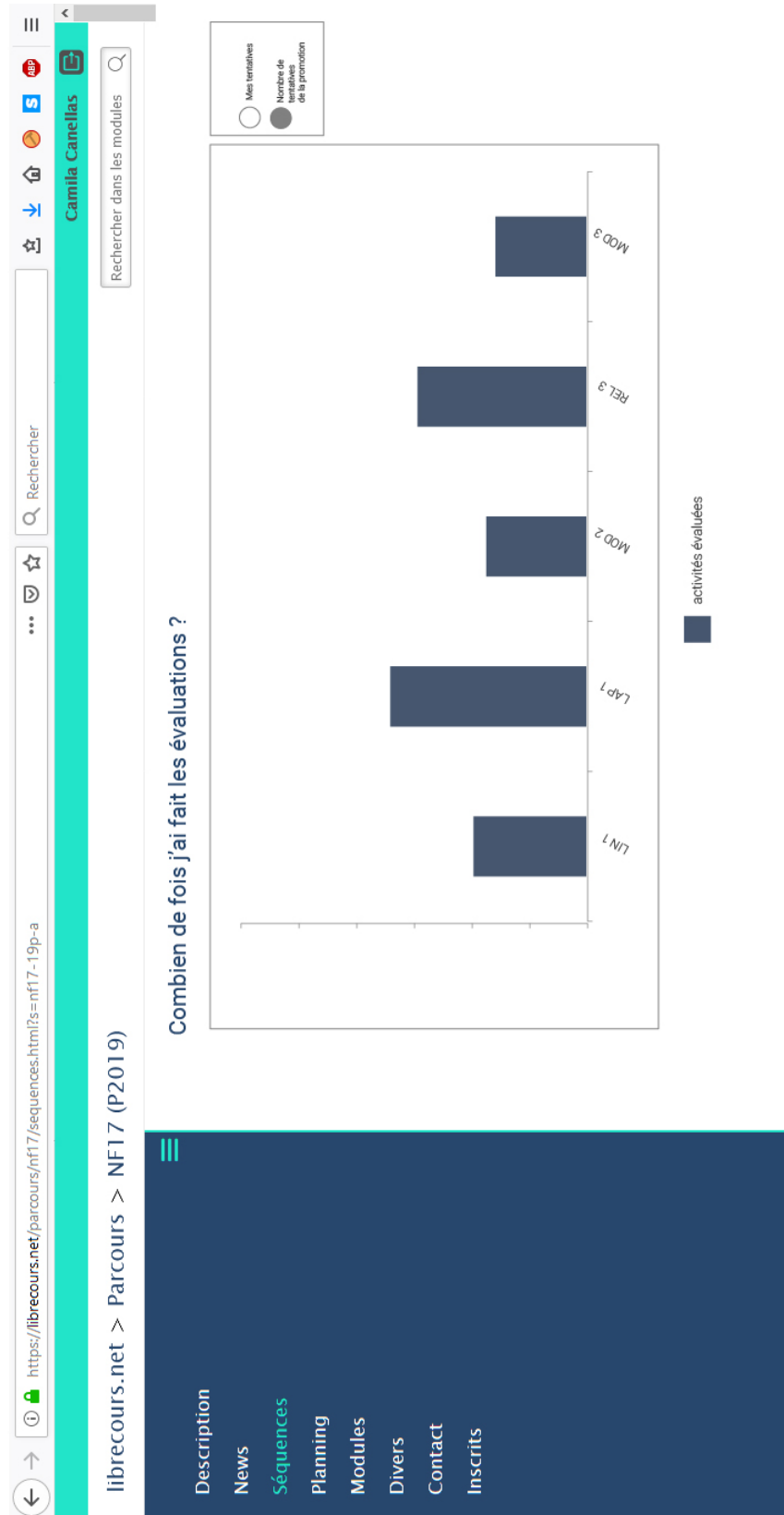


FIGURE B.9 – Étude auprès des utilisateurs — Maquette o8.



FIGURE B.10 – Étude auprès des utilisateurs — Maquette 09.

The image shows a web browser window displaying the 'librecours.net' website. The browser's address bar shows the URL 'https://librecours.net/module/gen1/'. The page title is 'librecours.net > Modules > Introduction générale aux bases de données'. The main content area is titled 'Quels contenus sont accédés plusieurs fois ?' and features a heatmap visualization of course content access frequency.

Navigation Menu (Left):

- Cours
- Introduction
- Notions fondamentales
- Base de données
- SGBD
- Application
- Donnée (en relationnel)
- Langage de données
- Approche de conception
- Exercice : Étapes de la conception d'une base de données relationnelle
- Méthodologie générale de conception d'une base de données
- Quatre étapes pour réduire la complexité
- Éléments pour l'analyse de l'existant et des besoins
- Médiation conceptuelle de

Heatmap Visualization (Center):

The heatmap is titled 'Quels contenus sont accédés plusieurs fois ?' and shows the following content items categorized by access frequency:

- Introduction:** Introduction (High access)
- Notions fondamentales:** Langage de données (High access), Données (en rela...) (High access), Application (High access), SGBD (High access), Base de données (High access)
- Approche de conception:** Éléments (High access), Quatre étapes (High access), Méthodologie (High access), Exercices (High access)
- Exercices:** Lab 1 (High access), Découverte (High access)
- Devoir:** Projet (High access), Pair à pair (High access), Caplaine Krik (High access)

Legend (Top Left):

- Mes accès (Dark blue circle)
- Moyenne d'accès de ma promotion (Light blue circle)

Color Scale (Bottom):

A horizontal color scale indicates access frequency, ranging from 'contenus les moins accédés' (light blue) to 'contenus les plus accédés' (dark blue).

Footer (Bottom Right):

- Aide
- Mentions légales

FIGURE B.11 – Étude auprès des utilisateurs — Maquette 10.



FIGURE B.12 – Étude auprès des utilisateurs — Maquette 11.

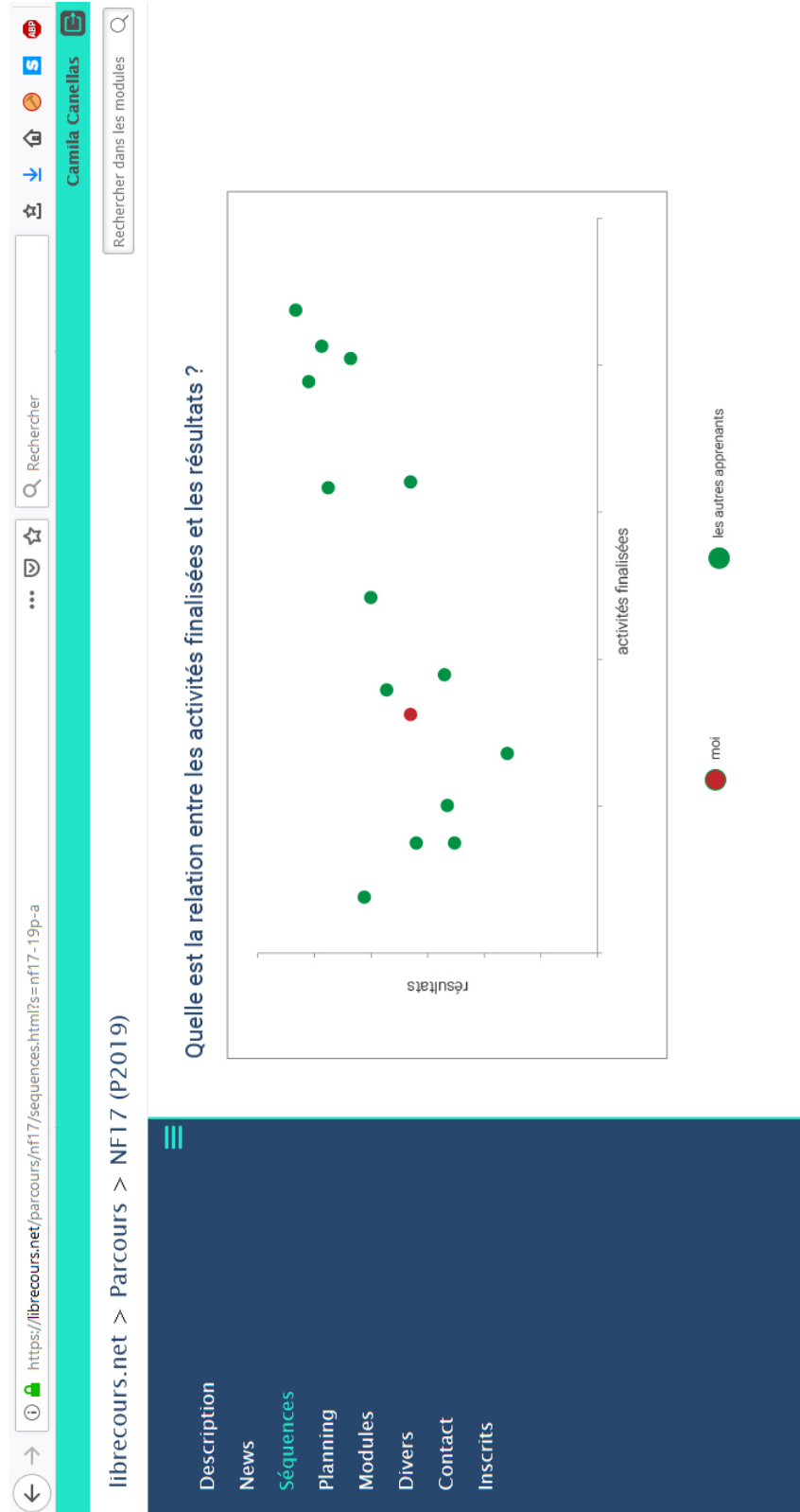


FIGURE B.13 – Étude auprès des utilisateurs — Maquette 12.



FIGURE B.14 – Étude auprès des utilisateurs — Maquette 13.

FORMULAIRE DE CONSENTEMENT DE L'ÉTUDE AUPRÈS DES MODÉLISATEURS

Formulaire de consentement libre, éclairé et exprès

Méta-modèle

Laboratoire Informatique de Paris 6 (Sorbonne Université) / Kelis Conseil et Développement

Je certifie avoir donné mon accord pour participer à une étude qui désire rencontrer certains modélisateurs la suite logicielle *Scenari* afin d'observer et d'enregistrer leurs commentaires, réactions, ressentis dans le cadre de l'utilisation de celui-ci.

J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette expérience, j'accepte que soient recueillies des données sur mes réponses, notamment par l'enregistrement audio des verbalisations. Je comprends que les informations recueillies sont strictement confidentielles et à usage exclusif des investigateurs concernés.

J'ai été informé que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'autorise que des extraits de ce que j'ai dit puissent apparaître dans des publications scientifiques à condition qu'ils soient anonymisés. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par l'Unité de recherche. J'ai bien noté que le droit d'accès prévu par la loi « Informatique et libertés » s'exerce à tout moment auprès de l'unité de recherche.

Date :

Nom du volontaire :

Signature du volontaire (précédée de la mention « lu et approuvé ») :

Nom de l'expérimentateur :

Signature de l'expérimentateur :

FIGURE C.1 – Étude auprès des modélisateurs — Formulaire de consentement.

QUESTIONNAIRE DE L'ÉTUDE AUPRÈS DES MODÉLISATEURS

1) Veuillez cocher la réponse qui reflète votre réponse spontanée à chaque affirmation. Ne réfléchissez pas trop longtemps à chaque affirmation, et veuillez à répondre à chaque affirmation. *

Mark only one oval per row.

	1. Pas du tout d'accord	2.	3.	4.	5. Tout à fait d'accord
Je pense que j'aimerais utiliser ce méta-modèle fréquemment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J'ai trouvé ce méta-modèle inutilement complexe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J'ai trouvé que le méta- modèle était facile à utiliser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Je pense que j'aurais besoin de l'aide d'un technicien pour pouvoir utiliser ce méta-modèle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J'ai trouvé que les différentes fonctions de ce méta- modèle étaient bien intégrées	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J'ai trouvé que ce méta- modèle était trop incohérent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE D.1 – Questionnaire de l'étude auprès des modélisateurs (1/3).

J'imagine que la plupart des gens apprendraient très rapidement à utiliser ce méta-modèle

J'ai trouvé ce méta-modèle très difficile à utiliser

J'ai utilisé ce méta-modèle avec beaucoup de confiance

J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir commencer à utiliser ce méta-modèle

2. Quand j'utilise ce méta-modèle je me sens... *

3. Avant de connaître ce méta-modèle, j'imaginai... *

4. Les composants du méta-modèle sont... *

FIGURE D.2 – Questionnaire de l'étude auprès des modélisateurs (2/3).

5. Ce que je préfère dans ce méta-modèle, c'est... *
-
6. Ce que je regrette avec ce méta-modèle, c'est... *
-
7. Si je pouvais changer quelque chose à ce méta-modèle, je... *
-
8. Ce méta-modèle serait surtout utile dans le contexte de... *
-
9. Si je devais réutiliser un indicateur conçu dans le cadre d'un modèle documentaire dans le contexte d'un autre modèle documentaire, cela serait... *
-
10. Si un client me demande de changer quelque chose dans un indicateur cela sera... *
-
11. Si je devais modéliser la plupart des indicateurs de Learning Analytics auxquels je peux penser avec ce méta-modèle cela serait... *
-

FIGURE D.3 – Questionnaire de l'étude auprès des modélisateurs (3/3).

-
- [1] Ángel F. AGUDO-PEREGRINA, Santiago IGLESIAS-PRADAS, Miguel Ángel CONDE-GONZÁLEZ et Ángel HERNÁNDEZ-GARCÍA. « Can we Predict Success from Log Data in VLEs? Classification of Interactions for Learning Analytics and their Relation with Performance in VLE-supported F2F and Online Learning ». In : *Computers in Human Behavior* 31.1 (fév. 2014), p. 542-550. DOI : [10.1016/j.chb.2013.05.031](https://doi.org/10.1016/j.chb.2013.05.031) (cf. p. 4, 37, 42).
- [2] Tuija ALASALMI. « Students Expectations on Learning Analytics : Learning Platform Features Supporting Self-regulated Learning ». In : *Proceedings of the 13th International Conference on Computer Supported Education*. SCITEPRESS - Science et Technology Publications, avr. 2021, p. 131-140. DOI : [10.5220/0010537101310140](https://doi.org/10.5220/0010537101310140) (cf. p. 115).
- [3] Mehrasa ALIZADEH. « Practicing the Scholarship of Teaching and Learning with Classroom Learning Analytics ». In : *7th International Congress on Advanced Applied Informatics (IIAI-AAI)*. Yonago, Japan : IEEE, 2018, p. 366-369. DOI : [10.1109/IIAI-AAI.2018.00079](https://doi.org/10.1109/IIAI-AAI.2018.00079) (cf. p. 110).
- [4] Naif Radi ALJOHANI et Hugh C. DAVIS. « Learning Analytics and Formative Assessment to Provide Immediate Detailed Feedback Using a Student Centered Mobile Dashboard ». In : *Seventh International Conference on Next Generation Mobile Apps, Services and Technologies*. Prague, Czech Republic : IEEE, 2013, p. 262-267. DOI : [10.1109/NGMAST.2013.54](https://doi.org/10.1109/NGMAST.2013.54) (cf. p. 46, 59, 109).
- [5] Naif Radi ALJOHANI, Ali DAUD, Rabeeh AYAZ ABBASI, Jalal S ALOWIBDI, Mohammad BASHERI et Muhammad Ahtisham ASLAM. « An Integrated Framework for Course Adapted Student Learning Analytics Dashboard ». In : *Computers in Human Behavior* 92 (2019), p. 679-690. DOI : [10.1016/j.chb.2018.03.035](https://doi.org/10.1016/j.chb.2018.03.035) (cf. p. 111, 116, 130).
- [6] ALYSSA FRIEND WISE et YEONJI JUNG. « Teaching with Analytics : Towards a Situated Model of Instructional Decision-Making ». In : *Journal of Learning Analytics* 6.2 (2019), 53-69. DOI : [10.18608/jla.2019.62.4](https://doi.org/10.18608/jla.2019.62.4) (cf. p. 111, 116, 129, 132).

- [7] Elaf Abu AMRIEH, Thair HAMTINI et Ibrahim ALJARAH. « Mining Educational Data to Predict Students' Academic Performance Using Ensemble Methods ». In : *International Journal of Database Theory and Application* 9.No. 8 (2016), p. 119-136. DOI : [10.14257/ijdta.2016.9.8.13](https://doi.org/10.14257/ijdta.2016.9.8.13) (cf. p. 152).
- [8] Thibaut ARRIBE. « Conception des chaînes éditoriales - Documenter l'activité et structurer le graphe documentaire pour améliorer la maîtrise de la rééditorialisation ». Thèse de doct. Université de Technologie de Compiègne, 2014, p. 255 (cf. p. 4, 19, 140).
- [9] Thibaut ARRIBE, Stéphane CROZAT, Bruno BACHIMONT et Sylvain SPINELLI. « Chaînes éditoriales numériques : allier efficacité et variabilité grâce à des primitives documentaires ». In : *Actes du colloque CIDE*. Tunis, Tunisie, 2012, p. 1-12 (cf. p. 20, 21).
- [10] Raheela ASIF, Agathe MERCERON, Syed ABBAS ALI et Najmi GHANI HAIDER. « Analyzing Undergraduate Students' Performance Using Educational Data Mining ». In : *Computers & Education* 113 (2017), p. 177-194. DOI : [10.1016/j.compedu.2017.05.007](https://doi.org/10.1016/j.compedu.2017.05.007) (cf. p. 56, 57).
- [11] David AZCONA, I-Han HSIAO et Alan F SMEATON. « Personalizing Computer Science Education by Leveraging Multimodal Learning Analytics ». In : *IEEE Frontiers in Education Conference (FIE)*. San Jose, CA, USA : IEEE, 2018, p. 1-9. DOI : [10.1109/FIE.2018.8658596](https://doi.org/10.1109/FIE.2018.8658596) (cf. p. 110, 123).
- [12] BMBF PROJECT PRIME (PROFESSIONAL REFLECTIVE MOBILE PERSONAL LEARNING. *Learning Context Project*. URL : <https://www.learning-context.de/text/1/Project> (cf. p. 48, 50).
- [13] Bruno BACHIMONT et Stéphane CROZAT. « Instrumentation numérique des documents : pour une séparation fonds/forme ». In : *Revue I3 - Information Interaction Intelligence* 4.1 (2004), p. 95 (cf. p. 20, 27).
- [14] Ryan S J D BAKER et Kalina YACEF. « The State of Educational Data Mining in 2009 : A Review and Future Visions ». In : *Journal of Educational Data Mining* 1.1 (2009), p. 3-16. DOI : [10.5281/zenodo.3554657](https://doi.org/10.5281/zenodo.3554657) (cf. p. 54, 56).
- [15] Armelle BALAS-CHANEL. « L'Entretien d'explicitation. Accompagner l'apprenant vers la métacognition explicite ». In : *Recherches & éducations* 1 (mar. 2002), p. 1-11. DOI : [10.4000/rechercheseducations.159](https://doi.org/10.4000/rechercheseducations.159) (cf. p. 173).

- [16] Duygu BEKTIK. « XIPIt : Updating the XIP Dashboard to Support Educators in Essay Marking at Higher Education ». In : *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. London, United Kingdom : Association for Computing Machinery, 2018, p. 1-4. DOI : [10.1145/3231644.3231696](https://doi.org/10.1145/3231644.3231696) (cf. p. [110](#), [116](#), [123](#)).
- [17] Jean BÉZIVIN. « On the Unification Power of Models ». In : *Software and Systems Modeling* 4.2 (2005), p. 171-188. DOI : [10.1007/s10270-005-0079-0](https://doi.org/10.1007/s10270-005-0079-0) (cf. p. [14](#)).
- [18] Sanyam BHARARA, Sai SABITHA et Abhay BANSAL. « Application of Learning Analytics Using Clustering Data Mining for Students' Disposition Analysis ». In : *Education and Information Technologies* 23.2 (2018), p. 957-984. DOI : [10.1007/s10639-017-9645-7](https://doi.org/10.1007/s10639-017-9645-7) (cf. p. [45](#), [54](#), [57](#), [149](#)).
- [19] Maria BRITO, Francisco MEDEIROS et Ed Porto BEZERRA. « An Infographics-based Tool for Monitoring Dropout Risk on Distance Learning in Higher Education ». In : *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)*. Magdeburg, Germany : IEEE, sept. 2019, p. 1-7. DOI : [10.1109/ITHET46829.2019.8937361](https://doi.org/10.1109/ITHET46829.2019.8937361) (cf. p. [110](#)).
- [20] Tom BROOS, Laurie PEETERS, Katrien VERBERT, Carolien VAN SOOM, Greet LANGIE et Tinne DE LAET. « Dashboard for Actionable Feedback on Learning Skills : Scalability and Usefulness ». In : *Learning and Collaboration Technologies. Technology in Education. LCT 2017*. Sous la dir. de Panayiotis ZAPHIRIS et Andri IOANNOU. T. 10296 LNCS. Springer, Cham, 2017, p. 229-241. DOI : [10.1007/978-3-319-58515-4_{_}18](https://doi.org/10.1007/978-3-319-58515-4_{_}18) (cf. p. [52](#)).
- [21] Tom BROOS, Katrien VERBERT, Greet LANGIE, Carolien VAN SOOM et Tinne DE LAET. « Multi-Institutional Positioning Test Feedback Dashboard for Aspiring Students : Lessons Learnt from a Case Study in Flanders ». In : *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. Sydney, New South Wales, Australia : Association for Computing Machinery, 2018, p. 51-55. DOI : [10.1145/3170358.3170419](https://doi.org/10.1145/3170358.3170419) (cf. p. [110](#), [116](#), [117](#)).
- [22] Susan BULL, Peter BRUSILOVSKY et Julio GUERRA. « Which Learning Visualisations to Offer Students? » In : *EC-TEL 2018*. January. Leeds, UK, 2018. DOI : [10.1007/978-3-319-98572-5_{_}40](https://doi.org/10.1007/978-3-319-98572-5_{_}40) (cf. p. [60](#)).

- [23] Rubiela CARRILLO, Yannick PRIÉ, Université JEAN MOULIN LYON, Élise LAVOUÉ et Jean MOULIN LYON. « Observing Learner Engagement on Mind Mapping Activities Using Learning Analytics ». In : *Transforming Learning with Meaningful Technologies. EC-TEL 2019*. Sous la dir. de Maren SCHEFFEL, Julien BROISIN, Viktoria PAMMER-SCHINDLER, Andri IOANNOU et Jan SCHNEIDER. Delft, The Netherland : Springer, Cham, 2019, p. 668-672. DOI : [10.1007/978-3-030-29736-7_{_}63](https://doi.org/10.1007/978-3-030-29736-7_{_}63) (cf. p. [111](#)).
- [24] Fabio CASSANO, Antonio PICCINNO, Teresa ROSELLI et Veronica ROSSANO. « Gamification and Learning Analytics to Improve Engagement in University Courses ». In : *Methodologies and Intelligent Systems for Technology Enhanced Learning, 8th International Conference. MIS4TEL 2018*. Sous la dir. de Tania Di MASCIO, Pierpaolo VITTORINI, Rosella GENNARI, Fernando De la PRIETA, Sara RODRÍGUEZ, Marco TEMPERINI, Ricardo Azambuja SILVEIRA, Elvira POPESCU et Loreto LANCIA. T. 804. Toledo, Spain : Springer International Publishing, juin 2019, p. 156-163. DOI : [10.1007/978-3-319-98872-6_{_}19](https://doi.org/10.1007/978-3-319-98872-6_{_}19) (cf. p. [110](#), [129](#)).
- [25] Mohamed Amine CHATTI, Anna Lea DYCKHOFF, Ulrik SCHROEDER et Hendrik THÜS. « A Reference Model for Learning Analytics ». In : *International Journal of Technology Enhanced Learning* 4.5-6 (2012), p. 318-331. DOI : [10.1504/IJTEL.2012.051815](https://doi.org/10.1504/IJTEL.2012.051815) (cf. p. [3](#), [42](#), [43](#), [65](#)).
- [26] Fang CHEN, Natalie RUIZ, Eric CHOI, Julien EPPS, M. Asif KHAWAJA, Ronnie TAIB, Bo YIN et Yang WANG. « Multimodal Behavior and Interaction as Indicators of Cognitive Load ». In : *ACM Transactions on Interactive Intelligent Systems* 2.4 (2012). DOI : [10.1145/2395123.2395127](https://doi.org/10.1145/2395123.2395127) (cf. p. [46](#)).
- [27] Qing CHEN, Zhen LI, Ting Chuen PONG et Huamin QU. « Designing Narrative Slideshows for Learning Analytics ». In : *IEEE Pacific Visualization Symposium*. T. 2019-April. IEEE Computer Society, avr. 2019, p. 237-246. DOI : [10.1109/PacificVis.2019.00036](https://doi.org/10.1109/PacificVis.2019.00036) (cf. p. [60](#)).
- [28] Florence CHERIGNY, Hassina EL KECHAI, Sébastien IKSAL, Marie LEFEVRE, Hugues LABARTHE et Vanda LUENGO. *L'analytique des apprentissages avec le numérique - Groupes thématiques de la Direction du numérique pour l'Éducation (DNE-TN2)*. Rapp. tech. Mar. 2020 (cf. p. [3](#), [45](#)).
- [29] Christophe CHOQUET et Sébastien IKSAL. « Usage Tracking Language : a Meta-Language for Modelling Tracks in TEL Systems ». In : *International Conference on Software and Data Technolo-*

- gies (ICSOFIT) (2006), p. 133-138. DOI : [10.5220/0001312701330138](https://doi.org/10.5220/0001312701330138) (cf. p. 67).
- [30] Christophe CHOQUET et Sébastien IKSAL. « Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un EIAH ». In : *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation* 14.1 (2007), p. 419-456. DOI : [10.3406/stice.2007.968](https://doi.org/10.3406/stice.2007.968) (cf. p. 54, 67, 68).
- [31] Chih Yueh CHOU, Shu Fen TSENG, Wen Chieh CHIH, Zhi Hong CHEN, Po Yao CHAO, K. Robert LAI, Chien Lung CHAN, Liang Chih YU et Yi Lung LIN. « Open Student Models of Core Competencies at the Curriculum Level : Using Learning Analytics for Student Reflection ». In : *IEEE Transactions on Emerging Topics in Computing* 5.1 (jan. 2017), p. 32-44. DOI : [10.1109/TETC.2015.2501805](https://doi.org/10.1109/TETC.2015.2501805) (cf. p. 59, 61).
- [32] Atika COHEN. « Manuel d'utilisation de Scenari Opale pour un cours de statistique ». In : *Statistique et Enseignement* 3.1 (2009), p. 73-101 (cf. p. 21).
- [33] Benoît COMBEMALE. *Ingénierie Dirigée par les Modèles (IDM) – État de l'art*. Toulouse, 2008 (cf. p. 11, 14, 15).
- [34] Benilda Eleonor V. COMENDADOR, Lorena W. RABAGO et Bartolome T. TANGUILIG. « An Educational Model Based on Knowledge Discovery in Databases (KDD) to Predict Learner's Behavior Using Classification Techniques ». In : *ICSPCC 2016 - IEEE International Conference on Signal Processing, Communications and Computing, Conference Proceedings*. Institute of Electrical et Electronics Engineers Inc., nov. 2016. DOI : [10.1109/ICSPCC.2016.7753623](https://doi.org/10.1109/ICSPCC.2016.7753623) (cf. p. 45).
- [35] Miguel Ángel CONDE, Ángel HÉRNANDEZ-GARCÍA, Francisco J. GARCÍA-PEÑALVO et María Luisa SÉIN-ECHALUCE. « Exploring Student Interactions : Learning Analytics Tools for Student Tracking ». In : *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. T. 9192. Springer Verlag, 2015, p. 50-61. DOI : [10.1007/978-3-319-20609-7_{_}6](https://doi.org/10.1007/978-3-319-20609-7_{_}6) (cf. p. 73).
- [36] IMS Global Learning CONSORTIUM. *Caliper Analytics | IMS Global Learning Consortium*. URL : <https://www.msglobal.org/activity/caliper> (cf. p. 51).
- [37] IMS Global Learning CONSORTIUM. *Initial xAPI/Caliper Comparison*. 2016. URL : <https://www.msglobal.org/initial-xapicaliper-comparison> (cf. p. 51).

- [38] Manfred CONSTAPEL, Dorian DOBERSTEIN, H. ULRICH HOPPE et Horst HELLBRUCK. « IKARion : Enhancing a Learning Platform with Intelligent Feedback to Improve Team Collaboration and Interaction in Small Groups ». In : *18th International Conference on Information Technology Based Higher Education and Training (ITHET)*. Magdeburg, Germany : IEEE, sept. 2019, p. 1-10. DOI : [10.1109/ITHET46829.2019.8937348](https://doi.org/10.1109/ITHET46829.2019.8937348) (cf. p. 61, 110).
- [39] Amélie CORDIER, Marie LEFEVRE, Pierre Antoine CHAMPIN, Olivier GEORGEON et Alain MILLE. « Trace-based reasoning - Modeling interaction traces for reasoning on experiences ». In : *FLAIRS 2013 - Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference* (2013), p. 363-368 (cf. p. 68-70).
- [40] Stéphane CROZAT. *Scenari, la chaîne éditoriale libre*. Eyrolles, 2007 (cf. p. 20).
- [41] Stéphane CROZAT. « Chaînes éditoriales et rééditorialisation de contenus numériques ». In : *Le document numérique à l'heure du web*. Sous la dir. de Lisette CALDERAN, Pascale LAURENT, Hélène LOWINGER et Jacques MILLET. ABDS, 2012, p. 179-220 (cf. p. 17-21).
- [42] Stéphane CROZAT. *Chaînes éditoriales*. 2016. URL : <https://stph.scenari-community.org/doc/ced.pdf> (cf. p. 23, 24).
- [43] Stéphane CROZAT. *Cours : Conception de bases de données*. 2020. URL : <https://librecours.scenari.ovh/module/nor1/nor1a1.xhtml?theme/theme/bdd02> (cf. p. 32).
- [44] Inès DABBEBI, Sébastien IKSAL, Jean-Marie GILLIOT, Madeth MAY et Serge GARLATTI. « Towards Adaptive Dashboards for Learning Analytic : An Approach for Conceptual Design and implementation ». In : *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU)*. Porto, Portugal, 2017, p. 120-131. DOI : [10.5220/0006325601200131](https://doi.org/10.5220/0006325601200131) (cf. p. 43, 60, 165).
- [45] Samba DIAW, Redouane LBATH et Bernard COULETTE. « État de l'art sur le développement logiciel basé sur les transformations de modèles ». In : *Techniques et sciences informatiques* 29.4-5 (2010), p. 505-536. DOI : [10.3166/tsi.29.505-536](https://doi.org/10.3166/tsi.29.505-536) (cf. p. 15, 16).
- [46] P. DILLENBOURG et P. TCHOUNIKINE. « Flexibility in Macro-Scripts for Computer-Supported Collaborative Learning ». In : *Journal of Computer Assisted Learning* 23.1 (jan. 2007), p. 1-13. DOI : [10.1111/j.1365-2729.2007.00191.x](https://doi.org/10.1111/j.1365-2729.2007.00191.x) (cf. p. 58).

- [47] John DUNLOSKY, Katherine A. RAWSON, Elizabeth J. MARSH, Mitchell J. NATHAN et Daniel T. WILLINGHAM. « Improving Students' Learning With Effective Learning Techniques ». In : *Psychological Science in the Public Interest* 14.1 (jan. 2013), p. 4-58. DOI : [10.1177/1529100612453266](https://doi.org/10.1177/1529100612453266) (cf. p. [155](#), [156](#)).
- [48] Anna Lea DYCKHOFF, Dennis ZIELKE, Mareike BÜLTMANN, Mohamed Amine CHATTI et Ulrik SCHROEDER. « Design and Implementation of a Learning Analytics Toolkit for Teachers ». In : *Educational Technology & Society* 15.3 (2012), p. 58-76 (cf. p. [106](#), [109](#), [116](#), [129](#), [162](#)).
- [49] Léonard Dumas Milne EDWARDS. « Conception de formes de relecture dans les chaînes éditoriales numériques ». Thèse de doct. 2016 (cf. p. [4](#)).
- [50] Aysu EZEN-CAN, Joseph F. GRAFSGAARD, James C. LESTER et Kristy Elizabeth BOYER. « Classifying Student Dialogue Acts with Multimodal Learning Analytics ». In : *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, LAK'15*. New York, NY, USA : Association for Computing Machinery, mar. 2015, p. 280-289. DOI : [10.1145/2723576.2723588](https://doi.org/10.1145/2723576.2723588) (cf. p. [57](#)).
- [51] A. F. ELGAMAL. « An Educational Data Mining Model for Predicting Student Performance in Programming Course ». In : *International Journal of Computer Applications* 70.17 (mai 2013), p. 22-28. DOI : [10.5120/12160-8163](https://doi.org/10.5120/12160-8163) (cf. p. [56](#)).
- [52] Jean-Marie FAVRE, Jacky ESTUBLIER et Mireille BLAY-FORNARINO. *L'ingénierie dirigée par les modèles : au-delà du MDA*. Lavoisier, 2006, p. 226 (cf. p. [13](#)).
- [53] Mingyu FENG, Neil T. HEFFERNAN et Kenneth R. KOEDINGER. « Predicting State Test Scores Better with Intelligent Tutoring Systems : Developing Metrics to Measure Assistance Required ». In : *Intelligent Tutoring Systems. ITS 2006. Lecture Notes in Computer Science*. T. 4053. Springer Berlin Heidelberg, 2006, p. 31-40. DOI : [10.1007/11774303_4](https://doi.org/10.1007/11774303_4) (cf. p. [56](#)).
- [54] Eduardo FERNANDES, Maristela HOLANDA, Marcio VICTORINO, Vinicius BORGES, Rommel CARVALHO et Gustavo Van ERVEN. « Educational Data Mining : Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil ». In : *Journal of Business Research* 94 (jan. 2019), p. 335-343. DOI : [10.1016/j.jbusres.2018.02.012](https://doi.org/10.1016/j.jbusres.2018.02.012) (cf. p. [55](#)).

- [55] Beatriz FLORIAN-GAVIRIA, Christian GLAHN et Ramon FABREGAT GESA. « A Software Suite for Efficient Use of the European Qualifications Framework in Online and Blended Courses ». In : *IEEE Transactions on Learning Technologies* 6.3 (2013), p. 283-296. DOI : [10.1109/TLT.2013.18](https://doi.org/10.1109/TLT.2013.18) (cf. p. 109).
- [56] Beatriz FLORIAN, Christian GLAHN, Hendrik DRACHSLER, Marcus SPECHT et Ramón FABREGAT GESA. « Activity-based Learner-models for Learner Monitoring and Recommendations in Moodle ». In : *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. T. 6964 LNCS. Springer, Berlin, Heidelberg, 2011, p. 111-124. DOI : [10.1007/978-3-642-23985-4_{_}10](https://doi.org/10.1007/978-3-642-23985-4_{_}10) (cf. p. 47).
- [57] Valentina FRANZONI, Paolo MENGONI et Alfredo MILANI. « Dimensional Morphing Interface for Dynamic Learning Evaluation ». In : *22nd International Conference Information Visualisation (IV)*. Fisciano, Italy : Institute of Electrical et Electronics Engineers Inc., déc. 2018, p. 332-337. DOI : [10.1109/iV.2018.00063](https://doi.org/10.1109/iV.2018.00063) (cf. p. 61, 110).
- [58] Meiyuzi GAO, Philip KORTUM et Frederick L. OSWALD. « Multi-Language Toolkit for the System Usability Scale ». In : *International Journal of Human-Computer Interaction* 36.20 (2020), p. 1883-1901. DOI : [10.1080/10447318.2020.1801173](https://doi.org/10.1080/10447318.2020.1801173) (cf. p. 172).
- [59] Inc. GARTNER. *Gartner Says Advanced Analytics Is a Top Business Priority*. 2014. URL : <https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority> (cf. p. 38, 39, 41, 112).
- [60] Dragan GAŠEVIĆ et Mykola PECHENIZKIY. « Let's Grow Together : Tutorials on Learning Analytics Methods ». In : *Journal of Learning* 3.3 (2016), p. 5-8. DOI : [10.18608/jla.2016.33.2](https://doi.org/10.18608/jla.2016.33.2) (cf. p. 3, 58).
- [61] Dragan GAŠEVIĆ, Yi Shan TSAI, Shane DAWSON et Abelardo PARDO. « How do We Start? An Approach to Learning Analytics Adoption in Higher Education ». In : *International Journal of Information and Learning Technology* 36.4 (2019), p. 342-353. DOI : [10.1108/IJILT-02-2019-0024](https://doi.org/10.1108/IJILT-02-2019-0024) (cf. p. 59).
- [62] E. GEBERS et S. CROZAT. « Chaînes éditoriales Scenari et unité ICS ». In : *Distances et savoirs* 7.3 (2010), 421-442. DOI : [10.3166/ds.7.421-442](https://doi.org/10.3166/ds.7.421-442) (cf. p. 19, 20).

- [63] Anouk GELAN, Greet FASTRÉ, Martine VERJANS, Niels MARTIN, Gert JANSSENSWILLEN, Mathijs CREEMERS, Jonas LIEBEN, Benoît DEPAIRE et Michael THOMAS. « Affordances and Limitations of Learning Analytics for Computer-Assisted Language Learning : a Case Study of the VITAL Project ». In : *Computer Assisted Language Learning* 31.3 (2018), p. 294-319. DOI : [10.1080/09588221.2017.1418382](https://doi.org/10.1080/09588221.2017.1418382) (cf. p. 110, 116).
- [64] Sten GOVAERTS, Katrien VERBERT, Erik DUVAL et Abelardo PARDO. « The Student Activity Meter for Awareness and Self-Reflection ». In : *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. Austin, Texas, USA : Association for Computing Machinery, 2012, p. 869-884. DOI : [10.1145/2212776.2212860](https://doi.org/10.1145/2212776.2212860) (cf. p. 47).
- [65] Arthur C. GRAESSER, Danielle S. McNAMARA et Jonna M. KULIKOWICH. « Coh-Metrix : Providing Multilevel Analyses of Text Characteristics ». In : *Educational Researcher* 40.5 (2011), p. 223-234. DOI : [10.3102/0013189X11413260](https://doi.org/10.3102/0013189X11413260) (cf. p. 58).
- [66] Jeff GRANN et Deborah BUSHWAY. « Competency Map : Visualizing Student Learning to Promote Student Success ». In : *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*. New York, New York, USA : ACM Press, 2014, p. 168-172. DOI : [10.1145/2567574.2567622](https://doi.org/10.1145/2567574.2567622) (cf. p. 109).
- [67] Wolfgang GRELLER et Hendrik DRACHSLER. « Translating Learning into Numbers : A Generic Framework for Learning Analytics ». In : *Journal of Educational Technology & Society* 15.3 (2012), p. 42-57. DOI : [10.2307/jeductechsoci.15.3.42](https://doi.org/10.2307/jeductechsoci.15.3.42) (cf. p. 42, 43, 64).
- [68] Anatoliy GRUZD, Ted ROGERS et Nadia CONROY. « Learning Analytics Dashboard for Teaching with Twitter ». In : *Proceedings of the 53rd Hawaii International Conference on System Sciences*. Maui, Hawaii, 2020, p. 2708-2717. DOI : [10.24251/HICSS.2020.330](https://doi.org/10.24251/HICSS.2020.330) (cf. p. 111, 116, 123).
- [69] Damien GUILLAUME, Stéphane CROZAT, Laurent RIVET, Manuel MAJADA et Xavier HENNEQUIN. *Chaînes éditoriales numériques*. 2015 (cf. p. 20-22, 33).
- [70] Rebecca GUILLOT, Jeremie SEANOSKY, Isabelle GUILLOT, David BOULANGER, Claudia GUILLOT, Vivekanandan KUMAR, Shawn N. FRASER et KINSHUK. « Assessing Learning Analytics Systems Impact by Summative Measures ». In : *IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. Mumbai,

- India : IEEE, août 2018, p. 188-190. DOI : [10.1109/ICALT.2018.00051](https://doi.org/10.1109/ICALT.2018.00051) (cf. p. 110, 128).
- [71] Brigitte GUYOT. *Introduction à l'ingénierie documentaire et aux sciences de l'information*. 2011. URL : https://archivesic.ccsd.cnrs.fr/sic_00665267 (cf. p. 18, 20).
- [72] Fatima HARRAK, François BOUCHET, Vanda LUENGO et Rémi BACHELET. « Towards Improving Students' Forum Posts Categorization in MOOCs and Impact on Performance Prediction ». In : *Proceedings of the 6th 2019 ACM Conference on Learning at Scale, L@S*. New York, NY, USA : Association for Computing Machinery, Inc, juin 2019, p. 1-4. DOI : [10.1145/3330430.3333661](https://doi.org/10.1145/3330430.3333661) (cf. p. 56, 57).
- [73] Davinia HERNÁNDEZ-LEO, Roberto MARTINEZ-MALDONADO, Abelardo PARDO, Juan A. MUÑOZ-CRISTÓBAL et María J. RODRÍGUEZ-TRIANA. « Analytics for Learning Design : A Layered Framework and Tools ». In : *British Journal of Educational Technology* 50.1 (jan. 2019), p. 139-152. DOI : [10.1111/bjet.12645](https://doi.org/10.1111/bjet.12645) (cf. p. 66, 67).
- [74] Christothea HERODOTOU, Martin HLOSTA, Avinash BOROOWA, Bart RIENTIES, Zdenek ZDRAHAL et Chrysoula MANGAFA. « Empowering Online Teachers Through Predictive Learning Analytics ». In : *British Journal of Educational Technology* 50.6 (nov. 2019), p. 3064-3079. DOI : [10.1111/bjet.12853](https://doi.org/10.1111/bjet.12853) (cf. p. 111, 130).
- [75] Dirk IFENTHALER et Chathuranga WIDANAPATHIRANA. « Development and Validation of a Learning Analytics Framework : Two Case Studies Using Support Vector Machines ». In : *Technology, Knowledge and Learning* 19.1-2 (mai 2014), p. 221-240. DOI : [10.1007/s10758-014-9226-4](https://doi.org/10.1007/s10758-014-9226-4) (cf. p. 43, 46).
- [76] ADL INITIATIVE. *Experience API (xAPI) Standard*. URL : <https://adlnet.gov/projects/xapi/> (cf. p. 49).
- [77] Jean-Marc JÉZÉQUEL, Benoît COMBEMALE et Didier VOJTISEK. *Ingénierie dirigée par les modèles - Des concepts à la pratique*. Paris : Ellipses, 2012, p. 122 (cf. p. 11-13, 63).
- [78] Ioana JIVET, Maren SCHEFFEL, Hendrik DRACHSLER et Marcus SPECHT. « Awareness Is Not Enough : Pitfalls of Learning Analytics Dashboards in the Educational Practice ». In : *Data Driven Approaches in Digital Education EC-TEL 2017*. Sous la dir. d'Élise LAVOUÉ, Hendrik DRACHSLER, Katrien VERBERT, Julien BROISIN et Mar PÉREZ-SANAGUSTÍN. Springer, Cham, 2017, p. 82-96. DOI : [10.1007/978-3-319-66610-5_{_}7](https://doi.org/10.1007/978-3-319-66610-5_{_}7) (cf. p. 132, 135).

- [79] Irish C. JUANATAS et Roben A. JUANATAS. « Predictive Data Analytics using Logistic Regression for Licensure Examination Performance ». In : *Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019*. Institute of Electrical et Electronics Engineers Inc., déc. 2019, p. 251-255. DOI : [10.1109/ICCIKE47802.2019.9004386](https://doi.org/10.1109/ICCIKE47802.2019.9004386) (cf. p. 56).
- [80] Asem KASEM, Siti Nur Asyiqin Mohd SHAHRIN et Au Thien WAN. « Learning Analytics in Universiti Teknologi Brunei : Predicting Graduates Performance ». In : *Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018*. Institute of Electrical et Electronics Engineers Inc., oct. 2018. DOI : [10.1109/ICACCAF.2018.8776690](https://doi.org/10.1109/ICACCAF.2018.8776690) (cf. p. 56).
- [81] Mohammad KHALIL et Martin EBNER. « Learning Analytics : Principles and Constraints ». In : *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*. June. Chesapeake, VA, USA : AACE, 2015, p. 1326-1336. DOI : [10.13140/RG.2.1.1733.2083](https://doi.org/10.13140/RG.2.1.1733.2083) (cf. p. 39, 41).
- [82] Barbara KITCHENHAM et Stuart CHARTERS. *Guidelines for Performing Systematic Literature Reviews in Software Engineering (Version 2.3)*. Rapp. tech. Keele University et University of Durham, 2007, p. 65. DOI : [10.1145/1134285.1134500](https://doi.org/10.1145/1134285.1134500). URL : <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.471&rep=rep1&type=pdf> (cf. p. 101).
- [83] Kirsty KITTO et Simon KNIGHT. « Practical Ethics for Building Learning Analytics ». In : *British Journal of Educational Technology* 50.6 (nov. 2019), p. 2855-2870. DOI : [10.1111/bjet.12868](https://doi.org/10.1111/bjet.12868) (cf. p. 54).
- [84] Kirsty KITTO, John WHITMER, Aaron E. SILVERS et Michael WEBB. « Creating Data for Learning Analytics Ecosystems ». In : *SOLAR Position Paper* September (2020), p. 1-43 (cf. p. 3, 6, 26, 51-54).
- [85] Anneke KLEPPE, Jos WARMER et Wim BAST. *MDA Explained - The Model Driven Architecture : Practice and Promise*. 1^{re} éd. Addison-Wesley Professional, 2003, p. 193 (cf. p. 15, 16).
- [86] Simon KNIGHT et Simon Buckingham SHUM. « Theory and Learning Analytics ». In : *Handbook of Learning Analytics*. Sous la dir. de Charles LANG, George SIEMENS, Alyssa WISE et Dragan GAŠEVIĆ. 1^{re} éd. SOLAR, 2017. Chap. 1, p. 356. DOI : [10.18608/hla17](https://doi.org/10.18608/hla17) (cf. p. 54).

- [87] Rustici Software LLC. *What is xAPI aka the Experience API or Tin Can API*. URL : https://xapi.com/overview/?utm_source=google&utm_medium=natural_search (cf. p. 50).
- [88] Carine LALLEMAND et Guillaume GRONIER. *Méthodes de design UX*. 2eme. Eyrolles, 2018, p. 712 (cf. p. 172).
- [89] Dictionnaire de français LAROUSSE. *Définitions : méthode*. URL : <https://www.larousse.fr/dictionnaires/francais/méthode/50965> (cf. p. 64).
- [90] Alexis LEBIS. « Capitaliser les processus d' analyse de traces d' apprentissage : modélisation ontologique & assistance à la réutilisation ». Thèse de doct. Sorbonne Université, 2019, p. 302 (cf. p. 73-75, 126, 166).
- [91] Philipp LEITNER, Mohammad KHALIL et Martin EBNER. « Learning Analytics in Higher Education — A Literature Review ». In : *Learning Analytics : Fundamentals, Applications, and Trends*. Sous la dir. d' Alejandro PEÑA-AYALA. T. 94. Studies in Systems, Decision and Control September. Cham : Springer International Publishing, 2017. Chap. 1, p. 1-23. DOI : [10.1007/978-3-319-52977-6](https://doi.org/10.1007/978-3-319-52977-6) (cf. p. 102).
- [92] Phil LONG et George SIEMENS. « Penetrating the Fog : Analytics in Learning and Education ». In : *EDUCAUSE Review* 46.5 (2011), p. 31-40. URL : <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education> (cf. p. 3).
- [93] Ouajdi MANAI, Hiroyuki YAMADA et Christopher THORN. « Real-Time Indicators and Targeted Supports : Using Online Platform Data to Accelerate Student Learning ». In : *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. Edinburgh, United Kingdom : Association for Computing Machinery, 2016, p. 183-187. DOI : [10.1145/2883851.2883942](https://doi.org/10.1145/2883851.2883942) (cf. p. 109).
- [94] Roberto MARTINEZ-MALDONADO. « A Handheld Classroom Dashboard : Teachers' Perspectives on the Use of Real-Time Collaborative Learning Analytics ». In : *International Journal of Computer-Supported Collaborative Learning* 14(3) (2019), p. 383-411. DOI : [10.1007/s11412-019-09308-z](https://doi.org/10.1007/s11412-019-09308-z) (cf. p. 111, 116).
- [95] Wannisa MATCHA, Noraayu Ahmad UZIR, Dragan GASEVIC et Abelardo PARDO. « A Systematic Review of Empirical Studies on Learning Analytics Dashboards : A Self-Regulated Learning Perspective ». In : *IEEE Transactions on Learning Technologies*

- 13.2 (avr. 2020), p. 226-245. DOI : [10.1109/TLT.2019.2916802](https://doi.org/10.1109/TLT.2019.2916802) (cf. p. 101).
- [96] Yoshiaki MATSUZAWA, Yoshiki TANAKA, Tomoya KITANI et San-shiro SAKAI. « A Demonstration of Evidence-Based Action Research Using Information Dashboard in Introductory Programming Education ». In : *Learning with and about Technologies and Computing. WCCE 2017*. Sous la dir. d'Arthur TATNALL et Mary WEBB. T. 515. Dublin, Ireland : Springer, Cham, 2017, p. 619-629. DOI : [10.1007/978-3-319-74310-3_{_}62](https://doi.org/10.1007/978-3-319-74310-3_{_}62) (cf. p. 109, 129).
- [97] Riccardo MAZZA. *Introduction to Visual Representations*. Springer London, fév. 2009. DOI : [10.1007/978-1-84800-219-7](https://doi.org/10.1007/978-1-84800-219-7) (cf. p. 59).
- [98] Riccardo MAZZA et Vania DIMITROVA. « CourseVis : A Graphical Student Monitoring Tool for Supporting Instructors in Web-based Distance Courses ». In : *International Journal of Human Computer Studies* 65.2 (fév. 2007), p. 125-139. DOI : [10.1016/j.ijhcs.2006.08.008](https://doi.org/10.1016/j.ijhcs.2006.08.008) (cf. p. 47).
- [99] J. McDONALD, R. J. BIRD, A. ZOUAQ et A. C.M. MOSKAL. « Short Answers to Deep Questions : Supporting Teachers in Large-Class Settings ». In : *Journal of Computer Assisted Learning* 33.4 (2017), p. 306-319. DOI : [10.1111/jcal.12178](https://doi.org/10.1111/jcal.12178) (cf. p. 109, 110, 133).
- [100] Brian J MCNELLY, Paul GESTWICKI, J Holden HILL, Philip PARLI-HORNE et Erika JOHNSON. « Learning Analytics for Collaborative Writing : A Prototype and Case Study ». In : *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Vancouver, British Columbia, Canada : Association for Computing Machinery, 2012, p. 222-225. DOI : [10.1145/2330601.2330654](https://doi.org/10.1145/2330601.2330654) (cf. p. 109, 129).
- [101] Carolina MEJIA, Beatriz FLORIAN, Ravi VATRAPU, Susan BULL, Sergio GOMEZ et Ramon FABREGAT. « A Novel Web-Based Approach for Visualization and Inspection of Reading Difficulties on University Students ». In : *IEEE Transactions on Learning Technologies* 10.1 (jan. 2017), p. 53-67. DOI : [10.1109/TLT.2016.2626292](https://doi.org/10.1109/TLT.2016.2626292) (cf. p. 109).
- [102] Agathe MERCERON. « Educational Data Mining / Learning Analytics : Methods, Tasks and Current Trends ». In : *Proceedings of DeLFI Workshops 2015 co-located with 13th e-Learning Conference of the German Computer Society (DeLFI 2015)*. Sous la dir. de Sabine RATHMAYER et Hans PONGRATZ. München, Germany : DeLFI 2015, 2015, p. 101-109 (cf. p. 56).

- [103] V.L. MIGUÉIS, Ana FREITAS, Paulo J.V. GARCIA et André SILVA. « Early Segmentation of Students According to their Academic Performance : A Predictive Modelling Approach ». In : *Decision Support Systems* 115 (nov. 2018), p. 36-51. DOI : [10.1016/j.dss.2018.09.001](https://doi.org/10.1016/j.dss.2018.09.001) (cf. p. 45, 58, 152).
- [104] Michael G. MOORE. « Editorial : Three Types of Interaction ». In : *American Journal of Distance Education* 3.2 (jan. 1989), p. 1-7. DOI : [10.1080/08923648909526659](https://doi.org/10.1080/08923648909526659) (cf. p. 46, 47, 52).
- [105] Arham MUSLIM, Mohamed Amine CHATTI, Memoona MUGHAL et Ulrik SCHROEDER. « The Goal-Question-Indicator Approach for Personalized Learning Analytics ». In : *CSEDU 2017 - Proceedings of the 9th International Conference on Computer Supported Education* 1 (2017), p. 371-378. DOI : [10.5220/0006319803710378](https://doi.org/10.5220/0006319803710378) (cf. p. 60).
- [106] Aneet Dharmavaram NARENDRANATH. « Raspberry Pi based learning center usage tracking system for optimal resource allocation ». In : *2018 IEEE Frontiers in Education Conference (FIE)*. San Jose, CA, USA : IEEE, oct. 2018, p. 1-5. DOI : [10.1109/FIE.2018.8659225](https://doi.org/10.1109/FIE.2018.8659225) (cf. p. 110).
- [107] Andres NEYEM, Juan DIAZ-MOSQUERA, Jorge MUNOZ-GAMA et Jaime NAVON. « Understanding Student Interactions in Capstone Courses to Improve Learning Experiences ». In : *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. Seattle, Washington, USA : Association for Computing Machinery, 2017, p. 423-428. DOI : [10.1145/3017680.3017716](https://doi.org/10.1145/3017680.3017716) (cf. p. 109, 123).
- [108] Roger NKAMBOU, Jacqueline BOURDEAU et Riichiro MIZOGUCHI, éd. *Advances in Intelligent Tutoring Systems*. T. 308. Studies in Computational Intelligence. Berlin, Heidelberg : Springer Berlin Heidelberg, 2010. DOI : [10.1007/978-3-642-14363-2](https://doi.org/10.1007/978-3-642-14363-2) (cf. p. 16).
- [109] OBJECT MANAGEMENT GROUP. *Meta Object Facility (MOF) 2.0 Core Specification*. 2006. URL : <https://www.omg.org/spec/MOF/2.0> (cf. p. 13).
- [110] Xavier OCHOA. « Multimodal Learning Analytics ». In : *The Handbook of Learning Analytics*. Sous la dir. de Charles LANG, George SIEMENS, Alyssa WISE et Dragan GAŠEVIĆ. 1^{re} éd. SOLAR, 2017. Chap. 11, p. 129-141. DOI : [10.18608/hla17.011](https://doi.org/10.18608/hla17.011) (cf. p. 46).

- [111] Taiwo Olapeju OLALEYE et Olufunke Rebecca VINCENT. « A Predictive Model for Students' Performance and Risk Level Indicators Using Machine Learning ». In : *2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020*. Institute of Electrical et Electronics Engineers Inc., mar. 2020. DOI : [10.1109/ICMCECS47690.2020.240897](https://doi.org/10.1109/ICMCECS47690.2020.240897) (cf. p. 56).
- [112] Katia OLIVER-QUELENNEC. « Vers une conception participative de tableaux de bord étudiants adaptatifs avec PADDLE ». In : *8e Rencontres Jeunes Chercheurs en EIAH*. Poitiers, France, 2020 (cf. p. 60, 117).
- [113] Zacharoula PAPAMITSIOU, Michail N. GIANNAKOS et Xavier OCHOA. « From Childhood to Maturity : Are We There yet? Mapping the Intellectual Progress in Learning Analytics during the Past Decade ». In : *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, LAK'20*. Frankfurt, Germany : Association for Computing Machinery, mar. 2020, p. 559-568. DOI : [10.1145/3375462.3375519](https://doi.org/10.1145/3375462.3375519) (cf. p. 59).
- [114] Abelardo PARDO, Feifei HAN et Robert A. ELLIS. « Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic Performance ». In : *IEEE Transactions on Learning Technologies* 10.1 (2017), p. 82-92. DOI : [10.1109/TLT.2016.2639508](https://doi.org/10.1109/TLT.2016.2639508) (cf. p. 42).
- [115] Abelardo PARDO, Jelena JOVANOVIĆ, Negin MIRRIAHI, Shane DAWSON, Roberto MARTINEZ-MALDONADO, Australia ROBERTO-MARTINEZ et Dragan GAŠEVIĆ. « Generating Actionable Predictive Models of Academic Performance ». In : *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. Edinburgh, United Kingdom : Association for Computing Machinery, 2016, p. 474-478. DOI : [10.1145/2883851.2883870](https://doi.org/10.1145/2883851.2883870) (cf. p. 109, 123).
- [116] Zachary A. PARDOS, Neil T. HEFFERNAN, Brigham ANDERSON et Cristina L. HEFFERNAN. « The Effect of Model Granularity on Student Performance Prediction Using Bayesian Networks ». In : *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. T. 4511 LNCS. Springer Verlag, 2007, p. 435-439. DOI : [10.1007/978-3-540-73078-1_60](https://doi.org/10.1007/978-3-540-73078-1_60) (cf. p. 56).
- [117] Michael Quinn PATTON. *Qualitative Evaluation and Research Methods*. 2nd. SAGE Publications, 1990, p. 532 (cf. p. 83, 171).

- [118] Daniel PEREZ-BERENGUER, Mathieu KESSLER et Jesus GARCIA-MOLINA. « A Customizable and Incremental Processing Approach for Learning Analytics ». In : *IEEE Access* 8 (2020), p. 36350-36362. DOI : [10.1109/ACCESS.2020.2975384](https://doi.org/10.1109/ACCESS.2020.2975384) (cf. p. [111](#), [116](#), [117](#)).
- [119] Ed de QUINCEY, Chris BRIGGS, Theocharis KYRIACOU et Richard WALLER. « Student Centred Design of a Learning Analytics System ». In : *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. Tempe, AZ, USA : Association for Computing Machinery, 2019, p. 353-362. DOI : [10.1145/3303772.3303793](https://doi.org/10.1145/3303772.3303793) (cf. p. [110](#), [116](#), [117](#), [123](#)).
- [120] Elvira G RINCÓN-FLORES, Juanjo MENA, Eunice LÓPEZ-CAMACHO et Omar OLMOS. « Adaptive Learning Based on AI with Predictive Algorithms ». In : *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*. Le[\]{o}n, Spain : Association for Computing Machinery, 2019, p. 607-612. DOI : [10.1145/3362789.3362869](https://doi.org/10.1145/3362789.3362869) (cf. p. [110](#), [123](#)).
- [121] Tobias ROHLOFF, Sören OLDAG, Jan RENZ et Christoph MEINEL. « Utilizing Web Analytics in the Context of Learning Analytics for Large-Scale Online Learning ». In : *IEEE Global Engineering Education Conference (EDUCON)*. Dubai, United Arab Emirates : IEEE, avr. 2019, p. 296-305. DOI : [10.1109/EDUCON.2019.8725118](https://doi.org/10.1109/EDUCON.2019.8725118) (cf. p. [110](#)).
- [122] Vicente-Arturo ROMERO-ZALDIVAR, Abelardo PARDO, Daniel BURGOS et Carlos DELGADO KLOOS. « Monitoring Student Progress Using Virtual Appliances : A Case Study ». In : *Computers & Education* 58.4 (2012), p. 1058-1067. DOI : [10.1016/j.compedu.2011.12.003](https://doi.org/10.1016/j.compedu.2011.12.003) (cf. p. [39](#), [43](#), [52](#), [53](#)).
- [123] Fatemeh SALEHIAN KIA, Stephanie D TEASLEY, Marek HATALA, Stuart A KARABENICK et Matthew KAY. « How Patterns of Students Dashboard Use Are Related to Their Achievement and Self-Regulatory Engagement ». In : *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. Frankfurt, Germany : Association for Computing Machinery, 2020, p. 340-349. DOI : [10.1145/3375462.3375472](https://doi.org/10.1145/3375462.3375472) (cf. p. [111](#)).
- [124] Clara SCHUMACHER et Dirk IFENTHALER. « Features Students Really Expect from Learning Analytics ». In : *13th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA)*. Mannheim, Germany : International Association

- for Development of the Information Society, 2016, p. 67-76 (cf. p. 40-42, 61, 76, 89, 96, 155).
- [125] Beat A. SCHWENDIMANN, Maria Jesus RODRIGUEZ-TRIANA, Andrii VOZNIUK, Luis P. PRIETO, Mina Shirvani BOROUJENI, Adrian HOLZER, Denis GILLET et Pierre DILLENBOURG. « Perceiving Learning at a Glance : A Systematic Literature Review of Learning Dashboard Research ». In : *IEEE Transactions on Learning Technologies* 10.1 (2017), p. 30-41. DOI : [10.1109/TLT.2016.2599522](https://doi.org/10.1109/TLT.2016.2599522) (cf. p. 42, 45-47, 52, 59, 60, 101).
- [126] John SCOTT et Philip NICHOLS. « Learning Analytics as Assemblage : Criticality and Contingency in Online Education ». In : *Research in Education* 98.1 (2017), p. 83-105. DOI : [10.1177/0034523717723391](https://doi.org/10.1177/0034523717723391) (cf. p. 109, 110, 128).
- [127] Shiva SHABANINEJAD, Hassan KHOSRAVI, Marta INDULSKA, Aneesh BAKHARIA et Pedro ISAIAS. « Automated Insightful Drill-Down Recommendations for Learning Analytics Dashboards ». In : *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, LAK '20*. January. Frankfurt, Germany : Association for Computing Machinery, 2020, p. 41-46. DOI : [10.1145/3375462.3375539](https://doi.org/10.1145/3375462.3375539) (cf. p. 42-44, 60).
- [128] Antonette SHIBANI, Simon KNIGHT et Simon Buckingham SHUM. « Contextualizable Learning Analytics Design ». In : *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. January. New York, NY, USA : ACM, mar. 2019, p. 210-219. DOI : [10.1145/3303772.3303785](https://doi.org/10.1145/3303772.3303785) (cf. p. 58, 134, 135).
- [129] George SIEMENS. *Learning and Academic Analytics*. 2011. URL : <https://www.learninganalytics.net/uncategorized/learning-and-academic-analytics/> (cf. p. 37).
- [130] George SIEMENS et Ryan S.J.D. BAKER. « Learning Analytics and Educational Data Mining : Towards Communication and Collaboration ». In : *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. LAK'12. Vancouver, British Columbia, Canada : Association for Computing Machinery, 2012, 252-254. DOI : [10.1145/2330601.2330661](https://doi.org/10.1145/2330601.2330661) (cf. p. 38, 54).
- [131] James M SNELL et Evan PRODROMOU. *Activity Streams 2.0*. 2017. URL : <https://www.w3.org/TR/activitystreams-core/> (cf. p. 48, 49).
- [132] James M SNELL, Martin ATKINS, Will NORRIS, Chris MESSINA, Monica WILKINSON et Rob DOLIN. *Activity Streams Working Group : JSON Activity Streams 1.0*. 2011. URL : <https://activitystreams/specs/json/1.0/> (cf. p. 47).

- [133] Pedro STRECHT, Luís CRUZ, Carlos SOARES, João MENDES-MOREIRA et Rui ABREU. « A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance ». In : *International Conference on Educational Data Mining (EDM)*. Sous la dir. d'International Educational Data Mining Society. Madrid, Spain : International Educational Data Mining Society, juin 2015 (cf. p. 45, 56).
- [134] Kaiwen SUN, Abraham H MHAIDLI, Sonakshi WATEL, Christopher A BROOKS et Florian SCHAUB. « It's My Data! Tensions Among Stakeholders of a Learning Analytics Dashboard ». In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland Uk : Association for Computing Machinery, 2019, p. 1-14. DOI : [10.1145/3290605.3300824](https://doi.org/10.1145/3290605.3300824) (cf. p. 110, 123).
- [135] Behnam TARAGHI, Martin EBNER, Laurens DE VOCHT, Selver SOFTIC, Benham TARAGHI, Erik MANNENS et Rik VAN DE WALLE. « Monitoring Learning Activities in PLE Using Semantic Modelling of Learner Behaviour ». In : *Human Factors in Computing and Informatics. SouthCHI 2013*. Sous la dir. d'Andreas HOLZINGER, Martina ZIEFLE, Martin HITZ et Matjaž DEBEVC. Maribor, Slovenia : Springer, Berlin, Heidelberg, 2013, p. 74-90. DOI : [10.1007/978-3-642-39062-3_5](https://doi.org/10.1007/978-3-642-39062-3_5) (cf. p. 109).
- [136] Harmid TARMAZDI, Rebecca VIVIAN, Claudia SZABO, Katrina FALKNER et Nickolas FALKNER. « Using Learning Analytics to Visualise Computer Science Teamwork ». In : *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*. Vilnius, Lithuania : Association for Computing Machinery, 2015, p. 165-170. DOI : [10.1145/2729094.2742613](https://doi.org/10.1145/2729094.2742613) (cf. p. 109).
- [137] Giles TEWKESBURY, Simon CHESTER, David SANDERS et Manish MALIK. « Monitoring Attendance and Its Impact on Engineering Students ». In : *Proceedings of the 6th Annual Symposium of the United Kingdom & Ireland Engineering Education Research Network*. Sous la dir. de Manish MALIK, Jane ANDREWS, Robin CLARK, Roger PENLINGTON et Rebecca BROADBENT. Portsmouth, United Kingdom : University of Portsmouth, 2018, p. 116-128. DOI : [9781861376695](https://doi.org/10.1145/9781861376695) (cf. p. 110).
- [138] Hendrik THÜS, Mohamed Amine CHATTI, Roman BRANDT et Ulrik SCHROEDER. « Evolution of Interests in the Learning Context Data Model ». In : *Design for Teaching and Learning in a Networked World*. Sous la dir. de Gráinne CONOLE, Tomaž KLO-

- BUČAR, Christoph RENSING, Johannes KONERT et Elise LAVOUÉ. Cham : Springer International Publishing, 2015, p. 479-484. DOI : [10.1007/978-3-319-24258-3_{_}43](https://doi.org/10.1007/978-3-319-24258-3_{_}43) (cf. p. 49).
- [139] Hao TIAN, Song LAI et Fati WU. « Does Time Play a Role? Prediction of Learning Performance with Time-use Habits in Online Assignments ». In : *Proceedings - International Joint Conference on Information, Media, and Engineering, IJCIME 2019*. Institute of Electrical et Electronics Engineers Inc., déc. 2019, p. 473-477. DOI : [10.1109/IJCIME49369.2019.00101](https://doi.org/10.1109/IJCIME49369.2019.00101) (cf. p. 56).
- [140] Saida ULFA, Izzull FATAWI, Ence SURAHMAN et Hayashi YUSUKE. « Investigating Learners' Perception of Learning Analytics Dashboard to Improve Learning Interaction in Online Learning System ». In : *2019 5th International Conference on Education and Technology (ICET)*. Malang, Indonesia : IEEE, oct. 2019, p. 49-54. DOI : [10.1109/ICET48172.2019.8987229](https://doi.org/10.1109/ICET48172.2019.8987229). URL : <https://ieeexplore.ieee.org/document/8987229/> (cf. p. 42, 110).
- [141] Vladimir L. USKOV, Jeffrey P. BAKKEN, Ashok SHAH, Timothy KROCK, Alexander USKOV, Jitendra SYAMALA et Rama RACHAKONDA. « Smart Learning Analytics : Conceptual Modeling and Agile Engineering ». In : *Smart Education and e-Learning 2018*. Sous la dir. de Vladimir USKOV, Robert J HOWLETT, Lakhmi C. JAIN et Ljubo VLACIC. T. 99. Gold Coast, Australia : Springer Science et Business Media Deutschland GmbH, juin 2019, p. 3-16. DOI : [10.1007/978-3-319-92363-5_{_}1](https://doi.org/10.1007/978-3-319-92363-5_{_}1) (cf. p. 110, 130).
- [142] Jonas VACLAVEK, Jakub KUZILEK, Jan SKOCILAS, Zdenek ZDRHAL et Viktor FUGLIK. « Learning Analytics Dashboard Analysing First-Year Engineering Students ». In : *Lifelong Technology-Enhanced Learning - 13th European Conference on Technology Enhanced Learning*. Sous la dir. de Viktoria PAMMER-SCHINDLER, Mar PÉREZ-SANAGUSTÍN, Hendrik DRACHSLER, Raymond ELFERINK et Maren SCHEFFEL. T. 11082 LNCS. Leeds, United Kingdom : Springer, Cham, sept. 2018, p. 575-578. DOI : [10.1007/978-3-319-98572-5_{_}48](https://doi.org/10.1007/978-3-319-98572-5_{_}48) (cf. p. 110).
- [143] Rémi VENANT, Kshitij SHARMA, Philippe VIDAL, Pierre DILLENBOURG et Julien BROISIN. « Using Sequential Pattern Mining to Explore Learners' Behaviors and Evaluate Their Correlation with Performance in Inquiry-Based Learning ». In : *Data Driven Approaches in Digital Education. EC-TEL 2017*. Sous la dir. d'Élise LAVOUÉ, Hendrik DRACHSLER, Katrien VERBERT, Julien

- BROISIN et Mar PÉREZ-SANAGUSTÍN. Tallinn, Estonia : Springer International Publishing, 2017, p. 286-299. DOI : [10.1007/978-3-319-66610-5{_}21](https://doi.org/10.1007/978-3-319-66610-5_{_}21) (cf. p. 109).
- [144] Pierre VERMERSCH et Maryse MAUREL. *Pratiques de l'entretien d'explicitation*. ESF, 1997, p. 263 (cf. p. 173).
- [145] Olga VIBERG, Mohammad KHALIL et Martine BAARS. « Self-Regulated Learning and Learning Analytics in Online Learning Environments : A Review of Empirical Research ». In : *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. Frankfurt, Germany : Association for Computing Machinery, 2020, p. 524-533. DOI : [10.1145/3375462.3375483](https://doi.org/10.1145/3375462.3375483) (cf. p. 42, 101).
- [146] Olga VIBERG, Mathias HATAKKA, Olof BÄLTER et Anna MAVROUDI. « The Current Landscape of Learning Analytics in Higher Education ». In : *Computers in Human Behavior* 89 (2018), p. 98-110. DOI : [10.1016/j.chb.2018.07.027](https://doi.org/10.1016/j.chb.2018.07.027) (cf. p. 55).
- [147] Alyssa Friend WISE et Jovita VYTASEK. « Learning Analytics Implementation Design ». In : *Handbook of Learning Analytics*. Sous la dir. de Charles LANG, George SIEMENS, Alyssa WISE et Dragan GAŠEVIĆ. First. SOLAR, 2017. Chap. 13, p. 151-160. DOI : [10.18608/hla17.013](https://doi.org/10.18608/hla17.013) (cf. p. 3, 42).
- [148] Martin WOLPERS, Jehad NAJJAR, Katrien VERBERT et Erik DUVAL. « Tracking Actual Usage : the Attention Metadata Approach ». In : *International Forum of Educational Technology & Society* 10.3 (2007), p. 106-121 (cf. p. 47, 48).
- [149] Mingming ZHOU, Yabo XU, John NESBIT et Philip WINNE. « Sequential Pattern Analysis of Learning Logs ». In : *Handbook of Educational Data Mining*. Sous la dir. de Cristobal ROMERO, Sebastian VENTURA, Mykola PECHENIZKIY et Ryan S J BAKER. CRC Press, 2010. Chap. 8, p. 107-121. DOI : [10.1201/b10274-10](https://doi.org/10.1201/b10274-10) (cf. p. 57).
- [150] *flow* — Wiktionnaire. URL : <https://fr.wiktionary.org/wiki/flow> (cf. p. 127).

COLOPHON

Cette œuvre est mise à disposition selon les termes de la Licence
Creative Commons Attribution 4.0 International.



<https://creativecommons.org/licenses/by/4.0/>