



HAL
open science

Médecine de précision : la voie ontologique - Application à la psychiatrie

Jacques Hilbey

► **To cite this version:**

Jacques Hilbey. Médecine de précision : la voie ontologique - Application à la psychiatrie. Intelligence artificielle [cs.AI]. Sorbonne Université, 2023. Français. NNT : 2023SORUS311 . tel-04300740

HAL Id: tel-04300740

<https://theses.hal.science/tel-04300740v1>

Submitted on 22 Nov 2023

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 Pierre Louis de santé publique

*Laboratoire d'Informatique Médicale et
d'Ingénierie des Connaissances en e-Santé*

Médecine de précision : la voie ontologique

Application au domaine de la psychiatrie

Par Jacques HILBEY

Thèse de doctorat d'informatique médicale

Dirigée par Jean CHARLET

Présentée et soutenue publiquement le 10 novembre 2023

Devant un jury composé de :

Nathalie	AUSSENAC-GILLES	Directrice de recherche	Présidente
Bruno	BACHIMONT	Professeur	Rapporteur
Jean-Arthur	MICOULAUD-FRANCHI	Professeur	Rapporteur
Marie-Odile	KREBS	Professeure	Examinatrice
Sandra	BRINGAY	Professeure	Examinatrice
Jean	CHARLET	Chercheur	Directeur de thèse

*À Mathilde.
À ma mère.
À feu mon père.*

Remerciements

Mes remerciements vont d'abord à Marie-Christine Jaulent et Jean Charlet pour m'avoir recruté comme ingénieur d'études au LIMICS à l'issue de mon Master.

Mes remerciements reviennent ensuite à Jean Charlet pour m'avoir confié ce sujet de thèse, m'avoir guidé tout au long de mon doctorat, m'avoir transmis chemin faisant sa connaissance théorique, pratique, écosystémique des ontologies.

Je remercie Bruno Bachimont et Jean-Arthur Micoulaud-Franchi d'avoir accepté d'être les rapporteurs de cette thèse, ainsi que Marie-Odile Krebs, Nathalie Aussenac-Gilles et Sandra Bringay d'avoir accepté d'en être membres du jury.

Je remercie à nouveau Marie-Odile Krebs et à travers elle mes interlocuteurs du GHU Paris psychiatrie & neurosciences.

Je remercie mes interlocuteurs de l'entreprise Fealinx : Philippe Boutinaud, Pierre-Yves Hervé, et particulièrement Amel Raboudi pour son aide précieuse aux moments opportuns.

Je remercie Gilles Guezennec († 2020) : ce travail porte sa marque, même s'il n'en a vu que les prémises. Je remercie Sylvie Després et Xavier Tannier pour la confiance qu'ils m'ont témoignée, et avec eux Jacques Bouaud pour les échanges enrichissants avec des chercheurs chevronnés.

Je remercie aussi mes compagnons de route pour les bons moments et la camaraderie : Perceval Wajsbürt qui m'a précédé, Akram Redjda et Chyrine Tahri avec lesquels j'ai navigué de conserve, Ons Aouina avec qui j'ai fructueusement collaboré et qui poursuit son travail tout en prolongeant le mien.

Je remercie enfin Élise Akan et Isabelle Verdier pour leur contribution à faire du LIMICS un laboratoire chaleureux et leur accompagnement des chercheurs dans leur développement.

«La base empirique de la science objective ne comporte donc rien d'«absolu». La science ne repose pas sur une base rocheuse. La structure audacieuse de ses théories s'édifie en quelque sorte sur un marécage. Elle est comme une construction bâtie sur pilotis. Les pilotis sont enfoncés dans le marécage mais pas jusqu'à la rencontre de quelque base naturelle ou "donnée" et, lorsque nous cessons d'essayer de les enfoncer davantage, ce n'est pas parce que nous avons atteint un terrain ferme. Nous nous arrêtons, tout simplement, parce que nous sommes convaincus qu'ils sont assez solides pour supporter l'édifice, du moins provisoirement. »

Karl POPPER, *Logique de la découverte scientifique*.

*All of old.
Nothing else ever.
Ever tried.
Ever failed.
No matter.
Try again.
Fail again.
Fail better.*

Samuel BECKETT, *Worstward Ho*.

Table des matières

Introduction	1
1 Une ontologie pour un projet de psychiatrie de précision	5
1 La psychiatrie de précision	5
1.1 La notion de « médecine de précision »	5
1.2 Abord de la psychiatrie de précision	9
1.3 Psychiatrie de précision et intelligence artificielle	13
2 Présentation du projet PsyCARE	15
2.1 Rappel sur les psychoses et la schizophrénie	15
2.2 Hypothèse clinique et objectifs du projet	17
2.3 Organisation en lots de travaux	18
3 Une ontologie modulaire	19
3.1 Les ontologies informatiques	20
3.2 Les applications des ontologies	31
3.3 L'architecture modulaire	32
4 Conclusion	33
2 La Petite Ontologie Fondationnelle	35
1 Les ontologies fondationnelles	35
1.1 Des ontologies informatiques <i>et</i> philosophiques	36
1.2 Les principales ontologies fondationnelles	38
1.3 Excursus épistémologique	41
2 Motivation de la Petite Ontologie Fondationnelle	43
2.1 La temporalité dans les ontologies fondationnelles	43
2.2 L'ergonomie cognitive d'une ontologie fondationnelle	46
3 Principes de constitution de la Petite Ontologie Fondationnelle	50
3.1 Principe d'individuation au regard de l'espace et du temps	50
3.2 Principe méréotopologique	53
3.3 La matrice de la Petite Ontologie Fondationnelle	55
3.4 Les relations	57
4 Conclusion	58

3	ontoDOME : une ontologie noyau des données médicales	61
1	Une ontologie noyau centrée sur les données	61
1.1	Les ontologies noyau	61
1.2	Les ontologies intermédiaires biomédicales	62
1.3	La place des données	64
2	Élaboration d'ontoDOME	66
2.1	Éléments médicaux	66
2.2	Éléments relatifs au système de santé	68
2.3	Éléments relatifs aux données	68
2.4	Éléments sociodémographiques	69
2.5	Éléments temporels	70
3	Conclusion	71
4	ontoDOPSY : une ontologie des données de psychiatrie	73
1	Une ontologie pour la psychiatrie	73
1.1	La psychiatrie en tant que spécialité médicale	73
1.2	Les ontologies de la santé mentale	76
2	Élaboration d'ontoDOPSY	78
2.1	Les signes cliniques	79
2.2	Les troubles psychiatriques	81
2.3	Les outils d'évaluation	85
2.4	Les événements vécus	87
2.5	Les domaines d'intérêt psychiatrique	88
2.6	ontoMEDPSY : un module pour les traitements médicamenteux	93
2.7	Excursus sur les abréviations	96
3	Conclusion	96
5	Modules spécifiques	99
1	Présentation des techniques et spécialités	99
1.1	Examens d'imagerie cérébrale	99
1.2	Examens de biologie médicale	101
1.3	Analyse de la parole	102
1.4	Evaluation des troubles de la dextérité et des fonctions cognitivo-motrices . . .	103
2	Élaboration d'un patron de conception ontologique	103
2.1	Éléments de description communs	103
2.2	Engagement ontologique antérieur	104
2.3	Les patrons de conception ontologique	105
2.4	État de l'art	105
2.5	Un patron de conception ontologique pour les paradigmes expérimentaux . .	106
3	Modules ontologiques de spécialité	107
3.1	ontoDOIMA : l'ontologie des données d'imagerie	107
3.2	ontoDOBIO : l'ontologie des données de biologie	108
3.3	ontoDOOMICS : l'ontologie des données de biologie moléculaire	109
3.4	ontoDODISC : l'ontologie des données d'analyse du discours	111
3.5	ontoDODEXT : l'ontologie des données d'évaluation de la dextérité	112
4	Conclusion	113

6 Déploiement d'ontoPSYCARE	115
1 Constitution d'ontoPSYCARE	115
1.1 Conception	115
1.2 Consolidation	117
2 Évaluation d'ontoPSYCARE	118
3 Intégration des données	119
3.1 Les données du projet PsyCARE	119
3.2 La plateforme PsyCare4.0	119
3.3 L'architecture commune ontoPSYCARE – BMS-LM	121
4 Annotation sémantique	124
4.1 L'annotation sémantique de documents de psychiatrie	124
4.2 Chaîne de traitement pour l'annotation	125
4.3 Développements en cours	127
5 Exploration des données dans un graphe de connaissances	128
5.1 Les graphes de connaissances	128
5.2 L'exploration des données	131
Conclusion	135
Références	139
Annexes	151
A Requête sur la psychiatrie de précision dans PubMed	151
B Petite Ontologie Fondationnelle – Mode d'emploi	152
B.1 Principes généraux	152
B.2 Les classes	153
B.3 Les relations	157
C Histoire externe, histoire interne	162
D Propriétés d'objet des relations de proximité entre personnes	163
E Arborescence de signes cliniques en psychiatrie	165
F Arborescence d'événements vécus	169
G Abréviations utilisées en psychiatrie	172
H Sources des données utilisées par le projet PsyCARE	183
H.1 Études antérieures	183
H.2 Études associées	184
H.3 Étude prospective	185
I Alignements entre classes d'ontoDOME et classes de bareBMSLM	186
Table des figures	188
Table des tableaux	191
Résumé	193

« Jeune homme, lui dit le premier eunuque, n'avez-vous point vu le chien de la reine? » »

Zadig répondit modestement : « C'est une chienne, et non pas un chien.

– Vous avez raison, reprit le premier eunuque.

– C'est une épagneule très-petite, ajouta Zadig; elle a fait depuis peu des chiens; elle boite du pied gauche de devant, et elle a les oreilles très-longues.

– Vous l'avez donc vue? dit le premier eunuque tout essoufflé.

– Non, répondit Zadig, je ne l'ai jamais vue, et je n'ai jamais su si la reine avait une chienne. »

VOLTAIRE, *Zadig*.

Position du sujet

La médecine de précision est une approche relativement récente en médecine qui vise une plus grande précision dans le diagnostic, la prévention, le pronostic, le traitement, en s'appuyant sur le développement de techniques permettant d'acquérir sur le patient beaucoup plus d'informations que ce n'était auparavant le cas. La médecine de précision est donc en elle-même, dans l'histoire longue de la médecine, une « voie ».

Évoquer une voie *pour* la médecine de précision renvoie donc plutôt au chemin qu'elle va suivre pour se développer en tant que discipline. La « voie royale » suivie par la médecine au cours des deux derniers siècles pour asseoir sa progression a été la voie des statistiques. Un jalon important de ce processus historique est l'introduction au XIX^e siècle de la « méthode numérique » par Pierre Louis, parfois considéré comme le père spirituel de la médecine fondée sur la preuve. La médecine de précision continue de recourir aux outils statistiques mais s'appuie aussi de façon toujours croissante sur les méthodes et les algorithmes en plein développement de l'apprentissage automatique, dessinant une voie nouvelle.

Les statistiques constituent une modalité d'administration de la preuve en médecine; l'apprentissage automatique apporte de nouvelles manières (fondées sur les statistiques) de décrire et de prédire, en se saisissant de volumes plus importants de données plus variées. Une « voie ontologique » propose un itinéraire différent, mais pas antinomique, qui concerne la manière d'intégrer

et de structurer ces masses toujours croissantes de données. Il sera donc question dans ce travail d'ingénierie des connaissances, et en particulier de modélisation et de formalisation ontologique, permettant la représentation coordonnée de domaines de connaissance divers.

La médecine de précision a été initiée en oncologie mais d'autres spécialités médicales adoptent cette approche, dont la psychiatrie. L'une des spécificités de la psychiatrie est la prise en compte d'informations qui ne se trouvent pas typiquement dans les bases de données, comme la qualité de vie, les données comportementales, les croyances et les pensées, l'usage de substances toxiques, les déterminants sociaux et environnementaux de la santé mentale. Ceci rend à notre sens cette voie ontologique particulièrement adaptée, à la fois pour identifier ces informations dans des documents en les annotant de façon sémantique et pour faciliter leur exploration ultérieure conjointement avec les données structurées.

Le terrain de notre recherche est le projet de recherche hospitalo-universitaire PsyCARE¹, mené par le GHU Paris psychiatrie & neurosciences, qui vise une détection et une intervention plus précoce en cas de psychose. L'un des moyens mis en œuvre par le projet est l'élaboration d'une plateforme de collecte de données *ad hoc*. L'enjeu premier de notre travail, dans ce contexte, est d'élaborer le modèle conceptuel permettant l'intégration des données produites au cours du projet dans cette plateforme.

Organisation du mémoire

Ce mémoire, outre une introduction générale qui l'ouvre et une conclusion générale qui le clôt, comporte six chapitres :

- le chapitre 1** expose le contexte de notre sujet de recherche : un projet de recherche hospitalo-universitaire en psychiatrie, qui s'inscrit dans une approche de médecine de précision, pour lequel nous proposons une ontologie modulaire;
- le chapitre 2** détaille les raisons d'être, les étapes d'élaboration et le positionnement philosophique de la Petite Ontologie Fondationnelle (ontoPOF);
- le chapitre 3** justifie l'intérêt de disposer d'une ontologie noyau de la médecine orientée vers les données médicales (ontoDOME) et en explore les conséquences en termes de modélisation et de ressources mises à contribution;
- le chapitre 4** propose une ontologie de domaine pour les données de psychiatrie (ontoDOPSY), un outil pour permettre son évolution, ainsi qu'une ontologie pour le sous-domaine des traitements pharmacologiques des patients de psychiatrie (ontoMEDPSY);
- le chapitre 5** rend compte de l'élaboration de plusieurs modules consacrés à des techniques et à des spécialités convoquées par la psychiatrie de précision (ontoDOIMA, ontoDOBIO, ontoDODISC, ontoDODEXT) en s'appuyant sur un patron de conception ontologique;
- le chapitre 6** revient sur la conception de l'ontologie modulaire dans son ensemble, rassemblée par un module de consolidation (ontoPSYCARE), et décrit ses utilisations actuelles et futures.

La figure 1 présente un schéma d'organisation du mémoire (les numéros des pastilles correspondant aux chapitres que nous venons d'exposer) en même temps qu'un schéma du mécanisme d'import entre les ontologies développées au cours de notre travail (les flèches noires sont des relations `owl:imports`).

1. <https://psy-care.fr/>

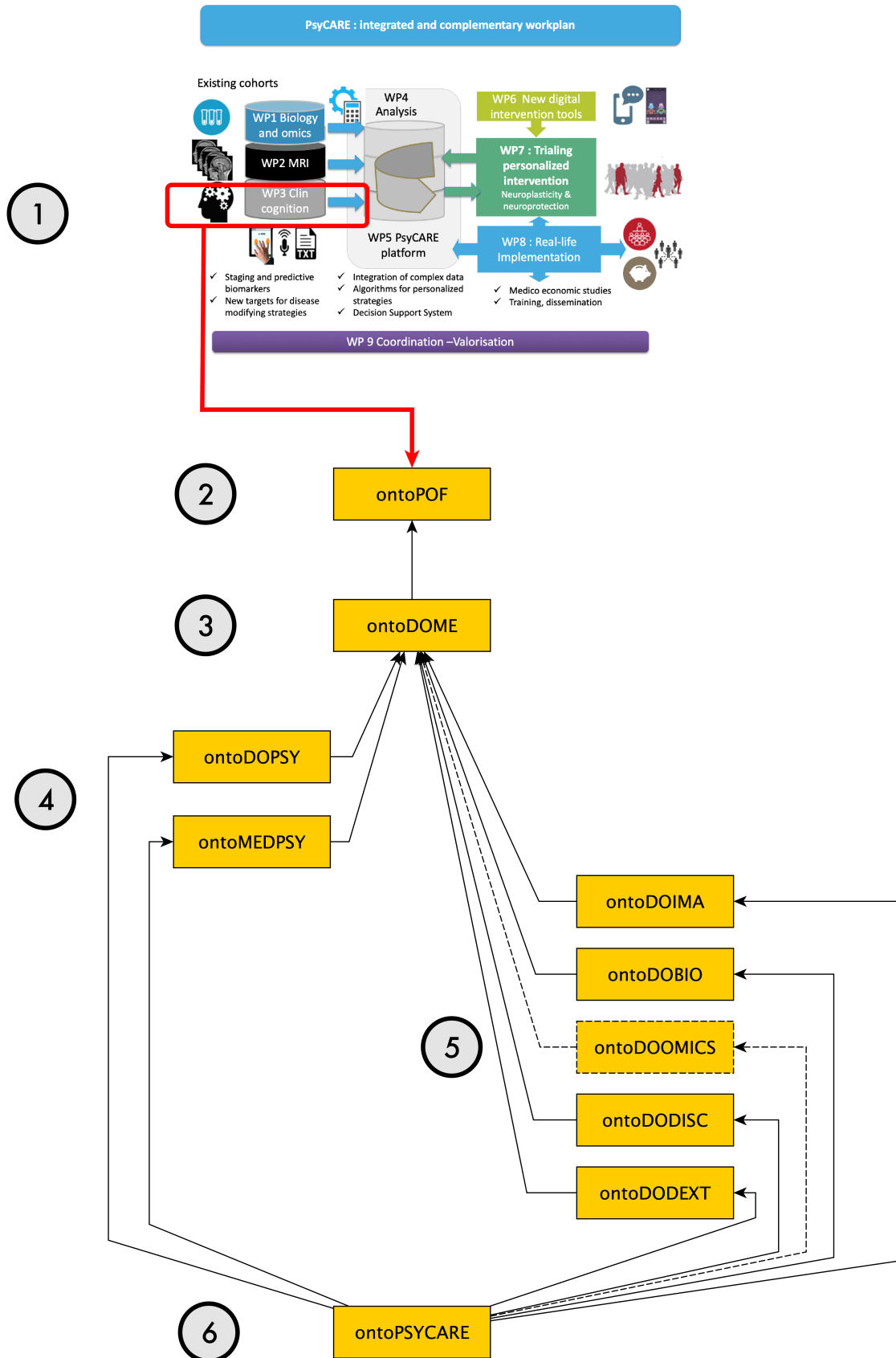


FIGURE 1 – Schéma d'organisation du mémoire.

Une ontologie pour un projet de psychiatrie de précision

« Une seule et même chose peut être dans le même temps bonne et mauvaise et aussi indifférente. Par exemple la Musique est bonne pour le Mélancolique, mauvaise pour l’Affligé; pour le Sourd, elle n’est ni bonne ni mauvaise. »

Baruch SPINOZA, *Éthique*.

Ce premier chapitre dresse la cartographie du terrain de notre recherche. Nous proposons une ontologie d’architecture modulaire utile à un projet de recherche hospitalo-universitaire qui s’inscrit dans l’approche émergente de la psychiatrie de précision. Mais qu’est-ce que la psychiatrie de précision? En quoi ce projet consiste-t-il? Et en quoi une ontologie informatique peut-elle lui être utile?

1 La psychiatrie de précision

1.1 La notion de « médecine de précision »

1.1.1 Plusieurs syntagmes proches

Le syntagme de « médecine de précision » apparaît au sein d’une constellation de syntagmes utilisés dans des contextes proches, comme « médecine personnalisée », « médecine stratifiée », « médecine individualisée », « médecine prédictive », « médecine génomique », voire « médecine 4P » (et plus récemment 5P ou 6P). Une recherche sur Google Trends¹ sur cinq de ces syntagmes (figure 1.1) met en évidence (a) un bond soudain qui met en tête l’expression « médecine de précision » suivi d’une tendance haussière, (b) une tendance baissière de l’expression « médecine personnalisée », (c) une augmentation assez lente de l’expression « médecine génomique », (d) une stagnation à un niveau bas des autres expressions (y compris celles qui n’apparaissent pas dans la figure).

1. L’axe des ordonnées indique des fréquences (de requêtes sur des termes) normalisées par la fréquence absolue maximum : il y a forcément une valeur au moins qui vaut 100.

Ce bond est à mettre en rapport avec l'annonce de la *Precision Medicine Initiative* le 30 janvier 2015 par le président Obama (USA).

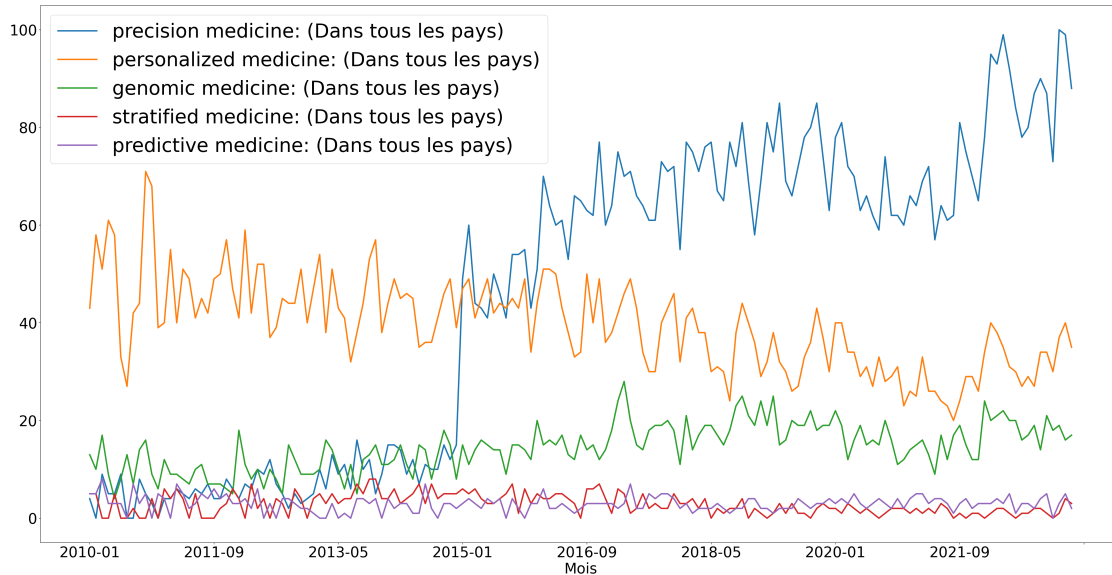


FIGURE 1.1 – Tendances d'emploi des termes exprimant la médecine de précision.

1.1.2 Définition de la médecine de précision

La *Precision Medicine Initiative* promeut la substitution d'une approche thérapeutique prenant en compte les différences individuelles en termes de gènes, d'environnements et de modes de vie, à l'approche *one-size-fits-all* traditionnelle, conçue pour le « patient moyen »². Quatre ans plus tôt, le rapport *Toward Precision Medicine* (National Research Council, 2011) avait pour sous-titre *Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (la nouvelle taxonomie évoquée se fondant sur des données de biologie moléculaire). Un déplacement a eu lieu, de la précision dans la classification à la précision dans le traitement, qui rapproche l'expression de celle de « médecine personnalisée ». Ces expressions sont désormais considérées comme synonymes aussi bien par le MeSH (qui définit la médecine de précision comme « *Clinical, therapeutic and diagnostic approaches to optimal disease management based on individual variations in a patient's genetic profile.* ») que par le Dictionnaire de l'Académie Nationale de Médecine, qui avait une entrée « médecine de précision (ou médecine personnalisée) »³ ainsi définie : « Terme qualifiant une approche médicale qui vise à proposer des mesures préventives ciblées, affiner les diagnostics et personnaliser les traitements en utilisant les données fournies par le séquençage du génome, mais également les données biologiques, environnementales ou sociales spécifiques au patient pour mieux connaître sa probabilité de développer des maladies. »

Quelque temps avant l'annonce du président Obama, l'article '*Personalized medicine*': *what's in a name?* (Pokorska-Bocci *et al.*, 2014) s'attachait à différencier ces termes en les rapportant à

2. <https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative> – consulté le 15 août 2023.

3. <http://dictionnaire.academie-medecine.fr/> – consulté le 10 mai 2023. La nouvelle édition (2023) est en cours.

leurs contextes d'apparition, constatait néanmoins des glissements de sens de l'un à l'autre, et considérait que le terme générique à privilégier était « médecine personnalisée ». Même s'il est difficile d'après la seule figure 1.1 d'attribuer la baisse de « médecine personnalisée » à la hausse de « médecine de précision », nous considérerons cette dernière expression comme le syntagme générique, et nous proposons de voir, dans ces différentes expressions, différentes perspectives sur un même phénomène, lié à l'espoir suscité par le développement de la génomique au tournant du millénaire, dont la première réalisation marquante a été l'achèvement du *Human Genome Project* en 2003 (Collins *et al.*, 2003). Ces différentes perspectives sont liées aux activités médicales (diagnostic, pronostic, traitement, prévention), aux différents acteurs impliqués (patient, médecin, laboratoires pharmaceutiques), aux différents moyens techniques mis en œuvre (génomique, imagerie, dispositifs portables).

Si l'on adopte ce point de vue, la médecine « personnalisée » insistera sur le patient dont les caractéristiques spécifiques (gènes, environnement, mode de vie) sont mieux prises en compte, quand la médecine « de précision » insistera sur le progrès médical qui est permis par cette prise en compte; la médecine « génomique » mettra l'accent sur le principal moyen technique pour obtenir ce surcroît de précision; la médecine « prédictive » sur l'identification et l'évaluation des risques de développer certaines maladies à forte composante génétique; la médecine « stratifiée » sur le discernement de sous-populations – pour une population diagnostiquée avec la même maladie – selon les réponses à différents traitements. La médecine « 4P » ajoutera quant à elle la notion de participation (à la prédiction, la prévention et la personnalisation), le patient étant mis plus à même, par les données dont il dispose sur lui-même, de s'investir dans sa propre santé.

Remarquons que l'expression « médecine personnalisée » apparaît à certains comme un pléonasmisme, la médecine étant depuis Hippocrate attachée à différencier les patients, tandis qu'elle apparaît à d'autres – qui y entendent une médecine capable de s'adapter à chaque individu, une « médecine individualisée » – comme une promesse intenable⁴. L'expression « médecine de précision », quant à elle, rejetterait indûment la médecine antérieure dans l'imprécision. Au-delà de ces arguties, on peut reconnaître un certain flou au moment d'établir une définition précise de la médecine de précision, ce dont témoigne la multiplicité des termes que nous venons d'évoquer. Ce flou tient selon nous aux mutations technologiques et sociales en cours sur lesquelles s'appuie cette redéfinition de la médecine : il provient du processus d'appropriation de ces changements par les différents acteurs, aussi bien scientifiquement (ce que permettent ces différentes techniques en termes de connaissance) que sociologiquement (redéfinition de rôles sociaux). Toutefois, l'objectif général d'un diagnostic plus précis (lié à l'élaboration de nouvelles nosographies), d'une prévention et de traitements mieux adaptés aux patients, d'une meilleure capacité pronostique, reste définitoire. Si cet objectif ne constitue pas en lui-même la révolution ou le changement de paradigme que certains voudraient voir dans la médecine de précision, les moyens mis en œuvre pour l'atteindre constituent néanmoins un changement majeur pour la médecine, comparable à celui de l'essor de l'anatomopathologie dans la médecine clinique au XVIII^e siècle.

Notons enfin que discerner des sous-types cliniques et discerner des réponses différenciées à un traitement n'aboutiront pas nécessairement à l'établissement des mêmes sous-populations dans une population diagnostiquée avec une certaine maladie (la première option s'apparentant à un *clustering* – sur des caractéristiques multiples – quand la deuxième serait plutôt une segmentation – sur une caractéristique unique). La place de visées étiologiques ou physiopathologiques en

4. Comme le remarquent justement Pokorska-Bocci *et al.* (2014), il peut être légitime de parler de médecine individualisée, par exemple dans le cas de thérapies développées à partir de cellules souches du patient lui-même, mais on est alors loin d'une pratique clinique de routine.

médecine, selon l'option retenue, pourrait se trouver modifiée. Et la redéfinition de la nosographie à partir des traitements constituerait bien un changement de paradigme.

1.1.3 Définitions annexes

L'approche de la médecine que constitue la médecine de précision a fait apparaître ou a modifié d'autres termes dont il est utile de préciser le sens.

Dans le droit fil de ce que nous venons de dire sur les relations éventuellement renouvelées entre diagnostic et traitement, la *théranostique*, mot-valise composé de « thérapeutique » et de « diagnostic », ambitionne de développer des couples traitement–test diagnostique, ce dernier recevant alors le nom de « diagnostic compagnon » : un test diagnostique permet d'identifier un biomarqueur dont la présence indique qu'une intervention spécifique sera efficace. Un exemple pharmacologique en est le traitement par un anticorps monoclonal (le trastuzumab) du cancer du sein HER2 positif. Mentionnons également, dans les approches centrées sur l'intervention, la *pharmacogénétique*, qui étudie les causes génétiques des variations interindividuelles dans les réponses aux thérapeutiques médicamenteuses, et la *pharmacogénomique*, qui dans la même direction s'intéresse au génome entier et pas seulement à des gènes singuliers.

La notion de *biomarqueur*, que nous avons mentionnée plus haut, a fait l'objet d'un groupe de travail des *National Institutes of Health* (NIH) états-uniens au tournant du millénaire (Biomarkers Definitions Working Group., 2001), qui définissent :

- un biomarqueur comme « *a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention* » (on pensera au taux de sucre élevé dans le sang pour diagnostiquer le diabète ou à la diminution de la taille d'une tumeur pour pronostiquer l'évolution d'un cancer),
- un critère clinique (« *clinical endpoint* ») comme « *a characteristic or variable that reflects how a patient feels, functions, or survives* » et
- un critère de substitution (« *surrogate endpoint* ») comme « *a biomarker that is intended to substitute for a clinical endpoint* ».

Tout biomarqueur ne sera pas nécessairement un bon critère de substitution, c'est-à-dire un bon prédicteur du bénéfice clinique (ou de la sûreté) d'un traitement, mais il pourra être retenu comme tel s'il est « raisonnablement susceptible » de prédire un bénéfice clinique (les essais contrôlés et les études observationnelles restant la manière privilégiée d'administrer la preuve du bénéfice clinique et de la sûreté d'une intervention). Un biomarqueur ou un ensemble de biomarqueurs constituant un critère de substitution peuvent également être déterminés rétrospectivement, dans le cas d'une intervention ayant déjà démontré un bénéfice clinique, à la condition que ce(s) biomarqueur(s) rende(nt) compte de tous les effets de l'intervention.

Enfin, nous avons pointé le lien entre médecine de précision et génomique, mais il faudrait ici parler plus largement de tous les « *omiques* » qui viennent compléter le « profil moléculaire » d'un patient :

- la protéomique, qui permet de comprendre les fonctions des différentes protéines et la manière dont elles affectent les systèmes biologiques,
- la métabolomique, qui étudie l'ensemble des métabolites dans l'organisme,
- la transcriptomique, qui examine l'ensemble des ARN messagers produits lors de la transcription d'un génome,
- la lipidomique, qui identifie les lipides membranaires et les molécules qui en dérivent, et étudie leur rôle dans les systèmes biologiques.

1.1.4 Caractérisation de la médecine de précision en tant que médecine

La médecine de précision, comme nous l'avons mentionné, entend rompre avec l'approche conçue pour le « patient moyen ». Or, ce patient moyen est celui qui est visé par une approche de santé publique et celui qui participe aux larges études cliniques de la médecine basée sur les preuves. On peut donc s'interroger sur d'éventuelles frictions entre la médecine de précision et la médecine de santé publique d'une part, la médecine basée sur les preuves d'autre part.

Parler de médecine basée sur les preuves, c'est d'abord évoquer une méthode d'administration de la preuve scientifique. Même si nous avons vu en parlant des biomarqueurs et des critères de substitution, mais aussi de la théranostique, qu'il pouvait y avoir dans la médecine de précision la tentation d'un circuit plus court entre diagnostic et thérapeutique, la méthode d'administration de la preuve reste l'essai contrôlé de la médecine basée sur les preuves, même si la hiérarchie des preuves peut s'en trouver modifiée (Tonelli & Shirts, 2017). En revanche, la visée – propre à la médecine de précision – d'affiner la nosographie en établissant des sous-populations de malades peut rendre plus difficile le recrutement, pour un essai contrôlé, du nombre suffisant de patients pour satisfaire aux exigences statistiques de la médecine basée sur les preuves⁵, et plus difficile aussi l'application en pratique courante des résultats obtenus lors d'essais contrôlés puisqu'il faut disposer d'informations génomiques sur les patients pour les mettre en œuvre.

Si personnalisation de la médecine et santé publique peuvent paraître s'opposer (Khoury *et al.*, 2007) en proposant des trajectoires différentes en termes d'investissement public ou en faisant craindre le développement d'une médecine élitiste, on peut également considérer que le progrès des connaissances scientifiques – induit par la prise en compte de la génomique – permettra de mieux délimiter ce qui, dans de nombreuses pathologies mettant en jeu des interactions entre gènes et environnement, ressortit des uns et de l'autre (la santé publique s'attachant plus spécifiquement aux aspects environnementaux) et permettra également des avancées en matière de détection précoce et de prévention.

Pour finir de caractériser la médecine de précision en tant que médecine, il nous faut noter que c'est une médecine qui s'appuie sur une très large quantité de données, qu'il s'agisse des données omiques⁶, des données d'imagerie, des données collectées en temps réel sur les états physiologiques ou psychologiques et les comportements des patients au moyen de dispositifs portables (dont le *smartphone* désormais extrêmement répandu), ou encore de bases de données reliant des données génétiques à des données cliniques (dont le réseau eMERGE⁷ est un exemple).

1.2 Abord de la psychiatrie de précision

Si l'oncologie a été la première spécialité médicale mise en avant pour montrer l'impact potentiel de la médecine de précision (Collins & Varmus, 2015) et reste celle où cette approche est la plus avancée, d'autres spécialités s'en sont emparées, comme l'immunologie, la cardiologie, la pédiatrie ou encore – c'est ce qui va nous intéresser ici – la psychiatrie.

5. Ce risque est contrebalancé par le développement de moyens de communication mettant en contact des patients qui se regroupent et échangent dans des forums.

6. Nous disposons déjà de beaucoup de connaissances sur les maladies « mendéliennes » (cf. le catalogue OMIM – <https://www.omim.org/>) à partir d'études menées sur des familles, mais l'intérêt se porte de plus en plus sur des maladies plus largement répandues, avec plus de gènes présentant une pénétrance moins forte, ou sur des études d'association pangénomiques (*Genome Wide Association Studies*, abrégées en GWAS) qui visent à établir des corrélations entre un grand nombre de variations génétiques et des traits phénotypiques.

7. <https://emerge-network.org/> – consulté le 15 août 2023.

1.2.1 Les sujets de la psychiatrie de précision

Requête PubMed. Nous recherchons les publications ayant les caractéristiques suivantes :

- article de journal,
- en anglais,
- disposant d'un résumé,
- comportant « *precision psychiatry* » dans le titre ou le résumé, ou comportant « *personalized psychiatry* » dans le titre ou le résumé, ou comportant à la fois « *psychiatry* » ou « *mental disorder* » (comme terme MeSH ou dans le titre ou le résumé) et « *precision medicine* » ou « *personalized medicine* » (comme terme MeSH ou dans le titre ou le résumé).

La requête utilisée se trouve dans l'annexe A.

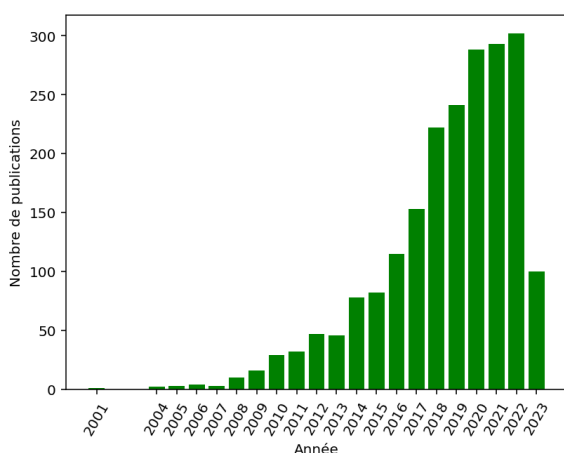


FIGURE 1.2 – Psychiatrie de précision / personnalisée : nombre d'articles par année de publication.

comme une approche émergente.

Modélisation des sujets (*Topic Modelling*). Afin de proposer une modélisation des sujets de la psychiatrie de précision, nous avons recueilli – quand c'était possible – les textes intégraux des articles repérés à partir de la requête. Nous obtenons 948 textes intégraux sur les 1 732 articles, soit un peu plus de 54%. Limiter nos analyses ultérieures aux articles pour lesquels nous disposons du texte intégral constitue évidemment un biais, mais dans la mesure où il s'agit ici d'identifier les sujets qu'aborde particulièrement la psychiatrie de précision et non pas d'évaluer l'importance respective de ces sujets dans le domaine, ce biais nous semble acceptable.

Les textes intégraux étant récupérés au format pdf, un premier traitement consiste à en extraire l'information. Nous utilisons le logiciel libre GROBID⁸ (pour *GeneRation Of Bibliographic Data*), particulièrement adapté aux publications scientifiques. Le résultat, pour chaque pdf, est un fichier au format XML suivant le schéma TEI. Des fichiers XML-TEI, nous extrayons particulièrement le corps de l'article.

Cette requête renvoie (à la date du 18 avril 2023) 1732 résultats. La figure 1.2 montre la répartition par année des publications. La requête ne spécifie pas de date, mais aucun résultat n'est antérieur à 2001. On constate une forte augmentation à partir de 2008, et une très forte augmentation à partir de 2015, suivie de 2020 à 2022 d'une augmentation beaucoup plus lente. Le compte au bout des premiers mois de 2023 peut laisser augurer d'une nouvelle augmentation cette année (ce qui suggérerait un lien entre le tassement constaté et la période du COVID).

Il faut toutefois remarquer que le nombre de publications par année est relativement faible. Cette double caractéristique d'un nombre d'articles qui reste en valeur absolue faible mais qui augmente fortement d'année en année met en évidence la psychiatrie de précision

8. <https://grobid.readthedocs.io/en/latest/> – consulté le 15 août 2023.

C'est à partir de ce corpus de corps d'articles que nous élaborons une allocation de Dirichlet latente (*Latent Dirichlet Allocation* – LDA), qui est un modèle de sujet probabiliste permettant la détection de thématiques (Blei *et al.*, 2003). Après le prétraitement des documents (normalisation, tokenisation, lemmatisation, suppression des mots vides), la question qui se pose est celle d'optimiser le nombre de sujets, celui-ci étant fixé *a priori*. L'heuristique choisie consiste à maximiser l'écart entre la cohérence de chaque sujet et le recouvrement moyen entre les sujets. La figure 1.3 montre le nombre optimal de sujets en suivant cette approche.

La librairie Python pyLDAvis propose une visualisation du résultat de la LDA⁹, comme le montre la figure 1.4, selon deux modalités : les regroupements de documents du corpus sont représentés d'une part sous forme de bulles et projetés en deux dimensions (selon les deux composantes principales de l'espace multidimensionnel dans lequel ils se situent); d'autre part sous forme de liste des mots les plus fréquents pour chaque groupe. Les mots qui apparaissent dépendent de la valeur choisie pour le paramètre λ : les mots les plus spécifiques du sujet si le paramètre est fixé à 0 et les mots de ce sujet les plus communs aux mots les plus fréquents du corpus si le paramètre est fixé à 1. La possibilité de faire varier ce paramètre permet de mieux appréhender, à partir de la liste de mots affichée selon chaque valeur, le sujet dont il est question.

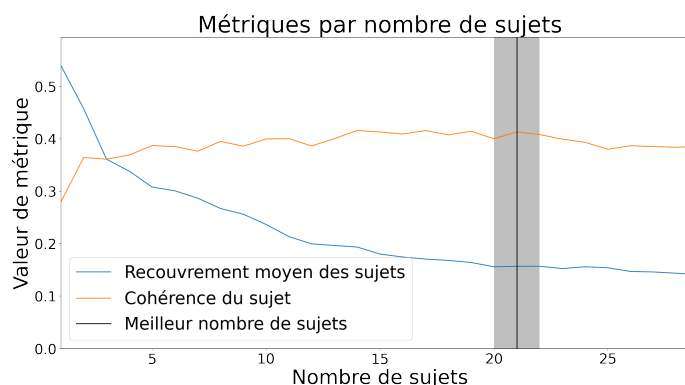


FIGURE 1.3 – Nombre optimal de sujets pour la LDA.

Sujets de la psychiatrie de précision.

Sujet 1 : **phénotypage digital** (*digital phenotyping, EMA, smartphone*) – à noter : le terme de *precision psychiatry* est assez haut

Sujet 2 : **prédiction de la réponse à un traitement de la dépression**

Sujet 3 : **biomarqueurs de la maladie d'Alzheimer**, principalement biologiques ($\alpha\beta$, tau, BACE) mais aussi d'imagerie (PET, *atrophy*)

Sujet 4 : **pharmacogénétique/pharmacogénomique** (*drug, gene, PGx*) **du métabolisme** (CYP, *poor metabolizer*) **des principaux médicaments psychotropes** (*antidepressant, antipsychotic*)

Sujet 5 : **machine learning** (model, parameter, distribution, dataset, algorithm, hyperparameter, classifier, clustering, autoencoder)

Sujet 6 : **génomique/génétique** (GWAS, PRS, SNPs) **de la schizophrénie**

Sujet 7 : **biomarqueurs épigénétiques** (*epigenetic, methylation, DNA methylation*) et **biologiques** (*biomarker, folate, fatty acid, cytokine*) **de la dépression et de la schizophrénie** (depression, MDD, SCH)

Sujet 8 : **trouble du déficit de l'attention avec ou sans hyperactivité** (ADHD) et symptômes associés (*anxiety, sleep, stress, fear*), du point de vue des régions du cerveau (vmPFC, OFC, HPA axis, amygdala) impliquées et des traitements associés (methylphenidate, PUFAs)

9. Cette visualisation est disponible à l'adresse : https://jacqueshilbey.frama.io/psy_prec_lda/lda.html.

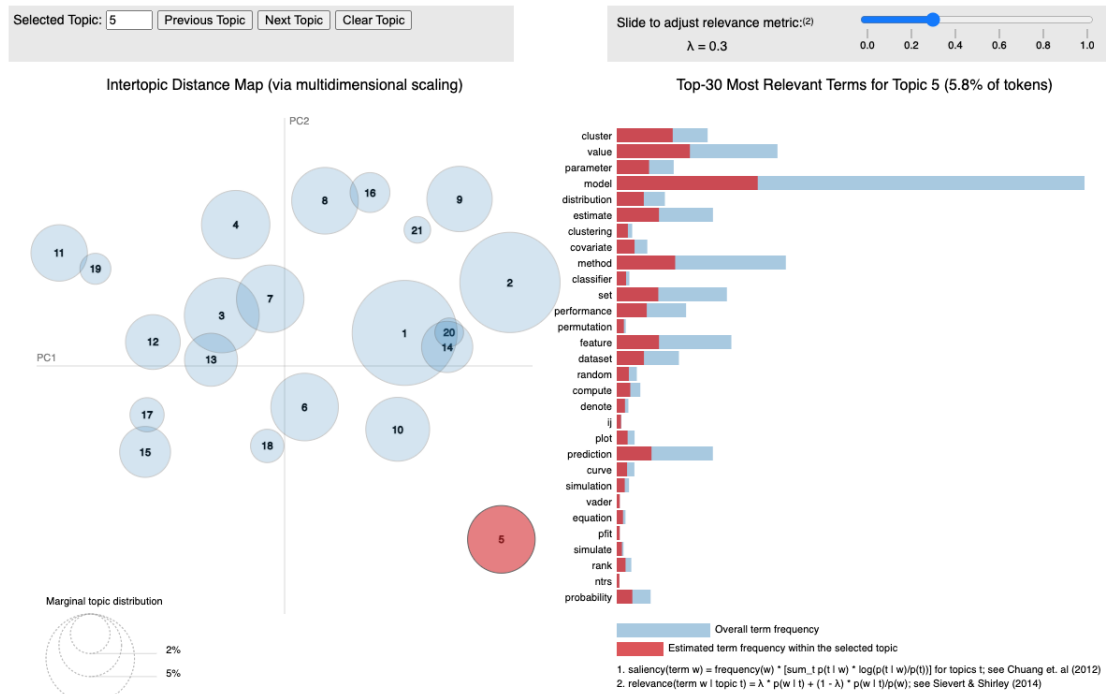


FIGURE 1.4 – Sujet n°5 de la LDA.

Sujet 9 : **réponses différenciées aux traitements** (naltrexone, acamprostate, doxazosine, buprenorphine) **des troubles liés à l'usage de substances** (*alcohol, drinking, craving, opioid*)

Sujet 10 : **neuroimagerie** (*neuroimaging, fMRI, scan, EEG, voxel*) **de la connectivité cérébrale** (*connectivity, network, region, brain*)

Sujet 11 : **génétique** (BDNF) **et thérapie génétique** (AAV) **de la neuroimmunité** (*immune, microglia, cell, astrocyte, inflammation*) – étudiée avec des modèles animaux (*mouse*)

Sujet 12 : **modèles expérimentaux du développement cérébral** (*neuron, model, organoid, iPSCs, stem cell*), notamment pour les troubles du spectre de l'autisme (ASD) – ici, le lien à la médecine de précision vient de l'espoir de mettre en évidence la contribution génétique aux troubles mentaux et des pistes thérapeutiques personnalisées

Sujet 13 : **facteurs de risque de la maladie d'Alzheimer** (AD) en premier lieu génétiques (APOE) mais pas uniquement (*menopause*)

Sujet 14 : **sous-types des troubles du spectre de l'autisme** (ASD, *subtype, subgroup, BvFTD, testosterone*)

Sujet 15 : **génétique des troubles du neurodéveloppement** (NDD, *gene protein, expression, mutation*) et notamment des troubles du spectre de l'autisme (ASD, SHANK)

Sujet 16 : **traitements des dysfonctionnements cognitifs** (*cognitive, memory, ChEI, primavanserin*) **en cas de psychose** (*psychosis, antipsychotic, schizophrenia*) **ou de troubles obsessionnels compulsifs** (OCD)

Sujet 17 : **génomique du cerveau** (EVs, miRNA, exosome, model, brain, receptor)

Sujet 18 : **généétique** (GRN) **des récepteurs du glutamate** (NMDAr, GluN) impliqués dans les **troubles neurodéveloppementaux** (*seizure, epilepsy, ASD*)

Sujet 19 : **rôle du microbiote intestinal** (*gut microbiota, probiotic, microbiome*) **dans la maladie d'Alzheimer** ($\alpha\beta$, AD)

Sujet 20 : **traitement de l'addiction nicotinique** (*smoker, smoking cessation, nicotine dependence*) **chez les patients souffrant de troubles de déficit de l'attention avec hyperactivité** (ADHD)

Sujet 21 : **sous-groupes de consommateurs d'alcool** (*reward drinker, relief drinker*) **selon des caractéristiques neuropathologiques** d'origine génétique (FLTD-TDP, GRN)

Comme on pouvait s'y attendre, on retrouve les termes évoqués à propos de la médecine de précision. On remarque toutefois certains alliages entre techniques mises en œuvre et troubles étudiés, semblant indiquer des affinités, des pistes explorées de façon privilégiée. Nous n'allons pas plus loin dans cette analyse qui déborde à la fois ce travail et nos compétences, mais cette exploration des sujets de la psychiatrie de précision est à comparer au travail de Gauld & Micoulaud-Franchi (2021). Les auteurs dégagent, de l'ensemble fortement hétérogène des concepts identifiés comme associés à la psychiatrie de précision, six dimensions (« la génétique, la pharmacogénétique, les approches computationnelles, le raffinement des essais thérapeutiques, les biomarqueurs et la stadification ») indiquant son orientation mécanistique. Ils interprètent ces résultats en termes d'ambition mécanistique manquant pour l'instant d'un cadre théorique intégratif.

Pour ce qui concerne plus proprement notre travail, nous constatons l'existence d'un sujet bien différencié des autres où n'apparaît pas précisément le type d'alliage que nous venons d'évoquer, mais qui présente tous les termes typiques du *machine learning*, un important courant de l'intelligence artificielle (ce sujet est à rapprocher de la dimension « approches computationnelles » identifiée par Gauld & Micoulaud-Franchi (2021)). Ce qui nous amène à nous poser plus généralement la question des algorithmes et modèles d'intelligence artificielle convoqués par la psychiatrie de précision.

1.3 Psychiatrie de précision et intelligence artificielle

Le terme d'« intelligence artificielle » (IA) peut être considéré comme un terme générique employé *largamano* pour désigner des choses très différentes, par exemple : une intelligence non-humaine « mythifiée »; un système artificiel permettant de mieux comprendre le fonctionnement du cerveau (ce qui correspond à l'origine cybernétique de l'IA); une discipline informatique s'efforçant de simuler au moyen de machines des traitements d'information considérés comme complexes pour un humain (Gerbaud, 2023). Même si l'on s'en tient à cette dernière définition, il convient de distinguer :

- (a) des *dispositifs* d'IA, qui accomplissent une certaine tâche,
- (b) des *modèles* d'IA, qui simulent une certaine capacité cognitive,
- (c) des *algorithmes* d'IA, qui permettent d'élaborer un modèle.

Les mêmes algorithmes peuvent être impliqués dans l'élaboration de modèles très différents (inversement, un même modèle peut être élaboré à partir d'algorithmes très différents), et les mêmes modèles peuvent être employés dans des dispositifs accomplissant des tâches très différentes (l'inverse étant là aussi possible).

L'objectif que nous poursuivons ici est de mettre en évidence les algorithmes et modèles d'intelligence artificielle auxquels recourt la psychiatrie de précision. Afin d'aborder cette question, nous nous appuyons sur les libellés en anglais associés aux classes d'une ontologie de l'intelligence

artificielle développée par Jean Charlet, en cherchant les mentions de ces libellés dans le corpus d'articles sur la psychiatrie de précision présenté plus haut. Les résultats obtenus sont présentés dans le tableau 1.1 (avec les libellés français).

Terme de PIA	Nombre de mentions
apprentissage	427
intelligence artificielle	265
statistique	218
apprentissage automatique	205
modèle de régression	187
régression logistique	139
régression linéaire	107
analyse en composantes principales	69
machine à vecteur de support	55
apprentissage supervisé	52
apprentissage profond	50
classification hiérarchique	37
arbre de décision	37
modèle linéaire généralisé	35
apprentissage non supervisé	33
traitement automatique du langage naturel	29
k-moyennes	28
analyse discriminante linéaire	22
vision et reconnaissance des formes	21
système multi-agents	21
apprentissage par renforcement	19
interaction homme-machine	14
classification ascendante hiérarchique	9
ingénierie ontologique	9
carte de Kohonen	8
reconnaissance de la parole	8
traitement d'images	8
classification descendante hiérarchique	7
méthode des k plus proches voisins	5
robotique	5
algorithmes de l'IA	4
analyse discriminante	4
algorithme génétique	3
raisonnement à partir de cas	2
algorithme évolutif	1
analyse morphologique	1
modèle log-linéaire	1
algorithme du raisonnement à partir de cas	0
programmation par contraintes	0
analyse textuelle	0
analyse grammaticale	0

analyse lexicale	0
réseau de neurones multicouches	0
apprentissage par transfert	0
intelligence artificielle numérique	0
intelligence artificielle symbolique	0
vie artificielle et complexité	0
intelligence artificielle neuro-symbolique	0
agents et robotique	0
représentation des connaissances et modélisation du raisonnement	0
modélisation cognitive et systèmes neuro-informatiionnels	0
planification et recherche heuristique	0
ingénierie des connaissances	0
décision et gestion de l'incertitude	0
aide à la décision symbolique	0

TABLE 1.1: Nombre de mentions des termes de l'intelligence artificielle.

Rappelons que ce corpus comporte un biais fondamental, puisqu'il s'agit d'articles pour lesquels le texte était en accès libre. D'autre part, les résultats obtenus dépendent des libellés recherchés, ce qui engendre à la fois du « bruit » (en raison des acronymes et des syntagmes ayant un autre sens dans la recherche en psychiatrie, comme « apprentissage » ou « apprentissage par renforcement ») et du « silence » (en raison de libellés qui auraient dû être associés à un concept mais ne l'ont pas été). Nous n'en tirerons donc que peu de conclusions positives : comme on pouvait s'y attendre, les méthodes issues des statistiques sont bien présentes, ainsi que des algorithmes et modèles d'apprentissage automatique¹⁰. Ce qui nous intéresse plus particulièrement est la conclusion négative concernant les ontologies : elles sont très peu mentionnées (le terme *ontology* est lui-même facteur de bruit – en raison du sens le plus courant du mot – mais l'adjectif *ontological* l'est beaucoup moins).

2 Présentation du projet PsyCARE

PsyCARE¹¹ est un projet de Recherche Hospitalo-Universitaire en psychiatrie dont l'objectif, exprimé le plus généralement, est d'améliorer l'intervention précoce en cas de psychose.

2.1 Rappel sur les psychoses et la schizophrénie

Les psychoses sont des troubles psychiatriques sévères, caractérisés sémiologiquement en distinguant : des signes positifs (au sens où ils représentent un ajout par rapport à la normalité, comme des délires ou des hallucinations), des signes négatifs (au sens où ils représentent un déficit par rapport à la normalité, comme le retrait social) et des signes de désorganisation de la pensée.

Les sujets touchés sont le plus souvent jeunes : l'âge habituel de début des troubles se situe entre 15 et 30 ans, c'est-à-dire pendant l'« adolescence cérébrale », période de maturation cérébrale

10. L'article *The potential of precision psychiatry: what is in reach?* (Kambeitz-Illankovic *et al.*, 2022) présente un point d'avancement des recherches en *machine learning* qui sont au plus proche de la pratique clinique réelle sur (a) le risque pronostique (*prognostic risk*), (b) la classification diagnostique et la stratification, (c) la prédiction de la réponse au traitement, de l'observance thérapeutique et de la rechute.

11. <https://psy-care.fr/>

qui se poursuit au-delà de l'adolescence. On compte en France 15 000 nouveaux cas par an. Ce trouble entraîne une réduction d'espérance de vie estimée à 15 années, pour différentes raisons (notamment cardio-vasculaires).

L'étude de la schizophrénie est fortement liée, historiquement, à la question du repérage des signes précoces (Krebs, 2015), voire du repérage de la vulnérabilité à la schizophrénie. Dès le milieu du XIX^e siècle, le psychiatre Benedict Morel, dans son *Traité des dégénérescences physiques, intellectuelles et morales de l'espèce humaine* (1857), explore les causes héréditaires des maladies et décrit un « terrain fragile » pour la démence précoce, caractérisé par une déviation pathologique et un abaissement du seuil de tolérance. A la fin du XIX^e, Emil Kraepelin remarque pour sa part les « personnalités frappantes » de parents de patients atteints de *dementia praecox* et suggère que les premiers ont la même « maladie principale » que les derniers. Peu après, le psychiatre suisse Eugen Bleuler – à qui l'on doit la paternité du mot « schizophrénie », en 1908 – fait des observations similaires sur les frères et sœurs des sujets malades chez lesquels il décrit « des anomalies qualitativement identiques » dans le comportement, mais dont les manifestations sont d'une moindre extension. Si l'expression de « vulnérabilité à la schizophrénie », appliquée à des individus qu'on juge plus particulièrement que d'autres susceptibles de développer une schizophrénie, n'apparaîtra que dans les années 1960, on voit qu'elle a été anticipée depuis la deuxième moitié du XIX^e siècle.

L'étude de Gottesman & Shields (1966) sur des paires de jumeaux monozygotes permettra d'appuyer l'hypothèse d'une base génétique de la schizophrénie, en montrant que le taux de concordance de la schizophrénie chez ces jumeaux est d'environ 50%¹², soit beaucoup plus que chez les jumeaux dizygotes. Avec un tel pourcentage, cette étude fait en quelque sorte d'une pierre deux coups : en même temps qu'elle met en évidence une composante génétique de la schizophrénie, elle invite à comprendre pourquoi le taux de concordance n'est pas de 100%, ou pour le dire autrement quels facteurs autres que génétiques expliquent la transition vers la schizophrénie. Le psychologue nord-américain Paul Meehl va s'inspirer des travaux du psychanalyste Sandor Rado, qui introduit dans les années 1950 le terme « schizotype » (mot-valise pour *schizophrenic phenotype*) afin de désigner une personnalité vulnérable à la schizophrénie. La schizotypie, chez Rado, est une forme d'organisation du psychisme résultant de l'interaction entre un certain type de patrimoine génétique et un certain environnement, et pouvant entraîner – en cas de décompensation – des symptômes schizophréniques. Meehl (1962) reprend cette notion de schizotypie, mais propose un modèle plus complexe : cette organisation particulière de la personnalité est elle-même le résultat de l'interaction entre un déficit d'intégration neuronale – que Meehl considère comme d'origine probablement génétique et nomme « schizotaxie » – et des facteurs environnementaux (environnement éducatif et renforcements sociaux). D'autres facteurs de risque, qui se trouveraient exprimés en situation de stress, créeraient les conditions d'une décompensation de la schizotypie aboutissant à la schizophrénie.

Zubin & Spring (1977) introduisent l'idée que la vulnérabilité puisse ne pas être liée seulement à des facteurs génétiques, mais être aussi acquise au cours du développement (par exemple lors de complications obstétricales). Ils avancent également l'idée d'une relation entre vulnérabilité et stress pour expliquer l'épisode psychotique (un stress faible subi par un individu à la vulnérabilité forte, aussi bien qu'un stress intense subi par un individu à la vulnérabilité moyenne ou faible, peuvent rendre compte d'un épisode psychotique) : le seuil de tolérance au stress est donc susceptible de diminuer si la vulnérabilité augmente. Comme le notent les auteurs, leur conception

12. On l'estime plutôt à 40% de nos jours selon le dossier en ligne de l'Inserm consacré à la schizophrénie, réalisé en collaboration avec Marie-Odile Krebs – <https://www.inserm.fr/dossier/schizophrénie/>

de la vulnérabilité se distingue par trois éléments : les nombreuses contributions au degré de vulnérabilité d'un individu; la tentative de repérer empiriquement les caractéristiques individuelles prédictives de la probabilité d'un épisode schizophrénique; la distinction nette entre la vulnérabilité vue comme un trait permanent et les épisodes schizophréniques, qui sont des états fluctuants.

Ciampi (1984) propose un modèle biopsychosocial distinguant une phase prémorbide au cours de laquelle des éléments génétiques, somatiques, constitutionnels et développementaux, en interaction avec un environnement psychosocial, culturel, familial et relationnel, aboutissent à la constitution d'une personnalité prémorbide vulnérable. Cette phase peut être suivie par une phase aiguë de déstabilisation à l'occasion d'épisodes de stress biologiques ou psychosociaux, entraînant une décompensation psychotique. Cette phase aiguë peut elle-même être suivie par une phase chronique de rémission totale, de trouble résiduel ou de chronicisation sévère.

L'article de Gourion *et al.* (2004) sur l'hypothèse neurodéveloppementale offre un modèle permettant de dégager les éléments principaux pour penser la vulnérabilité lors de la phase prémorbide. L'hypothèse neurodéveloppementale, comme son nom l'indique, cherche les facteurs étiologiques de la schizophrénie dans les phases de développement neurologique, ce qui la distingue d'autres recherches s'appuyant sur les données en faveur d'aspects neurodégénératifs, ou encore se focalisant sur certains systèmes neurologiques (hypothèses dopaminergique, glutaminergique). Ce modèle s'attache particulièrement à rendre compte de la diachronie de l'apparition des symptômes en différenciant deux périodes principales : (a) une période très précoce du développement menant à l'expression d'un éventuel phénotype de vulnérabilité (mettant en jeu l'intervention de facteurs génétiques, mais aussi de facteurs environnementaux comme des complications obstétricales ou périnatales, et des interactions entre les deux), (b) une période de maturation cérébrale (prépubertaire et jusqu'à 25-35 ans) durant laquelle l'état initial peut rester relativement stable ou évoluer vers l'émergence de symptômes et d'un premier épisode psychotique – sous l'influence de facteurs d'environnement précipitants et en l'absence de facteurs protecteurs – modulés par certains facteurs génétiques. L'article expose le « faisceau d'arguments cliniques, neuropsychologiques, biologiques et neuropathologiques » permettant de conforter cette hypothèse, tout en mentionnant les « nombreuses questions [qui] demeurent : par quels biais interviennent les facteurs de vulnérabilité? Comment passe-t-on de l'état de vulnérabilité à l'état de maladie? Existe-t-il des facteurs de résilience à la maladie? ». Les auteurs concluent sur l'enjeu majeur de « parvenir à une meilleure caractérisation du phénotype vulnérable ». C'est dans le cadre de cette hypothèse que s'inscrit le projet PsyCARE.

2.2 Hypothèse clinique et objectifs du projet

Le retard diagnostique en cas de psychose (la durée de psychose non traitée, ou *Duration of Untreated Psychosis* – DUP, en anglais) est estimé, en France, entre 1,5 et 2 ans. C'est cette durée que le projet entend réduire, en partant de l'hypothèse clinique que les psychoses, et notamment la schizophrénie, seraient les formes chroniques d'un trouble trouvant son origine dans des anomalies précoces du développement, et commençant à se manifester lors de l'adolescence cérébrale par des épisodes psychotiques déclenchés par des événements stressants. Le renforcement de la neuroprotection et de la neuroplasticité permettrait d'atténuer ou d'éviter l'apparition d'épisodes psychotiques, et donc d'éviter la chronicisation du trouble. Cette hypothèse est testée lors d'un essai clinique multicentrique à quatre branches sur 500 patients présentant des premiers signes de psychose.

L'objectif de réduction de la durée de psychose non traitée suppose une détection plus précoce des troubles, ainsi qu'une intervention plus précoce une fois qu'ils ont été détectés. Les démarches

mises en œuvre au sein du projet afin d’atteindre cet objectif sont les suivantes :

- l’identification de biomarqueurs,
- le développement d’une application d’entraînement cognitif,
- la mise en place d’une plateforme de collecte de données,
- l’élaboration d’un système d’aide à la décision thérapeutique,
- la diffusion des connaissances acquises.

2.3 Organisation en lots de travaux

La figure 1.5 présente les différents lots de travaux (*Work Packages* – WP) du projet PsyCARE.

- WP1 – Marqueurs biologiques** : identification des variants génétiques impliqués dans la vulnérabilité à la psychose, détection des anomalies métaboliques, validation d’un panel de biomarqueurs périphériques pour prédire le résultat clinique ;
- WP2 – Imagerie** : création d’un réseau étendu d’installations d’imagerie dédiées aux psychoses précoces, validation croisée d’un classificateur basé sur l’IRM pour prédire le résultat clinique ;
- WP3 – Phénotypage digital** : mise en place de la collecte de données, validation d’un nouveau dispositif d’évaluation des signes neurologiques mineurs, traitements sémantiques de données structurées et de corpus textuels phénotypiques, identification des marqueurs linguistiques dans le discours libre afin de prédire l’évolution clinique ;
- WP4 – Analyse multimodale** : score neurodéveloppemental composite, arbre de décision pour la stratégie thérapeutique, apprentissage par transfert à partir de grandes cohortes existantes ; intégration multimodale pour prédire le résultat cognitif et finalisation du système d’aide à la décision ;
- WP5 – Plateforme PsyCARE** : déploiement de la plate-forme (gestion des données et capacité opérationnelle), alignement des ontologies de domaine en vue de la gestion du cycle de vie des données, intégration des flux et gestion de la provenance de données, interfaces utilisateur, anonymisation des données ;
- WP6 – Outils digitaux** : test cognitif et application d’entraînement cognitif personnalisé sur appareils mobiles, application à destination des patients, des soignants et des gestionnaires de cas ;
- WP7 – Intervention précoce personnalisée** : obtention des autorisations réglementaires et éthiques ; lancement de l’étude, recrutement prospectif des patients et suivi, analyse statistique, analyse du rapport coût/efficacité ;
- WP8 – Implémentation** : exploration des attentes des patients, des aidants et des cliniciens ; sensibilisation du grand public et renforcement des capacités des professionnels de la santé ; impact socio-économique de l’amélioration de la pratique médicale apportée par PsyCARE ;
- WP9 – Coordination et valorisation** : diffusion des résultats scientifiques, gestion des droits de propriété intellectuelle et planification de l’exploitation, coordination scientifique, gestion de projet et administration.

Le travail que nous présentons est à l’articulation des WP3 et WP5. C’est au sein du WP3 que sont abordées les questions de modélisation et d’ontologies, mais les ontologies produites servent de modèle conceptuel à la plateforme développée dans le WP5. Les interactions principales au cours de cette thèse ont donc eu lieu d’une part avec les psychiatres de Sainte-Anne, afin d’élucider

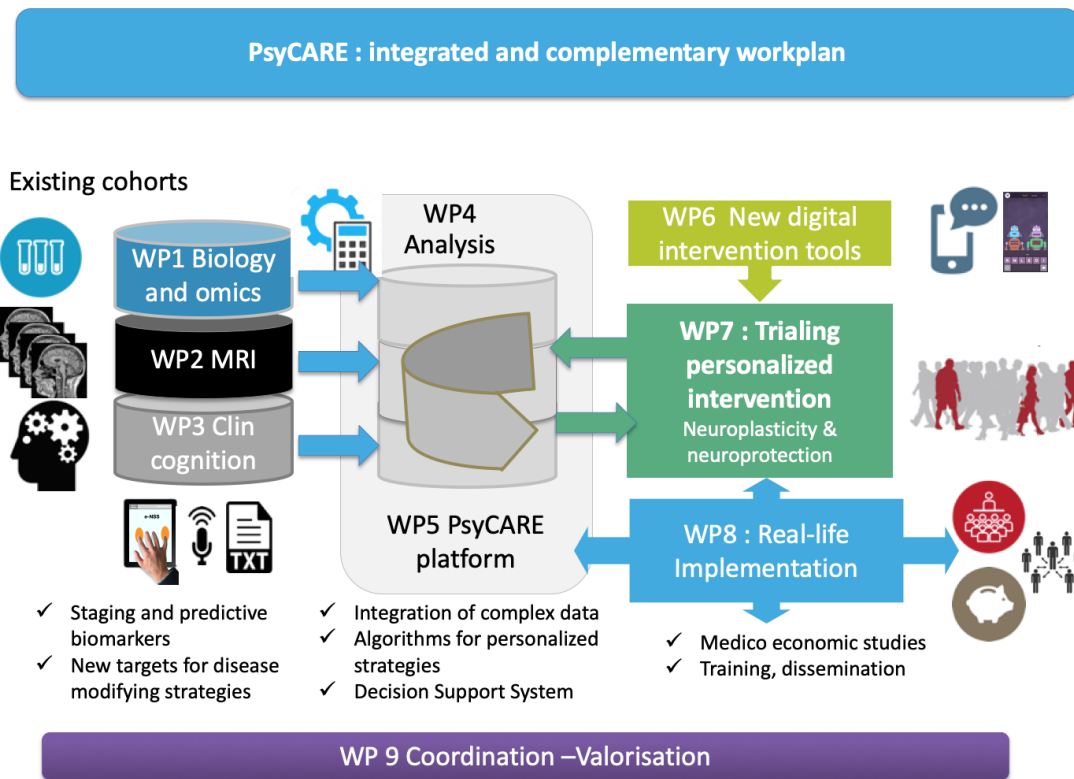


FIGURE 1.5 – Lots de travaux du RHU PsyCARE (doc. PsyCARE).

ou de trancher des questions relatives à la modélisation, et d'autre part avec l'industriel en charge de la mise en place de la plateforme, Fealinx. Lors d'un projet précédent, cet industriel a déjà été amené à développer une modélisation ontologique (Raboudi-Souilem, 2021) reflétant de manière assez étroite la structure de la base de données, orientée vers le *Product Lifecycle Management* (PLM). Une partie de cette ontologie, BMS-LM, est réutilisée dans notre contexte afin de faciliter l'intégration entre le modèle ontologique et le modèle de la base de données, et de fournir les outils ontologiques propres à la gestion du cycle de vie des données (maintien de la provenance des données, traces des traitements effectués sur les données, en se basant sur la *PROV Ontology* promue par le W3C¹³).

3 Une ontologie modulaire

Une ontologie informatique n'est pas un objet simple à appréhender, en raison de son rapport au langage naturel, de son recours au formalisme logique (et de son rapport à la théorie des modèles), de son rapport à l'information incomplète, de sa structure en réseau. Nous donnons ici des éléments pour appréhender ce qu'est une ontologie informatique, puis nous en exposons les principales utilisations, pour en venir finalement à un type particulier d'architecture des ontologies :

13. <https://www.w3.org/TR/prov-o/> – consulté le 15 août 2023.

l'architecture modulaire.

3.1 Les ontologies informatiques

3.1.1 Définition

Il n'est pas inutile de rappeler le sens philosophique du mot pour marquer la différence entre ontologie au sens philosophique et ontologie au sens informatique, même si nous verrons dans le chapitre 2 les raisons de cet emprunt de l'informatique au vocabulaire philosophique. L'ontologie *au sens philosophique* est d'abord une branche de la métaphysique qui traite de ce qui est, de « ce qu'il y a » (Quine, 1948), et ceci au niveau le plus général qui soit. L'ontologie *au sens informatique*, quant à elle, propose une *représentation d'un domaine de connaissance*, censée refléter les entités, les propriétés de ces entités, les relations entre les entités – la notion d'entité étant ici à prendre au sens le plus large. Lorsque nous parlerons par la suite d'ontologie, sans l'accompagner d'un adjectif, nous parlerons d'ontologie au sens informatique.

Pour définir plus précisément une ontologie, nous pouvons commencer par la définition la plus connue, la plus souvent citée, celle de Gruber (1993) (une « spécification explicite d'une conceptualisation »), amendée par Borst (1997) qui parle de « spécification formelle d'une conceptualisation partagée », puis par Studer *et al.* (1998) qui fusionnent les deux pour aboutir à parler de « spécification formelle et explicite d'une conceptualisation partagée ». Guarino *et al.* (2009) proposent une explication fouillée de cette définition, rappelant :

- qu'« une conceptualisation est une vue abstraite et simplifiée du monde que l'on souhaite représenter dans un certain but »,
- qu'elle est « formelle » au sens où elle est lisible par une machine,
- qu'elle est une spécification explicite au sens où elle exprime des contraintes d'interprétation du vocabulaire de l'ontologie (au moyen de « postulats de signification », ou « axiomes », qui relient entre eux – dans des formules – les classes et les relations de l'ontologie),
- que ce qui est partagé est l'engagement ontologique (qui peut être considéré comme un équivalent intensionnel de la signification).

Ajoutons qu'en parlant de représentation simplifiée ou abstraite, on parle d'une ligne de partage entre ce qui est important, signifiant ou simplement pertinent, et ce qui ne l'est pas. Cette exigence de pertinence – aboutissant à une simplification ou à une abstraction – est le régime normal du discours scientifique mais aussi du langage naturel. Ajoutons également qu'en parlant de conceptualisation partagée, on sort de l'esprit de celui ou de ceux qui conceptualisent pour faire intervenir le langage : ce qui est représenté dans une ontologie est une connaissance susceptible d'être partagée dans un langage, qu'il s'agisse du langage naturel ou d'un langage formel. Ces deux considérations nous amènent à insister sur l'abord des ontologies sous l'angle de leur capacité à exprimer, dans un certain langage précisément défini, une conception d'une partie de la réalité. Cet abord amène à inscrire les ontologies au sein d'autres efforts pour fixer la signification de termes. Sur cette dimension, on trouvera notamment, du moins formel (et du moins expressif) au plus formel (et au plus expressif), les vocabulaires contrôlés, les glossaires, les thésaurus, les taxonomies et les ontologies. Par expressivité, on entend la capacité à exprimer des relations complexes entre les termes. Alors qu'un thésaurus proposera des relations de plus ou moins grande généralité (« bouche » pouvant avoir comme terme générique « appareil digestif ») et qu'une taxonomie proposera des relations de subsomption (tout « descendant » dans la taxonomie possède les traits définitionnels de son « ascendant », auxquels s'ajoutent des traits qui le définissent en propre), une ontologie peut proposer un éventail beaucoup plus riche de relations et de spécification de ces

relations. Parce qu’une ontologie associe aux entités qu’elle définit des termes du langage naturel (des « libellés », *labels* en anglais) et qu’elle permet d’établir des liens entre ces entités, elle peut être rapprochée d’autres **systèmes d’organisation des connaissances** (SOC) fournissant le vocabulaire d’un domaine (Zacklad, 2010), dont les schémas de classification, les langages documentaires et les thésaurus destinés à la gestion de l’information.

Ce qui permet cette expressivité, au moyen d’une syntaxe claire et précise et d’une sémantique écartant les ambiguïtés, est le recours à un langage logique. Afin de concilier l’expressivité du langage logique utilisé et la calculabilité (voire la décidabilité) de la validité d’une proposition exprimée dans ce langage, c’est dans la famille des logiques de description (Baader *et al.*, 2017) qu’est choisi le langage logique d’une ontologie.

Une logique de description :

- permet plus d’expressivité que la simple logique propositionnelle, qui dispose d’un nombre réduit de connecteurs logiques pour former des propositions complexes à partir d’atomes propositionnels (non \neg , et \wedge , ou \vee , implique \Rightarrow , équivalent à \Leftrightarrow),
- sans poser les problèmes de décidabilité du calcul des prédicats (ou logique du premier ordre, *First Order Logic* en anglais) qui ajoute aux connecteurs déjà cités les quantificateurs universel (quel que soit \forall) et existentiel (il existe au moins un ... tel que \exists).

Afin d’éviter les problèmes de calculabilité et de décidabilité, les logiques de description se limitent aux relations unaires (à un seul argument) ou binaires (à deux arguments), recourent à un plus grand nombre de symboles (mais d’un usage plus restreint) et restreignent également les propositions qui peuvent être acceptées comme valides.

Destinées à la représentation des connaissances, elles parlent de *concepts*, de *rôles* et d’*individus* là où la logique du premier ordre parle de relations unaires, de relations binaires et de constantes. À partir de cette distinction entre concepts, rôles et individus, on parle de *Terminological Box* (TBox) pour la partie d’une base de connaissance qui exprime les connaissances sur les concepts et leurs relations, de *Relational Box* (RBox) pour la partie qui exprime les connaissances sur les relations, d’*Assertional Box* (ABox) pour la partie qui exprime les connaissances sur les individus.

Pour illustrer ce que signifie la modélisation de connaissances au moyen d’un langage logique nous pouvons tenter d’exprimer les propositions « tous les adultes sont responsables » et « il y a des pays qui ont une frontière commune avec la France » :

- pour exprimer que toute personne adulte est responsable, on pourra écrire : $\forall x : \text{Adulte}(x) \Rightarrow \text{estResponsable}(x)$ – si on écrit $\forall x : \text{Adulte}(x) \wedge \text{estResponsable}(x)$, on parle d’un monde uniquement composé d’adultes responsables (il n’y a pas d’autres entités que des adultes et donc pas de possibilité d’être responsable sans être une personne adulte);
- pour exprimer que certains pays ont une frontière commune avec la France, on pourra écrire : $\exists x \mid \text{Pays}(x) \wedge \text{frontiereFrance}(x)$ – si on écrit $\exists x \mid \text{Pays}(x) \Rightarrow \text{frontiereFrance}(x)$, on laisse la possibilité que des entités aient une frontière commune avec la France sans être des pays.

Une logique de description importante, parce qu’elle est la base de beaucoup d’autres, est *ALC* (*Attributive (Concept) Language with Complements*). Elle permet la création de classes complexes par intersection ($A \sqcap B$), par union ($A \sqcup B$), par complément ($\neg A$) de classes (y compris le complément de classes complexes), par l’utilisation des quantifications universelle ($\forall R.C$, avec R une relation et C une classe – qui désigne l’ensemble des instances telles que si elles sont dans une relation R à une autre instance, cette instance est instance de la classe C) et existentielle ($\exists R.C$ – qui désigne l’ensemble des instances qui sont dans la relation R avec au moins une instance de la classe C).

Citons également *SRIQ* (Horrocks *et al.*, 2006), qui est l’une des plus expressives des lo-

giques de description décidables et permet notamment la transitivité, la disjonction, la réflexivité, l'irréflexivité, l'antisymétrie des rôles; les restrictions qualifiées sur la cardinalité(\mathcal{Q})¹⁴; les propriétés inverses(\mathcal{S}); les classes définies comme des énumérations contenant un seul individu (\mathcal{O}); la réflexivité locale d'un rôle. A l'aide de ces langages formels, on peut établir un ensemble de formules qui constitueront un modèle.

De tels modèles peuvent être (et sont) envisagés sous l'angle de la théorie des modèles : on en étudiera les *interprétations* en s'appuyant sur la théorie des ensembles et sur une conception de la vérité héritée du logicien et philosophe polonais Alfred Tarski. Si le modèle dit que « les A sont des B qui sont dans une relation R asymétrique avec au moins un autre B », l'interprétation dans laquelle les B sont des personnes, les A sont des parents et la relation R est la relation aPourEnfant satisfera ce modèle (ainsi que l'interprétation dans laquelle les A sont à présent des employeurs et la relation R la relation aPourEmployé). Les ontologies informatiques sont donc des **théories** au sens de la théorie des modèles. Ces considérations permettent également de comprendre de quelle manière les ontologies peuvent rendre la signification, en quoi elles sont « sémantiques » : c'est en précisant *syntactiquement* les propositions d'une telle théorie qu'on peut restreindre les interprétations possibles (c'est-à-dire qui n'entraînent pas de contradiction) jusqu'à aboutir à une interprétation qui correspond à la signification que nous souhaitons traduire. C'est au niveau de la cohérence logique et dans un cadre ensembliste qu'une machine « comprend » les informations qu'elles traite.

Nous avons considéré la proximité des ontologies avec les systèmes d'organisation des connaissances et avec les théories de la théorie des modèles, mais dans quelle langue exactement sont exprimées les ontologies? Si cette question a pu recevoir différentes réponses au cours de l'histoire des ontologies informatiques, la famille de langages OWL (*Web Ontology Language*), promue par le *World Wide Web Consortium* (W3C) s'est imposée comme un standard de fait. Les différents langages OWL varient selon leur degré d'expressivité¹⁵. Le langage d'ontologie du Web OWL 2 est « compatible avec la logique de description SROIQ » (Motik *et al.*, 2012). En tant que langage, il est une extension du langage *Resource Description Framework Schema* (RDFS) qui permet de déclarer des classes et de les instancier (par la relation entre un individu et la classe à laquelle il appartient), de définir des sous-classes (relation de subsomption), de définir le domaine et la portée d'une relation (relation entre une classe « domaine de définition » ou « ensemble d'arrivée », et la relation considérée). En ce sens, RDFS permet déjà de définir des « ontologies légères ». RDFS est lui-même une extension du modèle de graphe *Resource Description Framework* (RDF), qui n'est pas un langage mais définit plutôt une structure de représentation, un modèle permettant l'échange de données sur le Web. Tout est susceptible de devenir une « ressource », à condition de disposer d'un identifiant unique (*Uniform Resource Identifier* – URI). Une information exprimée en RDF est un triplet sujet - prédicat - objet, qui définit une relation binaire (exprimée par le prédicat) entre une ressource (le sujet) et une ressource ou un littéral (l'objet, qui peut être lui-même une ressource mais aussi une valeur numérique, une chaîne de caractères, une date, etc., génériquement nommés « littéraux »).

Une ontologie fondée sur RDF (ce qui est le cas si elle est exprimée dans un langage OWL) peut donc être représentée sous la forme d'un graphe de connaissance dont les nœuds sont des ins-

14. Sur le sens des différentes lettres, qui correspondent à des extensions d'opérateurs, on peut consulter https://en.wikipedia.org/wiki/Description_logic

15. OWL 2 a trois sous-langages appelés « profils », correspondants à des finalités privilégiées : (a) OWL EL pour un raisonnement rapide sur un grand nombre de classes et de relations, (b) OWL QL pour une bonne capacité à requêter les données, (c) OWL RL pour le raisonnement sur les règles – <https://www.w3.org/TR/owl2-profiles/> – consulté le 15 août 2023.

tances, des classes ou des littéraux et les arcs des relations. Un tel graphe de connaissance peut être comparé à d'autres **bases de données orientées graphe** susceptibles d'accueillir des connaissances, comme les graphes attribués (*labeled property graphs* – LPG), qui présentent l'intérêt de pouvoir associer des attributs aux propriétés mais qui, par manque de standardisation, rendent plus difficile l'interopérabilité. Une comparaison plus étendue dépasserait les limites de cette présentation. Timón-Reina *et al.* (2021) proposent une comparaison entre LPG et graphes RDF au regard de leurs applications dans le domaine biomédical.

Deux autres caractéristiques importantes des ontologies exprimées dans le langage OWL sont à noter. La première, qui provient des logiques de description, est qu'elles ne font pas d'hypothèse par défaut sur d'éventuelles informations qui leur seraient inconnues. C'est ce qui est parfois appelé « hypothèse du monde ouvert » : le fait qu'une connaissance soit absente de l'ontologie ne signifie pas qu'elle soit fautive (Keet, 2013). Ainsi, si nous spécifions dans notre modèle qu'une personne doit avoir exactement un prénom (ni plus, ni moins) et que nous introduisons une instance de personne qui n'a pas de prénom, aucune erreur n'est relevée par un raisonneur : que le prénom de cette personne ne soit pas connu ne signifie pas qu'elle n'a pas de prénom. La deuxième est qu'elles rejettent l'« hypothèse du nom unique » : deux instances portant des noms différents peuvent, sauf mention explicite du contraire, se référer à une seule et même entité. Dans notre modèle personne-prénom, un instance de personne peut être associée à deux instances de prénom sans que le raisonneur n'y voie d'erreur logique, tant que nous n'avons pas explicitement signifié que les deux instances de prénom étaient disjointes. Ces deux caractéristiques distinguent nettement les ontologies (exprimées en OWL) des **bases de données relationnelles**. Dans celles-ci, c'est l'« hypothèse du monde fermé » qui prévaut (le vrai et le connu se superposent exactement) ; et elles souscrivent à l'hypothèse du nom unique (des noms différents dénotent des individus différents).

Remarquons que c'est ce recours aux logiques de description qui permet aux ontologies de contribuer (en leur adjoignant un moteur d'inférence, c'est-à-dire un logiciel mettant en œuvre un algorithme d'application de règles logiques) à l'élaboration de modèles symboliques d'**intelligence artificielle** permettant de simuler le raisonnement inférentiel déductif. On parle ici de modèle *symbolique* dans la mesure où l'information est ici traitée au moyen de symboles et de règles. Les ontologies informatiques, de ce point de vue, peuvent être situées dans l'arsenal des outils et techniques du champ de l'intelligence artificielle.

Les caractéristiques des ontologies que nous venons d'étudier, et dont nous commençons par dire qu'elles rendaient les ontologies difficiles à appréhender, en font aussi l'intérêt, notamment dans le domaine de la recherche biomédicale. Hoehndorf *et al.* (2015) met en avant l'usage d'identifiants uniques (pour l'intégration des données), les libellés associés (pour le traitement automatique du langage et les interfaces utilisateur), les métadonnées et les descriptions (pour une compréhension de la sémantique des classes), les axiomes et les définitions formelles (pour une consultation et une exploitation automatisées à l'aide de logiciels).

3.1.2 Conception et cycle de vie

Une fois précisées les caractéristiques des ontologies, il faut également les envisager comme l'aboutissement d'un processus de production et plus généralement comme ayant un cycle de vie. Les méthodes d'ingénierie ontologiques proposées sont le plus souvent issues de l'ingénierie logicielle. Ces méthodes sont nombreuses et rendre compte de leurs nuances dépasserait le cadre de cette section. Nous présentons d'abord METHONTOLOGY, souvent citée et qui comporte de nombreux éléments communs à ces méthodes, avant d'offrir un panorama plus large.

Visant à opérer la transition « de l'art ontologique à l'ingénierie ontologique », la méthodologie METHONTOLOGY (Fernández-López *et al.*, 1997) structure la construction d'ontologies selon les étapes suivantes :

1. la spécification, contenant au moins :
 - la finalité de l'ontologie,
 - le niveau de formalisme de l'ontologie,
 - le champ d'application (termes à représenter, caractéristiques, granularité),
2. l'acquisition de connaissances (activité indépendante du processus de développement de l'ontologie *per se*), c'est-à-dire par exemple :
 - des entretiens informels avec des experts,
 - de l'analyse informelle de textes,
 - de l'analyse formelle de textes,
 - des entretiens structurés avec des experts;
3. la conceptualisation, comprenant :
 - l'établissement d'un glossaire de termes,
 - la séparation des concepts et des verbes,
 - l'établissement d'un dictionnaire de données pour les concepts du domaine (signification, attributs, instances, etc.),
 - l'établissement de tables des attributs d'instance, des attributs de classe, des constantes, des instances, et d'arbres de classification des attributs,
 - l'établissement d'un dictionnaire des verbes et de tables des conditions préliminaires ou successives à l'exécution d'une action,
 - l'établissement de tables des formules et de tables des règles;
4. la formalisation¹⁶;
5. l'intégration, autant que possible, de définitions provenant d'ontologies déjà existantes;
6. l'implémentation, requérant un environnement de développement d'ontologie complet;
7. l'évaluation, qui porte sur :
 - la correction technique de l'ontologie, de ses environnements logiciels et de sa documentation,
 - la correspondance au système qu'elle est censée représenter;
8. la documentation, qui s'enrichit à chaque phase, pour contenir finalement :
 - un document de spécification des besoins,
 - un document sur l'acquisition de connaissances,
 - un document sur le modèle conceptuel,
 - un document sur la formalisation,
 - un document sur l'intégration,
 - un document sur l'implémentation,
 - un document sur l'évaluation.

Citant ce travail, Corcho *et al.* (2007) distinguent :

- les activités de gestion d'ontologie :
 - activité de planification,
 - activité de contrôle,

16. Nous ajoutons cette étape. Elle est indiquée dans l'étape sur la documentation mais ne fait pas l'objet d'un développement dans l'exposé de la méthodologie.

- activité d'assurance de la qualité;
- les activités orientées vers le développement d'ontologies :
 - (pré-développement) étude de l'environnement,
 - (pré-développement) étude de faisabilité,
 - (développement) activité de spécification,
 - (développement) activité de conceptualisation,
 - (développement) activité de formalisation,
 - (développement) activité d'implémentation,
 - (post-développement) activité de maintenance,
 - (post-développement) activité d'évolution (liée aux réutilisations);
- les activités de soutien :
 - activité d'acquisition de connaissance,
 - activité d'évaluation,
 - activité d'intégration,
 - activité de combinaison (*merging*),
 - activité d'alignement,
 - activité de documentation,
 - activité de gestion de la configuration (versions de la documentation et de l'ontologie).

Nous pouvons observer que si le « cœur » du processus de développement reste le même, la méthodologie s'est étoffée et fait plus de place aux activités liées au cycle de vie. Le cycle de vie lui-même s'est complexifié, comme l'indiquent les auteurs : « le cycle de vie original de METHONTOLOGY a été modifié récemment pour tenir compte du fait qu'un plus grand nombre d'ontologies sont disponibles dans des bibliothèques d'ontologies ou diffusées sur internet, de sorte que leur réutilisation par d'autres ontologies et applications s'est accrue. Les ontologies de domaine peuvent être réutilisées pour en construire d'autres d'une granularité et d'une couverture plus grandes, ou peuvent être fusionnées avec d'autres pour en créer de nouvelles » (Corcho *et al.*, 2007, p.50).

Keet (2018) propose une grille d'analyse des différentes méthodes et méthodologies de conception d'ontologies. Les méthodes se situant au niveau « macro » sont celles qui adoptent le point de vue des systèmes d'information et des technologies de l'information. Parmi celles-ci, l'auteur distingue les méthodes de conception « en cascade » (modèle linéaire et séquentiel) et les méthodes « agiles » (modèle itératif, avec livraisons successives intégrant les retours de l'utilisateur). Les méthodes se situant au niveau « micro » « se concentrent sur les directives visant à formaliser le domaine, c'est-à-dire sur la manière de passer d'une représentation informelle à une représentation basée sur la logique », ce qui « affecte [...] certains choix et étapes au niveau macro ».

Selon cette grille de lecture, METHONTOLOGY est une méthode macro, en cascade. M. Keet indique qu'elle a été supplantée par la méthodologie NeON¹⁷ (Suárez-Figueroa *et al.*, 2012), fondée sur des scénarios, qui propose « de nombreux itinéraires possibles composés de plusieurs étapes » et « comprend plus de détails pour la phase de spécification » (Keet, 2018, p.94-95). Dans cette catégorie de méthodologies, nous pourrions également citer UPON (De Nicola *et al.*, 2005) (pour *Unified Process for ONtology building*), dérivée de l'*Unified Software Development Process* et adoptant l'*Unified Modeling Language* (UML). Cette méthodologie, axée sur les cas d'usages, est considérée comme plus adaptée à des ontologies d'application qu'à des ontologies de domaine.

Parmi les méthodes agiles, nous pouvons citer SAMOD (pour *Simplified Agile Methodology for*

17. <https://oeg.fi.upm.es/index.php/en/methodologies/59-neon-methodology/index.html> – consulté le 15 août 2023.

Ontology Development) (Peroni, 2017) et AMOD (pour *Agile Methodology for Ontology Development*) (Abdelghany *et al.*, 2019).

Ces méthodes ne sont pas disertes sur la manière dont sont réutilisées les ressources extérieures à l'ontologie en développement. Ce point amène à une autre distinction méthodologique : entre les méthodes ascendantes et les méthodes descendantes. Les méthodes descendantes (*top-down* en anglais) s'appuient, pour faciliter le développement de nouvelles ontologies, sur les principes génériques de modélisation mis en œuvre dans des ontologies de plus haut niveau – typiquement des ontologies fondationnelles – que l'ontologie en développement va spécifier. Nous revenons dans la section 1.3 du chapitre 2 sur les bénéfices de l'utilisation d'ontologies fondationnelles pour guider le développement de nouvelles ontologies. Les méthodes ascendantes (*bottom-up* en anglais) permettent de surmonter le *knowledge acquisition bottleneck* (Keet, 2018, p.139) en réutilisant les données, les informations, les connaissances déjà existantes¹⁸.

Ainsi, la méthode TERMINAE, concrétisée dans une plateforme (Aussenac-Gilles *et al.*, 2008), décrit de la manière suivante les étapes du processus de développement d'une ontologie à partir de ressources textuelles et termino-ontologiques d'un domaine :

- constitution d'un corpus (mis sur support informatique) couvrant le domaine, ce qui suppose une expertise sur les documents à inclure ;
- analyse linguistique recourant à des techniques et outils de Traitement Automatique des Langues et aboutissant à l'extraction de termes et de relations ;
- normalisation, c'est-à-dire (a) tri parmi les termes et relations et choix de syntagmes préférés et (b) définition sémantique de ces termes et relations ;
- formalisation en une base de connaissances.

Citons également la méthode ARCHONTE (Bachimont *et al.*, 2002), qui associe trois descriptions à une primitive¹⁹ :

- « une description sémantique linguistique qui fournit à un utilisateur humain une compréhension non ambiguë d'un terme ;
- une description sémantique formelle qui fournit à un utilisateur humain un compte rendu mathématique et formel du niveau précédent ;
- une description computationnelle qui rend explicite le comportement prévu de l'ordinateur lors de la manipulation de cette primitive ».

Cette distinction guide les trois étapes de la méthodologie : normalisation sémantique, formalisation, opérationnalisation. L'étape de normalisation sémantique aboutit à une ontologie différentielle, c'est-à-dire qui suit certains principes différentiels exposés également par Bachimont (2000), permettant de déterminer la signification d'un nœud du graphe en fonction de ses parents et de ses frères :

- **le principe de communauté avec le père** – « Toute unité se détermine par l'identité qu'elle possède avec l'unité parente. Il faut expliciter en quoi l'unité fille est identique à l'unité parente. C'est, *mutatis mutandis*, le principe aristotélien de définition par le genre proche » ;
- **le principe de différence avec le père** – « Toute unité se distingue de l'unité parente, sinon il n'y aurait pas lieu de la définir. Il faut donc expliciter la différence qui la distingue de l'unité parente. C'est, *mutatis mutandis*, le principe aristotélien de définition par la différence spécifique » ;

18. La disponibilité de telles ressources, notamment sous forme textuelle, est en forte croissance dans la période récente, de même que les outils permettant de les exploiter.

19. Dans un langage formel, les primitives sont les symboles à partir desquels peuvent être construites des formules. Les primitives ici visées sont principalement les classes et les relations qui figureront dans l'ontologie, donc les primitives non logiques du domaine considéré.

- **le principe de différence avec les frères** – « Toute unité se distingue de ses frères sinon il n’y aurait pas lieu de la définir. Il faut donc expliciter la différence de l’unité avec chacune des unités sœurs. Ce principe n’est pas aristotélien, mais provient du paradigme différentiel choisi »;
- **le principe de communauté avec les frères** – « Toutes les unités filles d’une unité parente possèdent par définition un même trait générique, celui qu’elles partagent avec l’unité parente. Mais il faut établir une autre communauté entre les unités filles; c’est celle qui permet de définir des différences mutuellement exclusives entre les unités filles. Par exemple, l’unité parente est “être humain”, et les unités filles sont “homme” et “femme”. Ces unités partagent le fait d’être des humains. Mais cette propriété ne permet pas de définir en quoi sont différents les hommes et les femmes. On choisit alors comme principe de communauté la sexualité, où l’on peut attribuer à “homme” le trait masculin et à “femme” le trait féminin. Ces deux traits sont mutuellement exclusifs car ce sont deux valeurs possibles d’une même propriété. Le dernier principe n’a pas exactement la même utilité que les 3 autres. En fait, il n’intervient que pour rendre le 3^e principe possible. Il repose sur le fait qu’il ne suffit pas de poser des propriétés différentes pour caractériser les différences entre les unités filles, mais qu’il faut savoir en quoi elles sont différentes ».

La recherche visant l’apprentissage d’ontologies ou le peuplement d’ontologies existantes à partir d’un corpus de ressources informationnelles (*ontology learning*) s’oriente vers les méthodes automatisées ou semi-automatisées, que ce soit à partir de textes, à partir de données liées sur le *web*, par apprentissage de concepts (méthodes d’acquisition automatisée de schémas de données), par *crowdsourcing* (comme alternative au recours aux experts).

Une méthode de développement trouve difficilement sa place dans la dichotomie ascendant / descendant : les patrons de conception ontologique (*Ontology Design Patterns* – ODP). Inspirés par l’architecture²⁰, ils apportent des solutions légères et adaptables à des problèmes de modélisation récurrents. Keet (2018, p.152) y voit « une méthode intermédiaire de développement d’ontologies », dans la mesure où « il peut s’agir de fragments d’ontologie fondationnelle astucieusement modularisés qui servent de bribes de conception » mais que les ODPs « peuvent également être considérés comme une manière ascendante de trouver des patrons qui sont ensuite réutilisés dans l’ontologie ».

De nombreuses méthodes ont également été développées pour la mise en correspondance d’ontologies (*ontology mapping*)²¹, souvent en vue de combiner efficacement des ontologies (*ontology merging*).

3.1.3 Évaluation de la qualité

Comme nous venons de le voir, évaluer la qualité d’une ontologie peut être une étape de sa conception et de son cycle de vie. Cela peut également diriger le choix, parmi des ontologies existantes, de celle que nous réutiliserons pour une finalité spécifique. La question de l’évaluation s’impose aussi avec le développement du web sémantique et la multiplicité des acteurs qu’il implique. La formalisation de l’évaluation peut permettre une plus grande acceptabilité des ontologies hors du milieu de l’ingénierie ontologique, notamment par les industriels. Toutefois, les cri-

20. Alexander *et al.* (1977) proposent un *Pattern Language* pour l’urbanisme et l’architecture : « The elements of this language are entities called patterns. Each pattern describes a problem that occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice. »

21. Harrow *et al.* (2019) proposent une revue de ces méthodes pour le domaine biomédical.

tères classiques de qualité logicielle s'appliquent difficilement aux ontologies (qui ne sont pas des « fichiers exécutables »), ce qui a stimulé la recherche de critères propres.

Gómez-Pérez (2004) sépare deux aspects de l'évaluation d'une ontologie : (a) la *vérification* qui « consiste à construire l'ontologie correctement, c'est-à-dire à s'assurer que ses définitions mettent correctement en œuvre les exigences et les questions de compétence de l'ontologie²², ou qu'elles fonctionnent correctement dans le monde réel » et (b) la *validation* qui « consiste à déterminer si la signification des définitions de l'ontologie modélise réellement le monde réel pour lequel l'ontologie a été créée. L'objectif est de prouver que le modèle du monde (s'il existe et est connu) est conforme au monde modélisé formellement ».

Vrandečić (2009) reformule cette distinction : « *ontology verification answers if the ontology was built in the right way, whereas ontology validation answers if the right ontology was built* » et s'attache plus particulièrement aux critères de vérification, en faisant la synthèse de plusieurs publications (Gangemi *et al.*, 2005; Gómez-Pérez, 2004; Gruber, 1995; Grüninger & Fox, 1995a; Obrst *et al.*, 2007) :

- **exactitude (accuracy)** – concerne la conformité à l'expertise de plusieurs utilisateurs et à la représentation correcte du monde réel²³ ;
- **adaptabilité** – concerne l'anticipation des utilisations, des extensions, des petites modifications ;
- **clarté** – concerne la communication efficace de la signification des termes, les définitions (indépendantes du contexte, documentées), la compréhensibilité de l'ontologie (ou ergonomie cognitive, transparence, intelligibilité) ;
- **exhaustivité (completeness) / compétence** – concerne la couverture du domaine d'intérêt et les réponses aux questions de compétence, la présence de tous les concepts pertinents et de leurs représentations lexicales ;
- **efficacité informatique (computational efficiency)** – concerne l'application de raisonneurs à l'ontologie (rapidité de détermination de la satisfaisabilité, de classification des instances, d'obtention de résultats de requêtes) ;
- **concision** – concerne la pertinence et la non redondance des axiomes, le minimalisme de l'engagement ontologique (c'est-à-dire spécifiant la théorie la plus simple possible, ne définissant que les termes essentiels, ayant la théorie philosophique sous-jacente la plus simple possible) ;
- **cohérence** – concerne la cohérence logique des axiomes, la cohérence des descriptions formelles avec les descriptions informelles (documentation), la cohérence entre la connaissance et son encodage (visant un *minimal encoding bias*) ;
- **adaptation à l'organisation / efficacité commerciale** – concerne le déploiement de l'ontologie dans une organisation, les exigences légales qui s'appliquent à elle, le fait qu'elle soit facile d'accès, alignée à d'autres ontologies, partagée autant que nécessaire.

Sur le critère de cohérence, l'auteur renvoie à OntoClean (Guarino & Welty, 2004) pour l'adéquation de la structure taxonomique. OntoClean est une méthodologie permettant de mettre en évidence les subsomptions problématiques, de clarifier l'engagement ontologique, en fournissant un cadre formel pour éviter certains écueils courants de modélisation. Elle s'appuie sur les notions philosophiques suivantes :

- essence – « une propriété²⁴ d'une entité est essentielle à cette entité si elle doit être vraie

22. Les questions de compétence sont les questions des utilisateurs auxquelles l'ontologie doit permettre de répondre. La notion est introduite par Grüninger & Fox (1995a).

23. Ce premier critère concerne la validation plutôt que la vérification.

24. Précisons avec les auteurs que les « propriétés » sont ici à entendre « comme les significations (ou intensions)

dans tous les mondes possibles, c'est-à-dire si elle s'applique nécessairement à cette entité »;

- rigidité – « une forme particulière d'essentialité est la rigidité; une propriété est rigide si elle est essentielle à toutes ses instances possibles; une instance d'une propriété rigide ne peut cesser d'être une instance de cette propriété dans un monde différent » (par exemple, « être un humain » est une propriété rigide, quand « être un étudiant » ne l'est pas : dans un monde différent, par exemple dans le futur, une entité qui a la propriété d'être étudiant dans le monde actuel ne l'aura plus);
- identité – « les critères d'identité sont des conditions utilisées pour déterminer l'égalité (conditions suffisantes) et qui sont impliquées par l'égalité (conditions nécessaires) » (il s'agit ici de la capacité à reconnaître des entités comme étant les mêmes ou comme étant différentes : deux durées d'une heure sont toutes les deux une durée d'une heure, deux intervalles temporels d'une heure successifs sont des intervalles temporels différents);
- unité – « l'unité se réfère au problème de la description des parties et des limites des objets, de telle sorte que nous sachions en général ce qui fait partie de l'objet, ce qui n'en fait pas partie, et dans quelles conditions l'objet est *un tout* » (par exemple, un océan est un tout mais « de l'eau » n'est pas un tout).

De ces notions de base sont tirées un ensemble de métapropriétés (semi-rigidité, anti-rigidité, critère d'identité propre, etc.) à partir desquelles sont formulées des règles qui permettent d'évaluer la hiérarchie de classes (par exemple, une sous-propriété d'une propriété anti-rigide *doit* être anti-rigide²⁵). Les mésusages les plus courants de la subsomption repérés par Guarino & Welty (2002) sont :

- la confusion entre subsomption et instanciation (« humain » ne peut pas être une sous-classe d'« espèce animale », sans quoi une instance d'humain serait aussi une instance d'espèce animale; « humain » pourrait en revanche être une instance d'« espèce animale »²⁶),
- la confusion entre subsomption et relation partie-tout (« moteur » n'est pas une sous-classe de « voiture », une instance de moteur n'étant pas une instance de voiture),
- la confusion de la subsomption et de la disjonction (« moteur » n'est pas une sous-classe de « partie de voiture », le moteur restant un moteur s'il est placé dans un bateau; mais on pourrait dire que toute partie de voiture est soit un moteur, soit une roue, soit un volant, etc.),
- la subsomption ne permet pas de représenter la polysémie (le mot « livre » peut désigner à la fois un volume relié et une entité abstraite, mais dans une ontologie, instancier la classe « livre » oblige à choisir dans quel sens on entend le mot),
- la confusion entre subsomption et constituant (« océan » n'est pas une sous-classe d'« eau », pour les raisons d'unité exposées ci-dessus).

Dans une veine moins théorique, mais toujours sur le critère de cohérence, citons également *l'Ontology Pitfall Scanner!* (OOPS!) (Poveda-Villalón *et al.*, 2014) – disponible en ligne sous forme

d'expressions telles qu'être une pomme ou être une table, qui correspondent à des prédicats unaires dans la logique du premier ordre ». Chaque propriété, en ce sens, est donc associée à une *classe* de l'ontologie et ce qui est en jeu est bien la subsomption entre classes dans la taxonomie.

25. Pour reprendre notre exemple, « être humain » ne peut pas être une sous-propriété d'« être étudiant ». Si c'était le cas, les instances de la classe « humain » seraient aussi des instances de la classe « étudiant ». Par la rigidité de la propriété « être humain », ces instances ne pourraient plus dans aucun monde possible cesser d'être étudiant sans cesser d'être. En revanche, « être étudiant de première année » peut être une sous-propriété d'« être étudiant » qui peut être une sous-propriété d'« être humain ».

26. Qu'une entité puisse être à la fois une classe et l'instance d'une (méta)classe, ou *punning*, est d'ailleurs permis par le langage OWL2 DL (cf. https://www.w3.org/TR/owl2-new-features/#F12:_Punning – consulté le 15 août 2023).

de service web²⁷ – qui permet de détecter, dans une ontologie, des « écueils » (*pitfalls*) comme l'absence de disjonction, l'absence de domaine et de codomaine pour les propriétés, les erreurs dans les relations inverses, la définition d'une relation inverse pour une propriété symétrique, etc. (le catalogue compte actuellement 41 écueils étiquetés comme « mineur », « important », « critique »²⁸).

Pour en venir à la validation, Vrandečić (2009, p.295) remarque que si « la validation d'une ontologie est un élément important de l'évaluation de la qualité d'une ontologie et constitue généralement le seul moyen de garantir l'exactitude des connaissances encodées dans l'ontologie », « la plupart des approches de validation requièrent la coopération étroite d'experts du domaine et de l'ingénierie ontologique », ce qui a pour conséquence que « souvent, la validation ne peut pas être effectuée automatiquement ». Obrst *et al.* (2007) proposent toutefois un certain nombre d'approches :

- l'évaluation basée sur des tâches applicatives (formulation de requêtes, exactitude des réponses, degré d'explication fourni par le système, degré de réutilisabilité de l'ontologie, évolutivité de la base de connaissance) ;
- la comparaison de l'ontologie avec une source de données du domaine (ce qui pose les problèmes classiques de mise en correspondance des termes) ;
- l'évaluation humaine selon des critères (clarté de la documentation, cohérence et non-ambiguïté dans l'emploi des termes, procédures de mise à jour et versionnage, procédures de prise en compte des corrections et des ajouts des utilisateurs) ;
- l'impact dans des tâches de Traitement Automatique des Langues (comme l'extraction d'information, le *question answering*, le résumé) ;
- l'utilisation de la réalité comme référence (en examinant si les versions successives d'une ontologie reflètent les changements intervenus dans la réalité, dans la compréhension scientifique, dans l'appréciation de ce qui est pertinent à inclure dans l'ontologie) ;
- l'accréditation (ou la certification) d'une ontologie, d'une équipe de développeurs, d'une organisation.

En considérant une ontologie comme un « objet sémiotique », Gangemi *et al.* (2006) évitent la notion de validation telle que nous venons de l'exposer²⁹ et proposent de distinguer trois dimensions d'évaluation :

- la dimension *structurelle* qui « se concentre sur la syntaxe et la sémantique formelle, c'est-à-dire sur les ontologies représentées sous forme de graphes. Sous cette forme, les propriétés topologiques, logiques et méta-logiques d'une ontologie peuvent être mesurées au moyen d'une métrique sans contexte » (l'ontologie est ici vue comme un objet informationnel) ;
- la dimension *fonctionnelle* qui « est liée à l'utilisation prévue d'une ontologie donnée et de ses composants, c'est-à-dire leur fonction dans un contexte. L'accent est mis sur la conceptualisation spécifiée par une ontologie » (l'ontologie est ici vue comme un langage) ;
- la dimension « profilage pour faciliter l'utilisation » (*usability-profiling*) qui « se concentre sur le profil de l'ontologie (annotations), qui concerne généralement le contexte de communication d'une ontologie (c'est-à-dire sa pragmatique) » (l'ontologie est ici vue comme un méta-langage).

La première dimension recouvre beaucoup d'aspects de la « vérification ». La deuxième dimension s'affronte au *matching problem* que les auteurs définissent comme la mesure dans laquelle l'ontologie reflète l'expérience d'une communauté (expertise, compétence, tâche). Son évaluation ne

27. <https://oops.linkeddata.es/> – consulté le 15 août 2023.

28. <https://oops.linkeddata.es/catalogue.jsp> – consulté le 15 août 2023.

29. La notion de validation reflète à notre sens une approche « réaliste » des ontologies.

peut être complètement automatisée mais peut s'appuyer sur une confrontation aux jugements des experts ou à « un jeu de données considérées comme l'expression qualifiée du jugement d'experts ». La troisième dimension, consacrée aux annotations, se décline en trois niveaux : celui de la reconnaissance (donner l'information et les instructions nécessaires aux utilisateurs), celui de l'efficacité (faciliter l'intégration de l'ontologie dans l'organisation qui l'utilise), celui de l'interfaçage (faciliter la correspondance de l'ontologie avec des interfaces utilisateur). Nous ne détaillons pas toutes les « descriptions structurées de la qualité d'une ontologie (qoods) » développées par les auteurs, mais nous revenons en particulier sur l'une de celles-ci dans la section 2.2 du chapitre 2 : l'ergonomie cognitive.

3.2 Les applications des ontologies

Nous avons présenté de premier abord les ontologies comme des représentations formelles de domaines de connaissances particuliers et nous avons vu l'intérêt de telles représentations en termes d'expressivité et de capacité d'inférence. Pour aller au-delà de ce premier constat, on peut se demander ce que permettent les ontologies, une fois établies. Nous présentons ici les principales applications des ontologies et leurs pré-requis :

Intégration de données : lorsqu'on dispose de différentes sources de données hétérogènes, qu'on souhaite exploiter conjointement, une ontologie peut représenter une langue commune dans laquelle sont traduits les différents vocabulaires des sources, ce qui permet de les rendre interopérables. Cette utilisation suppose généralement en amont que la sémantique de ces vocabulaires soit explicitée, entre celui qui élabore l'ontologie et les producteurs de données. Le rôle que peuvent jouer les ontologies dans l'intégration et l'échange de données en permettant une interopérabilité sémantique est bien établi (Bodenreider, 2008).

Extraction de connaissances : si certaines de nos données ne sont pas structurées – au sens informatique du terme, c'est-à-dire si elles ne sont pas structurées d'une manière qui permette directement leur exploitation par une machine, par exemple dans le cas de documents et particulièrement de données textuelles – l'annotation sémantique des termes et des relations trouvés dans les documents permet d'en extraire des connaissances qui trouvent leur place dans l'ontologie. Cette phase d'annotation sémantique s'appuie sur une phase précédente de traitement automatique du langage permettant de repérer de façon non ambiguë, dans les documents, les termes correspondants aux expressions associés à une entité de l'ontologie.

Raisonnements inférentiels : la représentation formelle des connaissances permet la déduction de connaissances implicites (sur les instances, sur l'organisation des classes, sur des relations complexes) découlant des connaissances déjà introduites dans l'ontologie. Cette déduction est obtenue au moyen d'un moteur d'inférence.

Validation des connaissances : un premier aspect concerne la validation du modèle ontologique : l'utilisation d'un moteur d'inférence permet de s'assurer de l'absence de contradiction ou d'incohérence dans l'ontologie; un deuxième aspect concerne la validation de la base de connaissances établie selon le modèle ontologique. Sur ce deuxième aspect, il peut être nécessaire de compléter la modélisation ontologique d'un ensemble de règles spécifiant les propriétés que doit avoir le graphe produit (la sémantique de « monde ouvert » spécifique à l'ontologie ne permettant pas toujours par elle-même une telle validation).

Exploration des connaissances : l'effort consenti pour élaborer une représentation ontologique des connaissances – explicite sur sa sémantique et sur sa modélisation des données – per-

met d’alléger l’effort à consentir pour explorer ultérieurement ces connaissances. Dans le cas d’une base de connaissances étendue, comportant un grand nombre de triplets RDF, il peut être nécessaire de mettre en place une base de données spécifiquement conçue pour gérer le stockage pérenne et l’indexation de triplets RDF (un *triplestore*) afin de permettre un requêtage efficace au moyen du langage standard SPARQL.

Systèmes d’aide à la décision (et plus généralement systèmes « prescriptifs » – assistance, recommandation – proposant une action à mener à partir des informations à disposition) : ici, ce sont les propriétés déjà exposées des ontologies (modélisation expressive, intégration de données hétérogènes, possibilités inférentielles) qui peuvent être mises à contribution pour produire une aide à la décision explicable.

Web sémantique : si le Web peut être considéré basiquement comme un réseau dont les nœuds sont des documents et les arcs des liens hypertextes, le syntagme « web sémantique » (entre autres syntagmes proches comme « web de données » ou « données ouvertes liées »), « met en avant la possibilité d’échanger les schémas de nos données et la sémantique associée afin d’enrichir la gamme des traitements automatiques qui peuvent leur être appliqués » (Bachimont *et al.*, 2011). Les nœuds sont des identifiants de choses (“*Things, not strings*”, pour reprendre Singhal (2012) parlant du *Google Knowledge Graph*) et les arcs sont des relations éventuellement complexes. Les ontologies en sont une pierre angulaire.

Dans le travail que nous présentons, les ontologies sont mises en œuvre pour l’essentiel à des fins d’intégration de sources de données hétérogènes (pour les données structurées) et d’annotation sémantique (pour les données non structurées).

3.3 L’architecture modulaire

Comme son nom l’indique, une ontologie modulaire est une ontologie composée de modules reliés les uns aux autres, et s’oppose donc à une ontologie d’un seul tenant, « monolithique ». Les modules sont eux-mêmes des ontologies, sont autonomes et effectuent une partition qui a du sens au regard du domaine de connaissance considéré. Ce type d’architecture est à envisager lorsqu’une ontologie monolithique serait trop grande pour être facilement manipulée ou lorsqu’elle fait intervenir des collaborateurs ayant des domaines d’expertise très différents.

Un premier type de modularisation à mentionner est une modularisation par couches, telle qu’elle est envisagée par Guarino (1998), selon le niveau de généralité des différents modules. L’auteur distingue quatre sortes d’ontologies : la *top-level ontology*, dont dépendent *domain ontology* et *task ontology*, et l’*application ontology* qui dépend de ces deux dernières ; les relations de dépendance correspondent ici à des relations de spécialisation : chaque couche plus basse spécifie les concepts de la ou des couche(s) plus haut placée(s) dont elle dépend. La modularisation par couches qui s’est imposée par la suite envisage plutôt, dans une vision minimale, trois niveaux : celui de la *top-level ontology*³⁰, qui modélise les concepts et les relations les plus généraux ; celui de la *core ontology*, qui modélise un champ disciplinaire spécifique couvrant plusieurs domaines, modélisés par des *domain ontologies*.

C’est dans le fait de placer plusieurs ontologies de domaine sous une même ontologie-noyau (*core ontology*), ou plusieurs ontologies-noyau sous une même ontologie fondationnelle, qu’intervient un deuxième type de modularisation, qu’on pourrait dire « par silos » (nous évitons de recourir aux termes « horizontal » et « vertical », qui ne sont, ici, pas intuitifs), où chaque module spécifie pour un certain domaine le module de la couche supérieure dont il dépend. Notons que

30. ou *upper level / generic / foundational ontology*; nous traduirons par « ontologie fondationnelle ».

des architectures plus complexes peuvent être imaginées, comme dans l'ontologie modulaire SSN (Hitzler *et al.*, 2019).

Les principaux intérêts de l'architecture modulaire sont les suivants :

Réutilisabilité : chaque module, dans la mesure où il est autonome, peut être repris séparément et réutilisé dans un autre contexte, pour d'autres applications ;

Maintenance et évolutivité : la nécessité de faire évoluer un module, que ce soit pour des raisons de maintenance ou pour intégrer de nouvelles connaissances, ne crée pas une nouvelle version de l'ontologie tout entière mais seulement une nouvelle version d'un module ; de même, un module couvrant un domaine non auparavant traité peut être élaboré séparément puis ajouté ;

Collaboration : l'expert d'un domaine peut se focaliser sur l'ontologie qui le concerne ; si plusieurs développeurs contribuent aux différents modules, la gestion des versions s'en trouve également facilitée ;

Validation : qu'on considère la validation humaine ou la validation au moyen d'un raisonneur, travailler sur de plus petites ontologies facilite le processus.

Il faut toutefois noter qu'une architecture modulaire, malgré son intérêt, peut aussi générer des difficultés : (a) dans la conception (définir des modules en silos qui ne se recouvrent pas), (b) dans l'évolution (qui peut mettre en danger l'interopérabilité entre les différents modules), (c) dans la gestion de la complexité (la complexité initiale de l'ontologie monolithique est en partie reportée sur l'architecture elle-même).

4 Conclusion

A l'issue de ce tour d'horizon de notre terrain de recherche, nous avons pu d'abord caractériser cette approche récente de la médecine qu'est la médecine de précision et donner quelques aperçus sur l'émergence plus récente encore de la psychiatrie de précision. Nous avons également fait le constat que la rencontre entre psychiatrie de précision et intelligence artificielle s'articulait autour des méthodes et des algorithmes de l'apprentissage automatique.

Nous avons ensuite présenté les objectifs du projet PsyCARE et sa partition en lots de travaux, qui laisse apparaître qu'il s'agit pleinement d'un projet de psychiatrie de précision, la majeure partie des intitulés de ces lots de travaux pouvant être reliés aux sujets de la psychiatrie de précision. Au cœur de ce projet, un lot de travail en particulier retient notre attention : celui concernant l'établissement d'une plateforme de collecte de données. Notre caractérisation de la médecine de précision mettait en évidence qu'il s'agit d'une médecine requérant la mise en commun de données volumineuses et hétérogènes.

Nous avons enfin présenté les ontologies informatiques, leurs méthodes de conception et d'évaluation, leurs principales applications. Au moment de réunir sur une plateforme des données volumineuses et hétérogènes, la pertinence des ontologies informatiques en matière d'intégration des données est donc à considérer pour établir un modèle commun dont la cohérence pourra être validée automatiquement. Sont également pertinentes, dans le contexte d'un projet de recherche, les applications des ontologies en matière d'extraction des connaissances et d'exploration des données. Nous avons insisté sur un type particulier d'architecture ontologique, la modularité, pour son intérêt dans un contexte collaboratif, pour faire face à l'évolutivité dans le temps de l'ontologie et pour faciliter une réutilisation ultérieure.

Tout ceci légitime, de notre point de vue, l'utilité de la proposition que nous faisons d'une ontologie modulaire pour le projet de médecine personnalisée PsyCARE.

La Petite Ontologie Fondationnelle

« Le cuisinier Ting dépeçait un bœuf pour le prince Wen-houei. On entendait des houa quand il empoignait de la main l'animal, qu'il retenait sa masse de l'épaule et que, la jambe arqueboutée, du genou l'immobilisait un instant. On entendait des houo quand son couteau frappait en cadence, comme s'il eût exécuté l'antique danse du Bosquet ou le vieux rythme de la Tête de Lynx.

– C'est admirable! s'exclama le prince, je n'aurais jamais imaginé pareille technique!

Le cuisinier posa son couteau et répondit : Ce qui intéresse votre serviteur, c'est le fonctionnement des choses, non la simple technique. (...) Quand je rencontre une articulation, je repère le point difficile, je le fixe du regard et agissant avec une pudeur extrême, lentement je découpe. Sous l'action délicate de la lame, les parties se séparent dans un houo léger comme celui d'un peu de terre que l'on pose sur le sol. Mon couteau à la main, je me redresse, je regarde autour de moi, amusé et satisfait, et après avoir nettoyé la lame, je le remets dans le fourreau. »

Jean-François BILLETER, *Leçons sur Tchouang-Tseu.*

Le premier module de notre ontologie est une ontologie de haut niveau, ou ontologie fondationnelle, que tous les autres modules viendront spécifier. Dans ce chapitre, nous présentons ce que sont les ontologies fondationnelles et leurs principales représentantes, nous expliquons ce qui nous a poussé à proposer la Petite Ontologie Fondationnelle, et nous en exposons les principes de construction.

1 Les ontologies fondationnelles

Nous avons commencé la section 3.1.1 du chapitre 1 en marquant la différence entre une ontologie philosophique et une ontologie informatique, ce qui pouvait amener à s'interroger sur la

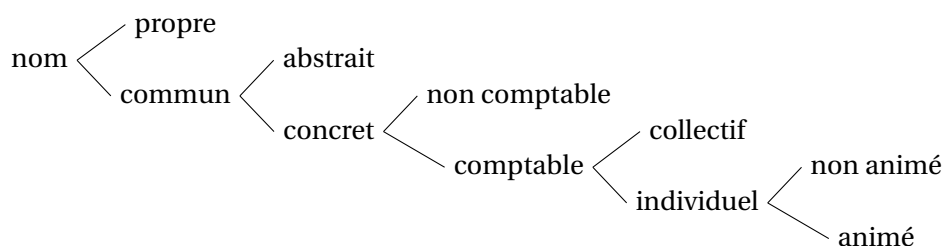
raison de cet emprunt de l’informatique au vocabulaire philosophique. Les ontologies fondationnelles (c’est-à-dire les *top-level ontologies* de la section 3.3 du chapitre 1), que nous avons décrites sommairement comme modélisant les concepts et les relations les plus généraux, permettent de comprendre cet emprunt et, cette fois, non plus les différences mais la proximité entre les ontologies informatiques et l’ontologie au sens philosophique.

1.1 Des ontologies informatiques et philosophiques

Depuis le début des années 1980, les chercheurs en Représentation des Connaissances et en Ingénierie des Connaissances se sont attachés à constituer des bases de connaissance reflétant le savoir des experts dans un domaine de connaissance. À mesure de l’augmentation du nombre et de la diversité des domaines à formaliser, et de l’augmentation concomitante du nombre de bases de connaissances, le besoin a surgi d’une montée en généralité, permettant une meilleure interopérabilité entre elles et une meilleure qualité des bases de connaissance constituées. C’est à ce besoin que répondent les ontologies fondationnelles, et cette montée en généralité rencontre le mouvement inverse de l’ontologie philosophique dès qu’elle sort d’un pur discours sur l’être, puisqu’elle est alors amenée à spécifier différentes modalités de l’être et des propriétés de l’être ou des relations existant entre les différentes modalités et propriétés de l’être (Aristote, avec ses *Catégories*, est ici le grand ancêtre : c’est en référence à sa philosophie qu’est forgé le mot « ontologie » au XVII^e siècle). C’est donc au niveau des ontologies fondationnelles que l’emprunt de l’informatique à la philosophie est le plus justifié et les problématiques les plus communes. Si certains ont marqué très tôt la différence entre l’Ontologie comme branche de la philosophie et les artefacts concrets que constituent les ontologies (Guarino & Giaretta, 1995), d’autres continuent de considérer la communauté philosophique comme une source d’expertise pertinente dans la formalisation des connaissances (Horrocks, 2013), tandis que d’autres encore voient, conformément à un processus maintes fois observé entre philosophie et sciences, l’ontologie appliquée (c’est-à-dire ici le développement d’ontologies informatiques) se constituer en discipline de plus en plus autonome vis-à-vis de l’ontologie philosophique (Barton & Rosier, 2016).

Idéalement, une ontologie fondationnelle devrait proposer des classes et des relations que toute autre ontologie viendrait spécifier. La question se pose alors de l’orientation que doit prendre la montée en généralité pour atteindre un tel but. Comme le soulignent Masolo *et al.* (2002), le travail d’élaboration des ontologies fondationnelles peut « tirer un immense profit des résultats et des méthodologies de disciplines telles que la philosophie, la linguistique et les sciences cognitives ». Ces trois disciplines sont autant de pistes crédibles (et non mutuellement exclusives) pour opérer la montée en généralité souhaitée. Nous présentons ci-dessous trois appuis possibles pour tenter de rendre une ontologie fondationnelle indépendante de tout domaine.

La considération de la structure de tout ce qui sera *dit* dans un domaine quelconque (qui formerait une ontologie implicite du langage) – On pourrait alors se tourner vers la **grammaire**, et en premier lieu vers la grammaire des noms, qui opère notamment les dichotomies suivantes :



La dichotomie entre nom commun et nom propre peut être mise en correspondance avec la séparation entre les classes et leurs instances. La dichotomie abstrait / concret invite à séparer les entités spatiotemporelles des entités non spatiotemporelles. La dichotomie non comptable (ou massif) / comptable met en jeu, pour le dire rapidement, ce qui fait ou pas une « chose ». La dichotomie collectif / individuel invite à réfléchir sur le groupe, l'agrégat. Enfin, la dichotomie animé / non animé recouvre une séparation facilement identifiable (faute d'être facilement définissable).

On pourrait en deuxième lieu se tourner vers la typologie grammaticale des adverbes :

- les adverbes de manière ou de qualité;
- les adverbes de temps;
- les adverbes de lieu;
- les adverbes d'ordre ou de rang;
- les adverbes de quantité;
- les adverbes de comparaison;
- les adverbes d'affirmation, de négation, de doute.

Du point de vue grammatical, les adverbes ajoutent une détermination à un verbe, un adjectif, un autre adverbe, voire une phrase. Mais ils permettent de mettre en évidence des catégories d'entité qui peuvent trouver leur place dans une ontologie : ainsi du temps, du lieu, de la qualité, de la quantité. La prise en compte de l'affirmation, de la négation et du doute peut en outre amener à s'interroger sur l'utilisation ou non d'une logique modale.

On pourrait en troisième lieu se tourner vers les verbes, par exemple vers la distinction entre verbes d'action et verbes d'état, au premier rang desquels l'auxiliaire être, substantivé dans le questionnement ontologique, et auquel Aristote donne plusieurs acceptions (*Métaphysique*, E, 1026a). Remarquons ici simplement ¹ l'usage du sens copulatif du mot être dans la relation de subsomption entre classes d'une ontologie informatique.

La considération des catégories cognitives les plus générales et les plus communément partagées, que pourrait mettre en évidence la **psychologie cognitive** – L'existence de telles catégories universelles est un sujet de questionnement pour Declerck *et al.* (2012), qui citent plusieurs travaux de psychologie cognitive relatifs à la mémoire sémantique en n'y trouvant pas matière à constituer une ontologie fondationnelle et concluent, en invoquant également des travaux d'anthropologie, à une position de relativisme culturel.

La prise en compte de l'héritage philosophique – La **philosophie** peut être mise à contribution de différentes manières. Elle peut permettre d'approfondir la réflexion sur une classe qu'on souhaite faire entrer dans une ontologie fondationnelle. Elle peut suggérer des candidats pour les classes les plus génériques de l'ontologie fondationnelle. Elle peut également orienter certains choix de conception d'une ontologie fondationnelle. Nous pensons ici notamment à l'ontologie formelle de Husserl : si la remontée en généralité visée par l'ontologie fondationnelle ressortit de l'ontologie matérielle, l'ontologie formelle est comme son nom l'indique formalisante plus que généralisante et propose des catégories valant de manière transversale, des « lois portant sur les structures à travers lesquelles nous pensons le monde » (Bachimont, 2000). C'est le cas par exemple des catégories de tout et de partie.

Ces trois pistes sont chacune à sa manière valides, jusqu'à un point seulement. La piste linguistique reste liée ici au contexte linguistique des langues occidentales. La piste cognitive s'affronte à la question du relativisme culturel (non sans lien peut-être avec le relativisme linguistique, si

1. Une abondante littérature philosophique traite du sens du mot « être » chez Aristote.

l'on suit l'hypothèse de Sapir-Whorf²). Enfin, Bachimont (2000) montre que l'engagement ontologique pris pour développer une ontologie référentielle permettant de formaliser des connaissances laisse nécessairement à cette dernière l'inflexion du domaine et de la tâche pour lesquels elle est conçue. Dès lors, si l'objectif d'une ontologie de haut niveau reste légitime, l'élaboration d'une ontologie fondationnelle universelle qui s'attacherait à « découper par espèces suivant les articulations naturelles, en tâchant de ne casser aucune partie, comme le ferait un mauvais boucher sacrificateur » (Platon, *Phèdre*, 265a, trad. L. Brisson), en ressentant une fois le travail accompli la satisfaction du cuisinier Ting de l'épigraphe de ce chapitre, semble un idéal non atteignable.

1.2 Les principales ontologies fondationnelles

La figure 2.1 montre la chronologie d'apparition des principales ontologies de haut niveau.

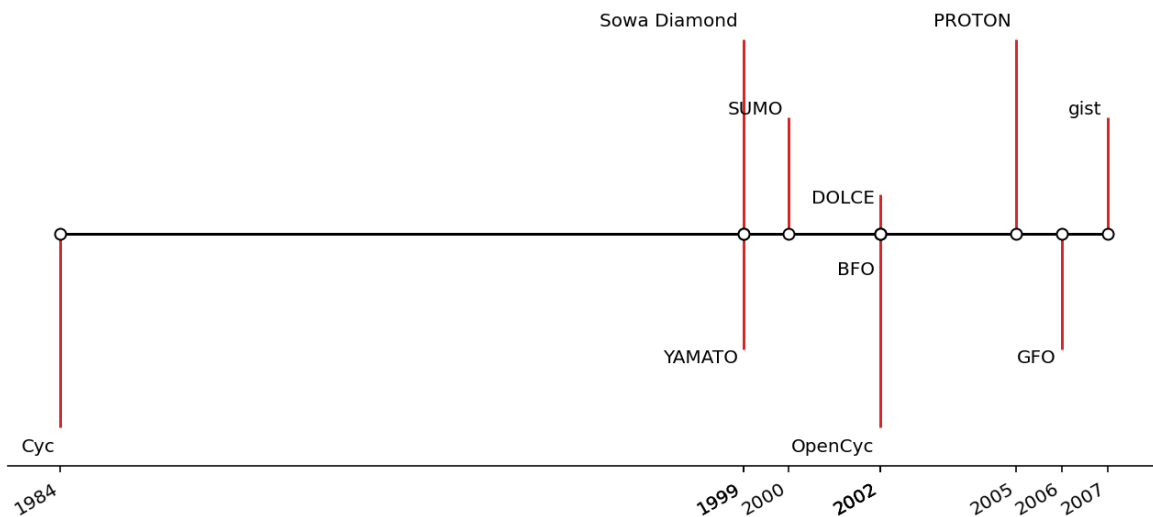


FIGURE 2.1 – Années de lancement initial des principales ontologies de haut niveau.

À l'exception de Cyc, qui fait figure de précurseur, c'est au tournant des années 2000 que sont élaborées la plupart des ontologies fondationnelles qui sont encore utilisées et discutées aujourd'hui.

Cyc vise à formaliser sous formes d'axiomes la connaissance commune, particulièrement quand elle est implicite; elle peut être vue selon son initiateur, Douglas B. Lenat, « comme un système expert dont le domaine s'étend à l'ensemble des objets et actions de la vie quotidienne » (Lenat, 1995);

La KR Ontology de Sowa (communément appelée « diamant de Sowa ») entend permettre l'intégration de toutes les bases de données et de connaissances et se veut plus un canevas de distinctions qu'une hiérarchie fixe de catégories³ – elle reconnaît une dette philosophique envers Charles Sanders Peirce et Alfred North Whitehead;

2. Cette hypothèse, du nom du linguiste et anthropologue Edward Sapir et de son étudiant Benjamin Lee Whorf, pose que notre vision du monde, nos représentations mentales, dépendent du langage et des catégories linguistiques avec lesquelles nous nous exprimons.

3. <http://www.jfsowa.com/ontology/toplevel.htm>

- YAMATO** (pour *Yet Another More Advanced Top-level Ontology*) est explicitement développée pour offrir une meilleure représentation des qualités, des « représentations » (au sens d'objets informationnels) et des processus et événements (Mizoguchi & Toyoshima, 2017), que les ontologies fondationnelles auxquelles elle fait référence : DOLCE, BFO, GFO, SUMO et CYC;
- SUMO** (pour *Suggested Upper Merged Ontology*) entend permettre le progrès continu du commerce électronique et de l'intégration des logiciels, et propose une méthode originale d'élaboration de l'ontologie à partir de la fusion de contenus ontologiques de haut niveau⁴, de l'élagage progressif des classes dispensables et de la résolution au cas par cas des incompatibilités (Niles & Pease, 2001);
- DOLCE** (pour *Descriptive Ontology for Linguistic and Cognitive Engineering*) vise à capturer les catégories ontologiques qui sous-tendent le langage naturel et la connaissance commune, à expliciter les hypothèses d'un domaine d'intérêt, et ne prétend pas être une ontologie de haut niveau « universelle » – elle s'ancre dans la philosophie mais n'entend pas défendre une position métaphysique par elle-même (Masolo *et al.*, 2002);
- BFO** (pour *Basic Formal Ontology*), élaborée par Barry Smith en vue de rendre compte (ontologiquement) de l'expérience quotidienne, suit le projet husserlien de logique formelle – les concepts de base sont donc distribués selon la théorie du tout et de la partie, de la dépendance ontologique, de la théorie de la frontière, de la continuité et du contact (Smith, 1998; Arp *et al.*, 2015);
- OpenCyc** est la version ouverte de Cyc, et son lancement visait à permettre aux chercheurs en intelligence artificielle de mesurer ce qui manquait, dans leurs ontologies, de la connaissance commune formalisée dans Cyc;
- PROTON** est conçue pour le requêtage et la gestion de vastes bases de connaissances, évite pour cette raison d'aller au-delà d'une hiérarchie de concepts et d'une définition de domaines et de codomaines pour ses propriétés, et se refuse notamment à une conceptualisation trop précise de l'espace et du temps (Kiryakov, 2006);
- GFO** (pour *General Formal Ontology*) est une composante d'un système intégré d'ontologies fondationnelles dont le champ d'application est la médecine, la biomédecine et la biologie (Herre *et al.*, 2006) – elle fait montre d'un fort ancrage philosophique (dans les philosophies d'Ingarden, de Chisholm, de Brentano par exemple);
- gist** vise à offrir la plus grande couverture des concepts d'une entreprise; elle se veut minimaliste (138 classes, 122 propriétés) et éloignée des abstractions philosophiques⁵.

Ces ontologies ont fait l'objet de comparaisons dans la littérature. Nous présentons quelques-unes de ces comparaisons, d'une portée plus ou moins descriptive et plus ou moins critique, selon différents critères.

Mascardi *et al.* (2007) comparent :

Cyc Sowa YAMATO SUMO DOLCE BFO OpenCyc PROTON GFO gist
selon les critères descriptifs suivants :

- page d'accueil,
- développeurs,

4. C'est-à-dire, à l'époque, les ontologies disponibles sur le serveur Ontolingua et à l'ITBM-CNR, l'ontologie de haut niveau de John Sowa, celle de Russell and Norvig, l'algèbre de James Allen, la théorie formelle des trous de Casati et Varzi, l'ontologie des délimitations (*boundaries*) de Barry Smith, la méréotopologie formelle de Nicola Guarino, et différentes représentations formelles des plans et des processus.

5. <https://www.semanticarts.com/gist/>

- dimensions de l'ontologie,
- langage(s) des représentation utilisés,
- modularité,
- applications,
- alignement avec Wordnet,
- licence.

Khan & Keet (2015) proposent avec ROMULUS (pour *Repository of Ontologies for MULTiple USes*) une comparaison très approfondie de :

Cyc Sowa YAMATO SUMO DOLCE BFO OpenCyc PROTON GFO gist
sur :

- les engagements ontologiques :
 - positionnement descriptif (cherchant à représenter les entités du langage naturel et du sens commun) ou réaliste (cherchant à représenter le monde tel qu'il est),
 - traitant d'universaux ou de particuliers⁶,
 - positionnement multiplicatif ou réductionniste (sur le nombre d'objets qui peuvent être localisés au même endroit au même moment),
 - positionnement endurantiste ou perdurantiste (cf. section 2.1.1),
 - positionnement actualiste ou possibiliste (sur l'existence actuelle de ce qui se trouve dans l'ontologie),
 - positionnement présentiste ou éternaliste (sur l'existence du seul présent, ou aussi du passé et du futur),
 - présence d'entités concrètes seulement ou aussi d'entités abstraites,
 - théorie méréologique utilisée,
 - aspects temporels fournis,
 - niveaux de granularité pris en compte,
 - possibilité de modéliser des propriétés et des valeurs,
 - modélisation de l'espace et du temps,
 - nombre de couches dans l'architecture de l'ontologie,
 - modélisation de situation et de « situoides »⁷ ;
- les langages de représentation utilisés ;
- les propriétés des ontologies (dimensions, modularité, licence, maintenance active) ;
- les domaines utilisant les ontologies ;
- les scénarios d'application (système d'information, web sémantique, recherche scientifique, traitement du langage naturel, etc.).

Jansen (2008b) porte un regard critique sur :

Cyc Sowa YAMATO SUMO DOLCE BFO OpenCyc PROTON GFO gist
en partant des *Catégories* d'Aristote et en s'attachant particulièrement à trois dichotomies :

- entités dépendantes / entités indépendantes,
- continuants / occurrents,

6. Point de désaccord d'Aristote avec Platon, célèbre querelle médiévale, débat toujours vif au sein de la philosophie analytique, nous n'entrerons pas plus avant dans cette question des universaux sinon pour rappeler que les « particuliers » désignent ici des choses singulières, que certaines de ces choses singulières ont beaucoup en commun, et que la question est de savoir si ce commun, cet « universel » (l'« équité » des chevaux par exemple), existe dans nos esprits seulement ou a une existence indépendante, voire une existence transcendante.

7. Les auteurs semblent se référer ici à la présence dans GFO de *topoids*, régions spatiales compactes et connexes, et de *chronoids*, intervalles temporels continus dont dépendent les instants et les régions temporelles, mais le terme de *situoid* ne figure pas dans GFO.

- universaux / particuliers.

Les critères sur lesquels l’auteur compare ces ontologies ont été explicités précédemment (Janzen, 2008a) comme des « règles pour des taxonomies de bonne qualité et utiles » et concernent :

- le fondement ontologique,
- la structure (en types et sous-types),
- la disjonction (entre classes ayant le même ascendant direct),
- l’exhaustivité (couverture complète du domaine considéré),
- l’absence d’ambiguïté,
- l’uniformité (dans le type de traits permettant de classer les objets),
- le caractère explicite et la précision,
- l’absence de méta-types (issus du processus de classification lui-même).

Des différentes critiques portées à l’égard du « diamant de Sowa », nous retiendrons ici qu’il ne contient pas et ne permet pas d’intégrer les lieux et les temps. Quant à OpenCyc, l’auteur montre qu’elle contrevient à nombre de règles de bonne classification qu’il promet.

Aux ontologies qui ont nourri notre réflexion, nous devons ajouter l’ontologie OntoMénélas, qui couvre tous les niveaux de généralité et dont le premier domaine d’application a été la chirurgie cardiaque mais qui a également été utilisée comme ontologie de haut niveau et ontologie noyau (Charlet *et al.*, 2012) pour plusieurs ontologies de domaines (ONTOLURGENCES⁸, ONTOPNEUMO⁹, ONTOPSYCHIA¹⁰, ONTOPARON¹¹, OntoDPN (Dhombres *et al.*, 2010)). Sa partie haute¹² a toutes les caractéristiques d’une ontologie de haut niveau.

1.3 Excursus épistémologique

L’article de Declerck *et al.* (2012) dégage cinq fonctions généralement attribuées aux ontologies fondationnelles :

- améliorer l’interopérabilité sémantique¹³,
- fédérer différentes ontologies de domaine,
- faciliter le processus de conception des ontologies de domaine,
- améliorer la qualité, la réutilisabilité, la maintenance et l’évolutivité de ces mêmes ontologies,
- expliciter les choix de conceptualisation implicites.

La considération de ces fonctions nous inspire un parallèle avec ce que Kuhn appelle d’abord, dans *La structure des révolutions scientifiques* (2018), un *paradigme*, puis dans sa postface de 1969 une *matrice disciplinaire*.

Une matrice disciplinaire se décline selon Kuhn :

- en généralisations symboliques (par exemple la mise sous forme mathématique d’une loi physique),
- en un corpus de croyances métaphysiques sur les entités fondamentales dont est composé l’univers et les relations qu’elles entretiennent entre elles,
- en un ensemble de jugements de valeurs sur la science (ce qui est scientifiquement valide, exact; la place sociale de la science)

8. <https://bioportal.bioontology.org/ontologies/ONTOLURGENCES>

9. <https://bioportal.bioontology.org/ontologies/ONTOPNEUMO>

10. <https://bioportal.bioontology.org/ontologies/ONTOPSYCHIA>

11. <https://bioportal.bioontology.org/ontologies/ONTOPARON>

12. <https://bioportal.bioontology.org/ontologies/TOP-MENELAS>

13. Les auteurs notent que le cas le plus crédible pour accomplir cette fonction est celui de l’usage de l’ontologie fondationnelle pour guider la création d’autres ontologies.

- en exemples typiques des problèmes qui peuvent se poser (par exemple aux étudiants lors de leur formation) et des solutions techniques qui peuvent y être apportées (ce que Kuhn appelle dans sa postface « paradigme », en un sens plus précis qu’initialement).

L’expérience menée par M. Keet sur l’utilisation d’ontologies fondationnelles dans le développement d’ontologies (Keet, 2011) illustre ce dernier point : l’usage d’ontologies fondationnelles permet des ontologies de meilleure qualité, plus interopérables, et génère moins de confusion entre *is_a* et *part_of*. On peut également citer ici l’expérience menée par Verdonck *et al.* (2019), comparant la modélisation conceptuelle traditionnelle à la modélisation conceptuelle guidée par des ontologies (avec ontoUML) ; les auteurs attribuent en partie la supériorité de cette dernière au fait que les primitives du langage ontoUML sont des patrons de conception ontologique (*Ontology Design Patterns* – cf. section 2.3 du chapitre 5). Ces derniers peuvent être comparés aux généralisations symboliques dont parle Kuhn.

Plus généralement, en définissant ses primitives, en choisissant d’intégrer dans sa taxonomie des distinctions qui pourraient être modélisées avec des relations (par exemple la dépendance ontologique), en décidant de la logique de description adoptée, une ontologie fondationnelle peut légitimement, nous semble-t-il, être comparée à une matrice disciplinaire.

Quelle est alors l’ontologie fondationnelle qui pourrait, dans le paysage actuel de l’ingénierie ontologique, revendiquer cette position « paradigmatique » ?

Si nous nous limitons au champ de la recherche biologique et biomédicale, la place de BFO comme ontologie de haut niveau dans l’*Open Biological and Biomedical Ontology Foundry* (OBO Foundry), qui rassemble plusieurs centaines d’ontologies, lui confère cette position. Slater *et al.* (2020), en mettant en évidence le nombre important de classes insatisfiables lorsqu’on combinait entre elles les ontologies OBO, ont montré que la seule utilisation d’une même ontologie de haut niveau ne permettait pas complètement d’atteindre l’interopérabilité sémantique. Des efforts ont été entrepris pour y remédier en mettant en place une suite logicielle de validation semi-automatique des métadonnées des ontologies de l’OBO Foundry et un tableau de bord pour les résultats obtenus (Jackson *et al.*, 2021).

Si nous étendons maintenant le champ à tous les systèmes d’information, l’adoption récente (2021) de BFO comme norme ISO (ISO/IEC 21838-2 :2021¹⁴) rend désormais difficile de démêler, dans le choix de BFO comme ontologie de haut niveau par des concepteurs d’ontologies, l’adhésion à l’ontologie sur des critères scientifiques de la conformité à une norme établie. De manière plus préoccupante, elle entrave – si elle ne les rend pas impossibles – le travail d’ajustement et les modifications qui, sans remettre en cause le dessin général, apportent des améliorations – ce que Kuhn appelle dans le domaine qu’il étudie la « science normale ».

D’autre part, sur ce périmètre très large, d’autres ontologies peuvent prétendre à la position de standard *de facto* : les ontologies du web sémantique que sont schema.org, l’ontologie de Wikidata ou encore l’ontologie de DBpedia. Haller & Polleres (2020) les comparent en se demandant « Est-il préférable de n’avoir qu’une seule ontologie sur le Web ? », et concluent sur le bénéfice qu’il y aurait à une unique ontologie en termes d’interopérabilité, en considérant comme meilleure candidate l’ontologie de Wikidata, malgré les défauts qu’elle présente encore (modularisation, accès, versionnage, qualité variable selon les domaines) et les succès de schema.org dans le domaine du commerce en ligne.

14. <https://www.iso.org/fr/standard/74572.html> – BFO a connu une deuxième version en 2015, BFO2, et la version adoptée comme norme est BFO-2020, qui intègre des *Object Properties* sensiblement différentes de celles qui étaient auparavant séparées dans l’ontologie des relations de BFO, RO Core.

2 Motivation de la Petite Ontologie Fondationnelle

La place de BFO dans le domaine biomédical explique que celle-ci ait été adoptée par le partenaire industriel responsable de la mise en place de la plateforme de collection des données du projet PsyCARE, Fealinx, au moment de proposer une modélisation ontologique de leur modèle de données (Raboudi-Souilem, 2021). Toutefois, il nous est apparu en travaillant sur la question de la représentation de la temporalité, que la division entre continuant et occurrent, fondamentale dans BFO (et parallèle à celle entre endurant et perdurant dans DOLCE), méritait d’être réexaminée. Nous avons donc été amenés à proposer une ontologie fondationnelle dont nous souhaitons qu’elle ne s’écarte pas de façon radicale de BFO – afin de conserver la possibilité d’un alignement – mais qu’elle articule de façon plus satisfaisante les continuants et les occurrents et soit, si possible, d’une appropriation plus aisée par des utilisateurs éloignés des questionnements philosophiques propres à l’élaboration d’une ontologie de haut niveau.

2.1 La temporalité dans les ontologies fondationnelles

La question de la représentation ontologique de la temporalité recouvre plusieurs choses : la représentation du changement, la représentation des entités temporelles (comme les instants ou les intervalles temporels), le raisonnement temporel que permet cette représentation ontologique. Ici, nous nous intéresserons principalement à ce que peut apporter l’engagement ontologique pris dès l’ontologie fondationnelle pour aider à la représentation du dynamisme des entités qui évoluent dans le temps. Le problème fondamental, de ce point vue, est de faire coexister tridimensionnalité et quadridimensionnalité dans une même ontologie.

2.1.1 Tridimensionnalité et quadridimensionnalité

Représenter dans une ontologie à la fois une chaise, un individu, le séjour de cet individu dans une ville étrangère, le match de football auquel il a assisté sur place, semble souhaitable. Toutefois, un débat encore vif au sein de l’ontologie (au sens philosophique) concerne la question de savoir si les objets¹⁵ qui nous entourent ont ou non des parties temporelles (Hawley, 2023). Les deux pôles principaux de ce débat sont l’endurantisme et le perdurantisme¹⁶, qui s’accordent tous deux sur le fait que les événements ont des parties temporelles, mais s’opposent sur la manière d’envisager les objets.

Pour l’*endurantisme*, les objets n’ont pas de parties temporelles et sont « tout entier présents » (Crisp & Smith, 2005) à chaque instant, par opposition aux événements qui se déploient dans le temps. Les objets sont tridimensionnels (ou 3D) et les événements sont quadridimensionnels (ou 4D). Pour le *perdurantisme*, les objets ont eux aussi des parties temporelles. Il n’y a donc que des entités quadridimensionnelles. La position endurantiste peut s’appuyer sur le langage naturel, le « bon sens ». A l’inverse, la position perdurantiste qui considère les objets sous l’angle des « vers spatiotemporels » que constituent le déploiement de leur existence, produirait des formulations aussi contre-intuitives que « une partie temporelle de Joe est entré dans une partie temporelle de la pièce » (Burek *et al.*, 2019).

15. Par objets, il faut entendre à la fois la chaise et l’individu, tout ce qui continue à travers le temps en conservant son identité. Nous emploierons par la suite le terme de « continuant temporel », qui sera expliqué plus avant dans la section suivante.

16. Il existe des positions mixtes que nous ne détaillerons pas ici.

La question qui peut se poser pour les ontologies fondationnelles est de savoir dans quelle mesure la co-présence, dans une même ontologie, d'entités de ces deux types risque de poser des problèmes de cohérence dans la modélisation.

Pour McCall & Lowe (2006), la controverse 3D/4D est une tempête dans un verre d'eau, les descriptions 3D et 4D du monde étant équivalentes et traduisibles sans reste. C'est peu ou prou la position de Grenon & Smith (2004), comme le rappelle Jaskolla (2011) qui montre pour sa part que la traduction du quadridimensionnalisme en tridimensionnalisme pose des problèmes non résolus. Pour aller dans ce sens, si l'on considère comme McCall et Lowe un enduring, par exemple un être vivant, comme un ensemble de particules 3D, et l'équivalent 4D qu'est l'événement de sa vie, la mort de cet être diminue d'une unité le nombre d'entités dans une description 4D du monde, alors que le nombre d'entités (de particules 3D) dans une description 3D n'est pas modifié. Cette raison de privilégier une description quadridimensionnelle du monde, qui concerne l'apparition ou la disparition d'un continuant temporel, ne doit pas nous faire ignorer les aspects d'invariance que présentent certains événements, qui donnent leur force à la description tridimensionnelle.

Un argument parfois invoqué pour justifier la position d'entités tridimensionnelles est le changement : pour qu'il y ait changement, il faut qu'il y ait quelque chose qui change. Un tel argument relève de la pétition de principe, puisque c'est parce que nous avons d'abord individué un continuant temporel que nous pouvons considérer des changements le touchant. Il dénote toutefois l'importance ou la « naturalité » des descriptions tridimensionnelles.

Si nous reprenons la distinction entre objets et événements que propose une modélisation recourant à des entités aussi bien tridimensionnelles que quadridimensionnelles : on a des événements auxquels participent des objets. Mais ces objets sont eux-mêmes les lieux d'occurrence d'un certain nombre d'événements auxquels participent des objets de granularité plus fine. Ce feuilletage est repéré par Galton & Mizoguchi (2009), qui voient dans l'objet une interface entre des processus¹⁷ internes et externes, abordent l'idée d'une hiérarchie descendante de granularités éventuellement infinie, et relèvent que la dépendance ontologique des événements aux objets qui y participent doit être complétée par la dépendance ontologique des objets aux événements qui leur permettent de persister. Cette dernière considération nous semble en faveur des descriptions quadridimensionnelles.

Des points qui précèdent, il ressort que le cadre quadridimensionnel est le plus à même de permettre une représentation exhaustive du monde. Si nous conservons une branche de l'ontologie consacrée aux continnants temporels, en se conformant ainsi à la position enduringiste de BFO (Grenon & Smith, 2004) ou de DOLCE (Gangemi *et al.*, 2002), ceux-ci ne sont pas des entités concrètes. Ce sont les « identités » – stables au cours du temps – de compositions très organisées de processus en interaction les uns avec les autres, susceptibles de participer comme un tout à des processus de granularité plus élevée. A ces continnants temporels doivent être associés les événements auxquels ils participent et les événements de granularité inférieure qui les constituent.

2.1.2 Conséquences sur la représentation du changement

En raison de l'intérêt que nous accordons à ce que nous individurons comme des continnants temporels, et en premier lieu les individus humains, il est nécessaire de tirer les conséquences du cadre que nous avons posé en matière de modélisation du changement, dans l'esprit des cas proposés par Borgo *et al.* (2022).

17. Galton et Mizoguchi distinguent les événements des processus selon la dissectivité; DOLCE selon la cumulativité; BFO n'envisage que des processus. Dans POE, nous ne considérons que des événements spatiotemporels, un processus défini avant et indépendamment de sa réalisation dans un événement concret étant considéré comme un projet.

Pour préciser la notion de changement, nous pouvons partir de la théorie des mouvements d'Aristote (2002). Précisons que les « mouvements » que distingue Aristote sont la génération et la destruction, l'augmentation et la diminution, l'altération, le changement selon le lieu. La langue moderne réserve le mot « mouvement » à ce dernier et emploie « changement » comme terme générique.

Le déplacement physique est un événement qui n'implique pas la notion d'une modification de l'entité qui connaît ce mouvement, et dans lequel cette entité est tout entière engagée. C'est le changement typiquement invoqué à l'appui des descriptions tridimensionnelles (Galton & Mizoguchi, 2009; McCall & Lowe, 2006). En ce sens, il est légitime de faire du continuant temporel un, voire le seul, participant à ce changement.

La génération et la destruction sont ignorées par une modélisation tridimensionnelle qui s'attache à l'identité à travers le temps de l'entité pendant son existence. Considérer l'événement de cette existence permet de modéliser la génération et la destruction comme les moments initiaux et finaux de celle-ci.

Les changements quantitatifs – on pense ici par exemple aux changements de volume, de poids, de température – font l'objet de mesures. Des propriétés d'un objet physique sont mesurées, et les valeurs de ces mesures sont susceptibles d'évoluer dans le temps. Être individué comme un objet physique est une propriété qui n'est perdue qu'avec la destruction de l'entité. De même, les propriétés descriptives de volume, taille, température attachées à un objet physique sont des propriétés qui ne sont pas susceptibles d'être perdues par une entité matérielle : elles sont « rigides » au sens de Guarino & Welty (2000). Elles peuvent donc être attribuées au continuant temporel.

Les changements qualitatifs nécessitent une prise en compte plus fine des propriétés impliquées. S'il s'agit d'une propriété rigide, on est ramené au cas précédent (la couleur d'un objet physique, par exemple). En revanche, si nous nous intéressons à des propriétés non-rigides, par exemple au fait d'être dans une *phase* (être un adolescent), d'occuper un *rôle* (être un médecin), ou d'être dans un *état* (être malade), propriétés qui ne sont pas essentielles à toutes leurs instances, elles ne touchent pas à l'identité à travers le temps de continnants temporels mais caractérisent bien plutôt certains moments de l'événement de leur existence.

Une phase peut être naïvement définie comme une tranche temporelle d'un parcours de vie, bien que la détermination de son début et de sa fin puisse être un casse-tête (Sawyer *et al.*, 2018). Cette détermination est liée à la manière dont on la caractérise, par exemple dans le cas des *stades* d'une maladie par des événements marquant l'entrée dans cette phase ou la sortie de cette phase. Quoiqu'il en soit, ce n'est pas le continuant temporel mais bien les événements qui lui sont associés qui sont ici en jeu.

Dans la mesure où un continuant temporel est tout entier engagé dans un rôle, on pourrait être tenté d'attribuer le rôle au continuant temporel qui le tient. Toutefois, en tant que propriété non-rigide, le rôle ne ressortit pas de ce qui assure la stabilité à travers le temps de ce continuant temporel. Ce dernier point nous engage à caractériser par le rôle les événements de l'existence du continuant temporel où ce rôle est présent. Il ne peut s'agir que d'événements auxquels le continuant temporel participe. Nous proposons donc de faire porter le rôle à une spécification de la relation de participation du continuant temporel à un événement.

Enfin, un état tel qu'« être malade » appelle à nouveau une autre modélisation. Être malade, c'est subir une évolution, la maladie, qui modifie tout ou partie des événements que nous

abritons ; ce qui n'exclut pas qu'il puisse y avoir des conséquences sur les événements auxquels participe la personne malade. La maladie en tant qu'évolution est donc un événement situé spatiotemporellement, et la situation spatiotemporelle de l'événement a pour extension spatiale tout ou partie de ce qui définit en propre un continuant temporel, c'est-à-dire une certaine portion d'espace. Un état associé à un continuant temporel s'inscrit donc dans l'ensemble des événements qui constituent ce continuant temporel.

Pour représenter le « changement », la modélisation que nous proposons associe donc à un continuant temporel les propriétés rigides qui le caractérisent, l'événement de sa vie entière, ainsi que deux ensembles d'événements de granularité différente : des événements « externes » auxquels le continuant temporel participe, en spécifiant éventuellement cette relation de participation par un rôle, et des événements « internes » dont la localisation spatiotemporelle a pour extension spatiale le continuant temporel.

2.2 L'ergonomie cognitive d'une ontologie fondationnelle

Nous avons mentionné dans la section 1.3 l'intérêt des ontologies fondationnelles pour élaborer des ontologies de domaine de meilleure qualité. Toutefois, en tant qu'artefacts symboliques concrets, les ontologies sont des outils intellectuels spécifiques, que des experts (dans d'autres domaines que la construction d'ontologies) doivent pouvoir s'approprier pour y intégrer leurs connaissances. Comme le remarque Keet (2011), ceci a un « coût » lié au temps passé à se familiariser avec une ontologie fondationnelle ». Même si elle considère que ce coût « par rapport à un départ de zéro a été plus que compensé par une meilleure qualité et une meilleure interopérabilité, et ce malgré la durée limitée de l'expérience », on peut souhaiter le diminuer en facilitant l'appropriation d'une ontologie fondationnelle – et ainsi étendre son utilisation (les participants de l'expérience de M. Keet étaient volontaires) – ou pour le dire autrement en améliorant son ergonomie cognitive.

L'ergonomie cognitive est citée par Gangemi *et al.* (2005) comme faisant partie des *Quality-Oriented Ontology Descriptions*, avec précisément pour objectif que l'ontologie soit « facilement comprise, manipulée et exploitée par les utilisateurs finaux ». Les auteurs associent les qualités suivantes à cet objectif :

- moins de profondeur,
- moins de largeur,
- moins d'« enchevêtrement » (*tangledness*)¹⁸,
- un ratio classes/propriétés élevé¹⁹,
- plus d'annotations (annotations lexicales et commentaires sur les éléments de l'ontologie, sujet de l'ontologie),
- plus d'interfaçage (annotations permettant la mise en correspondance de l'ontologie avec une interface utilisateur),
- plus de motifs (*patterns*)²⁰.

18. C'est la situation où des classes ont plusieurs surclasses, où l'ontologie est organisée en treillis plutôt qu'en arborescence, ce qui produit – dans une représentation graphique de la taxonomie – un enchevêtrement des liens entre classes. La formule proposée dans l'article fait le ratio nombre de nœuds du graphes/nombre de nœuds ayant au moins deux parents. Le ratio inverse ne poserait pas de problème de division par 0 (lorsqu'aucun nœud n'a plus d'un parent) et diminuerait quand l'enchevêtrement diminue (quand le ratio proposé augmente dans ce cas).

19. L'exemple du passage d'un ratio 1:6 à un ratio 1:1, suite aux modifications apportées à une ontologie, est présenté comme souhaitable.

20. Il s'agit ici du repérage de régions denses dans l'ontologie, c'est-à-dire fortement reliées par des relations ne tenant pas à la taxonomie.

Gavrilova *et al.* (2010) s'intéressent en premier lieu à l'évaluation de la conception visuelle et de l'ergonomie cognitive des ontologies dans un contexte d'enseignement. Les auteurs s'inspirent de la psychologie de la forme (*Gestaltpsychologie*) et de ses lois²¹, ainsi que du « rasoir d'Ockham » (loi de parcimonie). Ces lois sont reformulées dans le contexte de l'ingénierie ontologique, en vue de favoriser des ontologies « harmonieuses », c'est-à-dire caractérisées par l'« équilibre conceptuel » et la « clarté ». Les auteurs en tirent un processus d'évaluation en plusieurs étapes auxquelles sont associées des métriques²² :

- évaluation de la taille de l'ontologie :
 - nombre de nœuds dans le graphe de l'ontologie,
 - chemin le plus long de la racine à une feuille,
 - nombre de feuilles dans l'ontologie,
 - nombre de nœuds qui ont des feuilles parmi leurs enfants,
 - nombre d'arcs dans le graphe de l'ontologie ;
- erreurs critiques dans l'ontologie :
 - liées aux cycles :
 - **nombre de cycles** dans l'ontologie,
 - **nombre de nœuds appartenant à l'un des cycles divisé par le nombre de tous les nœuds du graphe,**
 - liées à l'enchevêtrement :
 - nombre de nœuds avec plusieurs parents,
 - ratio nombre de nœuds avec plusieurs parents/nombre total de nœuds,
 - nombre moyen de nœuds parents d'un nœud,
 - liées aux nœuds ayant des feuilles et des non-feuilles dans leurs enfants :
 - nombre de nœuds qui ont des feuilles dans leurs enfants,
 - nombre de nœuds qui n'ont que des feuilles dans leurs enfants,
 - **ratio nombre de nœuds qui ont des feuilles et des non-feuilles dans leurs enfants/nombre de nœuds qui ont des feuilles dans leurs enfants** (ratio parents de feuilles et non-feuilles/parents de feuilles) ;
- métriques de Yngve-Miller :
 - ratio nombre de nœuds de degré normal/nombre total de nœuds²³,
 - degré moyen d'un nœud,
 - degré médian d'un nœud,
 - degré de 90% des nœuds,
 - **écart-type du degré d'un nœud,**
 - **ratio écart-type du degré d'un nœud/degré moyen d'un nœud** (coefficient de variation

21. Fondée dans les années 1920 par Max Wertheimer, Wolfgang Köhler et Kurt Koffka, la psychologie de la forme est une théorie psychologique, d'abord des processus de perception et de représentation mentale, selon laquelle la « forme » (*Gestalt*) est perçue de façon globale et non par assemblage d'éléments. Les lois évoquées ici sont les lois de la bonne forme (la tendance à percevoir préférentiellement une bonne forme, c'est-à-dire aussi régulière, ordonnée, simple et symétrique que possible), de clôture (la tendance à compléter une forme), de similitude (la tendance à regrouper les éléments similaires pour percevoir une forme), d'inclusion (la tendance à percevoir la forme extérieure plutôt que la forme plus petite qui y est incluse).

22. Précisons que la « racine » est le nœud le plus élevé dans l'arborescence ; qu'une « feuille » est un nœud qui n'a pas d'enfant et donc une classe qui n'a pas de sous-classe (symétriquement, le parent d'un nœud est la surclasse d'une classe) ; qu'il y a un cycle – dans un graphe orienté – quand deux chemins différents permettent de rejoindre deux nœuds.

23. Le degré est le nombre de liens qui relie un nœud à d'autres nœuds. Un degré est normal s'il est proche du nombre de Yngve-Miller (7 ± 2).

- du degré);
- variété de types de relation :
 - nombre de types de relation différents,
 - ratio nombre de types de relation différents/nombre total de nœuds,
 - **ratio nombre de nœuds avec des liens sortants différents/nombre total de nœuds,**
 - **ratio nombre de nœuds avec des liens entrants différents/nombre total de nœuds,**
 - nombre moyen de liens sortants d'un nœud,
 - nombre moyen de liens entrants d'un nœud;
- profondeur de l'ontologie :
 - profondeur minimale,
 - profondeur maximale,
 - profondeur moyenne,
 - profondeur médiane,
 - profondeur de 90% des nœuds,
 - écart-type de la profondeur,
 - **écart-type de la profondeur/profondeur moyenne** (coefficient de variation de la profondeur);
- largeur de l'ontologie :
 - largeur moyenne,
 - ratio moyen de largeurs de niveaux adjacents,
 - **ratio maximal de largeurs de niveaux adjacents**²⁴,
 - ratio écart-type des relations de largeur des niveaux adjacents/relation moyenne de largeur des niveaux adjacents;
- dispersion de l'ontologie
 - nombre moyen d'enfants-feuille d'un nœud,
 - nombre maximal d'enfants-feuille d'un nœud,
 - nombre minimal d'enfants-feuille d'un nœud,
 - **écart-type du nombre d'enfants-feuille d'un nœud** (pour les nœuds ayant des enfants-feuille),

Les métriques indiquées en gras sont celles qui sont sélectionnées par les auteurs comme les plus importantes. On voit que ces métriques couvrent les aspects les plus structurels des critères de Gangemi *et al.* (2005).

Nous conservons la plus grande partie des métriques sélectionnées par Gavrilova *et al.* (2010), en omettant les métriques liées aux types de relation, peu adaptées à des ontologies fondationnelles. Nous sélectionnons en outre le **nombre de nœuds** et le **nombre de feuilles** pour des raisons descriptives, ainsi que **la profondeur minimale, la profondeur maximale, la profondeur moyenne**, qui nous semblent particulièrement intéressantes à considérer pour des ontologies fondationnelles.

La figure 2.1 permet de comparer les métriques retenues sur trois ontologies fondationnelles : BFO-2020, DOLCE-Lite et ontoPOF.

Les trois ontologies sont de tailles identiques ou très proches, avec un pourcentage plus important de feuilles pour ontoPOF. Aucune des trois ontologies ne comporte de cycle. ontoPOF obtient un meilleur ratio parents de feuilles et non-feuilles/parents de feuilles²⁵. Aucune des ontologies ne

24. La ratio nombre de nœuds à un niveau de profondeur/nombre de nœuds au niveau de profondeur précédent est considéré à chaque niveau de profondeur et le ratio maximal est retenu.

25. Nous verrons dans la section suivante que c'est la prise en compte de la temporalité qui fait qu'ontoPOF n'a pas un ratio nul ici.

	BFO-2020	DOLCE-Lite	ontoPOF
TAILLE			
Nombre de nœuds	36	37	36
Nombre de feuilles	20	20	27
ERREURS CRITIQUES			
Nombre de cycles	0	0	0
Ratio nœuds dans des cycles/nombre total de nœuds	0	0	0
Ratio parents de feuilles et non-feuilles/parents de feuilles	0,36	0,38	0,25
DEGRÉ			
Ratio de nœuds ayant un degré normal	1	1	1
Écart-type du degré d'un nœud	1,27	1,30	1,76
Coefficient de variation du degré	0,66	0,68	0,96
PROFONDEUR			
Profondeur minimale	2	2	2
Profondeur maximale	5	5	3
Profondeur moyenne	4,05	3,9	2,19
Coefficient de variation de la profondeur	0,26	0,22	0,18
LARGEUR			
Ratio maximal de largeurs de niveaux adjacents	3,5	3	5
RÉPARTITION			
Moyenne du nombre d'enfants-feuille d'un nœud	1,55	1,31	3,38
Écart-type du nombre d'enfants-feuille d'un nœud	1,04	0,48	1,60

TABLE 2.1 – Métriques d'ergonomie cognitive pour BFO-2020, DOLCE-Lite et ontoPOF.

comporte de nœud ayant un nombre de liens trop élevé. ontoPOF présente un écart-type et un coefficient de variation du degré d'un nœud plus élevés : c'est ici qu'est marquée la difficulté d'aperception propre à ontoPOF. Les mesures de profondeur que nous avons souhaité ajouter montrent la faible profondeur d'ontoPOF, ce qui est à corréluer au pourcentage de feuilles et est à nos yeux une qualité (les feuilles d'une ontologie fondationnelle étant généralement plus « lisibles » puisqu'elles seront spécifiées par les ontologies de niveau inférieur). Cette faible profondeur d'ontoPOF par rapport à BFO et DOLCE est compensée par une plus grande largeur. Les mesures de répartition mettent en évidence le niveau de détail (éventuellement insuffisant) au dernier niveau de profondeur. L'écart-type important pour ontoPOF est à mettre en regard de la moyenne plus de deux fois plus élevée par rapport à BFO et DOLCE. Quant au faible écart-type de cette dernière, il s'explique en partie par le ratio de parents de feuilles et de non-feuilles élevé.

Globalement, BFO-2020 et DOLCE-Lite obtiennent des résultats très similaires. ontoPOF présente un profil différent, avantageux en termes d'ergonomie cognitive à part peut-être sur les mesures relatives au degré. Les principes de constitution d'ontoPOF, exposés ci-après, permettront d'en juger et plus généralement de comprendre ce profil.

3 Principes de constitution de la Petite Ontologie Fondationnelle

Un mode d'emploi détaillé de la Petite Ontologie Fondationnelle est proposé dans l'annexe B, accompagné de figures présentant ses 36 classes (figure B.1) et ses 72 propriétés d'objet (figure B.2). Nous justifions ici les principes de sa constitution.

3.1 Principe d'individuation au regard de l'espace et du temps

La Petite Ontologie Fondationnelle est une ontologie d'individus, au sens où ce sont des individus qui viendront instancier ses classes²⁶. Le premier principe concerne donc ce qui nous fait discerner des individus, c'est-à-dire des entités formant une unité distincte.

Dans l'article *Four ontologies*, Zemach (1970) décrit ce qui lui apparaît comme les quatre ontologies (au sens philosophique) auxquelles recourt le langage naturel quand il traite d'entités spatio-temporelles. L'espace et le temps sont les deux dimensions dans lesquelles les entités spatio-temporelles peuvent être étendues, selon deux modalités : de façon limitée ou de façon continue. Une entité est *continue* dans une dimension si elle n'a pas de parties dans cette dimension; si elle change dans cette dimension, c'est l'entité entière qui est changée. Une entité est *limitée* dans une dimension si les différentes localisations où elle se trouve dans cette dimension contiennent ses parties.

On obtient donc quatre sortes d'entités, à la base des quatre ontologies éponymes de l'article :

- des entités limitées dans le temps et dans l'espace – que Zemach appelle des *non-continuant*s (NCs) ou des *événements*;
- des entités limitées dans l'espace mais continues dans le temps – les *continuant*s dans le temps (CTs) ou plus simplement les *choses*;
- des entités continues dans l'espace mais limitées dans le temps – les *continuant*s dans l'espace (CSs), que Zemach choisit d'appeler des *processus*;
- des entités continues dans le temps et dans l'espace – les *purs continuant*s (PCs), dans lesquels Zemach voit les *types*.

Pour Zemach, chacune de ces quatre ontologies est auto-suffisante, il est possible de les traduire l'une dans l'autre, et il serait également possible de se limiter à l'emploi d'une seule d'entre elles, mais les quatre sont employées dans la langue naturelle.

Dans la mesure où nous nous inspirons de Zemach sans le suivre totalement, nous allons préciser en quoi nous nous écartons de sa pensée.

On reconnaîtra dans les « choses » et les « événements », c'est-à-dire les entités des deux premières ontologies, les individus qui intéressent les ontologies tridimensionnelles et quadridimensionnelles dont nous avons parlé dans la section précédente. C'est la partition qui constitue la première division de BFO (Grenon & Smith, 2004) – entre continuant et occurrents, et se trouve également dans DOLCE – entre endurants et perdurants. Par « choses », il faut entendre ici à la fois les objets mais aussi les animaux ou les personnes : tout ce qui a des parties dans l'espace mais qui est à chaque instant tout entier. C'est selon Zemach celle qui nous vient le plus naturellement, celle qui est enracinée dans notre langage au point qu'on ne la voit pas comme une ontologie particulière. De fait, remarquons que c'est l'ontologie à la base de la plupart des *noms*, propres aussi bien que communs, que nous employons.

26. Les individus que nous considérons ont une existence spatio-temporelle. Ce choix, nous le verrons, n'exclut pas totalement la prise en compte de l'abstraction (notamment par l'utilisation de relations), mais s'éloigne du « réalisme platonicien » concernant les nombres, les idées, les vérités logiques – pour prendre les exemples les plus habituels de candidats à l'existence non spatio-temporelle.

Par « événements » sont désignées des entités qui ont des parties à la fois dans l'espace et dans le temps. On peut prendre l'exemple d'un match sportif. C'est cette ontologie qui est à l'œuvre chez les philosophes ou les physiciens lorsqu'ils parlent de *vers spatiotemporels*, indique Zemach. Pour bien saisir la différence entre les deux sortes d'entités que nous venons de distinguer, imaginons que l'on nous ait présenté une personne et qu'on l'évoque devant nous par la suite. Nous dirons que nous avons déjà vu cette personne, quand bien même notre rencontre n'aurait duré que quinze minutes. En revanche, si nous n'avons vu que les quinze premières minutes d'un match sportif, nous ne dirons pas que nous avons vu ce match. Une autre différence entre ces deux sortes d'entités, nous l'avons mentionné dans la section précédente, est qu'une chose – contrairement à un événement – est susceptible de changer au cours du temps, même si la représentation du changement est rendue difficile par la définition d'une chose comme continue dans le temps.

La troisième ontologie concerne des entités continues dans l'espace mais limitées dans le temps. Le rôle individuant que joue l'espace pour les objets est joué par le temps pour les processus. Zemach précise qu'elle est rarement utilisée, et dans des termes qui encouragent la confusion avec la première ontologie. Il en donne pour exemples « la Révolution Industrielle » ou « l'occupation allemande ». Selon lui, la Révolution Industrielle ou l'occupation allemande ont pu se dérouler à différents endroits au même moment, mais pas au même endroit à différents moments. Nous ne suivons pas Zemach ici : nous considérons la Révolution Industrielle ou l'occupation allemande, comme des entités de type « événement », même s'il faut bien entendu prendre en compte leur caractère composé : des événements qui sont une composition d'événements. En revanche, la définition d'entités qui n'ont pas de parties spatiales mais qui ont des parties temporelles, pour lesquelles c'est le temps qui joue un rôle individuant, nous paraît s'appliquer de manière très pertinente aux projets. Par projet, nous entendons une succession planifiée d'activités. Cela concerne donc le projet au sens managérial de la gestion de projets, mais aussi bien une procédure d'évacuation d'un bâtiment ou un algorithme de résolution d'un problème. Quand bien même la localisation spatiale peut avoir une place importante dans le projet, parce qu'il va porter sur certains objets en particulier (par exemple dans le cas d'un projet d'aménagement urbain ou de la planification d'une intervention chirurgicale sur un certain organe), nous considérons que ces objets doivent être vus comme des *intrants* du projet mais que ce dernier est défini avant tout temporellement, par l'organisation d'une succession d'étapes.

La quatrième ontologie qu'envisage Zemach traite d'entités non limitées dans l'espace ni dans le temps, les « types », qui pour lui ne sont ni des universaux, ni des entités abstraites, ni des formes, ni des classes. Les exemples qu'il en donne sont en premier lieu ce qui est défini par un « le » institutionnel, comme dans « le lion africain est féroce ». L'auteur critique la position de Frege et Russell visant à interpréter ces expressions comme des assertions quantifiées (« Tout ce qui est un lion africain est féroce »). Mais sa propre position est fragile, comme il le reconnaît lui-même, en affirmant que dire « la Chrysler est une bonne voiture » ne revient pas à dire que toutes les Chrysler sont de bonnes voitures, ou que dire « l'ennemi a pris la colline 69 » ne revient pas à dire que tout x qui est un ennemi a pris la colline 69 ou qu'une partie de l'ennemi a pris la colline 69. Nous pourrions considérer que cette assertion peut être quantifiée en « il existe au moins un x qui est un ennemi, qui a pris la colline 69 ». Il voit également des entités de ce genre dans les noms de masse (« l'eau », « rouge »). Pour prouver qu'une ontologie de purs continuants est possible, l'auteur s'attache à montrer – contre Strawson qui veut limiter les noms de masse à des entités non « sortales » – qu'on pourrait faire disparaître la différence entre « eau » et « chat » au prix de quelques aménagements théoriques (le nom de *type* incluant une taille minimale qui ne sera pas la même pour l'eau et pour le chat). Il envisage également la possibilité que des noms propres soient considérés comme des noms de types. Là non plus, nous ne suivons pas Zemach, dans la mesure où les

chats ne sont pas individués de cette manière dans le langage courant. Sur les noms de masse comme l'eau, le sang, le sable, son argumentation est plus forte. Mais il nous apparaît que c'est surtout l'exemple du « rouge », moins développé, qui est le plus convaincant. « Rouge » n'a pas de partie spatiale ni temporelle. Une chose qui est rouge peut cesser de l'être ou ne pas l'être sur toute sa surface visible, mais rien n'est « une partie de rouge ». Nous considérons donc que la manière d'individuer propre à cette quatrième ontologie s'applique de la façon la plus pertinente aux propriétés.

Dans la dernière partie de son article, Zemach s'interroge sur le statut des œuvres d'art, et se demande si le roman de Tolstoï *Guerre et Paix* est un type ou une chose, si la *Symphonie Héroïque* de Beethoven est un type ou un processus. D'autres entités, proches de celles-ci, sont marquées par l'ubiquité et le polymorphisme : les informations. Nous parlons d'ubiquité dans la mesure où c'est la même information, par exemple « le spationaute français est arrivé à l'ISS », qui est présente dans le journal que je lis et dans celui que lit mon voisin, et de polymorphisme dans la mesure où c'est toujours la même information présentée lors du journal télévisé ou dont je parle avec un collègue de bureau. Nous proposons de parler, pour ces entités qui peuvent occuper de multiples espaces-temps, de *pluricontinuants*, ce qui nous amène à ajouter une cinquième ontologie aux quatre précédemment dégagées.

Nous avons donc distingué cinq « mondes » :

1. le monde des objets,
2. le monde des événements,
3. le monde des projets,
4. le monde des propriétés et
5. le monde des informations.

Cette métaphore des « mondes » ne doit pas être prise dans le sens platonicien de mondes radicalement séparés, ni tout à fait dans le sens que Goodman attribue à ce mot dans *Manières de faire des mondes*, comme des versions correctes mais incompatibles du monde, mais plutôt, comme nous avons tenté de le montrer, comme des ensembles suffisamment séparés par le mode d'individuation au regard de l'espace et du temps des individus qui les composent.

Nous devons aussi ajouter une considération sur le niveau d'abstraction auquel se situent les individus de chacun de ces mondes :

Les **objets**, ou continuants dans le temps, sont ici considérés comme des identités. Ce qui signifie que chaque objet a une contrepartie dans le monde des événements, qui est sa durée de vie, conçue comme l'ensemble des espace-temps qu'il a occupés pendant son existence. Pour reprendre une division de Guarino et Welty (2004), les propriétés rigides qui caractérisent un objet peuvent lui être attribuées; en revanche, les propriétés non rigides ou même anti-rigides, par exemple être malade pendant un certain temps, peuvent être associées à une tranche temporelle (si tout l'individu est concerné), voire spatiotemporelle, de l'événement que constitue la durée de vie de cet objet. Il nous semble qu'on peut ainsi mieux prendre en compte la question du changement.

Les **projets** peuvent connaître plusieurs réalisations dans différents événements. Ils sont à ces événements dans la relation qu'un modèle logique entretient avec différents modèles physiques, dans la méthode MERISE.

Les **événements** sont les seuls éléments entièrement concrets.

Les **propriétés**, qu'on les considère comme dépendantes ontologiquement d'une substance ou comme assemblées en faisceau conformément à la théorie des tropes, peuvent être vues comme des abstractions par sélection opérées sur ce qui nous apparaît comme un individu.

Enfin, concernant les **informations**, le niveau d'abstraction auquel on les envisage dépend du choix de modélisation requis selon le cas d'usage. Le modèle conceptuel FRBR (*Functional Requirements for Bibliographic Records*²⁷ - entités du Groupe 1) distingue à propos des livres quatre niveaux. L'*item* est le livre que j'emprunte à la bibliothèque. La *manifestation* est l'édition particulière de cet exemplaire, qui se caractérise par un contenu mais aussi par un aspect et des circonstances de production. L'*expression* est le contenu textuel spécifique de ce livre. Enfin, l'*œuvre* est le contenu intellectuel commun aux différentes expressions. L'information telle que nous la comprenons, c'est-à-dire comme pluricontinuant, se situe au moins au niveau d'abstraction auquel se situe la manifestation dans le modèle FRBR, l'«individu informationnel» – donc l'instance – étant alors l'item.

3.2 Principe méréotopologique

Le deuxième principe concerne la manière dont nous divisons, délimitons, assemblons ou situons les individus appartenant à un même type.

Dans la section précédente, une notion s'est révélée importante dans le processus d'individuation : la notion de partie. C'est à partir du fait de posséder ou non des parties temporelles et des parties spatiales que nous avons distingué les différents types d'entités au regard de l'espace et du temps. L'étude des relations entre parties et tout est la méréologie. C'est *a minima* une relation d'ordre partiel ayant pour primitive la relation binaire « est une partie de », à partir de laquelle on peut définir les notions de disjonction, pour deux entités n'ayant pas de parties en commun, ou de recouvrement, pour deux entités ayant au moins une partie en commun. Comme le stipule Varzi (1994), une ontologie formelle pour le sens commun doit ajouter, à la méréologie, des concepts et des principes topologiques, par exemple pour exprimer la différence entre continuité et contiguïté, pour exprimer que quelque chose est à l'intérieur ou à l'extérieur d'autre chose, ou encore entourant autre chose. On peut également penser à la nécessité de pouvoir définir de façon topologique certains espaces, comme un lumen en anatomie. Ce qui nous amène à la question de la localisation d'un individu, qu'on pourra définir soit dans un espace de référence soit topologiquement, par rapport à d'autres individus du même type. Enfin, la relation entre partie et tout peut être envisagée selon deux modalités : dans la première, le « tout » considéré est un individu ontologiquement indépendant mais pas ses parties²⁸ (par exemple les parties du corps en anatomie); dans la deuxième, les parties sont des individus ontologiquement indépendants réunis en un tout (par exemple les musiciens membres d'un quartet).

Ce deuxième principe à l'œuvre dans notre ontologie fondationnelle distingue donc :

1. le tout,
2. la partie,
3. l'interface,
4. la composition et
5. la position.

27. Pour une traduction française par la BnF de ce modèle : https://multimedia-ext.bnf.fr/pdf/frbr_rapport_final.pdf

28. Ses parties sont des pseudo-objets, pour le dire comme OntoMénélas.

Ces considérations méréotopologiques peuvent apparaître assez consensuelles lorsqu'il s'agit de représentation spatiale, donc pour les continuants dans le temps. C'est évidemment dans la considération du monde des objets qu'elle s'ancre. Il nous semble possible d'étendre une telle partition aux individus des autres mondes que nous envisageons, au prix de quelques aménagements :

- à part les purs continuants, qui ne sont limités ni dans l'espace ni dans le temps, tous les autres types d'individus ont des parties temporelles et/ou spatiales; la relation partie-tout peut donc s'appliquer à eux, même si dans le cas de parties temporelles, il nous faudra plutôt nous tourner, pour trouver un équivalent de la topologie, vers l'algèbre des intervalles d'Allen (1983).
- le terme généralement utilisé pour désigner la limite de l'extension spatiale d'un continuant dans le temps est celui de « frontière »; afin de généraliser cette notion aux autres types d'individu, nous parlerons, pour ce qui permet de marquer la limite d'un individu avec d'autres individus du même monde ou sa limite dans l'espace dans lequel il se situe, d'interface.
- nous parlons de composition pour désigner le cas où les parties d'un individu sont également des individus de plein droit.
- enfin, nous parlons de position pour désigner les différentes manières possibles de situer un individu dans un espace de référence immatériel ou par rapport à d'autres individus du même type.

Reste à envisager le cas des propriétés, qui comme nous l'avons remarqué n'ont pas de partie et ne se prêtent donc pas aux considérations méréotopologiques. Les propriétés permettent de qualifier, de caractériser, et ce faisant de discriminer. Elles sont appelées à recevoir des valeurs. On peut donc se poser la question du type de valeurs susceptible de s'appliquer à un tout, à une partie, à une interface, à une composition. Il ne s'agit plus, comme dans les mondes dont les individus ont des parties, d'envisager ces individus sous l'angle de la méréotopologie, mais plutôt de distinguer les propriétés susceptibles de s'appliquer aux individus, à leurs parties, à leurs interfaces et à leur composition, selon les mêmes principes méréotopologiques. La démarche peut sembler relever d'un « esprit de système », mais permet de retrouver des types de propriétés assez classiques :

- pour qualifier un tout, on recourra à une *qualité*; par exemple, on distinguera un ballon d'un autre ballon de même forme par leurs couleurs.
- pour qualifier une partie, on recourra à une *quantité*; par exemple, on distinguera cette part de cette autre part de la même tarte par la portion qu'elle représente de la tarte entière.
- pour qualifier une interface, on précisera son rôle; rôle est à entendre ici en un sens large, comprenant aussi bien un rôle social qu'une fonction ou une propriété dispositionnelle (on retrouve ici les « entités réalisables » de BFO (Arp & Smith, 2008)).
- pour qualifier une composition, on précisera le lien à l'œuvre dans cette composition, par exemple la simple agrégation, la composition au sens spatial et presque pictural du terme, la structure (y compris celle d'une structure de données), voire la correspondance (pour un mapping de termes par exemple). Une précision s'impose ici : dans la mesure où une partie, une interface, une composition ont été constituées en tant qu'individus à partir des considérations méréotopologiques exposées précédemment, ils peuvent être considérés comme des totalités et on peut leur attribuer à ce titre des qualités. Un individu conçu comme un tout ne pourra se voir attribuer un rôle que si on envisage son interaction avec d'autres individus.
- enfin, comme nous l'avons remarqué, ces propriétés sont susceptibles de recevoir des valeurs; situer des propriétés dans un espace des possibles qui leur est propre revient donc à

leur attribuer une position dans une région des valeurs – nous nous inspirons ici explicitement de la distinction entre *quality* (la couleur d'une rose) et *quale* (la position d'une qualité individuelle dans un espace conceptuel) présente chez DOLCE (Gangemi *et al.*, 2002) et elle-même inspirée par Gärdenfors (2000).

Certaines distinctions, d'une importance philosophique de premier plan, ne se trouvent néanmoins pas dans la Petite Ontologie Fondationnelle. Il en va ainsi de la question de la dépendance ontologique, qui divise notamment les tenants de la théorie du substrat des tenants de la théorie des tropes et s'exprime généralement sous forme de relations de possibilité ou de nécessité. Cette notion apparaît très haut dans BFO pour distinguer les continuants selon qu'ils sont indépendants (et donc des substances), spécifiquement dépendants d'une substance (qualités ou entités réalisables), ou génériquement dépendants d'une substance (comme les informations). DOLCE ne fait pas de la dépendance ontologique un principe de division parmi les classes (à part pour les *dependent-place* comme le trou dans le fromage ou le dessous de la table) mais parmi les relations. Dans la mesure où nous considérons les propriétés et les informations comme individuées différemment des objets, nous considérons que les relations de dépendance sont appelées à apparaître, si elles le doivent, sous forme d'*object properties*.

Il en va également ainsi de la dichotomie entre matériel et immatériel (*material entity / immaterial entity* dans BFO, *physical / non-physical* ou *physical / abstract* dans DOLCE). A nouveau, il ne s'agit pas de dénier l'importance philosophique d'une telle distinction. Dans une perspective de proposer une ontologie fondationnelle plus propre à suggérer son usage, il apparaît que cette distinction, comme la précédente, retarde de façon dommageable l'apparition de classes dont le sens parle à celui qui n'est pas familier des débats d'ontologie philosophique. Que l'on considère :

- un *object* dans BFO, qui est une *material entity*, elle-même un *independant continuant*, lui-même un *continuant*, qui est une *entity*, ou
- un *physical-object* dans DOLCE, qui est un *physical-endurant*, qui est un *endurant*, qui est un *spatio-temporal-particular*, qui est un *particular*,

cinq niveaux de profondeur sont nécessaires pour atteindre la classe « objet ». En outre, cette distinction ne présente aucun problème dans de nombreux cas (par exemple entre un objet et la région spatiale qu'il occupe) mais en présente dans d'autres cas, qui sont susceptibles d'être tranchés de façon différente selon les contextes (la frontière entre deux pays appartient-elle aux deux?) ou d'appeler une modélisation plus avancée (pour exprimer par exemple qu'une bille soit *à l'intérieur* d'une gorge sans que les deux ne partagent pourtant aucune part).

3.3 La matrice de la Petite Ontologie Fondationnelle

En croisant les cinq mondes distingués avec les cinq catégories méréotopologiques, on peut représenter les vingt-cinq classes principales de la Petite Ontologie Fondationnelle sous la forme d'une matrice (table 2.2).

Une revue exhaustive des classes de l'ontologie est proposée dans l'annexe B. Nous nous contentons ici d'un éclairage sur certaines d'entre elles.

Les projets, en tant que planifications intentionnelles d'actions, d'instructions, requérant certains moyens et visant certains objectifs, ont pour interface ce que nous avons appelé ici intrants et extrants, ou pour le dire en anglais *input* et *output/outcome*. Nous considérons les intrants et les extrants au même niveau d'abstraction que le projet lui-même. La nature des intrants peut être très diverse, et peut notamment être constituée des extrants d'un autre projet (pipeline). Les extrants peuvent eux-mêmes être divers, mais on peut distinguer les produits prévus en sortie du projet (*output*) et les résultats visés par la mise en place du

	Tout	Partie	Interface	Composition	Position
Le monde des objets	Un objet	Une partie d'objet	Une frontière d'objet	Un objet composé	Une situation dans un espace
Le monde des projets	Un projet	Un élément de projet	Intrant / extrant	Un programme	Une situation dans le temps
Le monde des événements	Un épisode	Une partie d'épisode	Apparition / disparition	Une histoire	Une situation spatiotemporelle
Le monde des propriétés	Une qualité	Une quantité	Un rôle	Un lien	Une région de valeurs
Le monde des informations	Un énoncé	Un élément d'énoncé	Un code	Un jeu d'énoncés	Une région culturelle

TABLE 2.2 – Les principales classes de la Petite Ontologie Fondationnelle sous forme de matrice.

projet (*outcomes*). La temporalité attachée aux projets peut être définie de façon chronométrique, selon l'algèbre des intervalles d'Allen, voire de façon algorithmique dans le cas d'un programme informatique.

Dans le monde des événements, la question de la granularité se pose de façon plus aiguë que dans le monde des objets. Le terme d'événement ne porte pas en lui de considération de granularité. Nous l'abordons ici en empruntant au vocabulaire de la narration de l'événement dans un récit. C'est pourquoi nous proposons comme individu de base l'épisode, c'est-à-dire ce qui reçoit une unité d'un point de vue narratif, et l'histoire comme composition d'épisodes. Un examen clinique, la passation d'un test psychologique ou un séjour hospitalier peuvent constituer des épisodes, dans une histoire médicale.

Le monde des informations présente une difficulté particulière liée à la diversité de ce qui peut être nommé « information », ce qui est lié au polymorphisme que nous évoquons. Chaque domaine a développé séparément ses propres conceptions sur les individus qui le concernent. Afin de conserver un terme générique et compréhensible, nous avons choisi de parler d'« énoncé » pour désigner dans le monde des informations un individu dont les parties ne sont pas des entités au sens plein, et de « jeu d'énoncés », faute de terme qui soit plus générique sans indiquer le lien entre les énoncés. Ce terme fait écho au « jeu de données », une donnée étant un exemple de ce que nous visons comme énoncé. La question de l'interface, dans le monde des informations, renvoie à la question de ce qui rend un énoncé opérable, voire interopérable. On distingue généralement dans une information sa forme et son contenu (cf. le modèle conceptuel des représentations dans YAMATO (Mizoguchi & Toyoshima, 2017)). C'est cette notion de forme que recouvre le terme de « code », qui peut désigner le cas échéant un encodage dans une langue formelle ou une expression dans une langue naturelle, mais aussi un code comme le code Morse ou un code visuel comme celui des pictogrammes. La notion de région culturelle intervient pour permettre de situer un énoncé dans l'espace des énoncés possibles. Ce dernier est évidemment immense et appelle à être précisé. *L'Information Coding Classification* de Dahlberg (2008), avec ses 6 500 domaines de connaissance, est une initiative systématique remarquable qui va en ce sens.

3.4 Les relations

Si les classes des ontologies fondationnelles sont généralement fortement hiérarchisées, il n'en va pas de même des relations : dans la plupart sinon la totalité des ontologies, les *Object Properties*, par exemple, sont des sous-propriétés directes de `owl:topObjectProperty`. Elles sont susceptibles d'avoir elles-mêmes des sous-propriétés, mais il manque un « haut niveau » d'organisation. On peut toutefois souhaiter, pour les mêmes raisons que pour les classes, disposer d'un principe de classification des relations, tout en conservant un souci de parcimonie pour parvenir le plus rapidement possible à des relations signifiantes. D'autant que les possibilités offertes par les logiques de description sont étendues en ce qui concerne les propriétés, puisqu'aux possibilités de subsomption, d'indication d'équivalence, d'indication de l'inverse, d'indication du domaine et du codomaine, de chaînage de propriétés (depuis OWL2), de disjonction, d'inférence des subsomptions au moyen de raisonneurs, il faut ajouter la possibilité de spécifier les caractéristiques d'une propriété : symétrique, asymétrique, réflexive, irreflexive, transitive, fonctionnelle, fonctionnelle inverse. Pourtant, alors que la littérature est abondante sur l'élaboration d'une taxonomie de classes, elle est inexistante sur l'élaboration d'une taxonomie de relations. Nous faisons ici une proposition qui est fortement reliée à notre taxonomie de classes.

S'appuyer sur les caractéristiques des propriétés peut s'avérer périlleux, d'abord parce que cela contraint fortement dès l'ontologie fondationnelle l'utilisation ultérieure des relations, et ensuite parce que l'héritage de ces propriétés est dans des directions inverses selon la caractéristique considérée : *top-down* pour la symétrie, mais *bottom-up* pour la réflexivité, par exemple.

Une deuxième possibilité peut être de reprendre une typologie, voir une taxonomie des propriétés comme celle de Guarino & Welty (2000). La difficulté sera alors, pour l'utilisateur, de bien maîtriser cette classification pour ajouter une propriété ou seulement pour retrouver une propriété déjà existante. A cette difficulté s'ajoutera un ordonnancement des relations qui ne correspondra en rien aux classes qu'elles relie.

Une troisième possibilité, qui répond à ce dernier point, est de s'appuyer pour classer les relations sur les domaines et les codomaines. C'est le choix opéré par OntoMénélas, qui ajoute, pour organiser les propriétés, des surpropriétés dont le nom mentionne le produit cartésien du domaine et du codomaine les plus restreints contenant les domaines et codomaines des sous-propriétés. C'est une démarche *bottom-up* qui regroupe des propriétés ayant des similarités évidentes et permet d'indiquer quelles régions de la taxonomie sont mises en relation.

Notre démarche est assez proche de cette troisième option, tout en s'appuyant sur la matrice que nous venons d'exposer. Les relations les plus importantes dans ontoPOF peuvent être :

- entre entités de types différents :
 - participe à – relation entre un objet et un événement auquel il prend part,
 - a pour vie entière – relation entre un objet et l'événement qu'est son parcours de vie,
 - est réalisé par – relation entre un projet et un événement qui le concrétise,
 - qualifie – relation entre une propriété et un individu qu'elle qualifie,
 - porte sur – relation entre une information et un individu sur lequel l'information porte;
- entre entités du même type mais de catégories méreotopologiques différentes :
 - a pour partie – relation entre une partie et le tout dont il est une partie,
 - a pour frontière – relation entre une frontière et le tout dont il est une frontière,
 - est composant de – relation entre un tout et une composition dont il est un composant,
 - se situe à la position – relation entre une partie, un tout, une frontière ou une composition, et la position à laquelle il se trouve.

Dans ce deuxième cas, les relations sont ensuite spécifiées selon les types (a pour partie spa-

tiale, a pour partie temporelle, etc.). Dans une démarche plutôt *top-down*, nous avons choisi d'organiser les relations à partir des types mis en relation. La considération de propriétés mettant en relation de façon privilégiée certains types et de propriétés mettant potentiellement en relation un type avec tous les types (qualifie, porte sur) aurait pu amener à une tripartition (a) tous types – type, (b) type – type, (c) type – tous types, en spécifiant éventuellement cette tripartition (les relations objet – événement, les relations projet – événement). Mais cette solution séparait les relations inverses, sauf à n'avoir qu'une bipartition type – type et type – tous types, cette dernière n'indiquant plus le sens de la relation.

Pour cette raison, nous avons séparé en premier lieu les relations hétérotypiques et les relations homotypiques puis – à l'intérieur des relations homotypiques – les relations méréotopologiques des relations non méréotopologiques (cf. figure 2.2). Ces dernières sont peu représentées dans la Petite Ontologie Fondationnelle mais le sont beaucoup plus par exemple dans ontoDOME au moment de modéliser les relations familiales (cf. chapitre 3, section 2.4), où l'on a bien des relations entre entités du même grand type ontologique mais qui ne sont pas entre elles dans des relations méréotopologiques. Les seules relations de ce genre dans la Petite Ontologie Fondationnelle résultent de la partition initiale entre homotypique et hétérotypique, qui nous a obligé à scinder les relations « universelles » selon qu'elles concernent ou non une entité du même type. Pour les propriétés, on a donc les relations `qualifie` et `estMetaproprieteDe`; pour les informations, on a les relations `porteSur` et `estMetainformationSur`.

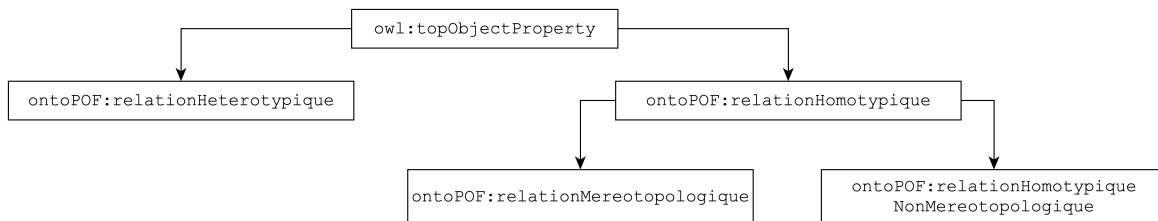


FIGURE 2.2 – L'organisation des relations de la Petite Ontologie Fondationnelle.

4 Conclusion

Notre objectif en nous engageant dans l'élaboration d'une ontologie fondationnelle était de parvenir à proposer une ontologie qui reste le plus largement compatible avec BFO tout en traitant mieux des questions de temporalité et en offrant une meilleure ergonomie cognitive.

Sur le premier point, notre travail sur le changement nous a amenés à considérer un engagement ontologique différent de celui d'une ontologie endurantiste comme BFO, et à voir dans les « objets » dont traite une ontologie endurantiste non pas des primitives, mais des entités complexes dont la représentation suppose de lier, à une primitive représentant l'identité à travers le temps de cette entité (marquée par l'identité à travers le temps de la portion d'espace qu'elle occupe), un certain nombre de primitives prises dans le monde des événements, au premier rang desquelles la vie entière de l'objet²⁹. Cette conception peut sembler s'éloigner du sens commun mais en conserve un substrat d'attribution de ce qui reste invariant et offre plus de souplesse dans

29. De la même manière, représenter par exemple un hôpital est difficilement imaginable sans associer au moins une localisation spatiale, une équipe, un ensemble de règles de fonctionnement

la représentation du changement : en se centrant sur les événements, on échappe à bon nombre de difficultés liées à la représentation des relation n-aires en logique de description³⁰.

Nous avons également été amenés à considérer les projets, les propriétés et les informations sur un pied d'égalité avec les objets et les événements. Outre les justifications philosophiques déjà apportées, la place accordée aux projets et aux informations pourraient aussi recevoir une justification d'ordre sociologique, liée à la manière dont est menée la science et la place qu'y ont pris les données.

Sur le deuxième point, la première division de la Petite Ontologie Fondationnelle reste marquée par le jargon philosophique, puisque les individus sont soit des continuants dans le temps, soit des continuants dans l'espace, soit des non-continuants, soit des purs continuants, soit des pluricontinuants. Mais dès le deuxième niveau de profondeur, on trouve vingt-cinq classes qui nous apparaissent suffisamment parlantes d'elles-mêmes pour faciliter leur appropriation. Nous aurions pu envisager d'inverser les deux dimensions à l'œuvre et de placer en premier la division entre tout, partie, interface, composition et position. Nous aurions ensuite retrouvé au deuxième niveau de profondeur, mais organisées différemment, les mêmes classes, faisant disparaître tout jargon philosophique. Cela aurait été, nous semble-t-il, au prix du « sens commun », l'appréhension en grands types d'entités l'emportant dans notre conception du monde sur les considérations méréotopologiques.

Enfin, comme son nom l'indique, la Petite Ontologie Fondationnelle a également une visée minimaliste : réduire l'engagement ontologique de cette ontologie fondationnelle amène à répartir plus graduellement cet engagement, à en reporter la charge sur les ontologies qui la spécifient, et en premier lieu une ontologie de niveau intermédiaire. C'est une telle ontologie que nous allons maintenant aborder.

30. On est alors assez proche du *Use Case 3* décrit dans <https://www.w3.org/TR/swbp-n-aryRelations/>.

ontoDOME : une ontologie noyau des données médicales

To be is, purely and simply, to be the value of a variable.

Willard V. QUINE, *On What There Is*.

Le deuxième module de notre ontologie est une ontologie noyau des données médicales. Après avoir rappelé le rôle d'une ontologie de niveau intermédiaire, présenté les principales ontologies de ce type dans le domaine biomédical et justifié le choix de se centrer sur les données, nous détaillons la manière dont elle spécifie l'ontologie fondationnelle pour préparer l'intégration des ontologies de domaine.

1 Une ontologie noyau centrée sur les données

1.1 Les ontologies noyau

Les ontologies de niveau intermédiaire font l'objet de beaucoup moins d'attention dans la littérature que les ontologies de haut niveau ou les ontologies de domaines. Comme nous le mentionnons dans la section 3 du chapitre 1, elles correspondent à cette couche ontologique spécifiant l'ontologie de haut niveau et spécifiée par les ontologies de domaine.

Dans différents processus de travail mettant en jeu une structuration hiérarchique, comme par exemple la gestion de projets, ou même dans le traitement humain de l'information (Lindsay & Norman, 1977), on distingue communément les approches *top-down* et *bottom-up*. À ces deux approches est venue s'ajouter l'approche *middle-out* (White, 1993), le développement à partir d'une couche intermédiaire rendant plus aisées à la fois la généralisation et la spécification.

Dans le domaine de l'élaboration des ontologies, l'approche *top-down* correspond classiquement à l'extension de la taxonomie et des relations à partir d'une ontologie de haut niveau quand l'approche *bottom-up* correspond classiquement à l'apprentissage à partir de documents (cf. section 3.1.2 du chapitre 1). L'approche *middle-out* permettrait quant à elle, selon Uschold & Grüninger (1996), de diminuer les efforts de conception en conservant le meilleur équilibre de niveau de détail, quand l'approche *top-down* risquerait d'imposer des catégories de haut niveau arbitraires, d'être par conséquent moins stable, de nécessiter plus de remaniement, et que l'approche *bottom-*

up risquerait d'être trop détaillée, demanderait plus d'effort, permettrait moins facilement la montée en généralité et augmenterait le risque d'incohérences. Il s'agirait, pour la mettre en œuvre, d'identifier les catégories les plus générales d'un domaine, desquelles les catégories de plus haut niveau émergeraient naturellement.

Rudnicki *et al.* (2016) offrent une image assez différente. Ils remarquent ainsi que « le développement d'ontologies de niveau intermédiaire se fait par extension vers le bas à partir de l'ontologie de niveau supérieur ». Ils continuent en rappelant les deux objectifs des ontologies de niveau intermédiaire, qui « contiennent une terminologie qui décrit des entités présentant un intérêt pour de nombreux groupes » et « permettent aux développeurs d'ontologies de niveau inférieur d'intégrer leur travail dans l'architecture sémantique commune », puisque « les nœuds racines de chaque ontologie de niveau inférieur sont des termes existants dans l'ontologie de niveau intermédiaire »¹. Ils ajoutent ensuite que « le développement d'ontologies de niveau inférieur contribue au développement d'ontologies de niveau intermédiaire, étant donné que les besoins terminologiques identifiés au niveau inférieur devront dans certains cas être satisfaits en ajoutant du contenu aux ontologies de niveau intermédiaire, qui peuvent servir de point de départ pour une population descendante ultérieure ». Le mouvement décrit est plus centripète que centrifuge et l'ontologie de niveau intermédiaire est *caught in the middle*, prise entre le marteau de l'ontologie de haut niveau qui impose ses catégories abstraites et l'enclume des ontologies de domaine qui présentent des exigences terminologiques.

Si l'architecture modulaire rend plus probable le tableau suggéré par Rudnicki *et al.* (2016), notamment lorsque l'ontologie de haut niveau est antérieure à l'élaboration de l'ontologie de niveau intermédiaire et que les ontologies de domaine sont développées en collaboration avec les experts du domaine, celui décrit par Uschold & Grüninger (1996) reste crédible dans le cas d'une ontologie monolithique et leurs recommandations de parcimonie et de stabilité restent valables.

1.2 Les ontologies intermédiaires biomédicales

1.2.1 OGMS, OBI et IAO

L'article de Rudnicki *et al.* (2016) – dont Barry Smith est l'un des coauteurs – préconise également, outre l'architecture à plusieurs niveaux de généralité, une modularisation par silos au niveau intermédiaire : « le domaine d'une ontologie de niveau intermédiaire doit être exprimable sous la forme d'une classe unique ou d'un énoncé composé de classes et d'une propriété d'objet au sein de cette ontologie ».

C'est ce qui est mis en œuvre par l'OBO Foundry, où la couche intermédiaire pour le domaine biomédical comprend trois ontologies de niveau intermédiaire : l'*Ontology for General Medical Science* (OGMS), l'*Ontology for Biomedical Investigations* (OBI) et l'*Information Artifact Ontology* (IAO).

OGMS² est la transposition sous forme ontologique de l'article de Scheuermann *et al.* (2009), qui vise à proposer un cadre cohérent de représentation des notions de base que sont la maladie, le diagnostic, le phénotype clinique. Il présente notamment une modélisation de la maladie

1. La mention de ces deux objectifs permet de comprendre qu'outre la dénomination d'« ontologie de niveau intermédiaire », qui correspond au dernier objectif, on rencontre aussi la dénomination d'« ontologie noyau », qui correspond au premier objectif : proposer un noyau de classes et de relations communes à un champ disciplinaire.

2. <https://obofoundry.org/ontology/ogms.html>

sous la forme du triptyque désordre (*trouble*) / disposition³ / cours de la maladie (*disease course*) : « nous défendons une vision de la maladie comme impliquant dans tous les cas une base physique au sein de l'organisme qui porte une disposition à l'exécution de processus pathologiques ».

IAO⁴ (Smith & Ceusters, 2015) est une ontologie des entités informationnelles. Elle spécifie la classe *generically dependent continuant* de BFO, qui désigne les entités qui sont ontologiquement dépendantes d'au moins un substrat, mais pas d'un substrat en particulier (cette dépendance générique d'une entité informationnelle au support physique qui la porte est à différencier de la dépendance spécifique, par exemple, de la masse d'une entité physique à cette entité). Elle vise la représentation des documents, des données et des bases de données, des images, et plus généralement de toutes les *Information Content Entities* (ICEs), définies comme des entités génériquement dépendantes d'une entité matérielle et entrant dans la relation d'*aboutness* (propriété d'être à propos de quelque chose) à une entité.

OBI⁵ (Bandrowski *et al.*, 2016) est une ontologie des études cliniques et biomédicales. Elle est issue de la *Functional Genomics Investigation Ontology* (FuGO), se place sous l'ontologie fondationnelle BFO et reprend des classes d'OGMS pour ce qui concerne les aspects médicaux (phénotype, maladie, diagnostic, traitement) et d'IAO pour ce qui concerne les objets informationnels. L'objectif d'OBI est de fournir un modèle ontologique exhaustif que toute étude biomédicale pourra venir instancier. Elle est donc très détaillée (plus de 4 000 classes), répond déjà à de nombreux cas d'usage et est appelée à croître pour en embrasser davantage.

1.2.2 COB

La *Core Ontology for Biology and Biomedicine*⁶ est une ontologie en cours de développement⁷ qui répond au manque d'ontologie noyau de haut niveau pour la biologie dans l'OBO Foundry. Elle entend s'ancrer dans BFO tout en masquant sa complexité pour les utilisateurs finaux, fournir un parent pour toute classe d'une ontologie OBO et inclure des axiomes logiques permettant de faire apparaître les éventuelles incohérences lors de l'application d'un raisonneur. Elle reprend l'IRI des classes de plus haut niveau des ontologies OBO comme IAO, OGMS, OBI, CheBI, ENVO (en adoptant éventuellement un libellé différent). Les classes les plus hautes de COB sont *material entity*, *immaterial entity*, *characteristic*, *process*, *information*. Outre quelques relations spécifiques à cette ontologie, les relations dans COB empruntent essentiellement à l'ontologie des relations RO Core, mais reprennent aussi la relation principale d'IAO (*is about*), ainsi qu'une relation de STATO permettant un lien plus direct entre un projet et son exécution sans passer par le document qui concrétise – en le notifiant – le projet (*executes* : realizes o concretizes \sqsubseteq executes).

Les points de développement en cours concernent (a) les phénotypes et les maladies (la modélisation préconisée par l'ontologie de haut niveau ne correspond pas à l'usage effectif des ontologies de phénotypes et de maladies) et (b) les mesures et les unités (une simplification du modèle actuel est proposée).

3. Disposition est à entendre ici au sens philosophique d'une propriété d'une entité non directement observable en l'état actuel et qui peut se manifester ou pas selon les circonstances.

4. <https://obofoundry.org/ontology/iao.html>

5. <https://obofoundry.org/ontology/obi.html>

6. <https://github.com/OBOFoundry/COB>

7. Nous nous appuyons ici sur la présentation de Chris Mungall lors de l'*Ontology Summit 2023* – https://ontologforum.com/index.php/ConferenceCall_2023_01_25 – consulté le 15 août 2023.

1.2.3 BioTop

BioTop⁸ se présente comme une « ontologie de niveau supérieur pour les sciences de la vie » (Schulz *et al.*, 2017), expression qui peut interroger au regard de la modularisation par couches de généralité que nous avons exposée. Reprenant l’argumentaire en faveur des ontologies de haut niveau (qui facilitent la maintenance, l’extension, la modularisation et la réutilisabilité), l’auteur ajoute que les ontologies de haut niveau les plus populaires ne sont pas complètement indépendantes de tout domaine (DOLCE visant les sciences cognitives et BFO les sciences naturelles). De ce point de vue, BioTop est donc une ontologie de haut niveau explicitant le domaine général qu’elle vise. Elle est également accompagnée d’un fichier d’alignement entre BTL2 et BFO2.

BioTop a connu plusieurs évolutions au cours du temps, aboutissant à des versions « allégées », BioTopLite puis BioTopLite2 (BTL2). Cette dernière porte une attention particulière au fait de proposer des libellés intuitifs pour les classes et comporte une hiérarchie de relations simplifiée. Les classes sont assorties d’axiomes contraignant leur extension (soit au niveau de la classe, soit en tant qu’axiomes généraux d’inclusion de la classe). Elle se déclare agnostique sur l’existence d’universaux et place comme classe la plus élevée *particular at some time*, en précisant que toutes les entités singulières sont dans un contexte temporel (qu’il s’agisse d’événements ou d’objets).

1.3 La place des données

Comme nous l’avons exposé dans le chapitre 2, notre ontologie fondationnelle accorde une forme de primauté à la prise en compte des événements. Toutefois, les informations y sont des « citoyennes de première classe », en tant qu’elles constituent un type spécifique d’entités. Parmi ces informations, certaines ont pris une place prééminente lors des dernières décennies, renforcée encore lors des dernières années : les données. Dans un ouvrage intitulé *The Fourth Paradigm – Data-intensive Scientific Discovery* (Hey, 2009), anthologie d’essais sur la science des données, les auteurs anticipaient une nouvelle manière de faire de la science, après la manière empirique datant de quelques milliers d’années (procédant par description de phénomènes naturels), après la manière théorique datant de quelques siècles (procédant par élaboration de modèles et généralisations) et après la manière informatique datant de quelques décennies (procédant par simulation de phénomènes complexes). Ce quatrième paradigme, celui de l’exploration des données, allait unifier la théorie, l’expérience et la simulation. C’était quelques années à peine avant les premiers succès marquants de l’apprentissage automatique, qui d’une certaine manière ont concrétisé cette promesse tout en faisant de l’abondance des données autant une opportunité qu’une nécessité. Concomitamment, de plus en plus de secteurs d’activité ont valorisé le fait de se fonder sur les données, au premier rang desquels l’économie. La médecine n’a pas échappé à ce mouvement et l’on parle à présent de *data-driven medicine* (Shah & Tenenbaum, 2012; Torkamani *et al.*, 2017). Comme nous l’avons vu dans le chapitre 1, la médecine de précision est une médecine qui s’appuie sur une quantité importante de données souvent hétérogènes, provenant de différentes sources. C’est à nos yeux une première raison d’orienter une ontologie noyau dans le domaine médical vers la prise en compte prioritaire des données.

Au-delà de ce constat de fait, consacrer la place accordée aux données dans une ontologie noyau du domaine médical représente une option concernant la destination des ontologies. À l’ambition maximaliste de voir les ontologies simuler le raisonnement humain se substitue l’objectif plus modeste d’ontologies permettant d’étendre la cognition humaine en offrant la possibilité de manipuler de grandes quantités d’informations ou de documents (cet objectif est convergent

8. <https://biotopontology.github.io/>

avec la situation d'abondance de données que nous venons de décrire). Pour les producteurs de données, le modèle qui a été défini permet la cohérence de ces vastes ensembles de données, cohérence que l'utilisation de raisonneur permet de valider (la capacité des ontologies à simuler le raisonnement déductif restant donc essentielle). Pour les utilisateurs finaux, l'accent est mis sur la capacité à explorer ces mêmes données de façon rapide et efficace, sans avoir à connaître le modèle de stockage sous-jacent (ici aussi, les capacités d'inférence déductive des ontologies sont mises à contribution). L'interopérabilité des données est assurée dans une étape intermédiaire, avec la collaboration des producteurs de données et des utilisateurs finaux, de constitution d'un modèle de données commun représenté dans l'ontologie. Nous sommes très proches d'un « type d'application où les ontologies ne sont pas tant la représentation de connaissances que la médiation entre expressions de connaissance, permettant à l'utilisateur de naviguer et travailler sur des expressions dont il maîtrise l'interprétation » (Bachimont, 2004). Cette option pourrait être qualifiée, pour emprunter au vocabulaire de l'épistémologie, d'« instrumentaliste » (Declerck & Charlet, 2014) : « cette approche, dont le point de départ est la reconnaissance de la nature prioritairement artéfactuelle des ontologies, conduit à reléguer au second plan leur fonction représentationnelle, pour mettre en avant leur fonction prothétique [...] les ontologies participent à la construction des domaines de connaissance qu'elles formalisent, et peuvent le cas échéant contribuer à structurer et organiser les pratiques de production de contenus et d'encodage de l'information ».

Ces éléments orientent la démarche de modélisation ontologique. Nous pourrions représenter le champ des possibles selon deux axes : l'un serait l'importance accordée à la gestion des données (c'est-à-dire à la modélisation des contextes de production, des flux de données, des traitements, des acteurs) ; l'autre serait l'« exposition » plus ou moins importante des données (l'effort de modélisation accordé à la représentation des connaissances contenus dans les jeux de données). Savoir où se positionner dans cet espace revient finalement à remotiver le rôle des *competency questions* (Grüninger & Fox, 1995b), mais selon deux dimensions et en prenant acte de la diversité possible d'utilisateurs finaux, voire d'utilisateurs intermédiaires. Notre position, qu'on peut induire de ce qui précède, est de favoriser la traçabilité de la provenance et des traitements appliqués aux jeux de données tout en « exposant » les événements ayant généré les données, les éventuels projets réalisés par ces événements, les objets (au sens large, en comprenant donc les personnes) impliqués dans ces événements, les variables enregistrées dans ces jeux de données.

L'importance de l'interopérabilité amène donc à considérer, outre les aspects strictement médicaux, le Modèle des Objets de Santé (MOS) préconisé par l'Agence du Numérique en Santé (ANS) et les nomenclatures associées⁹ mais aussi la norme *Fast Healthcare Interoperability Resource* (FHIR) d'HL7¹⁰, le modèle OMOP d'OHDSI¹¹, ainsi que les principes FAIR¹² (Wilkinson *et al.*, 2016).

9. qui constituent un vocabulaire de base pour le cadre d'interopérabilité des systèmes d'information en santé (CI-SIS) – <https://esante.gouv.fr/produits-services/mos-nos> – consulté le 15 août 2023.

10. qui se présente comme un standard d'interopérabilité pour l'échange de données de santé. – <https://fhir.org/> – consulté le 15 août 2023.

11. qui se présente comme *designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence.* – <https://www.ohdsi.org/data-standardization/> – consulté le 15 août 2023.

12. pour *Findability, Accessibility, Interoperability, and Reusability*, à propos de la gestion des données scientifiques.

2 Élaboration d'ontoDOME

Nous ne présenterons pas en détail les quelques centaines de classes et de relations d'ontoDOME, mais nous allons montrer – selon une organisation thématique – la manière dont elles spécifient les classes et les relations de l'ontologie de haut niveau ontoPOF. La médecine, entendue dans son acception la plus immédiate, s'articule autour des problèmes de santé rencontrés par des êtres humains, de la capacité à identifier ces problèmes et à proposer des moyens d'y remédier : c'est la médecine comme science et comme ensemble de pratiques. Ces pratiques s'inscrivent dans un ensemble de pratiques sociales, dans un système social dont elles forment un sous-système : c'est la médecine comme système de santé. Au sein de ce système, des données et des informations sont échangées entre les différents acteurs, ce qui constitue le point de vue que nous avons choisi de privilégier. Certaines de ces données, qui ne sont pas purement médicales mais qui sont pourtant utilisées de manière transversale par les différentes spécialités médicales, contribuent à l'accroissement de la science médicale et à l'amélioration de la pratique médicale : les données sociodémographiques relatives au patient. Enfin, ni la maladie ni le soin ne peuvent s'appréhender autrement que comme des processus dynamiques, comme des histoires, dont le déploiement dans le temps ne peut être ignoré.

2.1 Éléments médicaux

Les « objets »¹³ considérés sont ici tout d'abord la *personne physique* (aussi bien celle qui présente un problème de santé que celle qui le soigne) ; une *partie du corps* en général, une *structure anatomique* en particulier ; un *échantillon biologique* ; le *groupe humain* (qui peut être une population, un groupe dans une étude clinique, une équipe médicale) ; un *équipement* en général, un *dispositif médical* en particulier ; une *portion de médicament* (nous ne considérons pas une substance en général, mais toujours une portion individuée, quand bien même nous ne serions pas capables de la quantifier). Conformément à ce que nous avons dit des rôles sociaux dans la section 2 du chapitre 2, les rôles sont indiqués dans des propriétés d'objet (par exemple *aPatientParticipant*, *aMedecinParticipant*) – notamment tous les rôles de professionnels de santé et tous les rôles liés à la recherche clinique.

Les « projets » comprennent les *mesures de prévention* (distinguées en primaire, secondaire et tertiaire) ; les *projets d'intervention médicale en général*, de *traitement thérapeutique* en particulier (chirurgical, médicotechnique, médical – et notamment la *prescription de médicament*) ; les *programmes de soin* ; les *projets d'études de recherche*, les *protocoles de visite* lors d'une étude, les *protocoles d'exams médicaux*. Les intrants de projet sont spécifiés en *objectif de projet* et en *ressource de projet* (la ressource étant ici à entendre au même niveau d'abstraction que le projet lui-même, et donc elle-même susceptible d'être réalisée par exemple par un objet lors de la réalisation du projet en un événement). Les extrants sont quant à eux spécifiés en *résultat attendu* et en *impact attendu*¹⁴.

L'un des intérêts de modéliser les projets étant de pouvoir comparer un « événement » au projet qu'il réalise, on trouvera ici les événements réalisés : les différentes *études de recherche clinique*, les

13. Le mot est à prendre ici au sens générique : nous ne distinguerons pas ici selon la méréotopologie.

14. Cette distinction entre résultat et impact correspond à la différence entre un extrant produit par le projet, quantifiable, disponible à l'issue du projet, et un extrant évalué plus subjectivement et pas nécessairement à court terme. Par exemple, une procédure d'évacuation en cas d'incendie peut être considérée comme un projet, dans notre modélisation, projet visant à la préservation de la vie des occupants du bâtiment et non pas à produire quelque chose de nouveau. En anglais, on parlerait d'*output* et d'*outcome*.

visites réalisées lors de ces études, les *passations d'examens médicaux*. On trouvera aussi les *problèmes de santé*, et notamment les *maladies*, les *processus diagnostiques* (en tant qu'événements, distingués donc de la simple mention du résultat de ce processus), les *épisodes de soin*. On trouve ici aussi la *vie humaine* (spécifiant, pour une personne physique, la classe « vie entière »), ainsi que l'*histoire externe* (qui est l'ensemble des événements auxquels la personne participe) et l'*histoire interne* (qui est l'ensemble des événements dont la personne est la localisation) – conformément aux différents types d'événements mentionnés dans la section 2.1 du chapitre 2 comme devant être associés à un objet pour en offrir une représentation complète. L'annexe C présente un exemple simple de ce qui ressortit de l'histoire interne et de l'histoire externe, et comment les deux s'articulent avec l'objet dont elles sont les histoires.

On distingue généralement les signes et les symptômes selon celui qui les rapporte : observés par le clinicien pour les signes, expérimentés et décrits par le patient pour les symptômes. On parle aussi de signes cliniques objectifs (pour les signes) et subjectifs (pour les symptômes). Cette distinction peut se révéler complexe à faire dès lors qu'on s'intéresse par exemple à l'examen à l'entrée d'un compte rendu d'hospitalisation. On peut partiellement discerner ce qui n'a pu qu'être observé par le clinicien, ce qui n'a pu qu'être rapporté par le patient, mais l'attribution est souvent incertaine. Pragmatiquement, on peut accorder aux signes cliniques présents dans un document une valeur épistémique équivalente : signes ou symptômes, ils ont été passés au crible du clinicien¹⁵. Nous parlerons donc de *signes cliniques*, sans plus de précision. D'autre part, comme ces signes sont rapportés, soit par le patient au clinicien, soit par le clinicien dans un document, nous les considérons comme des « informations »¹⁶. Les *syndromes*, en tant que motifs fréquemment repérés de signes cliniques associés, sont également considérés comme des informations, de même que la *mention d'un diagnostic*. Sont également considérés ici toutes les *représentations visuelles, diagrammes, images ou séquences d'images* issues de l'imagerie médicale. Un outil de recueil d'information, particulièrement important dans les études cliniques et lui-même objet informationnel, est le *formulaire*, qui est composé le plus souvent de différents *items de formulaires*. C'est sous ce terme, qui n'est pas sans lien avec le *Case Report Form* (CRF) des études cliniques, que nous classons les différents questionnaires, échelles, entretiens structurés ou non, relevés de tests, historiques, qui peuvent intervenir dans des événements d'ordre médical pour recueillir des informations sur le patient. Toutes ces informations peuvent être situées dans des régions culturelles et notamment dans des *régions médicales*, que nous considérons ici les différentes *spécialités médicales* ou les *domaines d'intérêt* propres à une spécialité médicale auxquelles elles peuvent être rapportées (ce dernier point est explicité plus en détail pour le cas de la psychiatrie à la section 2.5 du chapitre 4).

Les « propriétés » couvrent à la fois des caractéristiques recueillies de façon transverse en médecine, comme des *variables biométriques*, mais aussi des *réponses à un item de formulaire* données par les patients – qui sont plus utiles à considérer comme des valeurs de variables que comme des énoncés. Ces valeurs peuvent être situées dans des régions de valeurs, dont les *systèmes de codes* de référence (ou les *listes de codes* constituées pour un cas d'usage plus spécifique à partir de

15. La définition du signe comme « a bodily feature of a patient that is observed in a physical examination and is deemed by the clinician to be of clinical significance » et du symptôme comme « a bodily feature of a patient that is observed by the patient and is hypothesized by the patient to be a realization of a disease », dans l'article de Scheuermann *et al.* (2009), indique un différentiel de valeur épistémique, en premier lieu : celui qui sépare un jugement clinique d'une hypothèse.

16. Ce choix présente un bénéfice secondaire : les signes cliniques pouvant se rapporter à des « objets » (« hématome »), des « processus » (« désorientation temporo-spatiale ») ou des « propriétés » (« pâleur »), leur classification dans l'ontologie à partir de ce qu'ils désignent les séparerait selon ces grands types ontologiques.

différents système de codes) constituent un cas particulier.

2.2 Éléments relatifs au système de santé

Nous nous intéressons ici aux moyens organisationnels et institutionnels mis en place au sein d'une société pour structurer l'exercice de la médecine, la dispensation des soins, et pour favoriser la santé. La modélisation du système de santé présente une complexité qui excède ce travail, mais la modélisation des *structures organisationnelles* est elle-même complexe à plusieurs titres. Une telle structure se caractérise au moins par le lieu qu'elle occupe, par l'équipe qui y travaille, par l'ensemble de règles qui régit son fonctionnement. D'autre part, la diversité des structures, les regroupements de structures (groupements hospitaliers, groupements de coopération sanitaire) ou les manières de les diviser (par exemple hôpital, pôle hospitalier d'activité, service, unité fonctionnelle), la multiplicité des acteurs, l'organisation des soins (soins hospitaliers, soins d'urgence, soins ambulatoires, santé publique, etc.) rendent la tâche plus ardue encore¹⁷. Une structure organisationnelle présentant néanmoins une identité dans le temps, nous pouvons *a minima* la modéliser comme un « objet ». Remarquons enfin que les modélisations en ce domaine sont très peu transposables d'un pays à l'autre.

Nous considérons également le *parcours de soin*, qui peut être vu comme un « événement » si l'on s'intéresse à celui qui est effectivement accompli par le patient (et pas au parcours de soins coordonnés, qui est une procédure normative). Il fait l'objet de nombreux travaux, en médecine pour améliorer le soin, mais aussi en intelligence artificielle pour l'identifier dans des documents (*process mining*), en optimiser la planification, en identifier les ruptures (Cardoso, 2019).

Le système de santé comprend aussi l'échange de nombreuses « informations », notamment sous la forme de *documents de santé* qui peuvent émaner du patient (*expression de la volonté du patient*) ou des professionnels de santé (*documents médicaux*, qui peuvent être des *comptes rendus* de divers types, des *lettres médicales*, des *feuilles de surveillance*).

2.3 Éléments relatifs aux données

Les données ressortissent de ce que nous avons appelé le « monde des informations », dans lequel l'individu de base est un énoncé. Une *donnée* est donc considérée comme un énoncé et un jeu de données comme un jeu d'énoncés.

La notion de *métadonnée* est employée de façon parfois ambiguë dans le langage naturel. Au cours du développement des bases de données relationnelles, dans les années 1970, elle a d'abord désigné les noms des champs (par exemple le champ « nom de l'étudiant » dans une base de données sur les étudiants d'une université). Quand les entités sur lesquelles portaient les données de la base de données étaient elles-mêmes des entités informationnelles, ces données aussi bien que les noms des champs ont été appelés « métadonnées ». Le *Resource Description Framework*, avec ses triplets sujet – prédicat – objet, permet de lever cette confusion puisque le nom de champ de la base de données y devient un prédicat (et donc dans une ontologie une relation, ne risquant plus d'être confondue avec une classe). La métadonnée peut donc être considérée simplement comme une donnée qui porte sur une donnée.

Les *jeux de données* considérés ici sont généralement produits par des événements (collectes de données pour les données brutes, chaînes de traitement pour les données dérivées), événe-

17. L'ensemble des établissements est recensé dans le Fichier national des établissements sanitaires et sociaux (FI-NESS), géré par le Ministère de la Santé et de la Prévention. Le Modèle des Objets de Santé de l'Agence du Numérique en Santé propose avec la nomenclature TRE_R66-CategorieEtablissement une catégorisation extensive des établissements.

ments qui obéissent généralement à des protocoles, c'est-à-dire à des projets. Cette structure (et les relations qui la soutiennent) est décrite plus en détail dans la section 2 du chapitre 5, avec une explicitation des différents niveaux de granularité auxquels elle se réplique.

2.4 Éléments sociodémographiques

Les caractéristiques sociodémographiques concernant les patients peuvent avoir pour finalité de devenir des variables indépendantes dans des études épidémiologiques mais permettent également, particulièrement en psychiatrie, d'évaluer la situation sociale générale du patient. C'est pourquoi on les trouve aussi bien dans des études épidémiologiques, dans le Système National des Données de Santé, dans le Modèle des Objets de Santé de l'Agence du Numérique en Santé, que dans dans les comptes rendus d'hospitalisation ou dans des questionnaires spécifiquement sociodémographiques utilisés dans le projet PsyCARE. Toutefois, les caractéristiques sociodémographiques retenues diffèrent selon les contextes. Nous avons retenu le périmètre suivant :

- *sexe* (qui peut être administratif ou de naissance),
- lieu de naissance (*commune, département, pays*) et date de naissance (*mois, année*),
- *localisation de l'enfance*,
- éducation (plus haut *diplôme* obtenu, *niveau scolaire* atteint, *situation scolaire*),
- type de *lieu de vie*¹⁸,
- type de *foyer*¹⁹ et nombre de personnes au foyer,
- *milieu territorial*²⁰,
- *pays de résidence*,
- *pays de nationalité*,
- *fratrie (rang de naissance, nombre de frères et de sœurs, nombre de demi-frères et demi-sœurs)*
- *langue parlée*,
- *statut marital*,
- *nombre d'enfants*,
- *situation judiciaire (antécédents judiciaires, séjour en prison)*,
- *situation juridique (mesure de protection, situation sur le territoire)*,
- *situation professionnelle (catégorie socioprofessionnelle, profession, situation relativement à l'emploi)*,
- *prestation ou allocation sociale perçue*,
- *habitude de vie (activité de loisir, consommation de toxiques)*,
- *âge*.

Ces caractéristiques sont modélisées comme des variables, afin de ne pas alourdir la modélisation et conformément à la finalité épidémiologique mentionnée. D'autres caractéristiques relatives à un patient concernent parfois ses proches (catégorie socioprofessionnelle des parents, niveau d'instruction des parents). Plutôt que de disposer pour un individu d'une variable « catégorie socioprofessionnelle » et d'une variable « catégorie socioprofessionnelle des parents », nous avons préféré considérer que ces variables peuvent s'appliquer à une personne de manière générale. Reste alors à représenter dans l'ontologie le lien de proximité entre deux individus.

Représenter les proches oblige à des choix de modélisation qui vont engager en amont la manière d'obtenir ces informations si elles sont extraites de données textuelles, et engager en aval

18. Par exemple : individuel, familial, hébergement, foyer, hôtel, etc.

19. Par exemple : seul, en couple, en famille, avec des amis, etc.

20. Par exemple : urbain, rural, etc.

l'utilisation qui peut être faite des données. Comme nous venons de l'indiquer, nous avons choisi de considérer que des personnes physiques entretenaient des relations de proximité, et que ces relations n'apparaîtraient donc pas dans la taxinomie (en créant des classes « mère » ou « oncle ») mais dans les propriétés d'objet. Nous avons considéré les relations les plus courantes :

- *lien amical*;
- *lien amoureux*;
- *lien familial* :
 - *conjoint*,
 - *parent / enfant*,
 - *frère / sœur* (et *demi-frère / demi-sœur*),
 - *oncle et tante / neveu et nièce*,
 - *grand-parent / petit-enfant*,
 - *cousin (germain et issu de germain)*;
- *lien juridique*.

Nous avons pris en compte ce qui pouvait avoir un intérêt dans le cadre d'une exploration génétique, en distinguant biologique et adoptif, vrais et faux jumeaux, et en indiquant pour les cousins et les grand-parents le lignage paternel ou maternel. Dans le même esprit, nous avons placé les relations par alliance à part. Nous avons également cherché à mettre en place des relations qui puissent être inversées, ce que l'indication du sexe empêche souvent : si x est grand-père de y, on ne peut pas en déduire ce que y est à x si on n'a que le choix entre petit-fils et petite-fille. Dans ce cas, il suffit de faire de *estGrandPereDe* et de *estGrandMereDe* des sous-propriétés de *estGrandParentDe*, et parallèlement *estPetitEnfantDe*, pour obtenir une relation inversable, ce qui permettra d'obtenir par application d'un raisonneur, si on sait seulement que x est grand-père de y, d'inférer que y est petit-enfant de x. Nous avons été amenés à employer le mot peu usité d'« adelphe », pour un nom neutre désignant un frère ou une sœur²¹ et à créer les relations *estOncle_tanteDe* et *estNeveu_nieceDe* (et pour les relations par alliance le peu usité *estBelEnfantDe* et le néologique *estBelAdelpheDe*). Pour les mêmes raisons, dans les relations fondamentalement symétriques (contrairement aux relations asymétriques que nous venons d'évoquer) nous avons créé les relations *estConjoint_eDe* et *estCousin_eDe*. Nous obtenons plus de 80 relations, qu'il est possible d'étendre si nécessaire (notamment pour les degrés de cousinage ou les relations avunculaires au sens large – grand-oncle et grand-tante, etc.), et qui ont soit une relation inverse, soit une surpropriété symétrique. Ces relations sont présentées dans l'annexe D.

2.5 Éléments temporels

La section 3 du chapitre 2 envisageait l'engagement ontologique relatif aux entités susceptibles de changement. Nous abordons ici plus concrètement la représentation des entités temporelles et d'une situation temporelle ou spatiotemporelle. Nous nous inspirons de la *Time Ontology* du W3C (Cox & Little, 2022) en définissant comme entités temporelles un *instant* et un *intervalle* et en nous appuyant sur les types de données XSD concernant les dates et les durées.

D'autre part, pour le cas où nous avons une variable mesurant une durée, la quantité *durée* est modélisée comme ayant une valeur numérique et une *unité de temps*, classes dont les instances sont au moins les différents niveaux de granularité permettant d'exprimer une *xsd:date* (seconde, minute, heure, jour, mois, année).

La *Time Ontology* considère qu'un instant a une extension temporelle nulle, contrairement à

21. Ce terme permet de traduire l'anglais *sibling*.

un intervalle temporel et prévoit donc les propriétés d'objet `time:hasBeginning` et `time:hasEnd` entre un intervalle temporel et deux instants, mais aussi les propriétés Type de données :

- `time:inXSDYear`,
- `time:inXSDYearMonth`,
- `time:inXSDDate`,
- `time:inXSDDateTimeStamp`,

puisqu'un instant est « dans » quelque chose qui a malgré tout une extension temporelle non nulle.

Nous proposons une autre approche, en considérant un instant comme quelque chose qui marque un repère dans une continuité temporelle (qui définit un avant et un après), alors qu'un intervalle temporel est quelque chose qui marque un début, possède une durée, marque une fin (il y a un avant, un après et un pendant), et ceci indépendamment de la granularité de temporalité avec laquelle l'instant ou l'intervalle temporel sont exprimés. En ce sens, si un événement n'est marqué que par une seule date, il sera considéré comme un instant (par exemple « tentative de suicide en 2020 »); en revanche, s'il est marqué par une date de début et une date de fin, ou l'une des deux accompagnée d'une durée, ou les trois, ou une durée accompagnée d'un repère temporel (par exemple « hospitalisation de trois mois en 2019 ») il sera considéré comme un intervalle temporel.

En élargissant la notion de date, qui désigne au sens le plus habituel une indication du jour, du mois et de l'année, mais que nous prenons ici comme l'indication d'un repère temporel, nous proposons les propriétés Type de données : `axSDDate`, ayant pour sous-propriétés `axSDDateDeDebut` et `axSDDateDeFin`, et `axSDDuree`. C'est le type de données XSD choisi, pour les dates, qui portera la granularité temporelle à laquelle est exprimé l'instant (horodatage à la seconde, à la minute, à l'heure, jour, mois, année). Pour les durées, le type de données `xsd:duration` permet une expression normalisée d'une durée, selon le format `PnYnMnDnHnMnS` (où P est nécessaire et indique la période; nY, nM et nD indiquent le nombre d'années, de mois, de jours; T indique le début d'une section horaire et est nécessaire si l'on veut indiquer nH, nM, nS, c'est-à-dire le nombre d'heures, de minutes, de secondes).

Afin de parvenir à intégrer dans une chronologie les mentions temporellement situées trouvées dans un document, des mécanismes extérieurs à l'ontologie doivent être mis en jeu :

- calcul simple du troisième élément si l'on dispose d'au moins deux éléments d'un triplet date de début – durée – date de fin;
- sinon, approximation par encadrement lorsque la durée est connue (« hospitalisation de trois mois en 2019 »), estimation de la durée lorsqu'elle est inconnue mais reconstituable par des informations extérieures (« cannabis : début à l'âge de 13 ans jusqu'en 2020 ») ou par des connaissances plus ou moins expertes (« hospitalisation de courte durée en 2020 »);
- sinon (nous sommes alors dans un intervalle ouvert puisque le début et/ou la fin ne sont pas approximables) : lorsque le début est approximable et l'événement en cours, la date du document peut être utilisée (« sous curatelle depuis 2020 »), et lorsque la fin est approximable mais pas le début, la date de naissance du sujet concerné peut être utilisée.

3 Conclusion

Comme `ontoPOE`, `ontoDOME` reste un module assez parcimonieux, au sens où il laisse aux modules de domaine qui viendront le spécifier la possibilité d'enrichir la modélisation sur les aspects médicaux ou d'accrocher aux classes qu'il comporte des branches ontologiques issues d'autres ontologies (ainsi, `ontoDOME` ne spécifie par d'ontologie de l'anatomie, mais contient la classe

StructureAnatomique). Cette démarche parcimonieuse, qui s'oppose par exemple à la démarche d'OBI, apparaîtra plus clairement encore et sera justifiée dans le chapitre 5. L'ontologie noyau que nous avons obtenue doit donc être jugée sur sa capacité à orienter la modélisation d'une ontologie du domaine médical et à en fournir les classes supérieures. En dehors du projet PsyCARE, nous avons eu l'opportunité d'utiliser ontoDOME comme ontologie noyau pour une ontologie de domaine consacrée à la néphrologie (ontoNEPHRO – qui avait été conçue originellement avec une autre ontologie noyau), avec des résultats satisfaisants, toutes les classes spécifiques à cette ontologie de domaine ayant pu trouver une classe qu'elles spécifiaient dans ontoDOME. Seule la répétition d'autres expériences du même type, au-delà de cette première expérience encourageante, permettra de tester et le cas échéant d'améliorer la stabilité d'ontoDOME au-delà de celle qu'elle présente dans le cadre de l'ontologie modulaire consacrée à PsyCARE.

Comme nous l'avons vu, le travail d'élaboration a puisé à de nombreuses sources. Nous voulons notamment citer ici le travail sur les facteurs sociaux et environnementaux des maladies psychiatriques accompli par Richard (2017) lors de son doctorat (cf. section 1.2 du chapitre 4), même si le choix que nous avons fait d'axer cette ontologie noyau sur les données n'a pas permis une reprise intégrale de son travail.

ontoDOPSY : une ontologie des données de psychiatrie

« Ceux pour qui certains mots ont un sens, et certaines manières d’être, ceux qui font si bien des façons, ceux pour qui les sentiments ont des classes et qui discutent sur un degré quelconque de leurs hilarantes classifications, ceux qui croient encore à des « termes », ceux qui remuent des idéologies ayant pris rang dans l’époque [...] – ceux-là sont les pires cochons. »

Antonin ARTAUD, *Le pèse-nerfs*.

Parmi les modules de domaine qui viennent spécialiser le module noyau que nous venons de présenter, un module présente une importance particulière dans notre travail : celui consacré à la clinique psychiatrique. Après avoir rappelé les spécificités de la psychiatrie en tant que spécialité médicale et les ontologies qui existent dans le domaine de la santé mentale, nous présentons les différentes branches du module permettant l’intégration et l’annotation des données de psychiatrie, ainsi que le module auxiliaire consacré aux traitements médicamenteux en psychiatrie.

1 Une ontologie pour la psychiatrie

1.1 La psychiatrie en tant que spécialité médicale

Dans le premier chapitre, intitulé « Comment parlons-nous de la santé mentale? », de son ouvrage *Mental Health Ontologies*, Hastings (2020) rappelle quelques questions lancinantes dans le débat public sur la santé mentale. Les troubles mentaux sont-ils *réels*? Les troubles mentaux sont-ils *des maladies*? Les troubles mentaux sont-ils *dans le cerveau*? Les deux premières questions sont liées : l’implicite est ici un questionnement sur le statut des troubles mentaux par rapport aux maladies physiques. La troisième prolonge d’une certaine manière les deux premières, puisqu’elle met en cause l’organicité des troubles mentaux, mais est susceptible d’engendrer plus de division, parmi les psychiatres pris dans leur ensemble, que les deux premières.

Le « problème corps-esprit » (*mind-body problem*) est un questionnement philosophique à l’histoire longue concernant la relation entre l’esprit et le corps (et particulièrement le cerveau

– on parle aussi de *mind-brain problem*). Si l'on peut raisonnablement faire l'hypothèse que peu de psychiatres affirmeraient un dualisme radical (thèse que la régulation de l'humeur par l'absorption de carbonate de lithium rend difficile à tenir), certains pourraient sans doute acquiescer à la fameuse métaphore de Bergson :

« Que nous dit en effet l'expérience? Elle nous montre que la vie de l'âme ou, si vous aimez mieux, la vie de la conscience, est liée à la vie du corps, qu'il y a solidarité entre elles, rien de plus. Mais ce point n'a jamais été contesté par personne, et il y a loin de là à soutenir que le cérébral est l'équivalent du mental, qu'on pourrait lire dans un cerveau tout ce qui se passe dans la conscience correspondante. Un vêtement est solidaire du clou auquel il est accroché; il tombe si l'on arrache le clou; il oscille si le clou remue; il se troue, il se déchire si la tête du clou est trop pointue; il ne s'ensuit pas que chaque détail du clou corresponde à un détail du vêtement, ni que le clou soit l'équivalent du vêtement; encore moins s'ensuit-il que le clou et le vêtement soient la même chose. Ainsi, la conscience est incontestablement accrochée à un cerveau mais il ne résulte nullement de là que le cerveau dessine tout le détail de la conscience, ni que la conscience soit une fonction du cerveau. Tout ce que l'observation, l'expérience, et par conséquent la science nous permettent d'affirmer, c'est l'existence d'une certaine *relation* entre le cerveau et la conscience. » (Bergson, 2017),

ou à l'une de ses formes, plus moderne et fonctionnaliste, rappelée par Hastings (2020), du logiciel (*software*) pour l'esprit et du matériel (*hardware*) pour le cerveau. L'une des manières de rendre compte du différentiel entre états mentaux et états physiques est de mettre en avant la dimension subjective et qualitative des états mentaux – les *qualia* de l'article de Nagel (1974) *What Is It Like to Be a Bat?* – que peut reprendre à son compte un psychiatre défendant une approche phénoménologique de la psychopathologie (Bin, 2008).

Cette considération sur les rapports entre corps et esprit ouvre une première piste pour caractériser la psychiatrie en tant que spécialité médicale. Quelle que soit l'autonomie accordée au psychique par rapport au somatique (ou au *reste* du somatique, au cas où aucune autonomie n'est reconnue), l'interaction entre les deux intéresse la psychiatrie, et au-delà de cette interaction celles de l'individu avec les autres individus, avec la société, avec sa culture, avec son environnement au sens le plus large. La prise en compte de ces interactions n'est pas éludée par la psychiatrie (et encore moins par l'addictologie) et fait du psychiatre un « marginal sécant »¹, selon l'expression de Jamous *et al.* (1969) citée par Danet (2013). Une forme de cette prise en compte est le modèle biopsychosocial d'Engel (1980), perspective holistique qui vise à dépasser le seul point de vue biomédical et à redéfinir la place de la subjectivité du patient, la relation entre médecin et patient, la maladie elle-même. Cette première piste concerne de façon majeure la psychiatrie de précision, qui s'attache notamment au recueil de données sociales et environnementales.

Une deuxième piste, découlant de la première, tient à la localisation du trouble mental. Lorsque celle-ci est établie, par exemple dans le cas d'une maladie neurodégénérative comme la maladie d'Alzheimer, on parlera plutôt de maladie neurologique. Mais la localisation d'un trouble dépressif majeur semble plus difficilement assignable, tant elle touche l'individu dans son intégralité. Quant à un trouble de la personnalité antisociale, il pourrait être localisé dans l'environnement social de l'individu (ou à la frontière entre l'individu et son environnement social) plutôt que dans ce dernier². Ce problème de localisation est une raison majeure pour laquelle Tabb & Lemoine

1. Marginal dans le système d'action dont il est partie prenante, il est sécant parce qu'en relation avec d'autres systèmes d'action. On peut penser ici au secteur social et au secteur juridique, notamment.

2. C'est d'ailleurs une caractéristique marquante de la psychiatrie que la souffrance à l'origine de la consultation

(2021) soutiennent que l'approche que constitue la médecine de précision n'est pas adaptée à la psychiatrie. Les auteurs commencent par distinguer les biomarqueurs statistiques – dont la présence est corrélée à un état pathologique mais dont la contribution au processus pathologique est inconnue – des biomarqueurs mécaniques (*mechanistic*) indiquant un mécanisme par l'intermédiaire duquel on peut intervenir sur le système considéré. Selon eux, la conversion – en oncologie – de biomarqueurs d'abord statistiques en biomarqueurs mécaniques, quand elle a lieu, est rendue possible par le caractère localisé de ce qui est pertinent pour la maladie (la tumeur, les cellules), ce qui réduit l'espace de recherche parmi les informations à prendre en compte. La psychiatrie aurait plus de mal à faire de même, faute de pouvoir déterminer l'« emplacement exact du dysfonctionnement mental dans l'organisme ».

Si le débat n'est pas terminé concernant la localisation des troubles psychiatriques, c'est aussi en raison des difficultés que rencontre la psychiatrie à mettre en évidence leur physiopathologie, ainsi que leur étiologie. Ces difficultés ressurgissent – c'est la troisième piste que nous voulions évoquer – sur la manière d'évaluer le seuil entre le normal et la pathologique. « Il existe des domaines de la connaissance médicale où les fondements objectifs de nombreuses valeurs seuils ne sont pas connus ou pas encore totalement compris - en particulier en psychopathologie », remarque Hucklenbroich (2014), « néanmoins, l'existence de valeurs seuils est fortement étayée par l'expérience clinique. Par conséquent, les cliniciens s'appuient généralement sur des scores et des indices qui constituent un substitut à cette connaissance objective ». Ceci permet d'illustrer l'importance que vont avoir les outils d'évaluation en psychiatrie et permet aussi de comprendre l'intérêt que présente pour la psychiatrie de précision le développement des techniques d'imagerie cérébrale ou de séquençage génétique, par exemple. Toutefois, il se pourrait que le problème ne soit ici que reporté : pour reprendre un exemple de Tabb & Lemoine (2021), malgré « la découverte d'une mutation complexe de C4 qui augmente le risque de schizophrénie », « la contribution du dysfonctionnement génétique à la maladie n'est que faiblement établie, et la voie mécanique qui mène de la pathologie génétique aux signes et symptômes de la maladie est encore inconnue ».

La connaissance encore partielle de l'étiologie et de la physiopathologie des troubles psychiatriques nous amène à la quatrième piste, qui concerne les classifications en psychiatrie. Selon Hucklenbroich (2014), la psychiatrie fait exception parmi les spécialités médicales en ceci qu'elle ne dispose pas d'une nosologie mature et complète, « la plupart des tableaux cliniques, des syndromes et des troubles psychiatriques n'[ayant] pas d'explication physiopathologique ou anatomopathologique ». La classification dominante reste celle proposée par le *Diagnostic and Statistic Manual* (DSM) de l'*American Psychiatric Association* (dont la cinquième version, le DSM-5, a été publiée en 2013 et traduite en français en 2015). Toutefois, la littérature reste abondante sur le sujet et plusieurs questions restent ouvertes sur les classifications en psychiatrie :

- une classification doit se fonder sur une théorie psychopathologique, sur les symptômes, sur la biologie? à quel point est-elle dépendante de l'âge? de la société et de la culture?³
- une classification doit-elle adopter une approche catégorique ou dimensionnelle? (Demaizeux, 2008)
- les troubles psychiatriques sont-ils des « espèces naturelles » (*natural kinds*)?⁴ (Zachar, 2006; Cooper, 2013)

puisse être celle de l'entourage. Ce qui rejaillit sur les modes d'hospitalisation en psychiatrie, avec l'« admission en soins psychiatriques à la demande du représentant de l'État » (ASPDRE) en cas d'atteinte grave à l'ordre public.

3. B. Falissard - ESCAP, Copenhagen, 29/06/2023 – https://www.youtube.com/watch?v=_ySezdZYotk&t=1s – consulté le 15 août 2023.

4. La réponse à cette question conditionne l'approche à adopter pour élaborer une classification.

— quel est l'effet de la classification sur les personnes classifiées?⁵

Le DSM a subi de nombreuses critiques depuis des décennies, émanant de nombreux acteurs. Il peut toutefois sembler à un observateur extérieur que ces critiques se focalisent désormais autour de la question de la *validité* des diagnostics établis selon le DSM, quand l'effort historique du DSM a été d'améliorer la *fiabilité* des diagnostics en psychiatrie⁶, ce dont témoigne par exemple l'annexe F du DSM-III (troisième révision du DSM, publiée en 1980), consacrée à la fiabilité inter-juges. Ces critiques se sont accompagnées, au cours des dernières années, de propositions de classifications concurrentes, comme celles présentées par Micoulaud-Franchi *et al.* (2019) et Dumas *et al.* (2019).

1.2 Les ontologies de la santé mentale

Dans un article récent sur les défis et les prochaines étapes de la santé mentale numérique, Smith *et al.* (2023) remarquent que « nous manquons d'ontologies efficaces des maladies mentales », alors que « des ontologies intégratives et interdisciplinaires sont [...] essentielles pour améliorer l'accessibilité et l'interopérabilité des preuves et réduire la fragmentation, tout en favorisant l'automatisation et la réduction des coûts » et qu'« un cadre ontologique partagé pour rendre compte des résultats dans le domaine de la santé mentale numérique permettrait [...] de se concentrer sur les caractéristiques essentielles à mesurer et d'utiliser un langage commun pour rendre compte des études ». Ce souhait peut toutefois rencontrer des interrogations de la part des experts du domaine des sciences comportementales, comme le montrent les questions relevées par Michie *et al.* (2023) lors de réunions de diffusion des ontologies que les auteurs promeuvent (sur le fonctionnement mental, les addictions, les interventions de changement de comportement, présentées plus bas) :

- les ontologies peuvent-elles représenter différentes perspectives et nuances d'idées?
- peut-on définir des concepts tels que ceux utilisés dans les sciences du comportement?
- les relations complexes impliquant des interactions dynamiques entre plusieurs entités peuvent-elles être appréhendées par des relations dyadiques?
- les ontologies peuvent-elles capturer des informations à des niveaux multiples, par exemple l'individu, le groupe, la société?
- les ontologies étouffent-elles la créativité et la diversité des points de vue?
- Les ontologies sont-elles nécessairement réductionnistes?

Après avoir répondu à ces questions (ainsi qu'à des questions plus spécifiques à l'implémentation des ontologies), les auteurs concluent sur l'intérêt des ontologies comme « système de représentation dans les sciences du comportement », notamment en termes de clarté et de cohérence (par rapport à des pratiques de communication scientifique non structurées) mais notent parmi les défis à relever la formation de spécialistes du comportement à l'utilisation des ontologies et la

5. I. Hacking pointe, notamment dans *Rewriting the Soul*, les « effets en retour » (*looping effects*) générés par la définition de types humains (*human kinds*) : « People classified in a certain way tend to conform to or grow into the ways that they are described; but they also evolve in their own ways, so that the classifications and descriptions have to be constantly revised. » (Les personnes classées d'une certaine manière ont tendance à se conformer aux manières dont elles sont décrites ou à se développer selon elles; mais elles évoluent également de leur propre manière, de sorte que les classifications et les descriptions doivent être constamment révisées [traduction libre]) (Hacking, 1998, p. 21).

6. Les notions de fiabilité (*reliability* en anglais) et de validité (*validity* en anglais) sont d'abord des critères de qualité d'une mesure : la fiabilité désigne la cohérence de la mesure, le processus de mesure aboutissant aux mêmes résultats lorsque le phénomène est le même; la validité désigne la capacité de la mesure à refléter ce qu'elle est censée mesurer. Ici, le processus de mesure est le processus diagnostique, et la fiabilité consiste par exemple à obtenir le même diagnostic si la même personne est évaluée par différents cliniciens, tandis que la validité consiste à aboutir à un diagnostic correspondant à une entité pathologique effective.

constitution d'un « groupe d'experts en ontologie » (comme il y en a « dans la conception d'essais randomisés ou dans la réalisation d'analyses qualitatives complexes »).

S'il n'y a pas à notre connaissance d'ontologie de la psychiatrie à proprement parler, nous présentons ci-dessous un certain nombre d'ontologies qui traitent de notre domaine d'intérêt.

La suite d'ontologies sur le fonctionnement mental de Larsen & Hastings (2018) comprend actuellement la *Mental Functioning Ontology* (MF) (Hastings & Schulz, 2012), l'*Emotion Ontology* (MFOEM) (Hastings *et al.*, 2011), la *Mental Disease Ontology* (MFOMD) (Hastings *et al.*, 2012) auxquelles s'est ajoutée plus récemment l'*Ontology for Phenomenological Psychopathology* (OPheP) (Larsen *et al.*, 2022). Ces ontologies participent de l'OBO Foundry et utilisent BFO comme ontologie fondationnelle. MF est l'ontologie de plus haut niveau dans cette suite et une ontologie de niveau intermédiaire pour tout le domaine psychologique. Elle vise à représenter le fonctionnement mental normal, ce qui « inclut, par exemple, des entités telles que la conscience, la perception, la pensée et la croyance, et met l'accent sur la perspective expérientielle en première personne du fonctionnement mental humain » (Larsen & Hastings, 2018). MFOEM spécifie MF dans le domaine des sciences affectives en s'attachant à représenter à la fois l'expérience émotionnelle subjectivement vécue, les changements physiologiques associés, l'évaluation cognitive de l'émotion, les expressions comportementales associées, sans adhérer à une théorie de l'émotion en particulier. MFOMD (dont Ceusters & Smith (2010) avaient posé les bases) suit globalement l'approche du DSM et de la CIM tout en pointant les défauts du cadre de classification du DSM et en restant ouverte à des approches alternatives, et réaffirme l'hypothèse d'un corrélât biologique sous-jacent au dysfonctionnement mental. Notons que le triptyque *disorder–disposition–disease course* d'OGMS est ici en remis en cause : ainsi le *major depressive disorder*, d'abord classé comme disposition (Larsen & Hastings, 2018, fig.3) est classé comme *disease course* dans la version la plus récente de MFOMD. Le projet général, en cours, est de surmonter l'opposition entre recherche et clinique ainsi que les dissensions théoriques, dans une attitude agnostique ouverte aux perspectives concurrentes et offrant la richesse du cadre ontologique pour établir des ponts entre elles.

ONTOPSYCHIA est l'ontologie développée par Richard (2017) lors de son doctorat sous la direction de Jean Charlet. La contribution qui intéresse le plus notre revue consiste en l'élaboration d'un module des facteurs sociaux et environnementaux des maladies psychiatriques à partir d'un corpus de textes en suivant la méthodologie hybride ToReuse2Onto.

BCIO (pour *Behaviour Change Intervention Ontology*) est une ontologie qui vise à offrir « un moyen systématique de décrire et de relier les entités dans le domaine des interventions de changement de comportement » (Michie *et al.*, 2020). L'ontologie est centrée sur la notion de scénario d'intervention, qui a notamment pour partie (en tant que processus) l'intervention et le comportement à l'issue de l'intervention, et qui est évalué par une étude d'évaluation. Cette ontologie se place sous BFO et reprend, outre des classes d'OGMS, IAO et OBI, un nombre important de classes d'autres ontologies de l'OBO Foundry comme l'*Environment Ontology* (ENVO) et l'*Ontology of Medically Related Social Entities* (OMRSE).

AddictO (ou *Addiction Ontology*) est une ontologie à l'attention des « chercheurs, praticiens et décideurs politiques » sur le thème de l'addiction (Hastings *et al.*, 2020) qui entend mettre en évidence l'intérêt de BFO pour apporter des distinctions masquées dans l'usage du langage naturel, par exemple en ce qui concerne l'addiction (disposition et processus), le *craving* (processus et qualité), l'arrêt de consommations de substances (processus ou fin d'un processus), le traitement (processus ou objet). L'ontologie comporte de nombreuses classes dé-

crivant les substances, reprises de l'ontologie *Chemical Entities of Biological Interest* (ChEBI) – ontologie de référence pour les composés chimiques, qui participe de l'OBO Foundry.

HP (pour *Human Phenotype Ontology*)⁷ est une ontologie visant à permettre la mise en relation de maladies héréditaires et de caractères phénotypiques (Robinson *et al.*, 2008). Robinson (2012) voit là un enjeu important pour la médecine de précision : celui du « phénotypage profond » (*deep phenotyping*), en postulant que la différenciation très fine des phénotypes peut permettre de différencier des sous-groupes de patients diagnostiqués avec la même maladie. Comme le rappelle Köhler *et al.* (2020), HP s'est d'abord focalisé sur les maladies rares avant de s'étendre au COVID puis à la santé mentale. Robinson (2012) explique que « dans le contexte médical, le mot “phénotype” est utilisé pour désigner un écart par rapport à la morphologie, à la physiologie ou au comportement normaux » (le terme utilisé dans la version actuelle de HP est « anomalie phénotypique »). Aussi utile que puisse être ce terme, il présente à notre sens une ambiguïté concernant son référent : ce qui est ou ce qui est connu. Dans l'option « réaliste » (où « phénotype » désigne ce qui est), il peut donc poser le même problème que celui signalé dans la section 2.1 du chapitre 3 à propos des signes cliniques à une ontologie de haut niveau qui souhaiterait subsumer HP et qui distingue par exemple les objets, les événements et les propriétés⁸. En outre, le fait que *schizophrenia* (trouble mental caractérisé par une variété de phénotypes, pour reprendre le terme), *self-neglect* (l'incurie, signe comportemental assez élémentaire), *psychosis* (ayant pour sous-classes *psychotic episodes* et *psychotic mentation*), *mania*, *lack of insight*, puissent tous être considérés comme des anomalies phénotypiques (de genre *atypical behavior*) illustre le caractère très extensif du terme et interroge la démarche de phénotypage profond (censée précisément permettre de distinguer les différents phénotypes de la schizophrénie).

COGAT (pour *Cognitive Atlas Ontology*) est la mise en forme ontologique du *Cognitive Atlas*⁹, un projet collaboratif mené par Russell Poldrack visant à développer une base de connaissances des sciences cognitives (Poldrack *et al.*, 2011). Quatre thèmes sont considérés : « concepts », « tâches » (des outils d'évaluation), « phénotypes » (répartis en « troubles », « traits de personnalité » et « comportements »), « collections » (réparties en « théories » et « batteries de tâches »).

2 Élaboration d'ontoDOPSY

L'élaboration d'ontoDOPSY répond à sa double finalité d'intégration des données dans le cadre du projet PsyCARE et d'annotation sémantique de documents de psychiatrie. Les éléments retenus en vue de l'annotation sont les éléments qui caractérisent la psychiatrie en tant que spécialité : les signes cliniques neuropsychiatriques, les troubles psychiatriques, les traitements et principalement les traitements médicamenteux, les outils d'évaluation. Les données à intégrer sont ici pour l'essentiel les résultats de l'usage de ces outils d'évaluation. Enfin, les domaines d'intérêt psychiatrique permettent en premier lieu de caractériser ce qu'évaluent ces outils, mais sont susceptibles d'une utilisation plus large.

7. La *Human Phenotype Ontology* était précédemment abrégée en « HPO ». Nous reprenons l'abréviation de la version la plus récente déposée sur BioPortal.

8. Nous pensons aux ontologies de plus haut niveau de l'OBO Foundry, à laquelle HP participe. C'est une difficulté non encore résolue pour COB (cf. section 1.2.2 du chapitre 3). Quant à OGMS, sa version la plus récente (2021) classe *phenotype* comme une *quality* et distingue *physical sign*, *processual sign* et *qualitative sign* (qui est une *quality* égale-ment).

9. <http://www.cognitiveatlas.org/> – consulté le 15 août 2023.

2.1 Les signes cliniques

Les signes cliniques ont une longue histoire. Hacking (2002) consacre un chapitre de son ouvrage *L'émergence de la probabilité* aux signes, en montrant comment le signe, concept majeur des « basses sciences » comme l'alchimie, la géologie ou la médecine – qui s'appuient sur l'opinion, par opposition aux « hautes sciences » comme l'astronomie, l'optique ou la mécanique qui visent un savoir démontrable – va voir son rôle évoluer au cours du XVII^e siècle. De témoignage lu dans le *Vrai Livre* (de la nature) qui rend une opinion « probable » (au sens d'« approuvée », et en l'occurrence par le meilleur des auteurs : celui de l'univers), le signe se transforme en évidence factuelle¹⁰ dont la fréquence rend un jugement plus ou moins digne d'approbation, « probable » en un sens nouveau. Ginzburg (1980) réinscrit quant à lui les signes, aux côtés des traces et des pistes, dans l'histoire très longue d'un « paradigme indiciaire » faisant se croiser un critique d'art cherchant les détails permettant de distinguer une toile originale d'une copie, un chasseur du néolithique observant des empreintes dans la boue et des branches cassées, Sherlock Holmes interprétant des cendres de cigarette, Freud, Cabanis (entre autres médecins) et, bien entendu, l'historien. Tous cherchent, à partir des effets, à remonter à la cause qui les a engendrés : ce que depuis C. S. Peirce on appelle le raisonnement abductif.

Dans le domaine psychiatrique, pour lequel l'examen clinique reste plus déterminant encore que pour d'autres spécialités médicales dans l'établissement du diagnostic, la sémiologie (c'est-à-dire la partie de la médecine qui étudie les signes relevés par le médecin) va donc être capitale. D'autre part, du point de vue ontologique qui est le nôtre, les signes cliniques observés par les cliniciens (ou rapportés par les patients) sont plus à même de permettre un consensus large à partir duquel proposer une modélisation.

Afin d'élaborer la branche de l'ontologie consacrée aux signes cliniques en psychiatrie, nous nous sommes appuyés pour l'essentiel, d'un point de vue théorique, sur le *Référentiel de psychiatrie et addictologie* (Collège national des universitaires en psychiatrie *et al.*, 2021), sur l'article *Le jardinier et le botaniste* (Micoulaud-Franchi *et al.*, 2015) et sur l'examen psychiatrique standardisé¹¹ proposé par l'Association pour l'Enseignement de la sémiologie psychiatrique (AESP). D'autre part, nous avons pu réutiliser le corpus de près de 9 000 comptes rendus d'hospitalisation semi-standardisés et anonymisés issus du GHU Paris psychiatrie et neurosciences, utilisé par Richard (2017).

Le travail sur ce corpus s'est déroulé en plusieurs phases. Dans une première phase, les documents du corpus ont été transcrits en un format textuel manipulable depuis Python. Ces documents étant semi-standardisés, nous avons repéré le nom des sections (la figure 4.1 présente la séquence canonique des sections dans des comptes rendus du corpus – les rubriques en gris sont des ajouts de notre part).

La section concernant l'examen à l'entrée étant la plus riche en signes cliniques, qui plus est exprimés selon un style télégraphique, avec souvent un signe clinique par ligne, c'est sur cette section que s'est porté notre intérêt (la figure 4.2 donne un exemple d'examen à l'entrée).

Une recherche floue (*fuzzy string matching*) à partir de l'expression canonique « EXAMEN A L'ENTREE » (paquetage Python *fuzzywuzzy*, `token_sort_ratio`, seuil de 70 déterminé par essais successifs en privilégiant la précision) a permis d'identifier différentes formes équivalentes. Un sous-

10. C'est le terme choisi par le traducteur pour traduire l'anglais *evidence*, dont il rappelle la polysémie : « preuve » chez certains auteurs (avec ce que ce terme implique de démonstratif, de certain), le terme est plus souvent proche de la « pièce à conviction » (pour employer un vocabulaire policier ou juridique), de l'élément d'évidence, de l'indice qui appuie un certain point de vue.

11. <http://www.asso-aesp.fr/wp-content/uploads/2023/01/examen-psychiatrique-standardise.pdf> – consulté le 15 août 2023.

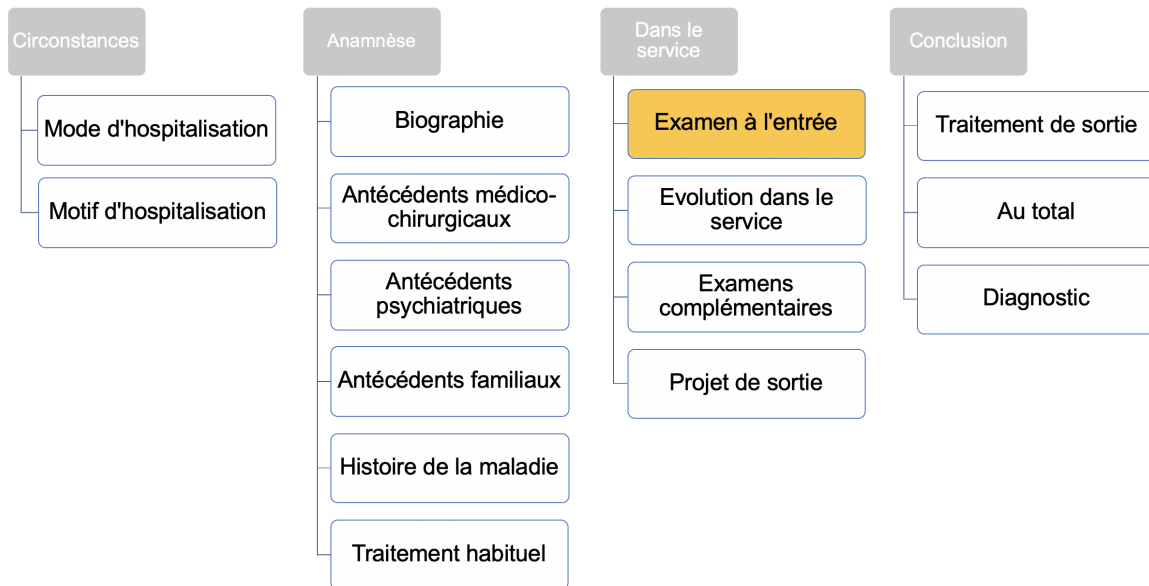


FIGURE 4.1 – Séquence canonique des sections dans un compte rendu d'hospitalisation en psychiatrie.

- Contact très immature
- Symptomatologie dépressive : tristesse, souffrance morale, ralentissement psychomoteur, aboulie, anhédonie, apragmatisme, insomnie, idées noires sans idée suicidaire actuellement.
- Peu d'éléments mélancoliques (pas de culpabilité, pas d'idée de ruine, pas de reconstruction négative du passé, mais sentiment d'incurabilité)
- Aucune critique du geste. Affirme la volonté qu'elle avait de mourir.
- Pas d'idée délirante
- Pas de dissociation
- Ambivalence et passivité dans les soins

FIGURE 4.2 – Exemple d'examen à l'entrée.

corpus des examens à l'entrée à été constitué. Après suppression des doublons, les examens à l'entrée ont été séparés en phrases (paquetage Python NLTK). Après nettoyage (notamment pour supprimer les différents types de « puces » des listes à puces : tirets, points, étoiles), un corpus de plus de 50 000 phrases a été constitué. Une analyse exploratoire a fait apparaître une loi de puissance très marquée : 6,5% des phrases uniques, apparaissant plus d'une fois, représentaient 29% du volume du corpus quand 93,5% des phrases uniques n'apparaissaient qu'une fois. En séparant les phrases selon les virgules, un ensemble de plus de 98 000 candidats-signes est obtenu, pour lequel 9,5% des candidats-signes uniques (au niveau *type*) représentent 45% des candidats-signes (au niveau *token*).

ExtraCandid est un outil d'extraction de candidats-terme développé par Antoine Cadiou et Vincent Jouve lors d'un stage de Master au LIMICS. Il est très proche de l'outil BIOTEX (Lossio-Ventura *et al.*, 2014), et recourt comme ce dernier à des méthodes statistiques et linguistiques pour extraire des termes d'un corpus de textes et en proposer un classement selon leur pertinence. Une spécificité d'ExtraCandid est d'utiliser, pour calculer l'*Inverse Document Frequency* dans une mesure TF-IDF, un corpus de textes généralistes (issus de notices Wikipédia), afin d'augmenter la pertinence des termes du corpus traité. Utilisé sur notre corpus dans son ensemble, ExtraCandid ne

permet pas d'obtenir des résultats satisfaisants. Afin d'améliorer les résultats obtenus, nous avons extrait du Référentiel de psychiatrie et addictologie, le plus détaillé sur les signes cliniques susceptibles d'être rencontrés, un ensemble de termes (ou parfois de racines, « dépressi » permettant de couvrir à la fois « dépression », « dépressif » et « dépressive ») à partir duquel nous avons constitué des sous-corpus de candidats-signes contenant le terme. En appliquant ExtraCandid (paramétré selon la fréquence) sur ces sous-corpus, nous avons pu associer aux termes issus du référentiel les candidats-signes les plus fréquents contenant ce terme.

Une deuxième phase, plus qualitative, a consisté à examiner les candidats-signes qui n'avaient été retenus dans aucun sous-corpus lié à un terme, ainsi que les candidats-signes retenus pour voir s'ils correspondaient au terme qui avait permis de les extraire. Le plus marquant étant la présence dans les comptes rendus de mentions de la normalité, qui n'était pas dans les termes du Référentiel mais étaient des catégories prévues dans l'Examen psychiatrique standardisé, nous avons opéré une fusion des deux. La structure de l'Examen psychiatrique standardisé (nous parlons de la partie « Entretien psychiatrique », ainsi que des Signes généraux de l'« Entretien physique minimal ») nous paraissant plus robuste pour notre contexte, nous l'avons reprise en la complétant au besoin (pour les signes neurovégétatifs) ou en la raffinant avec le répertoire de termes du Référentiel, plus riche. Nous avons également systématisé autant que possible, outre le normal (*eu-*), les dimensions quantitatives (*hypo-*, *hyper-*) et qualitative (*dys-*) associées aux signes cliniques. À l'issue de cette phase, une arborescence de signes cliniques était constituée et à chaque feuille (ou dans de rares cas à des nœuds plus élevés dans l'arborescence) étaient associées des expressions langagières fréquemment rencontrées dans le corpus de comptes rendus.

La troisième phase a été une phase de validation avec des psychiatres de l'équipe du professeur Krebs. Ce travail a permis d'éliminer des expressions langagières, d'en réattribuer certaines à d'autres feuilles de l'arborescence, et de modifier ponctuellement l'arborescence ou d'« élaguer » lorsque des divisions trop détaillées n'avaient pas de pertinence psychiatrique.

L'arborescence des signes cliniques finalement obtenue est reproduite dans l'annexe E. La même arborescence complétée des expressions langagières associées se trouve dans le wiki TIPSY¹² – dont la vocation première est expliquée dans la section 2.5.

2.2 Les troubles psychiatriques

Le Référentiel de psychiatrie et addictologie (Collège national des universitaires en psychiatrie *et al.*, 2021, p.28) rappelle les critères pour définir un trouble mental :

- des critères sémiologiques spécifiques,
- des critères d'évolution temporelle,
- des critères fonctionnels (détresse psychologique et/ou répercussion fonctionnelle et sociale),
- l'élimination des diagnostics alternatifs (autre trouble mental, pathologie non psychiatrique) par diagnostic différentiel.

Historiquement toutefois, l'avancée de la connaissance médicale psychiatrique peut amener à subdiviser des troubles, à raffiner le discernement entre différents troubles, à organiser la liste des troubles connus (ce sont autant d'enjeux possibles de la psychiatrie de précision). Cet effort nosologique peut aboutir à des nosographies psychiatriques. Nous avons évoqué dans la section 1.1 la situation actuelle des classifications en psychiatrie. Nous présentons plus en détail ci-dessous différentes approches pour discerner et organiser les entités pathologiques propres à la psychiatrie :

12. https://tipsy.limics.fr/index.php/Signe_clinique

le DSM-5, la Classification Internationale des Maladies (CIM) qui comprend un chapitre « Troubles mentaux et du comportement », le *Research Domain Criteria* (RDoC), la *Hierarchical Taxonomy Of Psychopathology* (HiTOP) et la théorie des troubles mentaux comme réseaux du projet *Psychosystems* mené par D. Borsboom.

Le DSM , Manuel Diagnostique et Statistique des troubles mentaux, est publié par l'Association Américaine de Psychiatrie. Sa première édition (1952) porte la marque de la psychobiologie d'A. Meyer, centrée sur la conception des troubles mentaux comme réactions d'adaptation à des événements organiques ou interpersonnels. La troisième édition (1980), sous la direction de R. L. Spitzer, revendique un athéorisme, cherche à établir un vocabulaire commun permettant d'améliorer la fiabilité inter-juges dans les diagnostics et met en place un système d'évaluation multiaxial. Le système multiaxial est systématisé dans la quatrième édition : l'axe I comprend les troubles cliniques (et d'autres situations pouvant faire l'objet d'un examen clinique), à l'exception des troubles de la personnalité et du retard mental qui figurent dans l'axe II, l'axe III enregistre les affections médicales générales pertinentes dans la compréhension ou la prise en charge, l'axe IV enregistre les problèmes psychosociaux (pouvant affecter le diagnostic, le traitement, le pronostic) et l'axe V évalue le fonctionnement global du sujet (au moyen de l'échelle *Global Assessment Functioning*). La classification du DSM correspond aux troubles mentaux mentionnés dans les axes I et II.

Le DSM-5 est la cinquième édition du DSM (2013). Il adopte une approche plus dimensionnelle des troubles de la personnalité, opère un regroupement des troubles du spectre autistique et un regroupement des troubles de l'usage de substance, simplifie la classification des troubles bipolaires et dépressifs.

La CIM est l'une des principales classifications de l'Organisation Mondiale de la Santé (OMS). Elle trouve son origine dans le rapport sur la classification des causes de décès présenté en 1893 à l'Institut international de Statistique, à Chicago, par le statisticien J. Bertillon et son comité. Cette classification sera révisée à chaque décennie environ jusqu'en 1939. Après-guerre, son évolution est confiée à l'OMS et la sixième révision, parue en 1949, intègre la morbidité et devient la CIM. En demandant aux États membres d'utiliser pour établir leurs statistiques de morbidité et de mortalité la dernière révision de la CIM, l'OMS peut établir des comparaisons à travers les pays et à travers le temps des données de morbidité et de mortalité.

La CIM-10 a été adoptée en 1990 pour rentrer en application en 1993. Elle inaugure un système de codification alphanumérique à une lettre suivie de deux à quatre chiffres. Les codes sont répartis en 22 chapitres, selon un principe essentiellement anatomique. Le chapitre V est consacré aux Troubles mentaux et du comportement et correspond au bloc de codes F00-F99. La classification suit à peu près le DSM mais son objectif de codification exhaustive des situations qui peuvent être rencontrées amène à une classification beaucoup plus extensive et systématisée. D'un pur point de vue classificatoire, ce souci d'exhaustivité oblige au recours problématique à des classes qui ne sont ni explicites, ni précises et dont l'extension est susceptible de changer au cours du temps (par exemple « Autres formes de schizophrénie » – F20.8). La CIM-10 n'est plus mise à jour par l'OMS depuis 2019¹³.

La CIM-11 a été adoptée en 2019 pour rentrer en application en 2022. Outre l'ajout de nouveaux chapitres, elle est fondée sur une représentation ontologique. De cette représentation peuvent être extraites des représentations simplifiées appelées « linéarisations ».

13. Des mises à jours postérieures à cette date ont pu exister au niveau national, par exemple en France dans la période du (et en lien avec le) COVID.

La CIM-11 rend possible la post-coordination des concepts (par exemple pour exprimer l'étiologie ou la latéralité) selon 40 axes. Le système de codification alphanumérique comporte désormais un chiffre (indiquant le chapitre) suivi d'une lettre et de deux ou trois chiffres. Le chapitre consacré aux Troubles mentaux, comportementaux ou neurodéveloppementaux est le chapitre 6.

Le projet RDoC a été initié par le *National Institute of Mental Health* états-unien et se fonde sur une remise en cause des systèmes diagnostiques en vigueur basés sur les signes et les symptômes, dont « les définitions actuelles ne reflètent pas de manière adéquate les systèmes neurobiologiques et comportementaux pertinents, ce qui entrave non seulement la recherche sur l'étiologie et la physiopathologie, mais aussi le développement de nouveaux traitements » (Cuthbert & Insel, 2013). Afin de pouvoir réviser ces classifications, RDoC promeut une approche dimensionnelle, fondée sur la neurobiologie et le comportement observable, intégrative, qui a en vue la psychiatrie de précision. Le projet s'articule autour de la définition d'une matrice¹⁴ dont les rangs sont des construits théoriques (*constructs*) hiérarchisés par grands domaines (*Negative Valence Systems, Positive Valence Systems, Cognitive Systems, Social Processes, Arousal and Regulatory Systems, Sensorimotor Systems*) et dont les colonnes sont les niveaux d'analyse de ces construits (*Genes, Molecules, Cells, Circuits, Physiology, Behaviors, Self-Reports, Paradigms*). Les gènes renvoient à l'étude des aspects génomiques liés au construit¹⁵, les circuits aux circuits cérébraux impliqués dans le construit, la physiologie aux réponses physiologiques, les *Self-Reports* aux questionnaires et les *Paradigms* aux tests, aux tâches, mais aussi à l'imagerie cérébrale. Au fur et à mesure des études transdiagnostiques fondées sur les construits, les cases de la matrice se remplissent.

La HiTOP émane d'un consortium de nosologues psychiatriques mené par la *Stony Brook University* (New York – USA) cherchant à « élaborer une classification dimensionnelle consensuelle qui soit plus informative sur le plan clinique que les systèmes de diagnostic traditionnels (DSM et CIM) » (Kotov *et al.*, 2018). HiTOP se présente comme une hiérarchie qui remonte de groupes de signes et de symptômes étroitement liés ou de traits inadaptés (au plus bas niveau) à des syndromes (qui correspondent le plus aux catégories traditionnelles du DSM et de la CIM) puis à des sous-facteurs, agrégats de syndromes fortement liés (le sous-facteur « peur » comprend phobie sociale, phobie spécifique, agoraphobie et trouble de la personnalité évitante, par exemple), jusqu'à des groupes de sous-facteurs appelés « spectres » (*Somatiform; Internalizing; Thought Disorder; Externalizing, Disinhibited; Externalizing, Antagonistic; Detachment*). Notons qu'un groupe de signes peut être lié à un ou plusieurs syndromes, que tout syndrome n'est pas nécessairement lié à un sous-facteur mais peut être directement lié à un spectre, qu'un sous-facteur peut être lié à plusieurs spectres (par exemple *Antisocial Behavior* à *Externalizing, Disinhibited* et à *Externalizing, Antagonistic*). HiTOP est également dimensionnelle : il s'agit d'établir un profil psychopathologique du patient selon la sévérité repérée pour les dimensions pertinentes, aux différents niveaux.

14. <https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/rdoc-matrix> – consulté le 15 août 2023.

15. Une note précise toutefois : « Depuis mai 2017, nous avons supprimé les références à des gènes spécifiques dans la colonne “Genes” de la matrice RDoC. Nous reconnaissons la pertinence évidente de l'étude des aspects génomiques des constructions et domaines RDoC. Toutefois, l'état actuel du domaine met l'accent sur la nécessité de disposer de preuves solides d'association, résultant généralement d'études d'association à l'échelle du génome suffisamment puissantes, par opposition aux approches fondées sur les gènes candidats. Au fur et à mesure que nous réévaluerons les informations à inclure dans cette rubrique, la matrice RDoC sera mise à jour en conséquence ».

HiTOP évoque plus une cartographie qu'une classification¹⁶ : les spectres, au plus haut niveau, correspondent aux groupements de symptômes présentés au plus bas niveau ; la référence aux syndromes est un rappel des classifications traditionnelles mais n'est pas vraiment opérante. Mis à part deux recouvrements, celui évoqué à propos de l'*Antisocial Behavior* et le recouvrement sur la *mania* des spectres *Internalizing* et *Thought Disorder*, les spectres sont bien séparés. Ce sont les dimensions, subdivisées jusqu'aux symptômes, du profil psychopathologique d'un patient. Cet aspect cartographique permet au clinicien de s'orienter parmi les traitements possibles, comme le montre le cas illustratif présenté par Ruggiero *et al.* (2019), et de coupler par exemple en première intention une intervention large dans un spectre et une intervention étroite dans un autre spectre, selon le profil du patient obtenu par une évaluation quantitative.

Le projet Psychosystems part du constat que des gens souffrent de symptômes qui s'agrègent selon des motifs non-arbitraires en troubles mentaux. Mais contrairement à la médecine somatique, les troubles mentaux ne peuvent être diagnostiqués indépendamment de leurs symptômes ni rapportés à une cause originaire. En revanche, il existe des relations causales entre symptômes (par exemple de l'auto-accusation à l'insomnie, de l'insomnie à la fatigue, de la fatigue aux problèmes de concentration). Il propose donc de considérer les troubles mentaux non pas comme la cause des symptômes mais comme le réseau que forment les symptômes. Un réseau générique peut indiquer des corrélations entre symptômes et un réseau personnel sera plus intéressant à décliner selon une série temporelle (Borsboom & Cramer, 2013). L'activation de certains symptômes par un événement extérieur peut se propager, activant d'autres symptômes, ce qui finit par former un réseau fortement connecté qui reste actif même après disparition de l'événement déclencheur. Ce modèle théorique rend également compte de la comorbidité, dans le cas de troubles mentaux partageant des symptômes (qui sont des « ponts » dans le réseau par lesquels peut se propager l'activation). Il n'est pas lié à un niveau d'explication particulier (biologique, social, environnemental). L'auteur reconnaît toutefois que la théorie proposée peut plus difficilement rendre compte des troubles à développement lent (troubles du spectre de l'autisme, troubles de la personnalité, schizophrénie) (Borsboom, 2017). Notons que cette théorie ne prévoit pas l'absence de symptômes, la santé mentale étant redéfinie comme l'état stable d'un réseau faiblement connecté. D'autre part, les travaux présentés s'appuient sur les catégories du DSM tout en ayant la même largeur de niveaux d'explication que le RDoC. Cependant, cette théorie ne propose pas de nosographie propre, la question restant donc ouverte de savoir si elle est susceptible de permettre d'en fonder une ou d'en abolir l'utilité.

Micoulaud-Franchi *et al.* (2015) reprennent du neurologue J. H. Jackson la métaphore du jardinier et du botaniste, pour caractériser les différentes logiques d'organisation d'une part du DSM et de la CIM, d'autre part de RDoC. Le psychiatre « jardinier » s'attache à définir les troubles de façon pragmatique, selon « leurs similitudes sémiologiques, leur évolution ou pronostic, leurs cibles thérapeutiques » quand le psychiatre « botaniste » s'attache à rechercher les fondements neurobiologiques sur lesquels appuyer une classification taxinomique. L'approche de D. Borsboom, même si elle ne donne pas la même place à la neurobiologie, peut s'apparenter à celle du botaniste : les progrès en matière de traitement restent un horizon souhaitable mais ne sont pas au cœur de la démarche. En revanche, HiTOP vise explicitement le progrès de la clinique. Pour filer la métaphore de Jackson, on parlerait pour les spectres de HiTOP de jardin à l'anglaise, attaché à l'équilibre des volumes, quand les axes du DSM et la symétrie de la CIM incarneraient le jardin à la française.

16. Ou alors une « classification hiérarchique ascendante » obtenue par *clustering*.

Le choix que nous avons opéré a été de constituer la branche consacrée aux troubles psychiatriques à partir du DSM-5 (*Troubles mentaux; Troubles des mouvements et autres effets indésirables induits par un médicament; Autres situations pouvant faire l'objet d'un examen clinique*), du chapitre *Troubles mentaux et du comportement* de la CIM-10 et du chapitre *Troubles mentaux, comportementaux ou neurodéveloppementaux* de la CIM-11. C'est un choix « conservateur » du point de vue psychiatrique, la classification du DSM restant à l'heure actuelle largement utilisée par les psychiatres. C'est un choix pragmatique du point de vue de l'objectif d'annotation sémantique : la CIM est la classification utilisée par le programme de médicalisation des systèmes d'information (PMSI) pour coder les diagnostics, et les codes de la CIM sont donc susceptibles d'apparaître dans des documents de psychiatrie. La sous-branche consacrée à la CIM-10 comprend environ 1 360 classes, celle consacrée à la CIM-11 environ 890 classes et celle consacrée au DSM-5 environ 760 classes (dont environ 550 pour les seuls troubles mentaux).

Le lien entre DSM et CIM est opéré par l'annotation des classes du DSM-5 (qui ne propose pas de codage) au moyen des annotations descriptives `code_CIM9` et `code_CIM10`. Quant au lien entre CIM-10 et CIM-11, il est opéré par l'annotation descriptive `icd10` des classes de la CIM-11.

2.3 Les outils d'évaluation

Lors d'un examen psychiatrique, l'évaluation purement clinique peut être complétée par l'utilisation d'outils d'évaluation comme des échelles, des tests, des questionnaires. Ces outils d'évaluation, ou outils de mesure¹⁷, sont également très utilisés lors des études de recherche qui ne visent pas tant à établir un diagnostic qu'à affiner l'observation de phénomènes d'intérêt et à quantifier les caractéristiques propres aux sujets d'étude.

Ces outils d'évaluation présentent une grande variété : l'évaluation peut concerner un comportement observé ou un comportement rapporté par le patient (on retrouve ici la même distinction qu'entre signe et symptôme), elle peut avoir une visée assez globale ou concerner un phénomène très précis; l'outil peut être spécifiquement conçu pour une tranche d'âge ou être adapté en différentes versions selon les tranches d'âge, avoir une finalité précise (diagnostic, *screening* ou suivi par exemple).

Une autre caractéristique de ces outils pris dans leur ensemble est leur nombre élevé. Le *National Institute of Mental Health* (NIMH), agence fédérale états-unienne pour la recherche sur les troubles mentaux, met à disposition des archives de données (*NIMH Data Archive – NDA*) désidentifiées recueillies lors de projets de recherche. Le site de la NDA¹⁸ propose un moteur de recherche qui permet notamment d'effectuer une recherche sur « plus de 2 000 instruments de collecte de données » (*data structures*) et affiche de fait plus de 4730 résultats en l'absence de tout filtre (même si l'on peut supposer la présence de nombreux doublons). Sur un domaine d'évaluation plus précis, Santor *et al.* (2006) remarque que « depuis 1918, plus de 280 mesures de la gravité de la dépression ont été élaborées et publiées ».

Enfin, si l'on s'intéresse plus en détail aux *data structures* de la NDA, c'est-à-dire en descendant au niveau des items des outils d'évaluation, on peut constater que s'il existe des outils standardisés, de nombreux projets de recherche adaptent des outils, ajoutent ou retranchent des items, en modifient l'énoncé ou la liste de réponses possibles associée.

Ces trois éléments, variété, multitude et faible standardisation, appellent à notre sens un effort de description de ces outils. C'est pourquoi nous avons ajouté un ensemble d'annotations descrip-

17. On parle en anglais d'*assessment tools*.

18. <https://nda.nih.gov/> – consulté le 15 août 2023.

tives (*annotation properties*) consacrées à ces outils¹⁹, reprenant des éléments descriptifs associés par le professeur Krebs et son équipe aux outils intervenant dans les études du projet PsyCARE ou dans des études antérieures :

- le domaine et le sous-domaine évalués par l’outil,
- la matériel supplémentaire éventuellement utilisé dans la passation,
- le nombre d’items de l’outil,
- les normes respectées par l’outil,
- la durée approximative de passation,
- l’agent de passation
- la tranche d’âge visée par l’outil,
- les valeurs seuil du score.

Le lien aux outils référencés dans NDA ou dans PhenX Toolkit²⁰ se fait au moyen de l’annotation `hasDbXRef`, empruntée à l’OBO Foundry. Une liste maniable et psychiatriquement fondée de domaines et de sous-domaines évalués a été établie par l’équipe psychiatrique. Nous proposons dans la section 2.5 une classification beaucoup plus étendue qui répond au départ à cette même problématique de définition des domaines évalués par un outil. Le manque de standardisation que nous avons pointé précédemment concerne aussi le nom des outils. Nous avons – en collaboration avec le partenaire industriel Fealinx et en particulier avec Amel Raboudi – proposé quelques règles pour harmoniser le nommage des outils dans l’ontologie :

- conserver le nom quand il est bien établi,
- suffixer par le nombre d’items en cas de versions concurrentes,
- suffixer par « `_SF` » pour une version courte (*short form*),
- suffixer par « `R` » pour une version révisée,
- utiliser la convention *snake case* pour les outils désignés par un syntagme nominal (par exemple « `Gnosies_digitales` »),
- reprendre la forme étendue d’une des lettres quand la confusion est possible²¹,
- préfixer par « `Q` » pour « questionnaire » (par exemple « `QConso` ») et par « `F` » pour « fiche » (par exemple « `FResume_clinique` »)

Nous ne prétendons pas que ces règles soient largement généralisables au-delà du présent travail mais il est certain que la normalisation des acronymes devient de plus en plus nécessaire à mesure que croît le nombre d’acteurs (notamment non psychiatres) amenés à les manipuler. Enfin, le rôle que peut jouer un outil a été modélisé avec la propriété d’objet `aRolePratiqueClinique` et la classe `RolePratiqueClinique` qui a comme sous-classes les rôles de bilan, de diagnostic, d’examen, de screening, de suivi, de test.

La modélisation que nous avons opérée est centrée sur l’événement de la passation de l’outil par le sujet. Cette passation est une réalisation, propre à un sujet particulier, d’un même protocole réutilisé d’un sujet à l’autre, ce protocole étant classé comme un projet. A cet événement de passation participent le sujet, le cas échéant le professionnel qui mène la passation (psychiatre, psychologue, etc.), ainsi que l’outil d’évaluation et l’instrument de capture du phénomène d’intérêt (il peut s’agir d’une réponse à une question, d’un dessin, etc.). L’outil d’évaluation est considéré

19. Dans la mesure où leur utilité peut concerner d’autres formulaires d’études cliniques, l’annotation `descriptionFormulaire` et ses sous-propriétés ont été intégrées à ontoDOME, mais c’est ici qu’elles trouvent leur justification première.

20. PhenX Toolkit est un catalogue d’outils, notamment en santé mentale (<https://www.phenxtoolkit.org/collections/view/1>), financé par le *National Human Genome Research Institute* (NHGRI) états-unien.

21. BIRCHWOOD_IS pour la *Birchwood Insight Scale*, quand BIS désigne plus couramment la *Barratt Impulsiveness Scale*.

comme un stimulus (ou un ensemble de stimulus) destiné à conditionner le phénomène d'intérêt, c'est-à-dire ce qui est évalué lors de la passation. La passation de l'événement génère un jeu de données, qui est modélisé comme tel. Ce jeu de données enregistre des variables qui qualifient le sujet (ces variables pouvant être les réponses elles-mêmes ou les scores qui en sont inférés).

Cette modélisation suit le patron de conception ontologique qui sera décrit dans la section 2 du chapitre 5. Il y est présenté parce qu'il est central dans la conception des modules spécifiques mais sa formalisation doit beaucoup au travail de modélisation des questionnaires. Notons également que cette modélisation retient de nombreux éléments de Bona *et al.* (2015) – notamment l'inscription dans l'histoire médicale du sujet de la passation de l'outil – ainsi que de Gonçalves *et al.* (2017), qui prolonge la proposition de Bona *et al.* (2015) en y ajoutant la description sémantique des questions, et également de Fabry *et al.* (2020), qui s'attache à la distinction entre réponse, capture de la réponse et stockage de la réponse.

2.4 Les événements vécus

La notion d'« événement vécu » n'est pas nécessairement explicite. L'expression traduit l'anglais *life event* et désigne un événement de la vie d'un individu suffisamment marquant, stressant, important, pour être pertinent dans la compréhension de son état psychologique. Il peut donc être intéressant de repérer et d'annoter sémantiquement la mention de tels événements dans des documents de psychiatrie. Si l'on considère la trilogie décès–divorce–déménagement, on conçoit néanmoins qu'un déménagement puisse avoir des retentissements émotionnels très différents selon les individus et les moments de leur vie, et qu'il ne s'agit pas d'un événement stressant en lui-même, ce qui justifie à nos yeux de parler d'événement vécu et pas d'événement stressant ou traumatisant. Un problème particulier au repérage de ce type d'événement tient toutefois à la capacité de lui attribuer une temporalité. Le décès du conjoint, le divorce, s'ils sont récents, peuvent permettre d'expliquer l'état émotionnel actuel d'un individu, une conduite impulsive de sa part : ils forment le contexte d'un comportement ou d'une attitude. S'ils sont plus anciens, ils peuvent éventuellement permettre d'expliquer une humeur triste, par exemple, mais ils rendent plus généralement compte d'une situation de vie. Cette notion temporelle est parfois présente dans les sources que nous avons utilisées.

La première source est une version étendue à 20 items, utilisée dans le projet PsyCARE, de la *List of Threatening Experiences* de Brugha *et al.* (1985) qui ne compte que 12 items. Les 12 items initiaux avaient été choisis parce qu'ils représentaient la grande majorité des événements de vie engendrant à long terme un sentiment de menace marquée ou modérée.

La deuxième source est l'inventaire d'événements de vie de Paykel *et al.* (1971)²² qui contient environ 60 événements de vie répartis en 10 catégories (travail, enseignement, finances, santé, deuil, déménagement, relation amoureuse et cohabitation, légalité, événement socio-familial, événement conjugal) dont la liste a été soumise à des sujets variés pour qu'ils en évaluent le degré de perturbation qu'ils entraînaient.

La troisième source est la *Social Readjustment Rating Scale* (SRRS) de Holmes & Rahe (1967)²³ qui comprend une liste de 43 événements de vie stressants auxquels sont associées des quantités d'« unités de changement de vie » allant de 11 pour une infraction mineure à la loi jusqu'à 100 pour le décès du conjoint. En faisant la somme des unités de changements de vie correspondant

22. Nous utilisons la traduction de J.M. Bastyns et J.Mendlewicz. Nous avons modifié « épouse », qui traduisait *spouse*, en « conjoint ».

23. Nous proposons une traduction à partir de l'inventaire présent sur le site de l'*American Institute of Stress*.

aux événements stressants vécus par un individu dans l'année écoulée, on obtient un score qui permet une évaluation du risque de maladie que court cet individu.

Ces listes d'événements ont beaucoup d'éléments communs, deux à deux ou toutes les trois ensemble, mais aussi de nombreux éléments qui diffèrent par le niveau de généralité. Dans un premier temps, nous avons procédé à des alignements exacts quand ils étaient possibles (par exemple : licenciement [LTE], renvoi professionnel [Paykel] , licenciement [Holmes&Rahe]) puis à une réorganisation hiérarchique selon ces niveaux de généralité (grossesse [Holmes&Rahe] venant s'intercaler entre Santé [Paykel] et les deux items : grossesse désirée [Paykel] et grossesse indésirée [Paykel]). Les catégories organisant la liste de Paykel ont été conservées et une catégorie d'« événements personnels » (au sens où ils concernent l'individu en propre plutôt que l'individu en interaction avec son milieu social, familial, professionnel) a été ajoutée.

À chacune des classes est associé un ou plusieurs libellés alternatifs annoté pour indiquer la ou les sources dont il provient, et dans le cas où une classe a des libellés de plusieurs sources, le libellé préféré est choisi selon l'ordre de priorité suivant : 1/ LTE 2/ Paykel 3/ Holmes et Rahe.

Certains des libellés associés aux classes de cette branche sont susceptibles d'apparaître tels quels dans des textes (« divorce », « mariage », « licenciement » par exemple), mais la plupart devront être spécifiés (« modification de la fréquence des querelles avec le conjoint », « séparation d'une personne importante », « vous avez perdu ou on vous a volé quelque chose de valeur »), voire associés à une liste d'expressions correspondantes susceptibles d'être effectivement rencontrées dans des documents (« problèmes juridiques complexes et inhabituels », « problème grave avec un ami proche, un voisin ou un apparenté », « changement dans les activités exercées dans le cadre de l'affiliation à une église »). La liste hiérarchisée des événements vécus qui constitue cette branche se trouve dans l'annexe F.

2.5 Les domaines d'intérêt psychiatrique

2.5.1 Motivation

Le travail sur l'intégration des outils d'évaluation dont nous avons parlé précédemment a fait surgir une difficulté dans la manière de classer ces outils. Une approche « naïve » aurait consisté à classer ces outils en fonction de ce qu'ils évaluent. Toutefois, cette approche se heurte au fait que certains outils évaluent une capacité très précise, d'autres des caractéristiques de natures très différentes, d'autres enfin de nombreuses caractéristiques. Pour prendre des exemples, une échelle comme la *Yale-Brown Obsessive-Compulsive Scale* évalue à la fois des pensées obsessionnelles et des comportements compulsifs et une échelle d'intelligence comme la WAIS-IV sera subdivisée en sous-tests consacrés (a) à l'arithmétique, (b) aux codes et symboles, (c) aux cubes et matrices, (d) aux similitudes, au vocabulaire et à l'information.

Dès lors, l'idée que nous avons proposée a été de construire une branche de l'ontologie qui organiserait tous les domaines qui sont susceptibles d'être évalués, parce qu'ils présentent un intérêt pour le psychiatre, en créant autant de branches que nécessaire, en les organisant selon les principes d'une taxonomie, et en formulant les classes de façon à ce qu'elles soient des caractéristiques de la personne évaluée (par exemple en parlant, plutôt que de fonction cognitive, de capacité cognitive).

Structurellement, cette branche de l'ontologie consacrée aux domaines d'intérêt psychiatrique (DIP) est donc une taxonomie (dont la racine est `DomaineDInteretPsychiatrique`), puisqu'elle est organisée en une hiérarchie de subsomption. Mais fonctionnellement, elle est un thésaurus, au sens où elle entend fournir des descripteurs pour les outils d'évaluation. Dans le domaine médical,

le thésaurus de référence est le *Medical Subject Headings* (MeSH). Mais celui-ci ne présente pas le degré de finesse ni de complétude que nous visons ici (on en jugera par exemple à l'arborescence qui se trouve sous le terme « *Memory* »²⁴ ou à l'absence du terme « *insight* ») et est organisé selon un treillis, un terme pouvant avoir plusieurs parents correspondant à des sens non univoques de ce terme (« *Theory of Mind* » est placé à la fois sous « *Psychological Theory* », au même niveau que « *Behaviorism* » ou « *Psychoanalytic Theory* », et sous « *Mental Processes* », au même niveau que « *Perception* » ou « *Executive Function* »). Plus proche que le MeSH, dans le périmètre et dans la géographie, le thésaurus SantéPsy²⁵ est un thésaurus de langue française spécialisé en psychiatrie qui couvre la psychiatrie, la psychologie, la psychanalyse, les soins infirmiers, la gestion administrative et hospitalière, le droit de la santé et l'addiction, et compte plus de 4 000 descripteurs. Un processus d'alignement automatique entre ce thésaurus et la branche des DIP ne retourne qu'une trentaine de correspondances, dont la moitié sont des troubles ou des syndromes. La couverture du thésaurus, plus large mais moins profonde que la branche des DIP, ne permet pas d'enrichir ceux-ci. Le thésaurus reste cependant utile pour étendre d'autres branches de l'ontologie.

Les outils d'évaluation les plus standards utilisés dans le projet PsyCARE ont donc pu être indexés par les classes de cette branche. L'intérêt de la structure taxinomique intervient dans un processus de recherche d'information, puisqu'on peut formuler une requête à différents niveaux de granularité : pour un domaine, voir tous les outils évaluant ce domaine ou ses sous-domaines ; pour un outil, remonter en généralité à partir des domaines les plus précis évalués.

On peut établir un parallèle entre les DIP et la hiérarchie de *constructs* et *subconstructs* de la matrice RDoC – ils ont d'ailleurs de nombreux éléments en commun – et entre les outils d'évaluation et les *self-reports* et les *paradigms*. Toutefois RDoC est plus ici une inspiration qu'un décalque, les domaines d'intérêt psychiatrique étant directement issus de l'étude des outils d'évaluation et leur structuration se fondant principalement sur la sémiologie psychiatrique.

2.5.2 Les domaines retenus

Plus de 400 domaines ont été retenus, en étudiant diverses sources : les articles académiques (séminaux ou ultérieurs) présentant les outils, les sites web commerciaux dans le cas d'outils propriétaires, des présentations des outils. Les classes les plus hautes de cette branche sont les suivantes :

- aspect de la vie quotidienne – concerne l'évaluation de la santé, de la qualité de vie, des fonctions instinctuelles ;
- caractéristique du discours – concerne l'évaluation du contenu (idées, pensées, croyances), du cours (débit, arrêts) ou de la forme (volume, logique) du discours ;
- dimension comportementale – concerne l'évaluation dimensionnelle du comportement social ;
- environnement – concerne l'influence de l'environnement prénatal, familial, social, professionnel ;
- état affectif – concerne les émotions, l'humeur, la motivation ;
- capacité cognitive – concerne les grandes capacités cognitives ;
- histoire – concerne l'histoire développementale, médicale, familiale, personnelle (biographie) ;

24. <https://meshb.nlm.nih.gov/record/ui?ui=D008568>, onglet *MeSH Tree Structures* – consulté le 15 août 2023.

25. https://santepsy.ascodocpsy.org/?lvl=cmspage&pageid=6&id_rubrique=125 – consulté le 15 août 2023.

- indication étiologique – concerne une hypothèse étiologique des comportements observés (pathologie, iatrogénèse médicamenteuse);
- perception – concerne les modifications pathologiques de la perception (notamment les hallucinations et les illusions);
- phase de vie – concerne la phase de vie du patient évalué;
- phase du trouble – concerne la phase du trouble évalué;
- réponse au traitement – concerne les effets recherchés ou indésirables du traitement, la résistance au traitement, ou dans des cas spécifiques l'expérience usager;
- risque – concerne l'évaluation d'un risque (de suicide, de transition psychotique);
- retentissement fonctionnel – concerne le retentissement personnel, professionnel ou social d'un trouble;
- signe comportemental – concerne l'évaluation de comportements en société;
- signe d'insight – concerne l'évaluation de la conscience des symptômes et l'adhésion au traitement;
- syndrome – concerne l'évaluation de syndromes spécifiques;
- tempérament – concerne l'évaluation des grands types de tempéraments (hyperthymique, déprimé, etc.);
- trait de personnalité – concerne l'évaluation des traits de personnalité (notamment les *Big Five*);
- traitement – permet de décrire le traitement évalué par l'outil;
- trouble – permet de décrire le(s) trouble(s) évalué(s).

Dans la mesure où, comme nous l'avons signalé, la sémiologie psychiatrique n'est pas étrangère à la structuration de cette branche, il est légitime de se demander s'il n'y a pas une forte redondance entre la branche des domaines d'intérêt psychiatrique et, notamment, la branche consacrée aux signes cliniques. De fait, des rapprochements sont évidemment possibles. Remarquons toutefois que le recueil des signes cliniques par un thérapeute s'inscrit dans un processus diagnostique et est orienté par le soin. La mise au point d'un outil psychométrique, quand bien même les mesures qu'il permettrait seraient « subjectives » et interviendraient *in fine* dans le choix d'une thérapeutique ou d'une prise en charge (Falissard, 2008), n'obéit pas à la même finalité. L'attention n'est pas nécessairement portée sur les mêmes phénomènes, ou à tout le moins pas de la même manière sur des phénomènes semblables. Le travail sur la hiérarchisation des descripteurs d'outils d'évaluation mérite donc à nos yeux une branche propre.

2.5.3 Le wiki TIPS Y

La matrice [Domaines d'Intérêt Psychiatrique x Outils d'évaluation] que nous avons décrite présente un enjeu de gestion de la connaissance. Elle constitue une base mais cette base est appelée dans un premier temps à être amendée et par la suite à évoluer, à la fois parce que de nouveaux outils apparaissent ou connaissent de nouvelles versions, et parce que ce qui constitue légitimement un domaine d'intérêt psychiatrique est susceptible de changer avec l'évolution de la connaissance psychiatrique. D'autre part, la connaissance synthétisée dans cette matrice ne peut que gagner à être le résultat des efforts d'une communauté, chaque praticien ayant une connaissance plus approfondie d'un sous-ensemble des domaines d'intérêt psychiatrique et une pratique plus quotidienne des outils qui permettent de les évaluer. Enfin, cette connaissance doit pouvoir être partagée pour que l'apport de chacun puisse profiter à l'ensemble de la communauté.

Cette triple exigence de travail collaboratif, évolutif et public, nous a orientés vers un outil de gestion de connaissance simple à mettre en place, connu, et d'un usage accessible à beaucoup,

le wiki. Notre choix s'est porté plus spécifiquement vers le moteur de wiki MediaWiki, qui est utilisé par les projets de la *Wikimedia Foundation* - dont le plus célèbre, WIKIPÉDIA. Les « bonnes propriétés » d'un wiki dans ce contexte sont son caractère centralisé, la possibilité de définir des rôles (simple utilisateur, contributeur pouvant ouvrir ou contribuer à une page de discussion, validateur actant une modification proposée, administrateur validant une demande de compte), la facilité d'utilisation (proche de celle d'un éditeur de textes), l'historicisation des débats dans les pages de discussion.

Nous avons donc mis en place, sur les serveurs du laboratoire LIMICS, le wiki TIPSYP (pour Taxonomie d'Indexation en PSYchiatry), ou dans une version anglophone *Topics of Interest in PSYchiatry*)²⁶ en reflétant dans les pages la structure matricielle : chaque domaine d'intérêt psychiatrique possède une page qui renvoie le cas échéant à ses sous-domaines et qui mentionne tous les outils permettant de l'évaluer (voir figure 4.3) ; chaque outil possède également une page mentionnant la hiérarchie des domaines qu'il évalue (voir figure 4.4). Le choix d'associer à chaque outil non seulement les domaines les plus précis qu'il évalue mais aussi les surdomaines de ces domaines est rendu possible par la structure taxonomique. Il est bien entendu envisageable d'enrichir les différentes pages, notamment les pages d'outils pour reprendre les différentes annotations descriptives associées aux outils d'évaluation dans la branche de l'ontologie qui leur est consacrée. De plus, il est également envisageable d'étendre au-delà des seuls outils d'évaluation l'usage d'un tel wiki : ainsi, les modèles animaux en psychiatrie et en addictologie sont censés permettre de modéliser des domaines d'intérêt psychiatrique, et il pourrait être utile de les décrire par les domaines qu'ils modélisent.

The screenshot shows the TIPSYP wiki page for 'Trouble psychotique'. The page layout includes a sidebar on the left with navigation links such as 'Accueil', 'Modifications récentes', 'Page au hasard', 'Aide concernant MediaWiki', 'Outils', 'Pages liées', 'Suivi des pages liées', 'Pages spéciales', 'Version imprimable', 'Lien permanent', and 'Informations sur la page'. The main content area has a search bar at the top right with the text 'Rechercher dans Tipsy'. Below the search bar, there are tabs for 'Page', 'Discussion', 'Lire', 'Voir le texte source', 'Voir l'historique', and a search input field. The main heading is 'Trouble psychotique'. Below the heading, there is a link 'Retour à la page Trouble'. The section 'Outils d'évaluation' contains a list of tools:

- TLC - Troubles du Langage et de la Communication
- SPQ_BR - Questionnaire de personnalité schizotypique (abrégé-32)
- SPQ_B - Questionnaire de personnalité schizotypique (abrégé-22)
- SPQ - Questionnaire de personnalité schizotypique
- SPI_A - Echelle de détection des symptômes de base de schizophrénie (SPI-A)
- SAPS - Echelle d'appréciation des symptômes négatifs
- SANS - Echelle d'appréciation des symptômes négatifs
- PQ16 - Questionnaire des prodromes (16 items)
- HCS - Echelle d'évaluation des hallucinations
- CGL_SCH - Impression clinique globale-version schizophrénie
- CDSS - Echelle de Dépression de Calgary
- CDS - Echelle de Dépersonnalisation de Cambridge
- CAPE42 - CAPE-42 (issu du PDI-21 et PDI-40)
- BPRS18 - Echelle abrégée d'appréciation psychiatrique (18 items)
- BABS - Echelle des croyances de Brown
- AHRS - Echelle d'évaluation des hallucinations auditives
- BPRS_24 - Echelle abrégée d'appréciation psychiatrique BPRS 24
- PANSS - Echelle Des Symptômes Positifs Et Négatifs
- SNS - Echelle Autoévaluation des Symptômes Négatifs
- SOS - Echelle d'évaluation de l'âge début
- CAARMS - Comprehensive Assessment Of At Risk Mental States

The 'Sous-catégories' section contains a list of sub-categories:

- Schizophrénie
- Trouble schizo-affectif
- Trouble schizotypique
- Trouble psychotique aigu transitoire
- Trouble délirant persistant

FIGURE 4.3 – TIPSYP : page d'un domaine d'intérêt psychiatrique.

26. Le wiki TIPSYP est consultable à l'adresse <https://tipsy.limics.fr/index.php/Accueil>.

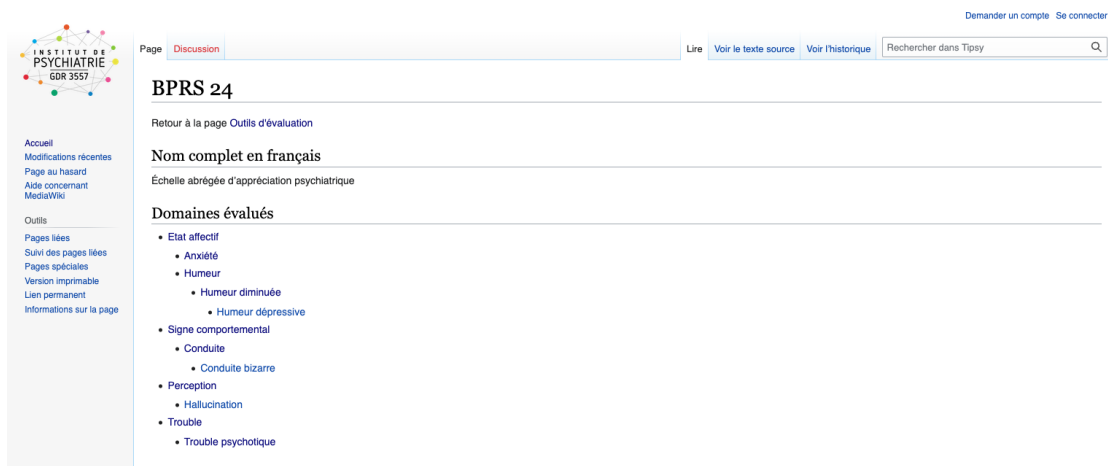


FIGURE 4.4 – TIPSy : page d'un outil d'évaluation.

Nous considérons que le wiki TIPSy pourrait constituer un commun numérique²⁷ de la connaissance, s'apparentant plutôt à un bien public dans la classification des biens de Ostrom & Ostrom (1977) (*non rival* au sens où la consommation du bien par l'un n'empêche pas la consommation par l'autre, et *non excluable* au sens où il est difficile d'empêcher qu'il soit consommé par l'autre), mais relevant néanmoins d'un commun au sens où la connaissance psychiatrique qu'il contiendrait constituerait une ressource produite et entretenue par une communauté de psychiatres. L'existence de divergences théoriques parfois fortes entre membres de la communauté française de psychiatres pourrait apparaître comme une entrave à l'extension du périmètre des contributeurs potentiels, mais il est toujours possible d'ajouter aux domaines d'intérêt psychiatrique de nouvelles branches ou sous-branches reflétant des approches théoriques différentes (phénoménologique ou psychanalytique par exemple). En l'état, la coexistence des trois domaines « dimension comportementale », « tempérament » et « trait de personnalité » reflète déjà une forme de coexistence de points de vue différents sur une caractéristique d'intérêt psychiatrique d'un patient.

Notre usage du conditionnel indique que le potentiel de ce wiki n'est pas complètement réalisé. Il a néanmoins permis des interactions avec des psychiatres et neuropsychologues du GHU Paris psychiatrie & neurosciences, aboutissant à un certain nombre de remaniements, et il a trouvé une utilité supplémentaire dans l'élaboration de la branche consacrée aux signes cliniques.

2.5.4 Lien entre ontologie et wiki

Afin de refléter dans l'ontologie les modifications apportées dans le wiki, ou de refléter dans le wiki des modifications du modèle ontologique, nous avons utilisé la librairie Python Pywikibot²⁸ et développé des scripts *ad hoc* permettant de peupler l'ontologie en extrayant les outils et leurs liens à des domaines ou de créer ou modifier des pages à partir d'un tableur contenant la matrice susmentionnée.

27. <https://communs.societenumerique.gouv.fr/> – consulté le 15 août 2023.

28. <https://pypi.org/project/pywikibot/>

2.6 ontoMEDPSY : un module pour les traitements médicamenteux

Les traitements médicamenteux se trouvent dans un module propre (ontoMEDPSY), mais l'élaboration de ce module s'est faite conjointement avec celui de la psychiatrie. Nous avons pu très largement nous appuyer sur le travail de modélisation des médicaments effectué par Grosjean *et al.* (2021) dans le cadre du projet ANR PSYHAMM²⁹, consacré à l'étude de l'usage des médicaments psychiatriques hors autorisation de mise sur le marché.

2.6.1 Trois vues sur le médicament : substance active, nom de spécialité et médicament virtuel

L'objectif du module ontoMEDPSY est l'annotation sémantique de documents psychiatriques et l'exploration ultérieure de ces documents. Il comporte trois branches qui correspondent à trois approches des médicaments : par la substance active qu'ils contiennent, par le nom de spécialité qui les caractérise, par le médicament virtuel auquel ils se rapportent.

Une branche reprend le système de classification anatomique, thérapeutique et chimique (*Anatomical Therapeutic Chemical Classification System* – ATC), géré par l'Organisation Mondiale de la Santé, dans lequel chaque médicament reçoit un code composé de 7 chiffres et lettres (selon l'ordre LCCLCC) où la première lettre désigne le groupe anatomique ou pharmacologique, les deux chiffres suivants le sous-groupe thérapeutique ou pharmacologique, les deux lettres suivantes le sous-groupe chimique, pharmacologique ou thérapeutique, et les deux dernières lettres la substance chimique. C'est donc la substance active du médicament qui est ici identifiée et inscrite dans un hiérarchie de classes pharmacologiques.

Une deuxième branche contient une liste à plat, non hiérarchisée, de racines pharmacologiques, c'est-à-dire de noms de spécialités, issue de la Base de données publique des médicaments produite par le Ministère de la Santé et des Solidarités. Le libellé (annotations `skos:prefLabel` et `rdfs:label`) est le nom de spécialité et l'annotation `code_ATC` permet de relier le médicament virtuel à la substance active qu'il contient.

Une troisième branche contient une liste à plat de médicaments virtuels, un médicament virtuel consistant en l'association d'un nom de spécialité et d'un dosage correspondant à une ou plusieurs formes commerciales. Les libellés alternatifs (annotation `skos:altLabel`) du médicament virtuel sont les différentes formes commerciales, et l'annotation `code_ATC` permet là aussi de relier le médicament virtuel à la substance active qu'il contient.

La figure 4.5 illustre les liens entre les trois branches. Les termes indiqués en rouge sont ceux qui sont susceptibles d'apparaître dans des comptes rendus d'hospitalisation. Par exemple, dans la section « Traitement de sortie », on trouvera plus généralement « Teralithe LP 400 » ou « Camcolit 400 », accompagné d'une posologie. Ces deux médicaments ont en commun d'être deux formes commerciales du médicament virtuel « lithium carbonate 400mg » et sont dans l'ontologie deux libellés alternatifs de ce dernier. Dans des parties plus anamnestiques, on pourra rencontrer des expressions comme « reprise du Teralithe » sans mention de dosage.

Les possibilités d'exploration d'un corpus de documents annotés sont donc supportées par la branche hiérarchisée selon la classification ATC. C'est à partir de celle-ci qu'on pourra rechercher les documents mentionnant, par exemple, le lithium ou un médicament ayant pour `code_ATC` le code du lithium, et de la même manière – en remontant dans les classes pharmacologiques – les documents mentionnant un antipsychotique, un psycholeptique, un médicament agissant sur le système nerveux.

29. <https://www.cismef.org/psyhamm/> – consulté le 15 août 2023.

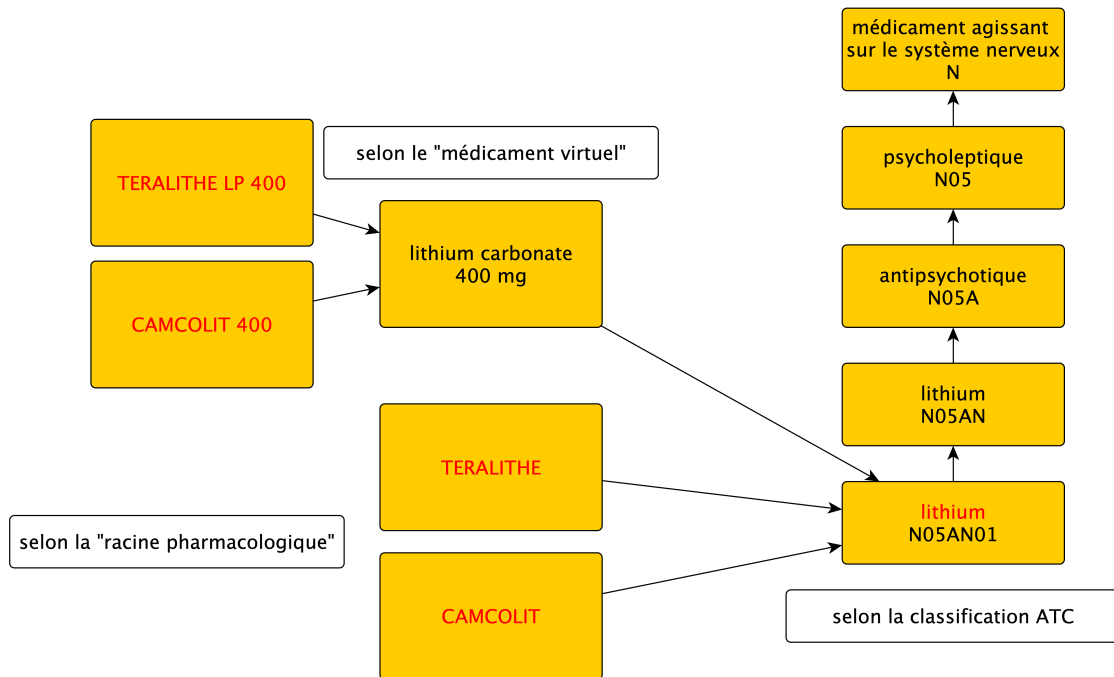


FIGURE 4.5 – Les trois branches de l'ontologie ontoMEDPSY.

Nous avons pris notre exemple dans le groupe N de la classification ATC, qui comporte quatorze groupes de premier niveau. Comme nous l'avons vu, il s'agit des médicaments agissant sur le système nerveux. C'est *a minima* le groupe qui doit figurer dans une ontologie sur les médicaments utilisés en psychiatrie, mais on peut se demander si le périmètre de l'ontologie doit se limiter au seul groupe N. Ce groupe compte plus de 700 classes (en comptant les classes intermédiaires et pas seulement celles qui représentent une substance) quand la classification ATC dans son ensemble en compte plus de 6 000. Deux documents classant les prescriptions médicamenteuses en psychiatrie selon la classification ATC (Capelle *et al.*, 2016; Weibel, 2012) convergent sur le fait que les quatre groupes A³⁰, B³¹, C³² et N couvrent plus de 90% des médicaments des patients de psychiatrie. En nous limitant à ce périmètre, et en ajoutant les noms de spécialité et les médicaments virtuels du même périmètre, nous obtenons environ 5 400 classes. Au-delà de ce critère de couverture purement quantitatif, ce périmètre pourrait être ponctuellement étendu pour des raisons de pertinence psychiatrique. Ainsi, les contraceptifs ne figurent pas dans l'état actuel de l'ontologie, alors que de nombreuses études pointent un lien entre contraception hormonale et dépression.

Notons pour finir qu'un travail d'élaboration d'un Référentiel unique d'interopérabilité du médicament mené par l'Agence du Numérique en Santé (et mis à disposition sur son Serveur Multi-Terminologies³³) est en cours. Ce référentiel « représente de manière univoque le médicament depuis les substances actives qui le composent jusqu'aux éléments dispensés (présentations et unités communes de dispensation), en passant par le concept de médicament prescriptible en dénomi-

30. Groupe des médicaments agissant sur le système digestif et le métabolisme.

31. Groupe des médicaments agissant sur le sang et les organes hématopoïétiques.

32. Groupe des médicaments agissant sur le système cardio-vasculaire.

33. https://smt.esante.gouv.fr/terminologie-ref_interop_med/ – consulté le 15 août 2023.

nation commune»³⁴. Ses sources primaires sont :

- les spécialités (codes CIS) et présentations (codes CIP) fournies par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM),
- les unités communes de dispensation (codes UCD) fournies par le Club Inter Pharmaceutique,
- les médicaments virtuels (codes MV) fournis par l'association Médicabase regroupant les 4 bases de données agréées par la Haute Autorité de Santé (HAS),
- les tableaux des médicaments en accès précoces ou compassionnels produits par la Direction de la Sécurité sociale (DSS).

La version la plus récente consultée (juillet 2023) lie les substances aux spécialités pharmaceutiques (relation substance active dans). Pour rester dans l'exemple que nous avons développé, lithium (carbonate de) est lié au TERALITHE 250 mg, comprimé sécable et au TERALITHE LP 400 mg, comprimé sécable à libération prolongée. Mais les médicaments virtuels ne sont pas encore intégrés. Toutefois, l'aboutissement de ce travail pourrait mener à une refonte de cette branche de l'ontologie dans le futur.

2.6.2 La *Neuroscience-based Nomenclature* (NbN)

Faisant le constat que les nomenclatures pharmacologiques en vigueur ne reflétaient pas la connaissance neuroscientifique sur les médicaments psychotropes, un consortium international³⁵ a lancé en 2008 un groupe de travail visant à établir une nouvelle nomenclature.

Cette nomenclature n'est pas fondée sur les indications des médicaments (antipsychotique, antidépresseur, hypnotique, anxiolytique, stabilisateur de l'humeur, stimulant, etc.) ou leur structure chimique, comme peut l'être l'ATC, mais d'abord (a) sur les cibles pharmacologiques (dopamine, sérotonine, noradrénaline, GABA, glutamate, histamine, mélatonine, etc.), c'est-à-dire le neurotransmetteur, la molécule ou le système dont l'activité est modulée, et (b) sur les modes d'action (antagoniste de récepteur, inhibiteur de la recapture, inhibiteur enzymatique, etc.) des médicaments sur les cibles pharmacologiques (Zohar *et al.*, 2015). Outre ces deux dimensions de pharmacologie et de mode d'action qui fondent la nomenclature, chaque médicament est décrit selon quatre autres dimensions :

- les indications approuvées (par la *Food and Drug Administration* ou l'*European Medicines Agency* notamment),
- l'efficacité et les effets secondaires (les plus courants et ceux qui sont potentiellement mortels),
- un résumé à visée pratique de la connaissance clinique du médicament et notamment les doses recommandées,
- la neurobiologie (avec une section préclinique et une section clinique).

Le bénéfice attendu d'une telle nomenclature, pour les médecins, est qu'elle permet de les guider s'il s'agit de changer de traitement, d'augmenter les doses, de combiner plusieurs médicaments; pour les patients, elle évite la confusion induite par le fait de se voir prescrire un antidépresseur suite à un diagnostic de trouble anxieux ou un antipsychotique de deuxième génération

34. https://smt.esante.gouv.fr/wp-json/ans/terminologies/document?terminologyId=terminologie-ref_interop_med&fileName=cgts_ref_medicament_fiche-detailllee.pdf - consulté le 15 août 2023.

35. L'*European College of Neuropsychopharmacology* (ECNP), l'*American College of Neuropsychopharmacology* (ACNP), l'*American College of Neuropsychopharmacology* (ACNP), l'*Asian College of Neuropsychopharmacology* (AsCNP) et l'*International Union of Basic and Clinical Pharmacology* (IUPHAR).

suite à un diagnostic de dépression. La deuxième édition révisée de la nomenclature, datant de 2019, est disponible sous forme d'une application³⁶ et comprend environ 130 substances actives (contre environ 500 pour la branche N de l'ATC). Pour ces raisons, nous n'avons pas ajouté, dans notre ontologie, de branche consacrée à cette nomenclature ni d'annotations issues de cette nomenclature aux autres branches. Cette initiative nous paraît toutefois suffisamment intéressante et novatrice pour avoir prévu dans les domaines d'intérêt psychiatrique consacrés au traitement pharmacologique, à côté des descripteurs des traitements vus selon le trouble (antidépresseur, antipsychotique, etc.), des descripteurs des traitements vus selon la cible pharmacologique (agit sur la sérotonine, agit sur les opioïdes, agit sur les enzymes, etc.) et des descripteurs des traitements vus selon le mode d'action (agent de libération, inhibiteur d'enzyme, inhibiteur de la recapture, etc.).

2.7 Excursus sur les abréviations

Dans la logique d'annotation sémantique qui nous a guidés au cours de l'élaboration d'ontoDOPSY, nous avons été amenés à établir une liste des abréviations utilisées en psychiatrie. Cette liste ne figure pas dans l'ontologie mais elle peut être utile pour le traitement automatique de documents de psychiatrie.

Une liste initiale a été compilée à partir des sources suivantes :

- *Neurologie, psychiatrie et addictologie* de Maillard & Jeantin (2019),
- *Manuel pratique des urgences psychiatriques* de Guedj-Bourdiau (2018),
- liste de sigles et abréviations fournie par l'association SERPSY (Soins Etudes et Recherches en PSYchiatrie)³⁷,
- liste de sigles et abréviations fournie par la Fondation Bon Sauveur³⁸

Cette compilation initiale a été dédoublonnée et les formes non abrégées ont été standardisées. Un criblage des formes abrégées effectivement rencontrées dans le corpus de comptes rendus d'hospitalisation a été effectué par Ons Aouina. La liste finalement obtenue est présentée dans l'annexe G. Elle constitue un premier jalon dans l'établissement d'une liste extensive d'abréviations et de sigles utilisés en psychiatrie (elle compte 494 abréviations, dont certaines correspondent à plusieurs formes non abrégées différentes), mais gagnera en pertinence à être révisée par un groupe de psychiatres.

3 Conclusion

Comme nous l'avons expliqué, la psychiatrie s'attache, plus que d'autres spécialités médicales, à une approche globale du patient : son esprit, son corps, ses relations sociales, son histoire, son environnement. Cette spécificité nous a confrontés à plusieurs reprises à la question du périmètre que devait considérer une ontologie de la psychiatrie, s'agissant par exemple de signes cliniques, des maladies, des traitements médicamenteux. Nous avons opéré des arbitrages au cas par cas, en élargissant plus ou moins au-delà du périmètre strict de la psychiatrie, mais ces arbitrages peuvent être sujets à réévaluation.

36. Il existe également un document pdf, suffisamment structuré pour en extraire l'information, mais il n'a pas été fait suite à notre demande de la nomenclature sous une forme plus directement manipulable informatiquement.

37. http://ancien.serpsy.org/psy_levons_voile/sigles.html – consulté le 15 août 2023.

38. <https://www.fondationbonsauveur.com/patient-usager/ressources-utiles/sigles-et-abreviations-334.html> – consulté le 15 août 2023.

Chemin faisant, nous avons également constaté un manque de normalisation ou de vocabulaire de référence, notamment sur les abréviations, les outils d'évaluation et leurs acronymes, les événements de vie, les signes cliniques. Outre les propositions que nous avons faites (placées en annexe), nous avons également indiqué (avec le wiki TIPSY) la piste d'un processus permettant d'aboutir à un consensus ne résultant pas seulement du travail d'un petit nombre d'experts (comme c'est généralement le cas pour les ontologies) ni des contributions d'un public très élargi (comme c'est le cas des folksonomies ou de l'initiative du *Cognitive Atlas*) mais d'une communauté d'experts la plus large possible. Cette démarche serait notamment intéressante, nous avons tenté de le montrer, pour une *taxonomie d'indexation* et plus largement pour permettre l'évolution d'une ontologie de référence pour la psychiatrie.

Modules spécifiques

There are only patterns, patterns on top of patterns, patterns that affect other patterns. Patterns hidden by patterns. Patterns within patterns. If you watch close, history does nothing but repeat itself. What we call chaos is just patterns we haven't recognized. What we call random is just patterns we can't decipher. What we can't understand we call nonsense. What we can't read we call gibberish. There is no free will. There are no variables.

Chuck PALAHNIUK, *Survivor*.

Outre les données issues de la clinique psychiatrique, et généralement en combinaison avec celles-ci, d'autres données émanant de différentes disciplines sont produites lors des études prospectives du projet PsyCARE ou l'ont été lors d'études antérieures dont les résultats sont réutilisés. Un module est dédié à chacune de ces disciplines. Nous présentons d'abord ces dernières, puis le patron de conception ontologique que nous avons développé afin de faciliter l'intégration des données dans les modules dédiés et leur exploration ultérieure, et nous présentons enfin les aménagements qui ont parfois été nécessaires selon les spécificités de chaque discipline.

1 Présentation des techniques et spécialités

De nombreux moyens d'acquisition et méthodes d'analyse issus de différentes disciplines sont convoqués par la psychiatrie de précision. Nous présentons ceux pour lesquels nous avons été amenés à développer des modules spécifiques dans l'ontologie du projet PsyCARE.

1.1 Examens d'imagerie cérébrale

L'imagerie cérébrale, ou neuroimagerie, comprend un ensemble de techniques d'imagerie médicale permettant d'observer l'encéphale¹ (nous parlerons par synecdoque de « cerveau », même

1. L'encéphale est la partie du système nerveux central enclose dans la boîte crânienne.

si le cervelet et le tronc cérébral sont aussi concernés), soit dans ses aspects anatomiques (neuroimagerie *structurelle*), soit pour suivre son fonctionnement (neuroimagerie *fonctionnelle*). Dans ce dernier cas, ce qui est observé est soit l'activité électrique, soit les flux sanguins dans le cerveau. Les principales techniques de neuroimagerie sont l'imagerie par résonance magnétique (IRM) anatomique ou fonctionnelle, la tomographie par émission de positons (TEP), l'électroencéphalographie (EEG) et la magnétoencéphalographie (MEG).

IRM

Les différents tissus d'un cerveau exposé à un rayonnement électromagnétique y réagissent différemment. On peut alors produire, dans le cas de l'IRM anatomique, une image en trois dimensions faisant apparaître des aspects morphologiques (taille et forme des différentes structures), ce qui permet de mettre en évidence des anomalies structurelles du cerveau (comme l'épaisseur du cortex), des altérations de structures normales (permettant d'évaluer le stade de la progression d'une maladie ou de poser un diagnostic), ou encore de planifier une opération. Les modifications structurelles peuvent également être mises en lien avec des données cliniques ou autres (biochimiques par exemple) concernant le sujet de l'examen. Dans le cas de l'IRM fonctionnelle, l'objectif est de mettre en évidence l'activité du cerveau. Si le dispositif technique est essentiellement le même, on obtient une séquence d'images. L'accent est mis sur la résolution temporelle, donc sur l'acquisition plus rapide d'images, éventuellement au détriment de la résolution spatiale (par rapport à l'IRM anatomique). Toutefois, l'acquisition ne peut être assez rapide pour capturer les signaux neuronaux eux-mêmes, c'est pourquoi on capture un indice indirect de ceux-ci, à savoir le contraste entre sang oxygéné ou non (c'est le signal *blood-oxygen-level-dependent*, ou BOLD), la variation d'oxygénation du flux sanguin indiquant l'activité des neurones qui le reçoivent. Ces séquences permettent d'identifier des réseaux de structures cérébrales fonctionnant de concert, et de les lier à une fonction psychologique ou à une pathologie.

TEP

La TEP vise également l'activité du cerveau, mais au niveau métabolique. Un traceur marqué par un atome radioactif est injecté dans le sang du sujet : c'est ce rayonnement γ qui est ensuite capté par le scanner, la distribution du traceur au cours du temps dans le cerveau indiquant son activité. La TEP permet d'étudier la liaison des neurotransmetteurs aux récepteurs neuronaux (par exemple pour diagnostiquer des démences neurodégénératives ou dans le processus de développement d'un nouveau médicament), le flux sanguin (par exemple pour diagnostiquer précocément des maladies cérébrovasculaires), de mettre en évidence des anomalies métaboliques (pour confirmer un diagnostic d'épilepsie en l'absence d'anomalies à l'EEG), des tumeurs ou des traumatismes crâniens.

EEG et MEG

L'EEG et la MEG permettent d'étudier l'activité du cerveau à partir des potentiels postsynaptiques, la première mesurant des champs électriques quand la seconde mesure des champs magnétiques. Ce sont des techniques non-invasives recueillant ces champs électriques ou magnétiques à la surface du scalp² présentant des différences notables en termes d'équipement : l'EEG ne nécessite qu'un casque portant des électrodes et n'entrave pas totalement la tête, quand la MEG nécessite un équipement lourd comportant des magnétomètres dans lequel la tête doit être immobile. Le potentiel d'un seul neurone étant trop faible, ce sont les potentiels de groupes de neurones alignés dans la même direction et ayant un flux synchrone qui sont captés (vraisemblablement les cellules pyramidales). Ces deux techniques

2. Mentionnons toutefois l'EEG intracrânienne, qui est parfois pratiquée dans des études animales.

présentent une excellente résolution temporelle mais présentent une résolution spatiale plus incertaine (l'EEG plus encore que la MEG).

Ces différentes techniques peuvent être comparées, comme nous venons de le voir, selon leur résolution spatiale et leur résolution temporelle, selon qu'elles sont invasives ou non, qu'elles permettent ou pas des mouvements de la tête, et selon la lourdeur de l'équipement qu'elles supposent.

L'application de l'IRM fonctionnelle dans le domaine de la psychiatrie, et plus spécifiquement dans le domaine de la schizophrénie (Martinelli & Shergill, 2015), a pu apporter des éléments sur la formation des hallucinations et la manière d'atténuer le symptôme au moyen de la stimulation magnétique transcrânienne, ainsi que sur le rôle du noyau caudé dans le délire paranoïaque. D'autre part, le croisement de données d'imagerie avec des données de génétique, en faisant l'hypothèse que les données de génétique seront mieux corrélées avec les systèmes neuronaux sous-jacents aux comportements qu'avec les comportements eux-mêmes, peut fournir des éléments sur l'héritabilité des structures cérébrales, permettre une détection plus précoce de certains troubles ou l'identification de personnes à risque. Enfin, l'utilisation d'algorithmes d'apprentissage automatique (*machine learning*) sur des données de neuroimagerie peut se targuer de certains succès dans le diagnostic de la maladie d'Alzheimer et dans le diagnostic différentiel entre maladie d'Alzheimer et d'autres formes de démence (Klöppel *et al.*, 2008), ou encore dans le diagnostic de l'autisme – à partir de l'épaisseur corticale (Ecker *et al.*, 2010).

1.2 Examens de biologie médicale

La biologie médicale est une spécialité médicale et pharmaceutique à part entière, dont l'activité consiste essentiellement en la réalisation d'examens permettant de mettre en évidence, de façon quantitative ou qualitative, la présence de certains constituants dans des échantillons biologiques prélevés sur le patient (le plus communément des liquides biologiques comme le sang ou l'urine).

Un examen de biologie médicale est défini par l'article L6211-1 du Code de la santé publique³ comme « un acte médical qui concourt à la prévention, au dépistage, au diagnostic ou à l'évaluation du risque de survenue d'états pathologiques, à la décision et à la prise en charge thérapeutiques, à la détermination ou au suivi de l'état physiologique ou physiopathologique de l'être humain, hormis les actes d'anatomie et de cytologie pathologiques, exécutés par des médecins spécialistes dans ce domaine ».

Il est décrit dans l'article L6211-2 comme se déroulant en trois phases :

- 1° « La phase pré-analytique, qui comprend le prélèvement d'un échantillon biologique sur un être humain, le recueil des éléments cliniques pertinents, la préparation, le transport et la conservation de l'échantillon biologique jusqu'à l'endroit où il est analysé » ;
- 2° « La phase analytique, qui est le processus technique permettant l'obtention d'un résultat d'analyse biologique » ;
- 3° « La phase post-analytique, qui comprend la validation, l'interprétation contextuelle du résultat ainsi que la communication appropriée du résultat au prescripteur et, dans les conditions fixées à l'article L. 1111-2, au patient, dans un délai compatible avec l'état de l'art. »

Le recours aux examens de biologie médicale, notamment comme « examens complémentaires », est transversal en médecine, et d'un usage très large : il est estimé qu'environ 60 à 70%

3. https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072665/LEGISCTA000006171482/

des diagnostics médicaux les font intervenir⁴. Ils interviennent également dans le suivi d'une pathologie ou la surveillance d'un traitement.

Le spectre des examens est très large et cette spécialité connaît des sous-spécialités⁵ :

- bactériologie et virologie,
- biochimie,
- hématologie,
- immunologie,
- parasitologie et mycologie,
- génétique moléculaire et cytogénétique.

Cette dernière sous-spécialité intéresse particulièrement la médecine de précision, surtout la génétique moléculaire (qui s'occupe de l'analyse des gènes, alors que la cytogénétique s'occupe de l'analyse des chromosomes). Les techniques mises en œuvre pour réaliser un examen de génétique moléculaire sont le séquençage (c'est-à-dire la lecture de séquences de bases azotées présentes dans l'ADN), les puces à ADN (*CGH array*) (qui procèdent par comparaison de l'ADN étudié à un ADN standard), le séquençage de nouvelle génération (*Next Generation Sequencing* – NGS) (qui permet de séquencer très rapidement de grandes quantités d'ADN). L'objectif est d'identifier des variations génétiques (appelées « mutations » si elles sont pathogènes).

Pour les autres omiques que nous avons évoquées dans la section 1.1.3 du chapitre 1, les techniques mises en jeu sont les puces à ADN pour la transcriptomique; la spectrométrie de masse pour la lipidomique et pour la protéomique (après avoir digéré à l'aide d'une enzyme les protéines de l'échantillon, pour disposer de peptides); de spectrométrie de masse et de spectroscopie de résonance magnétique nucléaire pour la métabolomique.

Le recours de la psychiatrie de précision à la biologie médicale vise l'identification de biomarqueurs biologiques ouvrant sur (a) la prévention de certaines pathologies (Chaumette, 2016), (b) leur dépistage au moyen de tests, (c) une caractérisation pour part biologique des pathologies, (d) un meilleur diagnostic différentiel entre troubles psychiatriques, (e) la suggestion de pistes thérapeutiques (Frajerman *et al.*, 2021), (f) la prédiction du résultat d'un traitement.

1.3 Analyse de la parole

Une part non négligeable de la sémiologie psychiatrique est consacrée aux signes relatifs au langage, et plus précisément (a) à la prosodie, (b) au rythme ou à la quantité de langage, (c) à l'intensité sonore de la voix, (d) à la syntaxe, (e) à la signification du langage.

Cette attention portée par la psychiatrie au discours croise celle d'autres disciplines comme la linguistique ou l'analyse de la parole (elle-même multidisciplinaire), toutes s'enrichissant du développement rapide, dans les années récentes, des techniques de Traitement Automatique du Langage Naturel (TALN) appuyées sur l'apprentissage statistique. Prolongeant la sémiologie psychiatrique, de nouvelles techniques peuvent donc être mises à contribution pour étudier du point de vue psychiatrique et de manière plus quantitative le discours dans ses aspects prosodiques émotionnels ou non-émotionnels (Lucarini *et al.*, 2020) – au moyen de logiciels développés à cette fin, mais aussi syntaxiques, lexicaux, sémantiques – par des techniques de TALN appliquées aux transcriptions.

Différentes modalités d'acquisition des données, plus ou moins « intrusives » sont envisageables :

- des enregistrements de sessions entre thérapeute et patient,

4. <https://sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/biologie-medicale/article/biologie-medicale>

5. <https://www.education.gouv.fr/bo/2003/32/MENS0301444A.htm>

- des évaluations écologiques momentanées (*Ecological Momentary Assessments* – EMA) lors d'appels téléphoniques au patient,
- des collectes de données passives à partir du micro d'un smartphone.

Plusieurs troubles psychiatriques, dont la schizophrénie, les troubles du spectre de l'autisme, la maladie d'Alzheimer, les troubles bipolaires et la dépression, sont susceptibles d'être caractérisés (d'un point de vue diagnostique) par un profil prosodique (Ding & Zhang, 2023) ou par des techniques de TALN (Le Glaz *et al.*, 2021). L'objectif peut également être l'évaluation de la sévérité des symptômes (notamment pour la dépression), la prédiction de la transition psychotique (Corcoran *et al.*, 2018), ou l'intervention « juste à temps » (dans le cas de la collecte passive).

1.4 Évaluation des troubles de la dextérité et des fonctions cognitivo-motrices

La sémiologie psychiatrique comprend également des signes relatifs à la psychomotricité dans son ensemble et à des signes neurovégétatifs (dont les tremblements). Lorsqu'une étiologie neurologique est suspectée, à partir de l'examen clinique, un examen plus spécifiquement neurologique peut être mené. Celui-ci peut inclure l'évaluation de la marche, de la station debout, de la force musculaire, du tonus musculaire, des réflexes, de la coordination motrice, de la sensibilité, des nerfs crâniens et des voies visuelles, des sphincters, de la vigilance et des fonctions cognitives⁶.

Cet examen peut permettre d'identifier des signes neurologiques révélant l'atteinte d'aires précises du cerveau. Lorsque les signes neurologiques identifiés ne peuvent être imputés à l'atteinte d'aires cérébrales spécifiques, on parle de « signes neurologiques mineurs » (SNM, ou en anglais *Neurological Soft Signs* – NSS).

Des tests spécifiques ont été développés pour évaluer les SNM, notamment pour évaluer le dysfonctionnement neurologique en cas de schizophrénie (Krebs *et al.*, 2000). Récemment, de nouveaux tests ont été développés, plus centrés sur la dextérité, à partir de tâches effectuées sur une tablette numérique (Rabah *et al.*, 2022) permettant une détection des psychoses précoces (Le Boerff *et al.*, 2023).

2 Élaboration d'un patron de conception ontologique

A partir du travail effectué sur les outils d'évaluation psychiatrique dans le module de clinique psychiatrique, nous avons élaboré un patron de conception ontologique (*Ontology Design Pattern*) qui a servi de base à la constitution des modules dédiés aux différentes sources de données.

2.1 Éléments de description communs

Les données à intégrer sont hétérogènes à la fois par leur contexte de production, par les modèles scientifiques qui les sous-tendent, par les usages de formatage et de transmission qui s'y appliquent, et même par le lexique développé par chaque domaine pour désigner ses objets et ses méthodes. Nous pouvons toutefois pointer les éléments communs à ces processus de production de données : **l'objet d'étude**, **la passation d'une expérience**, expérience qui obéit pour les différents objets d'étude à un même **protocole** et qui génère en premier lieu un **produit concret** à partir duquel sont effectuées des **mesures** ou des évaluations. Le protocole, antérieur à la réalisation des expériences, prévoit nécessairement un **instrument de captation** du produit concret, et implique éventuellement une mise en condition (stipulée dans des **instructions**) de l'objet d'étude ou de

6. <https://www.cen-neurologie.fr/fr/premier-cycle/examen-neurologique>

DOMAINE	Psychométrie	Imagerie	Biologie	Biologie moléculaire	Analyse du discours	Dextérité
EXPÉRIENCE (PASSATION)	questionnaire, entretien, test	séquence IRM, EEG, TEP	prélèvement en laboratoire	prélèvement en laboratoire	enregistrement	tâche de motricité
STIMULUS	question, dessin, etc.	son, image, etc.	NA	NA	question	images, sons
OUTIL DE CAPTATION	papier et crayon	appareil IRM, EEG, TEP	kit de prélèvement	kit de prélèvement	appareil enregistreur	tablette numérique
INSTRUCTION	au sujet, à l'agent de passation	paramétrage, agent de contraste	au sujet, de prélèvement	de prélèvement	NA	au sujet
PRODUIT DIRECT	réponse	jeu d'images	échantillon	échantillon	fichier audio	gestes
MESURE	scores	volumes, régions d'intérêt	valeurs, comparaisons à des seuils	id. de gènes, de variants, de <i>pathways</i>	valeurs de variables	valeurs de variables

TABLE 5.1 – Éléments communs selon les différentes sources de données de PsyCARE.

l'agent de passation permettant d'assurer la comparabilité des résultats obtenus, ainsi qu'un ou des **stimulus** permettant de conditionner le phénomène étudié pour en affiner l'observation.

Le tableau 5.1 montre comment les différents éléments communs retenus sont déclinés selon les différentes sources de données.

2.2 Engagement ontologique antérieur

Certains engagements ontologiques pris lors de l'élaboration d'ontoPOF (cf. chapitre 2) et d'ontoDOME (cf. chapitre 3) influent sur la conception du patron que nous présentons.

Nous avons vu la place accordée par ontoPOF aux projets, susceptibles d'être réalisés dans un ou plusieurs événements. Les différents protocoles qui sont ici à modéliser sont considérés comme des projets, le même protocole étant le plus souvent réalisé dans plusieurs passations d'expériences. D'autre part, ontoPOF accorde une grande importance à la représentation de la dynamique temporelle des entités, à la représentation du changement (Hilbey *et al.*, 2022).

Nous avons également vu qu'ontoDOME s'attachait à modéliser les connaissances telles qu'elles sont échangées ou décrites dans des documents médicaux, plutôt que telles qu'elles existeraient dans l'esprit des experts médicaux.

Ces engagements ontologiques nous amènent, dans l'élaboration du patron de conception ontologique, à porter une attention particulière :

- au maintien de la structure [protocole – événement réalisant le protocole – données générées par l'événement] (figure 5.1), et ce aux différents niveaux de granularité des événements : collecte de données, examen, examens dans un même domaine, visite, étude (les résultats correspondant à un même sujet ou à un même groupe d'étude sont également regroupés, mais ne sont pas considérés comme générés par un événement obéissant à un protocole),
- au positionnement, entre l'événement et les mesures auxquelles il donne lieu, des données générées, dans la mesure où ce sont elles qui sont effectivement transmises entre les différents acteurs.

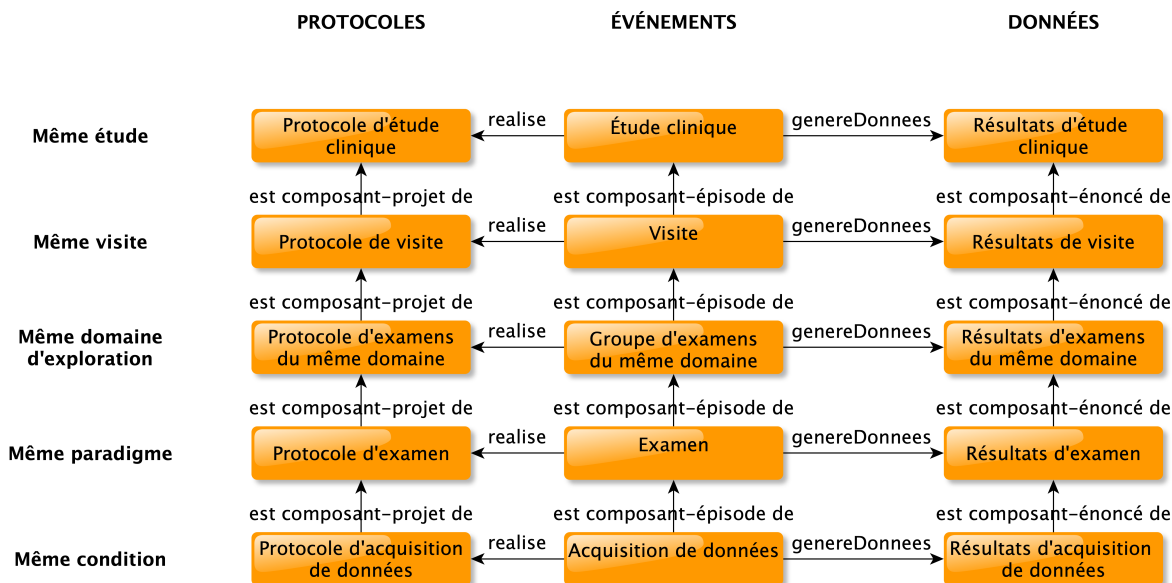


FIGURE 5.1 – Les différents niveaux de granularité des événements d'acquisition de données.

2.3 Les patrons de conception ontologique

Un patron de conception ontologique est une modélisation qui peut servir pour résoudre un problème de conception ontologique qui se pose de façon récurrente (Hitzler *et al.*, 2016). Dégager un tel patron repose donc sur l'observation d'invariants dans les données, les objets, les processus, les relations.

Les patrons de conception ontologique (ou *Ontology Design Patterns* – OP) peuvent être de plusieurs types (Gangemi & Presutti, 2009) :

- les *Structural OPs*, qui subsument (a) les *Logical OPs* résolvant un problème d'expressivité et (b) les *Architectural OPs* contraignant la constitution générale de l'ontologie ;
- les *Reasoning OPs* qui orientent le raisonnement du moteur d'inférence ;
- les *Correspondence OPs*, qui subsument (a) les *Reengineering OPs* transformant un modèle source en modèle ontologique et (b) les *Mapping Ops* pour exprimer des correspondances ;
- les *Presentation OPs* pour les conventions de nommage et les schémas d'annotation ;
- les *Lexico-syntactic OPs* qui associent des formes linguistiques à une signification ;
- les *Content OPs* qui proposent des patrons pour le contenu d'un domaine particulier.

Cette dernière catégorie correspond très précisément à ce que nous proposons ici.

2.4 État de l'art

Une recherche par mots-clés sur le site web de l'*Association for Ontology Design & Patterns* (ODPA)⁷ ne donne pas de résultat pour le sujet que nous traitons. En revanche, quelques ontologies proposent des modélisations d'expériences scientifiques ou d'examen clinique.

EXPO est une ontologie s'appuyant sur une description formelle des expériences scientifiques afin de les annoter (Soldatova & King, 2006). Elle utilise SUMO comme ontologie fondationnelle et entend couvrir tout le champ de la science expérimentale, en termes de types d'hypothèse, de

7. <http://ontologydesignpatterns.org/>

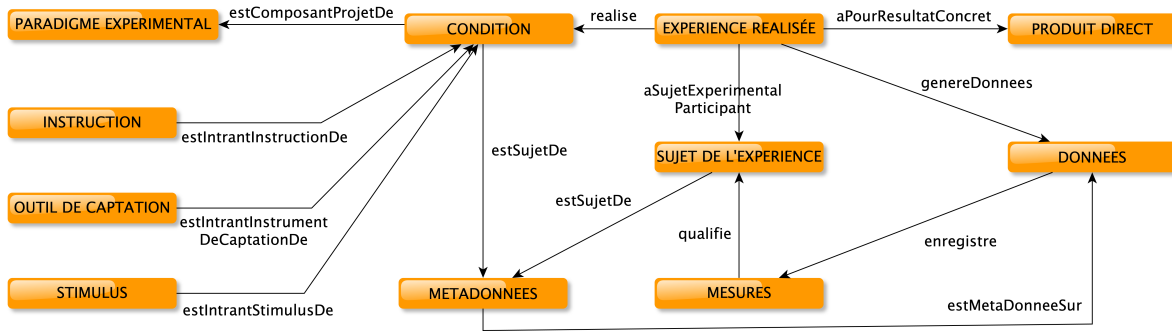


FIGURE 5.2 – Un patron de conception d’ontologie pour les paradigmes expérimentaux.

modèle, d’expérience, de variable, de résultats. Elle se présente comme un ontologie de niveau intermédiaire entre SUMO et des ontologies de domaine consacrées à un champ d’étude particulier. Sa finalité est donc en premier lieu la caractérisation d’expériences scientifiques, ce qui permet de les situer méthodologiquement dans le champ de la science. En ce sens, cette ontologie ne répond pas à notre cas d’usage.

CogPo est une ontologie qui entend décrire les paradigmes de psychologie cognitive (Turner & Laird, 2012). On y trouve la plupart des éléments avancés dans la section 2.1 : instructions, stimulus, paradigme, condition (qui correspond à ce que nous appelons ici « protocole d’expérience »). Toutefois, comme pour EXPO, l’enjeu est plus une caractérisation d’un paradigme dans un champ scientifique qu’une modélisation d’une expérimentation en action, qui génère des données.

Ontology for Biomedical Investigations (OBI) est une ontologie des études cliniques et biomédicales que nous avons présentée dans la section 1 du chapitre 3 et dont nous rappellerons ici qu’elle vise à permettre d’intégrer toute étude biomédicale possible. En proposant un patron de conception ontologique facilitant l’élaboration de modules selon les domaines, nous opérons un choix différent : nous proposons une structure minimale permettant d’intégrer les données et de les explorer, et laissons aux modules de domaines une modélisation plus poussée du domaine. D’autre part, les niveaux de granularité des événements de l’étude, des données qu’ils génèrent et des éventuels protocoles correspondants, ne sont que marginalement pris en considération ; cette critique concerne toutefois plutôt IAO, voire BFO, qu’OBI en propre.

2.5 Un patron de conception ontologique pour les paradigmes expérimentaux

Un paradigme expérimental peut être considéré comme étant dans une relation d’abstraction par généralisation à différentes conditions expérimentales, qui en seraient les spécifications. Cette conception amènerait à une modélisation des conditions comme autant de sous-classes d’un paradigme expérimental. En prenant en compte la pratique expérimentale effective, qui mêle au sein d’une même session (unité temporelle) la passation de différentes conditions expérimentales obéissant au même paradigme, nous choisissons de considérer une relation méréotopologique de composition entre le paradigme et les conditions, reflétant la relation méréotopologique entre les événements (temps de la session, qui a pour parties les temps des différentes conditions).

La figure 5.2 présente les résultats au niveau de la condition. C’est à ce niveau que sont précisés, pour le protocole de la condition, quels sont les stimulus, instruction et outil de captation prévus en tant qu’intrants – le protocole est un projet qui est réalisé lors de chaque passation de l’expérience par un sujet.

Le patron de conception tient essentiellement dans les relations (*Object Properties*) qui sont indiquées. Les éléments de description communs que nous avons dégagés dans la section 2.1 ne doivent pas être conçus comme des classes nécessairement présentes dans l'ontologie.

Si certains éléments (nous nous plaçons ici au niveau des instances) présentent une proximité ontologique forte, comme les sujets, les événements de passation, les protocoles, ce qui peut amener à prévoir des classes (éventuellement des classes définies) pour les représenter dans la partie taxonomique de l'ontologie, d'autres peuvent recouvrir une forte diversité ontologique, comme les stimulus, les produits directs, les outils de captation ou les instructions. Qui plus est, ces derniers éléments peuvent selon les contextes apparaître sous différents aspects (une portion de sang prélevé peut faire l'objet d'analyses ou d'un don de sang; un son servant de stimulus peut entrer dans une composition musicale) et leur position dans la taxonomie dépendra de leur nature intrinsèque, mais c'est une relation à d'autres entités qui permettra d'indiquer leur fonction ou leur rôle dans le contexte qui nous intéresse. D'autre part, la position des éléments de description communs dans la taxonomie peut tenir à l'engagement ontologique pris en amont, comme nous l'avons montré dans la section 2.2.

Aux éléments communs spécifiés plus haut s'ajoutent les métadonnées, qui portent sur le protocole et le sujet de l'expérience, ce qui permet d'établir un lien plus direct entre les données et les informations importantes de leurs conditions de production : le sujet et le protocole spécifique de la condition expérimentale.

Certaines relations peuvent sembler redondantes. Elles permettent de raccourcir le chemin dans le graphe de connaissance, selon le mode d'exploration souhaité. Si l'on se centre sur le sujet de l'expérience, on peut accéder rapidement à l'étude dans laquelle l'expérience a été réalisée et aux mesures qui le qualifient, ou inscrire l'événement de la passation dans une chronologie des événements connus de la vie du patient; si l'on se centre sur les données, on peut accéder rapidement à celles-ci et à leurs métadonnées; si l'on se centre enfin sur l'événement de la passation de l'expérience (qui dans le graphe présenté est le nœud qui a la plus grande centralité de proximité aux autres nœuds), on peut l'inscrire dans une chronologie des événements de l'étude.

3 Modules ontologiques de spécialité

Le patron de conception ontologique que nous avons présenté n'est pas complètement une solution clé-en-main. Selon les sources de données, certains des éléments qui le constituent seront présents ou pas, certaines spécificités dans l'acquisition de données devront être prises en compte. Ce sont ces spécificités que nous abordons ici.

3.1 ontoDOIMA : l'ontologie des données d'imagerie

Le jeu d'images résultant de la passation d'une expérience d'imagerie est le plus souvent numérisé et peut être considéré comme un jeu de données (par ailleurs très volumineuses et faisant l'objet d'un stockage spécifique dans le cadre du projet PsyCARE). Toutefois, de notre point de vue, ces images sont le résultat concret de la passation. Quand nous considérons les données d'imagerie, nous faisons donc bien plutôt référence soit aux résultats d'analyses de ces images identifiant par exemple des régions d'intérêt ou des volumes spécifiques, soit aux données associées aux images et qui peuvent concerner leur contexte technique de production (auquel cas elles se trouvent dans des fichiers associés aux images, selon les standards utilisés, DICOM ou BIDS) aussi bien que leur contexte clinique de production (auquel cas elles se trouvent dans des cahiers d'observation électronique).

Les données indiquant le contexte technique de production peuvent concerner l'appareil utilisé pour l'acquisition, son fabricant, la modalité d'acquisition et ses spécifications techniques (antenne utilisée, force du champ magnétique, description de la séquence d'impulsion par exemple). On parlera ici de résultats d'une séquence d'imagerie.

Les données indiquant le contexte clinique de production peuvent concerner l'état du patient lors de la passation (heure de la dernière prise d'anxiolytiques, de caféine, de tabac) ou en amont (existence de contre-indication à la passation), ainsi que la gestion de l'examen (séquences réalisées ou pas, complétude de l'examen). On parlera ici de relevé de passation d'examen d'imagerie.

Le niveau le plus bas d'acquisition de données retenu est celui de la séquence d'imagerie. Le niveau de granularité supérieur est l'examen d'imagerie, qui est défini par la modalité d'acquisition des images (examen IRM, examen EEG, examen TEP, etc.). Lors d'une même passation d'un examen, plusieurs séquences sont généralement acquises.

La figure 5.3 illustre l'acquisition d'un jeu de données de carte du temps de relaxation longitudinale (T1map) lors d'un examen IRM.

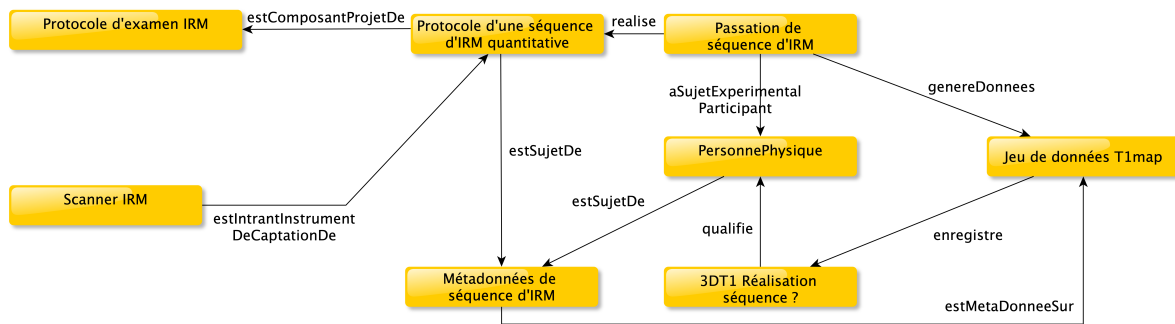


FIGURE 5.3 – Application aux données de neuroimagerie.

3.2 ontoDOBIO : l'ontologie des données de biologie

Les données de biologie médicale proviennent de Centres de Ressources Biologiques (CRB), qui réalisent les analyses sur des prélèvements opérés sur les sujets⁸. Ce sont ces échantillons prélevés sur les sujets qui sont ici les produits concrets de l'acte de prélèvement, qui est considéré comme une expérimentation. Le recours à un prestataire externe fait que nous disposons de peu d'informations hors les résultats d'analyse.

En biologie médicale, on nomme généralement « examen » un dosage particulier (par exemple le dosage de la vitamine C) ou une détection de présence de constituants spécifiques dans un échantillon. Nous avons choisi de considérer, comme niveau le plus basique de collecte de données, des groupes d'examen orientés par une visée diagnostique (« bilans »), en nous basant sur les informations des cahiers d'observation électronique (eCRF) de l'étude principale de PsyCARE.

Les différents bilans, sont les suivants :

- dosage de la bêta-hCG,
- bilan de l'hémostase,
- bilan hépatique,
- bilan inflammatoire,

8. Des *Annotation Properties* sont prévues pour décrire, à propos d'une étude, quel est le CRB, quel est le type d'échantillon prélevé et le type de préparation qui lui est appliqué.

- bilan métabolique et endocrinien,
- bilan phosphocalcique,
- bilan urinaire,
- ionogramme,
- numération de la formule sanguine,
- test sérologique et recherche d'anticorps.

Les bilans ne sont pas des entités présentes dans l'ontologie, mais plutôt leurs protocoles, leurs passations, leurs résultats. Le dosage de la bêta-hCG offre le cas particulier d'un bilan qui n'est constitué que d'un seul examen biologique.

Au niveau de granularité supérieur (ce que nous appelons « examen médical »), on trouve le bilan biologique, qui regroupe ces différents bilans.

La figure 5.4 illustre l'acquisition de données correspondant à un bilan métabolique et endocrinien, avec en exemple de variable le taux de vitamine C.

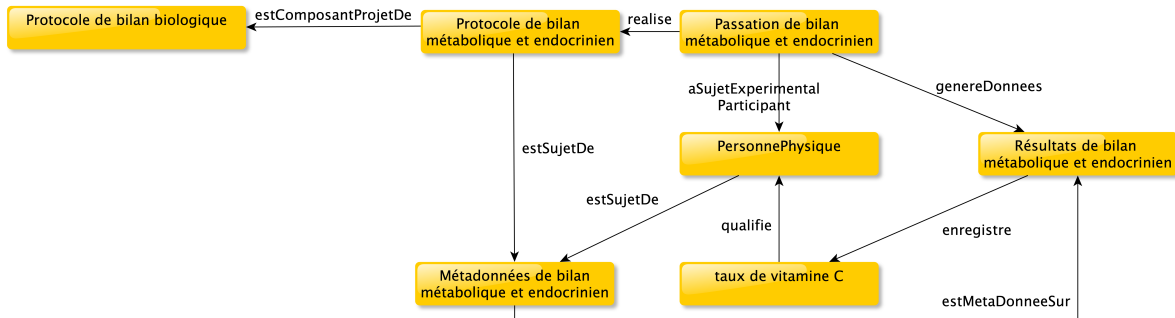


FIGURE 5.4 – Application aux données de biologie médicale.

3.3 ontoDOOMICS : l'ontologie des données de biologie moléculaire

Nous présentons ici des travaux préparatoires sur le module de biologie moléculaire. Au moment de ce travail, les conditions n'ont pas été réunies pour permettre l'élaboration complète de ce module. Nous indiquons toutefois les pistes qu'ont suivies nos réflexions lors des travaux de préparation de ce module.

Comme nous l'avons déjà remarqué, la biologie moléculaire est une discipline dans laquelle les ontologies ont trouvé leur place comme ressources de référence (ou dans laquelle des ressources de référence ont été mises sous forme ontologique). Nous présentons ci-dessous quelques-unes de ces ontologies.

OMIM (pour *Online Mendelian Inheritance in Man*) est soutenue par le *National Human Genome Research Institute* états-unien. C'est d'abord une base de données qui propose une compilation des connaissances scientifiques sur les phénotypes génétiques et les gènes qui y sont associés. Nous considérons ici la forme ontologique, issue du méta-thésaurus *Unified Medical Language System* (UMLS). Notons que l'IRI de l'ontologie (et de ses classes) respecte la recommandation de l'*OBO Foundry* d'un *OBO Foundry persistent URL* (PURL).

C'est une ontologie volumineuse (plus de 100 000 classes et autant d'instances, les classes étant également des instances) qui comprend de manière non hiérarchisée un ensemble de maladies (e.g. *AUTISM*), de syndromes (e.g. *CK SYNDROME*; *CHROMOSOME 22q11.2 DELETION SYNDROME*, *DISTAL*), de prédispositions génétiques à des maladies (e.g. *AUTISM*,

SUSCEPTIBILITY TO, 3), de gènes (e.g. *SHANK3, PRO141ALA*), de modes de transmission (e.g. *Autosomal Recessive*) ainsi que des « manifestations » (signes cliniques) classées (dans la taxonomie) selon l'anatomie (e.g. *Above-average height* sous *height* sous *GROWTH*). Selon les connaissances disponibles, les gènes sont reliés à un *locus* (position sur le chromosome) par la propriété de Types de données *GENELOCUS* et à une forme abrégée par la propriété de Types de données *GENESYMBOL* (le locus est exprimé sous forme d'une chaîne de caractères); les maladies sont reliées à leurs manifestations par la propriété d'objet *has_manifestation*, à leur mode de transmission par la propriété d'objet *has_inheritance_type*, à des loci (*GENELOCUS*) et à des gènes (*GENESYMBOL*). Les relations de variation allélique sont indiquées par la propriété d'objet *has_allelic_variant*.

GO (pour *Gene Ontology*) est une ontologie développée par le *GO Consortium*⁹, dont la mission est de développer un modèle informatique complet des systèmes biologiques, à toutes les échelles du vivant (du niveau moléculaire au niveau de l'organisme) et pour toutes les espèces.

L'ontologie comporte plus de 51 000 classes et 14 000 instances (les IRI respectent également la recommandation PURL). Elle décrit les connaissances du domaine biologique sous trois aspects disjoints qui sont les trois racines de l'ontologie : (a) la fonction moléculaire (activités au niveau moléculaire réalisées par les produits géniques), (b) le composant cellulaire (emplacements relatifs aux structures cellulaires dans lesquelles un produit génique remplit une fonction), (c) le processus biologique (processus plus vastes ou « programmes biologiques » accomplis par de multiples activités moléculaires). Il n'y a pas de représentation *per se* des gènes dans GO : ils apparaissent dans des annotations associées aux classes (de fonction moléculaire, composant cellulaire ou processus biologique). Les principales propriétés d'objet sont la relation *part_of* (entre composants moléculaires ou entre processus biologiques) et la relation *regulate* (lorsqu'un processus affecte la manifestation d'un autre processus).

HGNC-NR (pour *HUGO Gene Nomenclature Committee* et *New Releases*) est l'ontologie produite par le comité de nomenclature de la *Human Genome Organisation* (HUGO), qui « est chargé d'approuver des symboles et des noms uniques pour les loci humains, y compris les gènes codant pour des protéines, les gènes d'ARNnc¹⁰ et les pseudogènes¹¹, afin de permettre une communication scientifique sans ambiguïté. »¹²

L'ontologie compte près de 49 000 classes (les classes des gènes respectent le codage IRI <http://identifiers.org/hgnc/> tandis que les IRI de l'ontologie et des classes hautes respectent la recommandation PURL). Elle est produite automatiquement à partir d'une archive régulièrement mise à jour¹³. Elle ne comporte aucune autre propriété d'objet ni aucune propriété Type de données (ce qui est conforme à la notion de nomenclature) et la hiérarchie des classes est très simple. Sous la racine *gene* se trouve cinq classes :

- *protein_coding_gene*,
- *non_coding_rna*,
- *pseudogene*,
- *other*,
- *withdrawn*.

9. <https://geneontology.org/docs/introduction-to-go> – consulté le 15 août 2023.

10. C'est-à-dire l'ARN non codant.

11. C'est-à-dire les gènes ne conduisant pas à l'expression d'une protéine.

12. <https://www.hugo-international.org/standards/> – consulté le 15 août 2023.

13. <https://www.genenames.org/download/archive/> – consulté le 23 août 2023.

Les gènes (non « abandonnés ») sont ensuite répartis entre les quatre premières de ces classes. Chaque gène a de nombreuses annotations descriptives permettant d'associer des libellés différents, des symboles différents, le locus (`chromosomal_location`), des références dans d'autres bases de données de gènes (relation `hasDbXref` avec des références dans OMIM, Ensembl, Entrez, NCBI Gene, MGI, etc.).

Ces trois ontologies apparaissent comme complémentaires et présentent suffisamment de relations et d'annotations pour permettre des alignements. Elles permettent notamment de représenter les gènes, leur localisation sur le chromosome, les différents variants génétiques (délétion, mutation), l'expression des gènes, différentes modifications épigénétiques (méthylation, déméthylation), ainsi que les phénotypes cliniques éventuellement associés. Comme nous l'avons remarqué, elles sont très volumineuses. Ce qui peut pousser à mettre en place des mécanismes d'enrichissement *ad hoc* de l'ontologie d'application afin qu'elle contienne finalement tous les gènes, variants, etc. mentionnés dans un jeu de données particulier sans avoir nécessairement à contenir d'avance l'intégralité de ces ontologies. Un tel mécanisme s'avèrerait particulièrement intéressant à mettre en place dans les études d'association pangénomiques (GWAS), dans lesquelles on ne peut pas préjuger des polymorphismes nucléotidiques (SNPs) qui apparaîtront.

Si nous nous plaçons dans une optique de modélisation du domaine à partir du patron de conception ontologique, les éléments que nous venons d'évoquer seront enregistrés dans des variables permettant de qualifier le patient. D'autres données importantes pour interpréter les résultats de l'examen, comme la méthode d'obtention des SNPs (séquençage, génotypage), pourront être une caractéristique du protocole d'acquisition des données.

3.4 ontoDODISC : l'ontologie des données d'analyse du discours

Dans le cas de l'ontologie consacrée à l'analyse du discours, la spécificité que nous avons dû prendre en compte est le fait que le produit concret de la passation de l'expérience est l'enregistrement de la conversation entre le patient et le thérapeute. A partir de cet enregistrement, différentes mesures sont effectuées qui concernent d'une part la prosodie :

- l'étendue de tons (*pitch range*),
- la fréquence fondamentale (f_0) et ses variations,
- l'intensité acoustique (*loudness*) et ses variations;

et d'autre part des signaux conversationnels :

- les temps de paroles de l'interviewer et du participant,
- les recouvrements de parole,
- les silences mutuels,
- les tours de parole,
- les pauses,
- les *backchannels* (signaux d'engagement, d'attention, de compréhension, de continuation émis par l'auditeur),
- les pauses remplies (signaux d'hésitation pendant la production langagière),
- les rires.

L'enregistrement est ensuite transcrit et des techniques de Traitement Automatique du Langage Naturel sont appliquées à cette transcription pour en tirer des valeurs d'autres variables, qui concernent :

- des comptes de mot (et leurs variations)
 - par catégorie grammaticale (noms, verbes, adjectifs, adverbes, pronoms aux différentes personnes) ou

- par portion de discours (total, par proposition) ;
- la densité lexicale (rapport mots lexicaux / mots grammaticaux) ;
- les diffluences (qui se manifestent sur un fond de discours fluide) ;
- les relations entre termes ou entre parties du document (au moyen de l'analyse sémantique latente).

Aucune instruction n'est mentionnée, ni aucune information sur l'outil de captation. Nous n'avons pas à considérer différents niveaux de granularité dans les événements.

Afin de simplifier la modélisation en conservant néanmoins le lien à l'événement générateur, l'enregistrement aussi bien que sa transcription sont considérés comme deux produits concrets de la passation de l'enregistrement. Les données générées à partir de l'un et de l'autre sont rassemblées et liées à la passation. Une modélisation plus complexe serait évidemment envisageable.

La figure 5.5 rassemble les éléments présents.

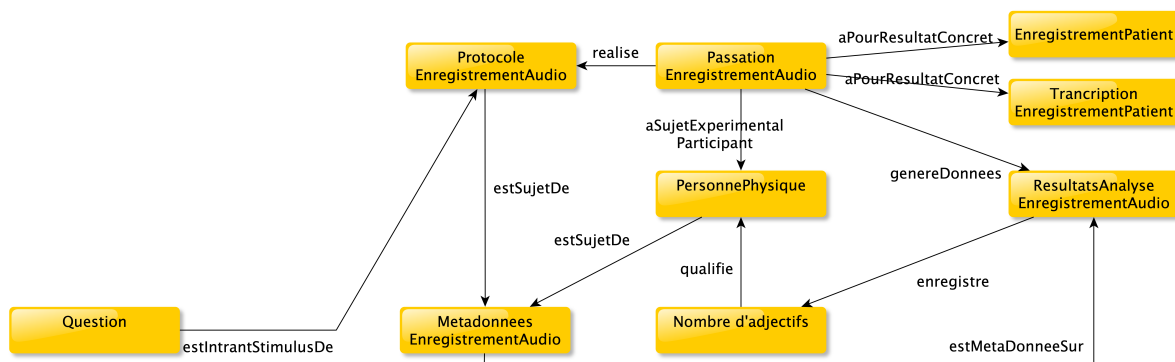


FIGURE 5.5 – Application aux données d'analyse de discours.

3.5 ontoDODEXT : l'ontologie des données d'évaluation de la dextérité

Les tâches d'évaluation neuropsychiatrique de la dextérité proposées au sujet expérimental dans ce module correspondent à six paradigmes expérimentaux déclinés pour la plupart en différentes conditions. Nous avons conservé le terme anglais *tapping* pour désigner l'action de taper de façon répétée. Ces paradigmes sont les suivants :

Reconnaissance des doigts, paradigme spécifié selon la manière dont est présentée l'image de main stimulant la réaction attendue du sujet :

- présentation en miroir,
- présentation inversée (le bas en haut),
- présentation transversale ;

Tapping multidoigts, spécifié selon le nombre de doigts mis en jeu :

- mettant en jeu un seul doigt,
- mettant en jeu deux doigts ;

Rythme de tapping, spécifié selon le rythme qu'il s'agit de suivre :

- fréquence de 1 Hz,
- fréquence de 2 Hz,
- fréquence de 3 Hz ;

Tapping maximal : cette tâche n'est pas spécifiée en différentes conditions ;

Suivi de ligne, spécifié selon qu’une autre tâche cognitive (décompte numérique mental) est ajoutée à la tâche principale :

- tâche simple,
- tâche double;

Tapping séquentiel, spécifié selon la présence ou pas d’un amorçage initial :

- sans feedback,
- avec feedback.

Les mesures effectuées concernent pour la plupart des tâches les échecs, les réussites, le temps de réaction, les éventuelles interruptions, la fréquence des gestes.

Nous présentons dans la figure 5.6 une application de notre patron de conception ontologique au module consacré à la dextérité et à la psychomotricité, sur la condition Inverse du paradigme de reconnaissance des doigts.

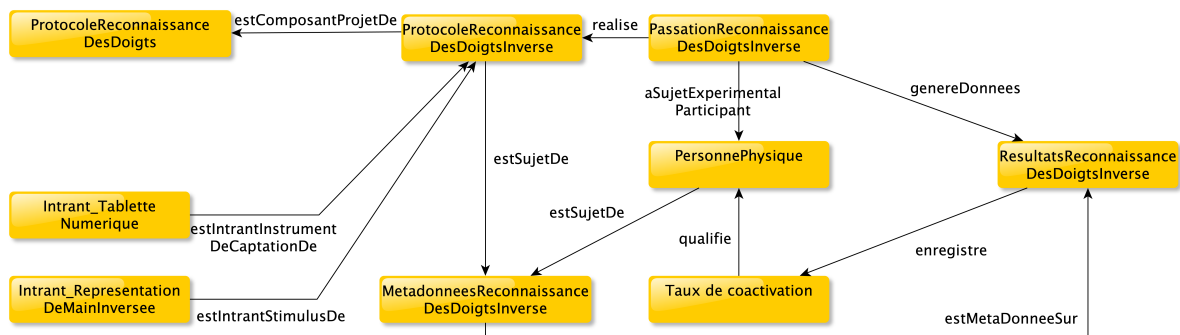


FIGURE 5.6 – Application aux données de dextérité.

Le stimulus est donc une image de main inversée. L’instrument de captation des réactions psychomotrices du sujet est ici une tablette numérique, mais pourrait être – pour des paradigmes du même domaine – un *manipulandum* comportant des pistons reliés à des capteurs enregistrant la force exercée par les doigts. La variable est le taux de coactivation, qui mesure si d’autres doigts bougent en même temps que le doigt ciblé.

L’instrument de captation étant conçu pour traduire immédiatement en mesures le produit direct (le geste du patient), celui-ci n’est pas conservé. D’autre part, aucune instruction n’est mentionnée dans le protocole, qui ne comporte donc pas ce type d’intrant.

Dans certains cas, les résultats issus de différentes conditions sont combinés. Par exemple, pour le paradigme de reconnaissance des doigts, par soustraction des résultats obtenus de la condition en miroir aux résultats obtenus de la condition inverse, ou pour le paradigme de tapping séquentiel, par moyenne des résultats obtenus avec et sans feedback. Dans ce cas, les résultats ne peuvent être liés à la passation de l’une de ces conditions, mais peuvent l’être, au niveau de granularité supérieur, à la passation du paradigme.

4 Conclusion

Les modules que nous venons de présenter sont « spécifiques » à la fois parce qu’ils permettent de représenter différentes techniques et disciplines spécifiques, mais aussi parce qu’ils ont spécifiquement été développés pour le projet PsyCARE, même si nous considérons qu’ils sont d’une application plus large. Nous avons vu que leur conception est intimement liée au patron de concep-

tion ontologique que nous avons présenté. Si d'autres modélisations ontologiques peuvent présenter un grand intérêt pour certaines de ces disciplines – nous pensons ici principalement à On-toNeuroLOG pour l'imagerie cérébrale (Gibaud *et al.*, 2011) – l'unité de la modélisation est ici un facteur-clé à la fois pour l'intégration des données et pour leur exploration ultérieure. En revanche, il reste toujours possible de lier les classes des ontologies développées à d'autres classes d'ontologies de référence ou à des vocabulaires standard. Nous avons engagé un travail dans ce sens pour annoter les classes d'ontoDOBIO consacrées aux examens de biologie avec les codes de la terminologie de référence LOINC¹⁴, qui n'a pu aboutir dans les limites de notre travail.

14. <https://loinc.org/>

Déploiement d'ontoPSYCARE

« En bref, il était impossible d'obtenir la répétition d'aucune manifestation préalablement observée. Constamment, il semblait qu'on fût sur le point de déchiffrer la masse grandissante des indices enregistrés ; n'avait-on pas construit à cette intention des cerveaux électroniques d'une capacité d'information pratiquement illimitée, tels qu'aucun autre problème n'en avait exigé jusqu'alors ? A vrai dire, on obtint des résultats. »

Stanislas LEM, *Solaris*.

Les différents modules que nous avons présentés prennent leur place dans l'ontologie modulaire ontoPSYCARE, qui vise en premier lieu l'intégration des données produites lors du projet PsyCARE, en deuxième lieu l'annotation de documents psychiatriques, et doit permettre en troisième lieu la constitution d'un graphe de connaissance – suite à l'intégration des données – dont nous évoquons l'exploration. Ce sont ces trois utilisations d'ontoPSYCARE que nous abordons ici, après avoir présenté la constitution de cette ontologie et l'évaluation que nous pouvons faire de sa qualité.

1 Constitution d'ontoPSYCARE

Nous revenons dans cette section sur la manière dont a été conçue ontoPSYCARE du point de vue méthodologique puis nous expliquons la manière dont elle est constituée du point de vue de l'ingénierie ontologique.

1.1 Conception

La découverte progressive des méthodes et méthodologies de conception d'ontologies ayant été concomitante de la conception des modules ontologiques, nous ne développerons pas ici un exposé d'une pièce sur le choix et l'application d'une méthodologie constituée. Nous présentons un regard rétrospectif porté sur le chemin effectivement suivi plutôt que sur un chemin tracé à

l'avance pour aboutir à un résultat ou sur la recherche d'un tel chemin¹. Nous emprunterons néanmoins à la grille d'analyse présentée dans la section 3.1.2 du chapitre 1 pour tenter d'affiner ce regard.

Le choix d'une architecture générale modulaire s'est imposé d'emblée, à la fois par l'anticipation de la taille à venir de l'ontologie, par la diversité des données à intégrer, par le souhait de favoriser la réutilisation des modules développés (en premier lieu ontoPOF et ontoDOPSY). La prise en considération des principes différentiels de la méthode ARCHONTE et des règles formulées par la méthodologie Ontoclean a également été précoce dans notre travail. Remarquons d'ailleurs que la première est présentée comme une méthode de construction et la seconde comme une méthode d'évaluation mais que la séparation entre construction et évaluation s'estompe ici, et que le respect des principes différentiels énoncés par ARCHONTE peut avoir sa place dans un travail d'évaluation tandis que les règles d'OntoClean peuvent être intégrées dans le processus de construction.

La présentation générale de notre travail, qui va du module le plus général vers les modules les plus spécifiques, peut suggérer une méthode globalement descendante. Au-delà d'un simple ordre de présentation, nous avons vu que le recours à une ontologie fondationnelle pouvait être caractéristique d'une méthode descendante. Il nous semble toutefois que le retour sur chaque module peut amener à préciser ce point de vue. Quant à la caractérisation du développement en termes de linéarité ou d'agilité, elle n'a pas été l'objet d'un choix mais s'est imposée, en faveur de l'agilité, en raison de la multiplicité de nos interlocuteurs (partenaire industriel et experts de domaines) et de la nature de nos interactions avec eux.

Le développement d'une ontologie fondationnelle a certes un départ très ouvert² mais on imagine mal développer une ontologie fondationnelle sans avoir la moindre prénotion des « feuilles » qu'on souhaite atteindre ou des relations qui pourront relier les instances. Dire cela ne préjuge pas d'éventuelles surprises, d'ajouts ou de retraits de telles feuilles, nous avons pu l'expérimenter lors du développement d'ontoPOF, mais amène à considérer un mouvement ascendant depuis les feuilles et un mouvement descendant depuis la racine³. Le mouvement descendant est toutefois plus fortement marqué, à la fois parce qu'il est souvent guidé par des principes philosophiques et parce que le mouvement de spécification allant, les conséquences d'une remise en question des choix opérés au plus haut sont de plus en plus lourdes.

Ce double mouvement est aussi lisible dans le développement de l'ontologie noyau ontoDOME : mouvement descendant depuis l'ontologie fondationnelle, mouvement ascendant depuis les ontologies de domaine, les ressources terminologiques externes utilisées, la ressource ontologique bareBMSLM (cf. section 3).

Le même double mouvement a été également à l'œuvre lors du développement des ontologies de domaine, même si le mouvement ascendant était beaucoup plus fort que le mouvement descendant. L'influence de l'ontologie fondationnelle, redoublée de celle de l'ontologie noyau (sur l'orientation vers les données), était présente mais posait une exigence moins forte sur le développement des modules que les données à intégrer. Dans le cas d'ontoDOPSY, la branche sur les signes cliniques en particulier a été l'occasion d'expérimenter une méthode de construction fondée sur les textes.

1. L'étymologie du mot « méthode » renvoie à ces deux sens – <https://www.universalis.fr/encyclopedie/methode/> – consulté le 15 août 2023.

2. Le langage utilisé pour décrire l'ontologie peut néanmoins poser des contraintes théoriques, par exemple l'adhésion à la théorie des ensembles.

3. Sur ces « conflits de direction » entre l'arbre comme plante et l'arbre comme métaphore pour caractériser une structure logique, on peut consulter l'article *The Medieval Tree of Porphyry : An Organic Structure of Logic* (Verboon, 2014).

Ce que nous décrivons ici rejoint nos considérations de la section 1.1 du chapitre 3 sur les ontologies de niveau intermédiaire, pour lesquelles nous décrivions un développement qui nous apparaissait *centripète*. Nous pourrions reprendre ce terme pour caractériser à la fois le développement de chaque module (même si le mouvement centripète est différemment exprimé selon le module) et celui de l'ontologie modulaire dans son ensemble (ce qui était notre considération initiale et concernait l'ontologie de niveau intermédiaire).

On peut penser que c'est ce qui justifie la proposition de développement « radial » (*middle-out*). C'est supposer résolu ce qui pose le plus de difficulté dans le développement d'une ontologie (son « milieu »), sauf à disposer d'une structure intermédiaire adaptée au niveau supérieur et adaptable aux niveaux inférieurs de l'ontologie. L'élaboration du patron de conception ontologique présenté dans la section 2 du chapitre 5 nous a permis de disposer d'une telle structure intermédiaire, qui s'est révélée très efficace dans le développement des modules.

1.2 Consolidation

En suivant Aimé *et al.* (2016), « le terme de *consolidation* est considéré ici dans sa définition comptable, avec le sens d'agrégation et d'opérations comme s'il s'agissait d'une seule entité ». Il sert à la définition d'un *module de consolidation* comme « module non spécifique qui importe un ensemble de modules spécifiques ou de consolidation », cette importation permettant « de spécialiser des concepts issus de modules spécifiques en ajoutant des libellés, des relations avec des concepts ou des propriétés ».

Les auteurs préconisent de respecter *a minima* un certain nombre de règles dans l'élaboration d'un tel module, réaffirmées par Cardoso *et al.* (2018) :

- « tout module de consolidation possède son propre espace de nommage »;
- « tout module de consolidation répond à une problématique (i.e. un besoin de modélisation d'un point de vue donné) »;
- « tout module de consolidation ne peut hériter que d'un ou de plusieurs modules spécifiques, ou (non exclusif) d'un ou de plusieurs modules de consolidation »;
- « toute relation d'objet d'un module de consolidation a pour domaine et codomaine des concepts dans des modules différents (sinon elles doivent être créées au sein du module concerné) ».

En outre, « un module de consolidation permet la spécialisation de concepts en provenance de modules spécifiques par l'ajout de labels, de relations avec des concepts ou d'attributs ».

De ce point de vue, ontoPSYCARE peut d'abord être considéré comme un module de consolidation : il comporte son propre espace de nommage, le besoin de modélisation auquel il répond est l'intégration des données du projet PsyCARE et il importe les modules de domaines ainsi que bareBMSLM (cf. section 3). C'est dans ontoPSYCARE que sont alignées les classes d'ontoDOME et de bareBMSLM, mais ontoPSYCARE ne contient pas d'autres classes ni d'autres relations que celles importées.

L'import d'une ontologie par une autre s'exprime par la relation owl : imports dans l'en-tête du document décrivant l'ontologie « importatrice » (en-tête qui contient un certain nombre d'informations sur l'ontologie elle-même⁴). La relation owl : imports a pour domaine et codomaine la classe owl : Ontology, et elle est transitive. C'est pourquoi il suffit d'importer dans le module de consolidation les ontologies de domaines, à la condition que celles-ci importent l'ontologie noyau, qui doit elle-même importer l'ontologie fondationnelle.

4. Pour une description de ces éléments, voir <https://www.w3.org/TR/owl-ref/#Header> – consulté le 15 août 2023.

Nous avons précisé qu'ontoPSYCARE pouvait *d'abord* être considéré comme un module de consolidation : il est possible (voire souhaitable pour faciliter son usage par des acteurs extérieurs) de fusionner le module de consolidation et toutes les ontologies qu'il importe – directement ou indirectement – en un ensemble unique. Du point de vue de l'architecture ontologique (cf. section 3.3 du chapitre 1), on revient alors à une ontologie monolithique.

Au regard de l'ontologie modulaire que nous avons présentée, ontoPSYCARE est un *module de consolidation* qui importe les autres modules et n'est pas affecté par les éventuelles modifications apportées à ces derniers (sauf modification majeure). Mais ontoPSYCARE désigne aussi une *version fusionnée*, monolithique, figée à un certain état qui est utilisée dans le processus d'intégration des données du projet PsyCARE.

2 Évaluation d'ontoPSYCARE

Comme nous l'avons vu dans la section 3.1.3 du chapitre 1, un premier aspect de l'évaluation concerne la *vérification* que l'ontologie est bien construite. De ce point de vue, le premier point à vérifier est la satisfiabilité de l'ontologie, c'est-à-dire vérifier l'absence d'axiomes et de formules aboutissant, quelle que soit l'interprétation de la formule logique, à l'ensemble vide. La satisfiabilité peut être évaluée au moyens de logiciels appelés « raisonneurs »⁵. Cette vérification a lieu régulièrement au cours du développement des ontologies et s'avère particulièrement utile lorsqu'on introduit des axiomes et des formules complexes. Nous avons utilisé régulièrement les raisonneurs Hermit, Pellet et FaCT++, qui peuvent être installés comme des *plugins* de l'éditeur d'ontologies Protégé.

Nous avons mentionné l'outil OOPS! dans la section 3.1.3 du chapitre 1. L'usage de cet outil, sur ontoPOF en particulier, nous a amenés à disjoindre les cinq grands types d'entités qui constituent le plus haut niveau de l'ontologie et à corriger des erreurs dans les domaines et codomaines des propriétés d'objet. D'une manière générale, en appliquant l'outil sur les différents modules de l'ontologie modulaire, nous avons rarement rencontré d'« erreurs critiques » (*Critical*) mais plus fréquemment des erreurs importantes (*Important*) ou mineures (*Minor*). Les erreurs les plus fréquemment rencontrées sont :

- *Missing disjointness*,
- *Missing annotations*,
- *Missing labels*,
- *Using different naming conventions in the ontology*,
- *Several classes with the same label*,
- *Untyped class*,
- *No license declared*.

L'erreur *Untyped class* semble provenir de la nature modulaire de l'ontologie : c'est une limite de l'utilisation de l'outil sur des modules séparés mais l'évaluation d'ontologies trop volumineuses aboutit à une erreur sur le site (ainsi, nous n'avons pas pu évaluer le module ontoDOPSY). A cette limitation près, OOPS! est un outil qui (a) permet de corriger les erreurs critiques et (b) amène à réfléchir, pour les erreurs importantes ou mineures, aux choix de modélisation qui les ont engendrées, qu'on décide finalement de les corriger ou pas. Notons également que cet outil est complémentaire des raisonneurs, OOPS! permettant de détecter des erreurs qui peuvent mener à des incohérences logiques (par exemple *Swapping intersection and union*).

5. L'Université de Manchester propose une liste fournie de raisonneurs, accompagnées de descriptions – <http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/> – consulté le 15 août 2023.

Nous avons porté une attention particulière, comme nous l'avons vu dans la section 2.2 du chapitre 2, à l'évaluation de l'ergonomie cognitive d'ontoPOE. Cette dimension d'une ontologie présente à la fois des aspects de vérification (comme le nombre de cycles dans l'ontologie, ce qui est également évalué par OOPS!) et des aspects de validation. Sur ces derniers aspects, nous avons opéré par comparaison avec d'autres ontologies mais l'expérimentation avec des utilisateurs serait évidemment intéressante à mener. Un élément essentiel de la validation, encore à venir au moment où nous écrivons, est l'usage en situation réelle d'ontoPSYCARE pour l'intégration de données. Toutefois, le travail opéré avec le partenaire industriel (cf. section 3) nous donne bon espoir de ce point de vue.

Enfin, l'évaluation d'une ontologie modulaire peut poser des défis particuliers. Kondylakis *et al.* (2021) proposent l'outil d'évaluation en ligne DELTA, qui n'est pas actuellement en service. Quant à l'évaluation de la modularisation en elle-même, d'Aquin *et al.* (2009) en ont proposé une grille d'analyse.

3 Intégration des données

Dans la section 2 du chapitre 1, nous avons présenté le projet PsyCARE et mentionné à cette occasion notre collaboration avec le partenaire industriel chargé de mettre en place la plateforme de collecte des données. C'est de cette collaboration dont nous rendons compte ici, la motivation principale d'ontoPSYCARE étant de faciliter l'intégration des données dans cette plateforme en proposant un modèle conceptuel englobant les données provenant de différentes sources, pour permettre leur exploration ultérieure.

3.1 Les données du projet PsyCARE

Les données utilisées par le projet PsyCARE proviennent de différentes sources :

- réutilisation de données issues de projets de recherche clinique antérieurs au projet,
- utilisation de données de projets de recherche clinique associés au projet,
- utilisation de données de l'étude prospective menée au cours du projet.

L'annexe H présente une liste détaillée des études mentionnant :

- le nom de l'étude,
- sa description succincte,
- le promoteur de l'étude,
- le statut d'avancement de l'étude,
- les types de données générées.

Elle est issue du plan de gestion des données (*Data Management Plan – DMP*) établi par Fealinx en concertation avec les différents producteurs de données au sein du projet. La figure 6.1 permet de schématiser les flux de données, depuis la production des données jusqu'à la plateforme PsyCare4.0.

3.2 La plateforme PsyCare4.0

Fealinx est le partenaire industriel du projet chargé du déploiement de la plateforme. Plus précisément, ses missions, telles que définies dans le WP5 (cf. section 2.3 du chapitre 1) sont de :

- spécifier et déployer la plateforme PsyCARE avec les données initiales (à partir des jeux de données du consortium),

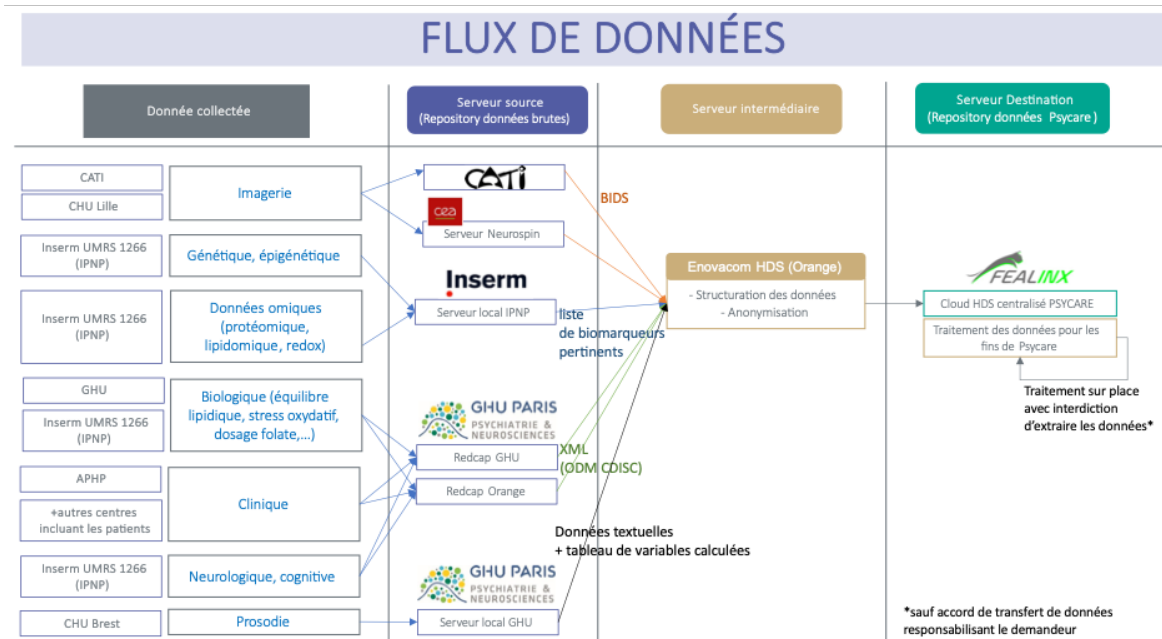


FIGURE 6.1 – Flux de données dans le projet PsyCARE (doc. DMP PsyCare4.0).

- assurer le maintien en conditions opérationnelles de la plateforme pendant toute la durée du projet et envisager le suivi technique après le projet,
- intégrer la classification définie par les ontologies du domaine,
- intégrer les flux de données provenant des autres partenaires du projet avec une gestion de la provenance des données brutes et dérivées,
- fournir des interfaces utilisateurs pour permettre l'accès aux données du projet pour les praticiens médicaux,
- assurer la gouvernance des données, en garantissant la sécurité et la confidentialité.

Sa participation au projet PsyCARE lui permet d'étendre son expertise dans l'intégration, l'organisation et la gestion de données biomédicales ainsi que dans la traçabilité de la provenance et des traitements appliqués à ces données. La solution logicielle développée, industrialisée et commercialisée par Fealinx est la solution SWOMed.

SWOMed a été développée à l'issue du projet BIOMIST⁶, en collaboration avec le Groupe d'Imagerie Neurofonctionnelle (GIN) de l'Institut des Maladies Neurodégénératives de l'Université de Bordeaux. La gestion des données s'appuie sur le logiciel de *Product Lifecycle Management* (PLM) Teamcenter⁷, édité par *Siemens Industries Software*. Son modèle de données est spécifiquement conçu pour la gestion des données de recherche biomédicales au long du cycle de vie d'une étude, comme le montre la figure 6.2.

La plateforme PsyCARE4.0 est une instance de SWOMed. La figure 6.3 présente de façon détaillée l'architecture de la plateforme. Elle est mise en œuvre sur le cloud HDS d'Enovacom⁸. Elle comporte six machines virtuelles, quatre dans la « zone de production » et deux dans la « zone de

6. <https://www.gin.cnrs.fr/fr/projets/biomist/> – consulté le 15 août 2023.

7. <https://plm.sw.siemens.com/fr-FR/teamcenter/>

8. <https://www.enovacom.fr/> – anciennement *Orange Healthcare*.

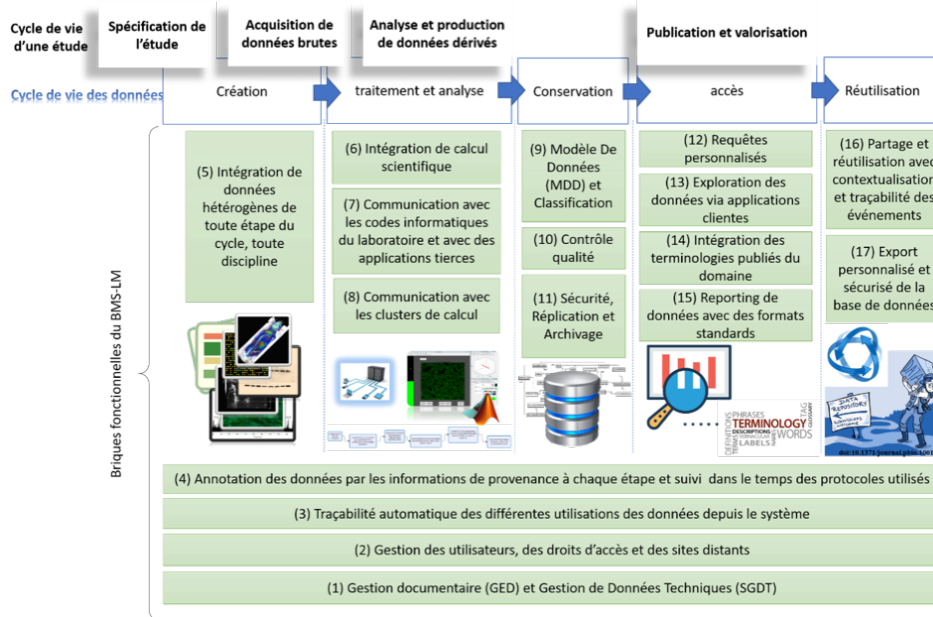


FIGURE 6.2 – SWOMed : cycle de vie des données (doc. Raboudi-Souilem (2021)).

pré-production ».

Chacune des zones est composée de trois niveaux : un niveau géré par le fournisseur (désigné ici comme *web-front*), le niveau *front* et le niveau *back*. Le niveau *back* se trouve derrière un deuxième firewall pour plus de sécurité et est utilisé pour accéder aux serveurs de base de données. L'accès en tant qu'administrateur est rendu possible par un service « bastion » avec authentification par mot de passe à durée de vie limitée, qui relaie les sessions de bureau (ou ssh) dans le navigateur, ce qui permet de créer un canal sécurisé de connexion vers les machines. L'accès des utilisateurs se fait par connexion à l'interface web de la plateforme PsyCare4.0 via https ou sur des serveurs SFTP.

Fealinx Konnexia est un serveur de *web services* comportant notamment une interface de programmation d'application respectant l'architecture REST (API REST) permettant des échanges de données, des transferts et des téléchargements de fichiers.

SWO360 est un composant de visualisation et tableaux de bords (*dashboarding*) de la suite SWOMed. Il propose une interface web avec des tableaux de données configurables et des tableaux de bord.

3.3 L'architecture commune ontoPSYCARE – BMS-LM

Un autre objectif de Fealinx lors de sa participation au projet PsyCARE est d'enrichir le modèle de données et les ontologies déjà utilisées par la plateforme SWOMed.

BMS-LM est l'ontologie noyau proposée lors de son doctorat par Amel Raboudi (Raboudi-Souilem, 2021). Elle se place dans une architecture à quatre niveaux :

- une couche de haut niveau (classes de l'OBO Foundry),
- une couche intermédiaire (et noyau) directement liée au modèle SWOMed,
- une couche « domaines » qui réutilise les ontologies de domaine publiées et reconnues

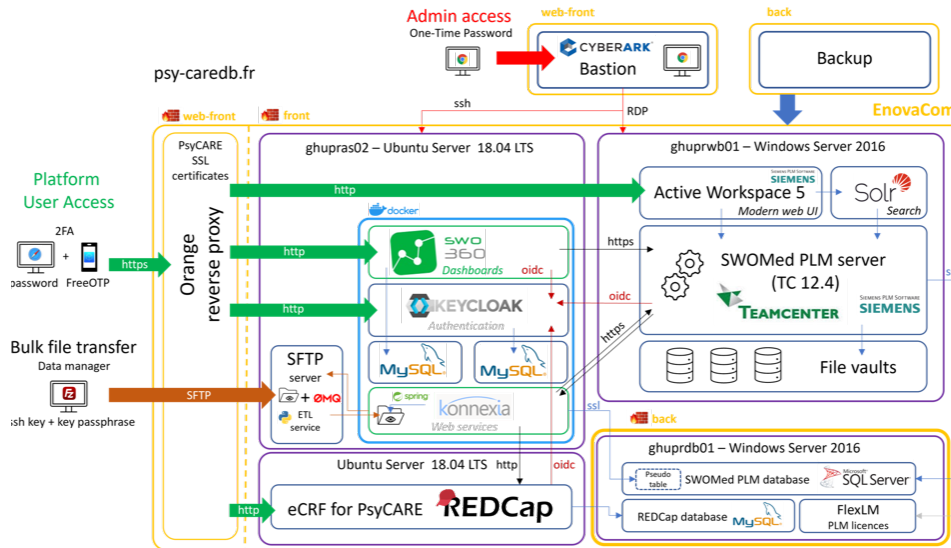


FIGURE 6.3 – Schéma de l’infrastructure de la plateforme PsyCare4.0 (doc. Fealinx).

dans un domaine particulier,

- (d) une couche « locale » qui intègre les termes vernaculaires d’un groupe de personne ou d’un domaine particulier.

En ce sens, BMS-LM enrichit le le modèle de données de SWOMed – proche de celui d’une base de donnée relationnelle – et permet de modéliser de manière générique le processus de production de données du domaine biomédical.

La figure 6.4 présente les concepts de la taxonomie de BMS-LM en indiquant les concepts parents dans BFO (*material entity*; *role*; *object aggregate*; *temporal region*; *process*; *process profile*), dans OBI (*protocol*) et dans IAO (*information content entity*, ici abrégée en « i-c-e »).

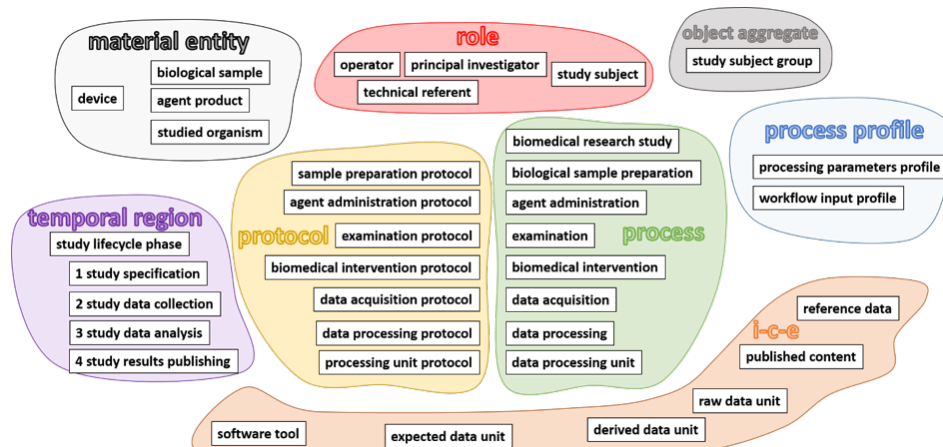


FIGURE 6.4 – Concepts de l’ontologie BMS-LM (doc. Raboudi-Souilem (2021)).

Comme nous l’avons mentionné dans la section 2 du chapitre 2, la perspective de l’alignement

avec une ontologie placée sous BFO a orienté dès le départ la conception d'ontoPOF, même si notre travail a mené à des différences notables. Au cours de la collaboration sur un rythme soutenu (hebdomadaire d'abord, puis bimensuel) avec Fealinx et en particulier avec Amel Raboudi (docteur en Génie Industriel), nous avons pu bénéficier du regard « extérieur » porté sur l'élaboration des différents modules. Certaines classes et relations d'ontoDOME en rapport avec les données ont ainsi été ajoutées pour faciliter l'alignement. Le patron de conception ontologique pour les paradigmes expérimentaux présenté dans le chapitre 5 a également fait l'objet de discussions avec Fealinx. L'ontologie BMS-LM a également été adaptée afin de pouvoir produire un alignement entre BMS-LM et ontoDOME : la version bareBMSLM ne comporte plus de références aux classes de l'*OBO Foundry*. L'annexe I présente les alignements trouvés entre bareBMSLM et ontoDOME ainsi qu'un schéma du regroupement opéré (fig. I.1).

Sur cette base, une stratégie d'intégration des données et une conception de l'architecture ontologique générale, intégrant ontoPSYCARE et bareBMSLM, a été construite. L'ontologie ontoPSYCARE a été figée en juillet 2023 en une version ontoPSYCARE-V1. Les différents modules peuvent être mis à jour et ajustés en fonction des nouveaux types de données ou des nouveaux standards à inclure, ce qui pourra donner lieu – après validation par les équipes de Fealinx et l'investigateur principal de l'étude – à une nouvelle version figée. L'ontologie ontoPSYCARE-V1 est enrichie par Fealinx des classes correspondant aux outils d'évaluation (outils, sections, items, scores, liste de choix pour les réponses) utilisés dans les différentes études et peuplée avec les instances issues des données intégrées. Cette nouvelle version, nommée « ontoPsyncareEXT », est ensuite projetée dans l'espace de la solution logicielle SWOMed et convertie vers un modèle de données générique – aligné avec les classes propres à BMS-LM – associé à un arbre de classification qui spécifie chaque concept générique. Les opérations que nous venons de décrire sont effectuées à l'aide d'un script Python qui convertit l'ontologie en classification et en objets clés pour le modèle de données générique. En amont de ces opérations, une configuration est effectuée pour assurer la cohérence avec les données intégrées et les contraintes techniques de SWOMed.

L'importation de données en vue de les rendre exploitables suppose plusieurs étapes de configuration et de mise en correspondance. La figure 6.5 permet de situer ces étapes dans le processus d'intégration et d'exploitation des données :

- ① analyse ontologique des données afin d'enrichir les annotations d'ontoPsyncareEXT et les convertir dans la classification de SWOMed;
- ② après propagation des annotations dans toutes les couches de modélisation des données (cf. plus haut), construction d'un mapping⁹;
- ③ génération de fichiers XML d'entrée pour SWOMed par un module d'agrégation (à partir de l'emplacement des données, des mappings correspondants et de la structure final d'accueil dans SWOMed) ;
- ④ import générique dans SWOMed (tout le paramétrage ayant été effectué dans les étapes précédentes)
 - l'exportation de données depuis SWOMed a plusieurs finalités :
 - ⑤ exploration des données / analyses simples via le composant SWO360,
 - ⑥ exploration des données via le composant KG,
 - exportation des données dans un format standard (CDISC, CSV, etc) pour des analyses hors-plateforme (cette voie, soumise à des contraintes réglementaires strictes et ne relevant pas du fonctionnement par défaut de l'outil, sera mise en place à l'avenir).

L'étape ⑥ appelle une remarque supplémentaire. Elle correspond au développement en cours

9. Un outil *ad hoc* a été développé chez Fealinx à cette fin. L'emploi d'un outil d'ETL plus générique est possible.

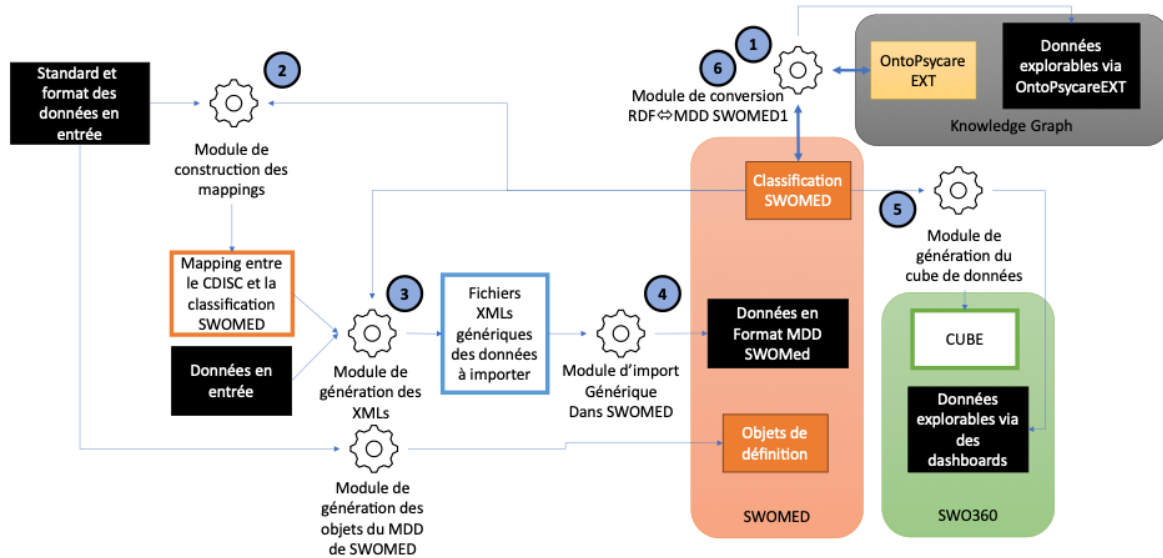


FIGURE 6.5 – Intégration des données dans SWOMed (doc. PsyCare4.0).

par Fealinx d'un composant KG (pour Knowledge Graph) permettant l'exploration des données – hébergées sur un *triplestore*¹⁰ – en se fondant sur l'ontologie. C'est dans cette perspective que l'alignement présenté plus haut entre BMS-LM et ontoDOME prend tout son sens.

4 Annotation sémantique

Le processus d'intégration des données que nous venons de décrire s'applique à tous les types de données. Toutefois, on peut les répartir selon la facilité à les exploiter informatiquement en raison de leur organisation, de leur format, éventuellement de leur stockage (on parle parfois de données *structurées* et de données *non structurées* mais une approche plus dimensionnelle que catégorique serait légitime). Dans notre cas d'usage, un enregistrement audio, une séquence de neuroimagerie, un fichier texte contenant la description clinique d'un patient, seront considérés comme non structurés, c'est-à-dire appelant des traitements supplémentaires pour être exploités. Dans le cas d'un document textuel, l'annotation sémantique peut être une manière de structurer les informations qu'il contient. C'est l'autre motivation d'ontoPSYCARE, après l'intégration de données : être une ressource ontologique pour un processus d'annotation sémantique de documents psychiatriques. Nous rendons ici compte du travail de doctorat (en cours) d'Ons Aouina pour développer une chaîne de traitement permettant l'annotation sémantique fondée sur une ontologie.

4.1 L'annotation sémantique de documents de psychiatrie

Rappelons d'abord la définition d'une annotation donnée par Bringay *et al.* (2004) : « Une **annotation** est une note particulière attachée à une **cible**. La cible peut être une collection de documents, un document, un segment de document (paragraphe, groupe de mots, mot, image ou

10. Il s'agit d'une base de données ne contenant que des triplets RDF.

partie d'image, etc.), une autre annotation. À une annotation correspond un contenu, matérialisé par une inscription, qui est une trace de la représentation mentale que l'annotateur se fait de la cible. Le contenu de l'annotation pourra être interprété à son tour par un autre lecteur. Nous appelons l'**ancree** ce qui lie l'annotation à la cible (un trait, un passage entouré, etc.) ».

À partir de cette définition, nous pouvons envisager différents types d'annotations attachées aux cibles. Nous nous intéressons ici aux annotations qui sont des classes ou des relations d'une ontologie. On parle alors d'*annotation sémantique fondée sur une ontologie*. Dans la mesure où les cibles sont, dans le cas qui nous occupe, de nature textuelle, le repérage de l'ancree supposera la mise en œuvre de techniques du traitement automatique des langues. Une fois les associations effectuées, le corpus peut être parcouru selon les annotations sémantiques ajoutées, que ce soit pour l'explorer ou pour en extraire des informations.

C'est un travail assez différent de l'intégration des données structurées dans la mesure où ces dernières sont conçues pour être manipulables informatiquement. Toutefois, si les textes qui nous intéressent ne sont pas structurés *du point de vue de l'informatique*, ils présentent la structure en réseaux du langage, sur laquelle nous pouvons nous appuyer.

4.2 Chaîne de traitement pour l'annotation

Le corpus utilisé pour l'établissement des règles et l'évaluation du processus d'évaluation est le même corpus de comptes rendus d'hospitalisation (CRH) que celui que nous avons mentionné dans la section 2.1 du chapitre 4.

L'ontologie de départ, ici, est une version fusionnée des modules ontoDOPSY, ontoMEDPSY, ontoDOME et ontoPOF. De cette ontologie est extraite une sous-ontologie ne contenant que les branches d'intérêt. À partir de cette base, un schéma d'annotation est construit. Une branche de l'ontologie dédiée à la structure des comptes rendus d'hospitalisation est ajoutée pour lier les concepts à leur contexte d'apparition dans le document, c'est-à-dire la section dans laquelle ils sont repérés (« Histoire de la maladie », « Traitement de sortie », etc.).

Nous détaillons ci-dessous les différentes étapes de la chaîne de traitement d'annotation sémantique, qui reflète le travail d'Aouina *et al.* (2023), en suivant les numéros indiqués sur la figure 6.6.

① Pré-traitement des comptes rendus médicaux

Cette première étape se décompose elle-même en plusieurs sous-étapes :

- tokenisation (qui permet d'obtenir une segmentation selon les mots, la ponctuation, etc.),
- segmentation en phrases,
- étiquetage morpho-syntaxique (*part of speech tagging* ou *POS tagging*),
- lemmatisation (pour normaliser toutes les formes fléchies selon une forme canonique),
- extraction de groupes nominaux (*noun phrase chunking* ou *NP chunking*). Ces opérations sont effectuées au moyen du *GATE Corpus Pipeline* de la boîte à outils logicielle GATE, auquel ont été ajoutés (a) un *tokenizer* adapté au français, (b) le POS tagger *Tree-Tagger* pour le français, (c) une adaptation d'*OpenNLP* (Zhang & Elhadad, 2013) au français pour obtenir une liste de groupes nominaux à partir du résultat en sortie du *Tree-Tagger*. Les signes, les symptômes, les maladies, les troubles et les événements cliniques sont ici considérés comme des groupes nominaux.

② Pré-traitement de l'ontologie

Les mêmes étapes qui viennent d'être décrites sont appliquées pour obtenir un index (*Gazetteer*) composé des concepts de l'ontologie, de leurs libellés pré-traités et des URI.

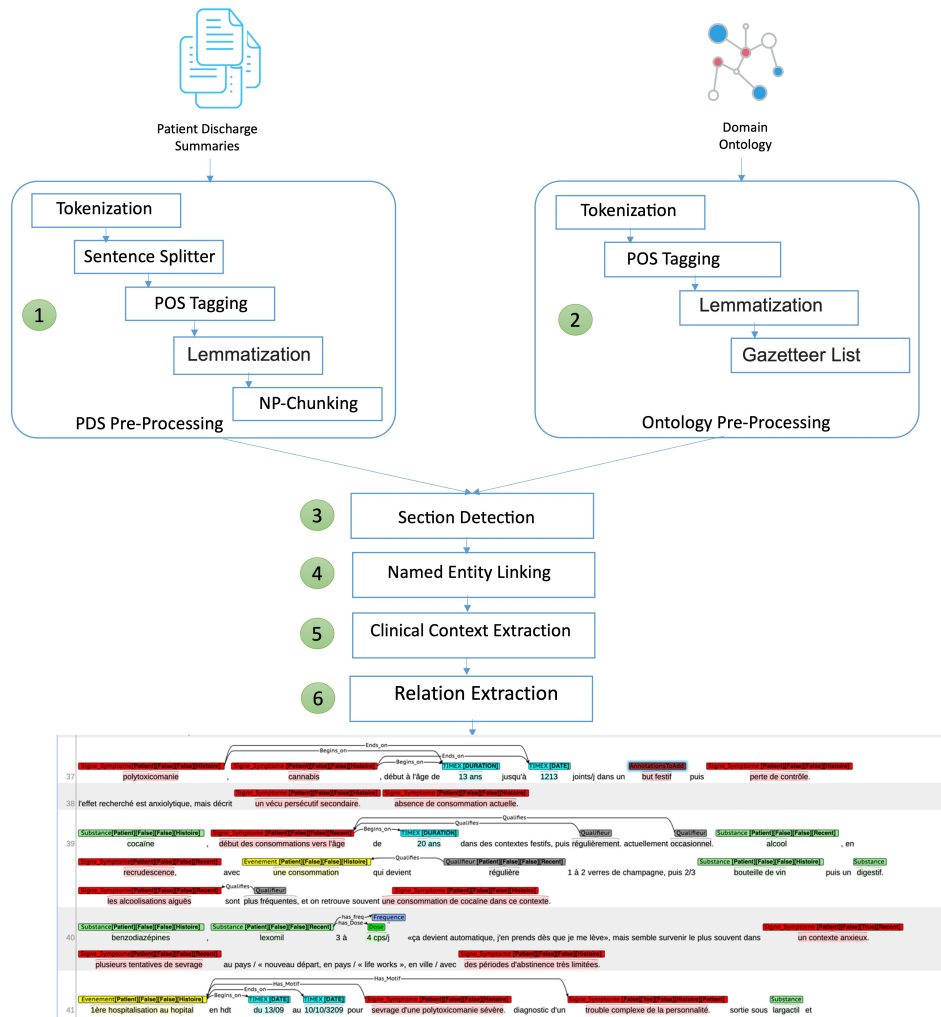


FIGURE 6.6 – Processus d'annotation (doc. O. Aouina).

③ Détection des sections

La structure discursive des comptes rendus d'hospitalisation, telle que nous l'avons décrite dans la figure 4.1 du chapitre 4, peut permettre d'améliorer la qualité des annotations, en différenciant par exemple une maladie ou un traitement passés (dans une section *Histoire de la maladie*) d'une maladie ou d'un traitement présents (au moment de la rédaction du compte rendu). Toutefois, l'identification de cette structure discursive demande une attention particulière : nous avons présenté sa forme « canonique », qui est sujette à une importante variabilité (intitulés de sections, fusion de sections, etc.). Pour rendre possible cette identification, des règles JAPE (Yacoubi *et al.*, 2011) ont été établies et une correspondance floue (*fuzzy matching*) est opérée entre une expression canonique du nom des sections et les candidats repérés par règles.

④ Reconnaissance d'entités et établissement des liaisons

Les candidats groupes nominaux obtenus lors du pré-traitement des textes sont classés selon leur proximité à des concepts de l'ontologie (système à base de règles complété

par une classification fondée sur des correspondances floues). Selon un seuil d'acceptation défini, un URI leur est associé.

Pour les entités temporelles, un *plugin* de GATE utilisant la bibliothèque Java *Stanford Temporal Tagger* (SUTime) permet d'annoter de façon normalisée les documents (avec des balises TIMEX3) et de repérer des dates, des durées et des fréquences.

⑤ Repérage des modalités (négation et hypothèse)

La détection de modalités comme la négation ou l'hypothèse est capitale pour obtenir des annotations fiables. Si « épisode maniaque franc » et « pas d'épisode maniaque franc » sont tous deux annotés comme « épisode maniaque », l'extraction d'information ou l'exploration des données qui suivent l'annotation s'en trouvent faussées.

⑥ Extraction des relations

Au-delà de l'extraction d'entités nommées, l'extraction de relations va permettre une compréhension approfondie des interactions et des événements complexes mentionnés dans des documents de psychiatrie. En ce sens, l'identification de relations sémantiques comme l'association entre un gène et un trouble, l'interaction entre une protéine et une substance chimique, la réaction indésirable à un médicament, mais aussi comme la relation entre une hospitalisation et la date à laquelle elle est intervenue, sa durée, ce qui l'a motivée, sont des éléments d'une grande valeur informative. En outre, ces relations permettent une représentation sous forme de graphe beaucoup plus élaborée que celle, radiale, reliant toutes les entités nommées au document ou à la partie de document où elles ont été trouvées. La détection de relations, historiquement, s'est fondée sur l'usage de règles. L'élaboration de ces règles est très chronophage et les règles elles-mêmes risquent d'être dépendantes du domaine et du contexte dans lequel elles ont été formulées. Des méthodes plus récentes d'extraction de relations biomédicales, à base d'apprentissage, peuvent offrir une meilleure généralisabilité.

Une évaluation du processus d'annotation sémantique des entités nommées et de leur modalité (négation, hypothèse) a été menée sur 50 comptes rendus d'hospitalisation choisis de façon aléatoire. Les deux évaluateurs présentaient un accord inter-annotateur de 0,88. Les résultats sont excellents : si l'on considère les résultats globaux, la précision est de 0,9674 ; le rappel de 0,9780 ; la mesure F de 0,9727. Quelques limites peuvent être toutefois évoquées, dont la généralisabilité à d'autres documents que les comptes rendus d'hospitalisation très structurés sur lesquels les tests ont été menés.

4.3 Développements en cours

Le schéma d'annotation qui a été utilisé dans les travaux ci-dessus constitue une simplification de l'ontologie. Il ne rend pas compte de la richesse des arborescences qui se trouvent sous *signe clinique* ou sous *trouble mental*, par exemple (même si cette richesse est utilisée dans la phase d'identification des entités nommées). Il ne rend pas non plus compte du fait que notre collaboration avec O. Aouina a pu stimuler notre réflexion sur des parties de l'ontologie et nous a amenés à ajouter des classes ou des relations (par exemple pour mieux caractériser une hospitalisation ou donner une représentation des événements présents dans un compte rendu d'hospitalisation).

Prendre en compte cette richesse pose toutefois de nombreux défis. Une piste pour les relever consiste à utiliser un outil comme Knowtator (Ogren, 2006) ¹¹, qui peut être installé comme un

11. Un Knowtator 2.0 est disponible depuis 2022 – <https://github.com/UCDenver-ccp/Knowtator-2.0> – consulté le 15 août 2023

plugin dans l'éditeur Protégé. La figure 6.7 rend compte des travaux actuels en ce sens. Les annotations produites lors du processus d'annotation sont exportées vers Knowtator. L'interface proposée facilite la correction ultérieure, par des experts, des annotations produites automatiquement, dans la mesure où ils peuvent se référer à l'ontologie.

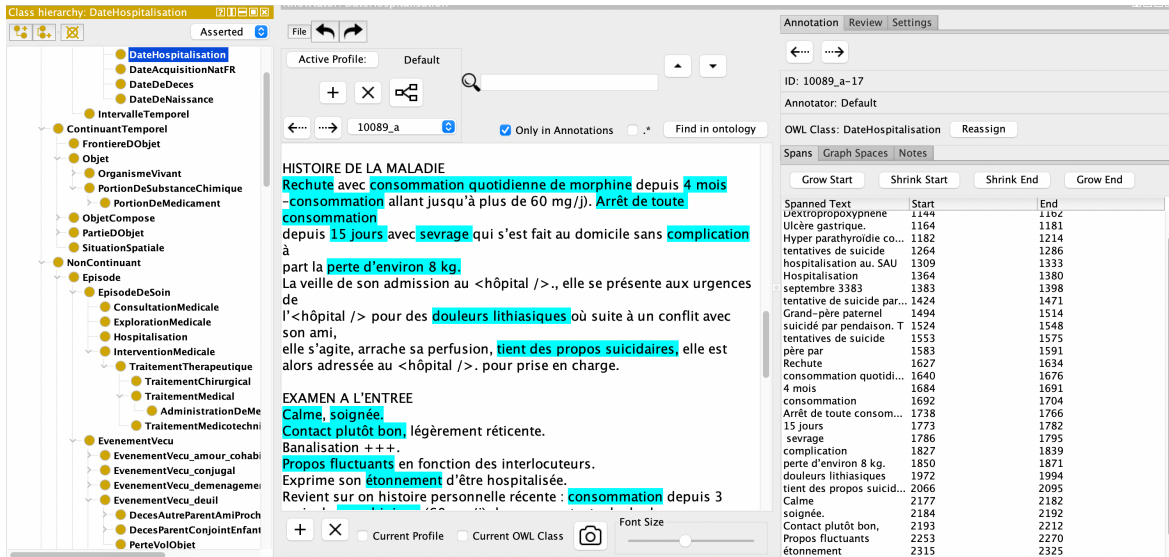


FIGURE 6.7 – Annotation avec Knowtator (doc. O. Aouina).

D'autre part, outre l'annotation fondée sur des règles exposée plus haut, le recours à des algorithmes de *deep learning* est en cours de développement, avec la perspective de repérer plus facilement les relations, à la condition de disposer d'un corpus annoté d'une taille suffisante.

5 Exploration des données dans un graphe de connaissances

Une base de connaissances (au sens d'une ontologie instanciée par des données de vie réelle) représentée sous forme de graphe peut donner lieu à de nombreuses analyses. Nous voudrions nous attacher ici plus particulièrement aux possibilités ouvertes en termes d'exploration des données. Cette section aura toutefois une tonalité programmatique. Nous avons vu dans la section 3 que du point de vue technique, la possibilité – pour le partenaire industriel – de produire un graphe de connaissances à partir des données intégrées est un futur proche. En revanche, la période du COVID et ses conséquences ont durablement ralenti l'avancée du projet PsyCARE en général (dont la date de fin a été repoussée de 24 mois) et notamment la mise en place de l'étude prospective centrale. Nous ne disposons donc pas, au moment de ce travail, des données permettant de faire des essais, de mettre en place des preuves de concept ou d'obtenir des résultats à présenter.

5.1 Les graphes de connaissances

Dans une revue critique de l'emploi et des définitions du mot *Knowledge Graph*, Ehrlinger & Wöß (2016) considèrent qu'il s'agit principalement d'un « mot à la mode » relancé par Google en 2012 (Singhal, 2012) et après avoir fixé ce qu'ils entendaient par ontologie et par base de connaissances (en l'occurrence une seule et même chose), les auteurs proposent la définition suivante :

« un graphe de connaissances acquiert et intègre des informations dans une ontologie et applique un raisonneur pour dériver de nouvelles connaissances ». Les auteurs pointent également le lien entre graphes de connaissances et web sémantique, concluant leur analyse terminologique en remarquant que « le web sémantique pourrait être interprété comme le graphe de connaissances le plus complet, ou – à l'inverse – un graphe de connaissances qui parcourt l'ensemble du web pourrait être interprété comme un web sémantique autonome ».

Nous pouvons également considérer, avec Bachimont *et al.* (2011), que nous avons là « des appellations différentes pour de multiples facettes » et que « le “gigantesque graphe global” (*Global Giant Graph*) remet en perspective ces milliers de liens entre données distribués sur le Web et le fait que, à travers les points de jointure que sont les URI, la structure de données qu'ils tissent est un graphe d'envergure mondiale » tandis que « le “web sémantique” (*semantic web*) met en avant la possibilité d'échanger les schémas de nos données et la sémantique associée afin d'enrichir la gamme des traitements automatiques qui peuvent leur être appliqués ».

Dans l'usage le plus courant de *Knowledge Graph*, il n'est pas fait référence à un formalisme particulier. Parmi ces vastes graphes de connaissances qui ont partie liée avec le web, on peut notamment citer les réalisations suivantes :

- *The Google Knowledge Graph* se présente d'abord comme une amélioration sémantique du moteur de recherche (Singhal, 2012), sans fournir de détail sur les technologies ou les ontologies utilisées. Google propose une API pour rechercher des entités dans le graphe de connaissances¹² Le graphe de connaissances comprenait, en 2019, 70 milliards de triplets pour 1 milliard d'entités (Noy *et al.*, 2019) ;
- *Facebook Open Graph* est le graphe de connaissances lié au réseau social. *Open Graph Protocol*¹³ permet aux utilisateurs de faire de leur page personnelle un nœud dans le graphe de connaissances en ajoutant des balises (sur le modèle de RDFa). Facebook propose également l'API *Graph*¹⁴ en lecture et écriture sur le réseau. Le graphe de connaissances contenait, en 2019, 500 millions d'assertions pour 50 millions d'entité primaires (Noy *et al.*, 2019) ;
- DBpedia (Bizer *et al.*, 2009) « est un effort communautaire destiné à extraire des informations à partir des pages de Wikipédia afin de les transformer en données disponibles sur le Web dans des formats structurés, les rendant aisément utilisables par des logiciels ou services en ligne » (Gandon *et al.*, 2016). Tirant parti du multilinguisme de Wikipédia, il a un « chapitre français », *DBpedia.fr*, qui compte plus de 185 millions de triplets. DBpedia a sa propre ontologie¹⁵, d'abord calquée sur les *infoboxes* de Wikipédia, puis étoffée par un effort collaboratif ;
- YAGO 4 (pour *Yet Another Great Ontology*), considérant que la hiérarchie de Wikidata peut être source de confusion, « recueille les faits concernant les instances de Wikidata, mais il les oblige à s'inscrire dans une hiérarchie de types rigoureuse avec des contraintes sémantiques. La taxonomie complexe de Wikidata est remplacée par la taxonomie plus simple et plus propre de *schema.org* » (Pellissier Tanon *et al.*, 2020). De plus des règles SHACL sont ajoutées aux classes. Le résultat est une base de connaissances de très grande ampleur (2 milliards de triplets) sur laquelle il est possible d'appliquer un raisonneur OWL2 DL ;
- Wikidata « est la plateforme centrale de gestion des données de Wikipédia » (Erxleben *et al.*, 2014) : elle est la source de données pour toutes les versions multilingues d'une même page de Wikipédia. Elle possède comme cette dernière un aspect collaboratif. Le graphe

12. <https://developers.google.com/knowledge-graph?hl=fr> – consulté le 15 août 2023.

13. <https://ogp.me/> – consulté le 15 août 2023.

14. <https://developers.facebook.com/docs/graph-api>

15. <https://www.dbpedia.org/resources/ontology/> – consulté le 15 août 2023.

de connaissances contenait en 2022 près de 1,7 milliard de triplets (Haller *et al.*, 2022). Elle a sa propre ontologie¹⁶.

D'importants graphes de connaissances ont également été constitués dans le domaine biomédical, parmi lesquels :

- PGxLOD, qui est un graphe de connaissances de la pharmacogénomique (Monnin *et al.*, 2019) agrégeant différentes sources d'information (*Pharmacogenomics Knowledgebase* ou PharmGKB; ClinVar; Uniprot; DrugBank; DisGeNET; CTD; SIDER). PGxLOD comporte un petit nombre de classes (gène, variant, médicament, maladie, phénotype) et un grand nombre de relations issues de PharmGKB et de la littérature (PubMed);
- SPOKE (pour *Scalable Precision medicine Open Knowledge Engine*), qui « est un graphe de connaissances qui relie des informations provenant de 41 bases de données spécialisées, structurées en 21 types de nœuds différents et 55 types d'arêtes, allant de la biologie moléculaire et cellulaire à la pharmacologie et à la pratique clinique » (Morris *et al.*, 2023). Elle contient plus de 27 millions de nœuds et plus de 53 millions d'arcs. Ses applications incluent « le repositionnement de médicaments, la prédiction de maladies et l'interprétation de données transcriptomiques ».
- PrimeKG entend « fournir une vision holistique et multimodale des maladies » en couvrant « largement les entités biomédicales, y compris les protéines, les gènes, les médicaments, les maladies, l'anatomie, les processus biologiques, les composants cellulaires, les fonctions moléculaires, les expositions, les phénotypes des maladies et les effets secondaires des médicaments » (Chandak *et al.*, 2023). Les auteurs notent que SPOKE ne couvrent que 137 maladies quand PrimeKG en couvre plus de 17 000 (obtenus par regroupements à partir des plus de 22 000 concepts de maladies de la *Monarch Disease Ontology* (MONDO)). PrimeKG s'appuie sur 20 ressources (des ontologies comme Uberon, *Disease Ontology*, *Gene Ontology*, *Human Phenotype Ontology*, mais aussi UMLS, *DrugBank*, DisGeNET, Reactome, SIDER). Le graphe de connaissances contient près de 130 000 nœuds et plus de 4 millions d'arcs.

Les graphes de connaissances que nous venons de présenter sont très volumineux, ce qui peut rendre difficile leur traitement par des algorithmes. Selon les cas d'usage, par exemple pour évaluer des similarités sémantiques entre nœuds, il peut être efficace de plonger le graphe (*graph embedding*) dans un espace vectoriel de moindre dimension. La recherche récente est très active dans ce domaine et de nombreuses méthodes ont été proposées pour les plongements de graphe :

- méthodes translationnelles comme TransE, TransH, TransR ou TranEdge;
- méthodes basées sur la rotation comme RotatE ;
- *graph convolutional networks* comme R-GCN, TransGCN;
- méthodes basées sur la marche comme DeepWalk ou RDF2Vec.

Une application de ces techniques peut être la complétion de graphes, par prédiction de triplets ou seulement du sujet, du prédicat, de l'objet du triplet. Un cas particulier est la classification d'entité (prédiction de l'objet dans un triplet dont le prédicat est `rdf:type`).

Notre ambition est plus modeste en souhaitant établir le graphe de connaissances d'une étude clinique, même si le nombre de triplets augmente rapidement avec le nombre de patients. Toutefois, parler de graphe rend plus immédiatement intuitif le fait qu'on peut extraire de ce graphe un sous-graphe de taille plus limitée, à partir d'un découpage *ad hoc* dans le graphe initial, mais aussi ajouter des sous-graphes au graphe initial, ou encore connecter ce graphe à d'autres graphes. Ce sont ces caractéristiques qui différencient le graphe de connaissances (et nous parlons ici d'un

16. https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology – consulté le 15 août 2023.

graphe RDF) d'une base de données relationnelles (dont le schéma est plus difficile à modifier et dans lequel les « découpes » que nous évoquons peuvent requérir des nombreuses jointures) que nous voulions mentionner avant d'envisager l'exploration des données dans un tel graphe.

5.2 L'exploration des données

Les études de recherche clinique donnant lieu à un test statistique (et obéissant au cadre méthodologique en vigueur) peuvent être schématiquement décrites comme (a) débutant par des observations, (b) se poursuivant par la définition d'une hypothèse nulle et d'une hypothèse alternative, ainsi que par le choix d'un risque α et d'un risque β puis (c) donnant lieu à un recueil des données et (d) se concluant par le test proprement dit de l'hypothèse nulle, au moyen de la p-value.

Ce que nous avons appelé ici « observations », dans une tonalité très empiriste, recouvre bien plutôt une diversité de situations. Nous sommes renvoyés aux sources de l'émergence d'une question – voire d'un problème – scientifique. Parmi ces sources, nous pouvons citer :

- les intuitions qui peuvent germer dans l'esprit du chercheur à partir de la connaissance empirique qu'il a de son domaine d'étude,
- une question pas ou peu renseignée qu'il veut investiguer,
- des résultats contradictoires entre lesquels il souhaite trancher ou dont il souhaite mieux cerner la nature contradictoire,
- un fait nouveau, qui appelle de nouvelles études.

Sans dénier la valeur du cadre méthodologique que nous avons évoqué plus haut pour guider la décision, par exemple pour trancher entre deux traitements, nous pouvons remarquer qu'il ne reflète pas nécessairement le point de vue du chercheur (pour lequel l'hypothèse nulle et l'hypothèse alternative ne sont pas nécessairement équiprobables), ni la manière de penser du chercheur (l'hypothèse alternative, celle qui motive la recherche, n'est jamais testée mais sera automatiquement acceptée si l'hypothèse nulle est rejetée), ni éventuellement l'objectif du chercheur (s'il souhaite se forger une conviction devant plusieurs hypothèses permettant de rendre compte des données qu'il observe).

Ce cadre est utile une fois qu'une question scientifique est suffisamment circonscrite. La formulation de cette question est elle-même contrainte par les informations dont on dispose sur le phénomène étudié, par l'étude qu'il est techniquement possible et éthiquement acceptable de concevoir pour y répondre, par l'utilité qu'on peut espérer de la réponse obtenue. Mais comme le rappelait Tukey (1980), « la science [...] ne commence pas par une question précise. Elle ne se termine pas non plus par une réponse toute faite ». Pour l'auteur, une telle *analyse confirmatoire des données* doit être précédée de beaucoup d'*analyse exploratoire des données* (Tukey, 1977), qu'il définit plaisamment ainsi : « c'est (1) une attitude et (2) une souplesse et (3) du papier millimétré (ou des transparents, ou les deux) ». Cette exploration concerne les données passées aussi bien que les données présentes, avant, pendant et après analyse. Elle met en jeu des techniques souvent graphiques (la « boîte à moustaches » est précisément de J. Tukey).

Ces techniques sont désormais très faciles à mettre en oeuvre de façon automatisée et le papier millimétré a cédé la place aux *dashboards*. Le risque peut alors être de réduire l'*exploratory data analysis* à une étape préliminaire, une prise de contact avec un jeu de données, donnant l'aperception suffisante pour savoir quelles techniques d'analyse lui appliquer et quels écueils sont à redouter (jeu de données déséquilibré, *outliers*, etc.).

La perspective que nous souhaitons illustrer ici est que la facilité de mettre en place des *dashboards* peut être combinée avec la possibilité offerte par les graphes de connaissances de formuler

des requêtes complexes, pour permettre au chercheur – dans une situation d'abondance de données – de retrouver l'« attitude » prônée par J. Tukey.

La richesse sémantique des relations reliant les entités dans un graphe de connaissances permet d'en extraire un sous-graphe répondant à des conditions complexes. Par exemple, on peut vouloir sélectionner tous les patients qui ont été diagnostiqués avec une maladie spécifique, ont présenté certains signes cliniques, ont suivi un traitement particulier, et ont réagi à ce traitement de manière spécifique avec la requête SPARQL suivante :

```
SELECT ?patient
WHERE {
  ?patient rdf:type ex:Patient ;
           ex:aMaladie ex:MaladieCIM-F20 ;
           ex:aSigneClinique ?signe ;
           ex:aSuiviTraitement ?traitement .
  ?traitement ex:aEffetTraitement ?reponseTraitement .
  FILTER (?signe = ex:Incurie &&
         ?reponseTraitement = ex:ResistanceAuTraitement)
}
```

Cette requête, encore assez simple mais qui pourrait être compliquée à l'envi, nécessiterait déjà, dans une base de données relationnelle, de nombreuses jointures dans des tables différentes.

Il est ensuite possible, à partir du sous-graphe obtenu, de sélectionner par exemple des variables ou des caractéristiques d'intérêt des patients et d'obtenir des représentations graphiques de la distribution de ces caractéristiques¹⁷. La capacité est ainsi offerte au chercheur de sonder une masse importante de données pour informer son intuition, voire se mettre en situation de sérendipité, afin d'aboutir à de nouvelles hypothèses de recherche.

Remarquons en outre qu'à la condition d'employer des ontologies de référence ou de produire les alignements nécessaires, le graphe de connaissances propre à une étude clinique ou à patient pourrait être lié à d'autres graphes et ainsi accroître les informations disponibles.

Au-delà de l'analyse exploratoire des données, il faudrait aussi évoquer le vaste champ de recherche autour de l'exploration ou la fouille de données (*data mining*), qui s'intéresse en particulier aux tâches de :

- prédiction (en généralisant à partir d'exemples de couples variables explicatives–variable expliquée connus) et notamment de :
 - classification (variable expliquée binaire, de type oui/non),
 - régression (variable expliquée numérique),
 - régression (ou classification) sur des séries temporelles;
- découverte de connaissances (*Knowledge Discovery*) (quand on ne dispose pas d'assez d'informations pour la prédiction) et notamment de :
 - détection d'anomalies (suspicion de valeurs aberrantes),
 - segmentation (subdivision en sous-groupes sur une caractéristique),
 - partitionnement de données (*clustering*) (agréger des cas similaires sur plusieurs caractéristiques),
 - recherche de règles d'association, analyse de séquences, analyse de liens,
 - résumés et visualisation de l'information,

17. Les *Default Views* proposées par le *Query Service* de Wikidata peuvent être ici une inspiration – https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples/fr – consulté le 15 août 2023.

- fouille de textes.

Là encore, et surtout pour les tâches de découverte de connaissances, il peut s'agir de combiner la capacité de constituer des sous-graphes à partir de requêtes complexes, avec des méthodes éprouvées d'analyse de données, pour multiplier les points de vue sur le phénomène d'intérêt et faire surgir de nouvelles hypothèses de recherche.

Conclusion

Arrivés à la fin de ce mémoire, nous souhaitons d’abord porter un regard rétrospectif sur le travail que nous avons présenté et sur les perspectives proches qu’il ouvre. L’enjeu principal en était l’élaboration d’une ontologie destinée à être le modèle conceptuel dans un processus d’intégration des données du projet PsyCARE. Comme nous l’avons expliqué, la validation de cet enjeu est en cours. Un deuxième enjeu était pour nous de proposer une première ontologie complète de la psychiatrie en langue française qui puisse servir de ressource pour l’annotation sémantique de documents de psychiatrie. Les travaux en cours allant dans ce sens montrent des premiers résultats prometteurs. Si nous regardons plus en détail les différents modules présentés, nous pouvons relever ce qui nous apparaît comme des contributions de notre travail dans le domaine de l’ingénierie des connaissances.

La prise en compte de la question de la temporalité dès l’ontologie fondationnelle ontoPOF permet à notre sens de réconcilier les positions philosophiques tridimensionnaliste et quadridimensionnaliste. Cette articulation a un coût : un objet, au sens le plus courant, ne peut plus être pris comme primitive de l’ontologie. En revanche, en articulant ce qui fait l’identité d’un objet à travers le temps avec un ensemble d’événements qui lui sont associés, et en se centrant sur ces événements, on surmonte la difficulté classique de l’indexation temporelle dans des logiques de description qui n’admettent que des relations unaires ou binaires. Outre la représentation du changement, ontoPOF favorise également la représentation de la réalisation d’un projet : dans les deux cas, il s’agit d’être mieux à même de représenter ce qui devient. Nous avons évoqué la réutilisation d’ontoPOF comme ontologie fondationnelle pour une ontologie de néphrologie. Les suites prévues de cette réutilisation pour de l’annotation sémantique de documents de néphrologie permettront de tester la robustesse d’ontoPOF au-delà du contexte dans lequel elle a été initialement conçue

Le choix de proposer une ontologie noyau des données médicales ontoDOME peut faciliter une intégration plus avancée, voire une mise en forme ontologique, du Modèle des objets de Santé de l’ANS. Ce modèle propose lui-même des alignements avec le standard d’échange de données médicales FHIR. C’est dans cette direction que nous souhaiterions faire évoluer ontoDOME, alors même que paraît un *Unofficial Draft* du W3C pour une *FHIR OWL Ontology*¹⁸. La branche de propriétés d’objet d’ontoDOME consacrée aux relations de proximité entre personnes nous paraît en elle-même une contribution ontologique intéressante. Son extension doit être éventuellement re-

18. En date du 31 août 2023 : <https://w3c.github.io/hcls-fhir-rdf/spec/ontology.html>

vue pour qu'elle corresponde le plus précisément possible aux besoins des chercheurs amenés à prendre en compte les relations familiales dans des travaux fondés sur la génomique.

L'ontologie des données de psychiatrie ontoDOPSY fait plusieurs propositions que nous souhaitons ici mettre en avant : une branche de signes cliniques neuropsychiatriques et une branche d'événements vécus toutes deux orientées vers l'annotation sémantique ; une branche de domaines d'intérêt psychiatrique constituant une taxonomie d'indexation ; un modèle descriptif pour les outils d'évaluations. D'autre part, le wiki TIPSY, proposé pour faire évoluer la branche de domaines d'intérêt psychiatrique, s'il ressortit plus clairement de la gestion que de l'ingénierie des connaissances, porte néanmoins en lui l'idée d'une branche d'une ontologie – et pourquoi pas d'une ontologie entière – élaborée par une communauté large d'experts. Cette manière de progresser dans l'élaboration d'une ontologie de référence pour la psychiatrie, en implémentant en outre une extension sémantique (*Semantic MediaWiki*) pour le wiki actuel, nous semble mériter d'être expérimentée. Comme nous l'avons signalé, l'ontologie consacrée aux médicaments en psychiatrie, ontoMEDPSY, pourrait voir sa structure profondément remaniée à l'aune des travaux récents d'élaboration d'un modèle du médicament, concrétisé en une ontologie, par l'ANS. D'autre part, une étude menée conjointement avec des structures hospitalières pourrait aider à préciser son périmètre.

Le patron de conception ontologique pour les paradigmes expérimentaux (et ce qui peut y être assimilé) peut être utile pour toutes les techniques et les spécialités convoquées par la médecine de précision, même s'il n'est pas une solution clé en main. La poursuite des discussions entamées lors de son dépôt sur le site de l'ODPA permettra de renforcer ses assises théoriques ; l'extension de son usage pourrait amener à l'ajuster afin de renforcer l'utilité pratique que nous lui avons déjà trouvée. L'ontologie des données d'imagerie ontDOIMA est appelée à s'étendre prochainement aux EEG. Nous considérons comme éminemment souhaitable de pouvoir associer, dans l'ontologie des données de biologie ontoDOBIO, les codes LOINC aux examens de biologie. L'ontologie des examens de biologie moléculaire ontoDOOMICS reste à mettre en place. L'ontologies des données d'analyse du discours ontoDODISC et l'ontologie des données d'évaluation de la dextérité ontoDODEXT sont les ontologies les plus directement élaborées sur mesure pour le projet PsyCARE. L'une comme l'autre sont susceptibles d'être ouvertes à d'autres formes, respectivement d'analyse du discours ou d'évaluation de la dextérité, en ajoutant les paradigmes et les variables correspondants.

Enfin, nous voulions mentionner un travail en phase d'achèvement qui n'est pas présenté dans cette thèse mais qui en est à bien des égards un reflet. Il s'agit d'un article sur la notion de disposition ancré d'une part dans la psychiatrie, d'autre part dans la philosophie et enfin dans l'ingénierie des connaissances. En philosophie, la disposition, ou propriété dispositionnelle, est une propriété assertée alors qu'elle n'est pas observée chez son porteur mais donne lieu dans certaines circonstances à une manifestation¹⁹. L'exemple le plus trivial est la solubilité du sucre, propriété assertée du morceau de sucre qui n'est pas actuellement (et ne sera peut-être jamais) plongé dans l'eau. Dans le domaine médical, la disposition permet de concevoir une vulnérabilité constatée à une maladie – en l'occurrence dans ce travail la vulnérabilité à la schizophrénie – chez des patients qui ne développent pourtant pas la maladie. Dans le domaine des ontologies, la représentation de la disposition est une question repérée comme difficile. Nous avons vu dans la section 1.2.1 du chapitre 3 que l'ontologie OGMS incluait la disposition dans sa modélisation de la maladie, ce qui a généré un ensemble de propositions sur la représentation ontologique des dispositions. Ce

19. Ce qui peut appeler une mise en forme logique par des conditionnels contrefactuels : si P *était* vrai alors q *serait* vrai.

travail est donc l'occasion de voir comment ces trois approches se répondent et l'état actuel de la représentation ontologique des dispositions. Ce triple ancrage est aussi, nous semble-t-il, une caractéristique du travail que nous avons présenté dans ce mémoire.

Nous souhaitons aussi porter un regard plus résolument prospectif, orienté vers la question de la causalité. Dans l'article *A Second Chance to Get Causal Inference Right* de Hernán *et al.* (2019), les auteurs présentent une trichotomie des tâches en science des données :

- la description, c'est-à-dire un « résumé quantitatif de certaines caractéristiques du monde »,
- la prédiction, qui « consiste à utiliser des données pour mettre en correspondance certaines caractéristiques du monde (les entrées) avec d'autres caractéristiques du monde (les sorties) »,
- la prédiction contrefactuelle, qui « consiste à utiliser des données pour prédire certaines caractéristiques du monde comme si le monde avait été différent, ce qui est nécessaire dans les applications d'inférence causale »²⁰.

Notons avec les auteurs que l'inférence causale n'aboutit pas à identifier des mécanismes causaux mais à quantifier des effets causaux, de même que « les essais randomisés identifient des effets causaux même si les mécanismes causaux qui les expliquent sont inconnus ». La différence entre prédiction et inférence causale tient au rôle de la connaissance des experts et plus particulièrement de leur connaissance de la structure causale du système étudié. C'est cette connaissance qui permettra notamment d'identifier des facteurs de confusion. Pour reprendre des exemples donnés par les auteurs, une mauvaise prise en compte des facteurs de confusion par des algorithmes prédictifs peut mener au « paradoxe du poids de naissance » (la mortalité parmi les bébés avec un faible poids à la naissance est moindre si leur mère est fumeuse); d'autre part, les algorithmes prédictifs n'aident pas nécessairement à la décision (« une hospitalisation antérieure peut être identifiée comme un prédicteur utile de la mortalité [pour des patients souffrant d'insuffisance cardiaque], mais personne ne suggérerait d'arrêter d'hospitaliser les gens pour réduire la mortalité »). Au contraire, la logique contrefactuelle au cœur de l'inférence causale amène à se poser des questions de type « et si? » (par exemple « comparer le profil bénéfice-risque de la transplantation cardiaque par rapport au traitement médical chez les patients présentant une certaine sévérité d'insuffisance cardiaque »).

La connaissance par les experts de la structure causale de leur domaine est donc essentielle. Elle est difficile à acquérir et souvent incomplète, notamment (mais pas seulement) en psychiatrie. D'autre part, la notion de causalité doit elle-même être interrogée. Des fameuses « quatre questions » du biologiste N. Tinbergen²¹, beaucoup de disciplines scientifiques n'ont conservé que les notions de causalité distale et de causalité proximale.

Pour prendre un exemple où la logique contrefactuelle est à nouveau à l'œuvre, le déclencheur d'une disposition est dans une relation de causalité proximale avec la manifestation de la disposition (la mise en contact avec l'allergène est la cause proximale de la crise allergique, chez la personne allergique). Remarquons que nous avons pris soin de dire « la mise en contact avec l'allergène » et pas simplement « l'allergène ». Si l'on souscrit au modèle déductif-nomologique de C. G. Hempel, la causalité est affaire d'événements.

20. Au début du XX^e siècle, les fondateurs des statistiques modernes ont fermé la porte à l'inférence causale utilisant des données observationnelles. La seconde chance dont il est ici question est la renaissance actuelle de la sciences des données, si celle-ci ne se cantonne pas aux tâches de prédiction.

21. Ces questions concernent les causes d'un comportement animal envisagées du point de vue (a) de la fonction adaptative, (b) de l'évolution phylogénétique, (c) des mécanismes causaux, (d) du développement ontogénétique – https://en.wikipedia.org/wiki/Tinbergen%27s_four_questions – consulté le 15 août 2023.

Si la causalité est une relation entre des événements, notre intérêt se porte souvent de manière préférentielle sur ce qu'il advient – au cours d'enchaînements causaux – aux participants à ces événements. Ces participants peuvent se situer à différentes échelles du vivant, des atomes aux populations en passant par les molécules, les cellules, les tissus, les organes, les organismes. Il n'y a pas, à notre connaissance, d'échelles correspondantes pour les événements dans lesquels sont impliqués spécifiquement des atomes, des molécules, etc. Elles seraient pourtant nécessaires à l'analyse de la causalité à plusieurs niveaux.

Dans le livre *Behave*, consacré aux processus biologiques influençant le comportement humain, Sapolsky (2017) expose la pluralité des explications à considérer, selon différentes échelles temporelles. Pour rendre compte d'un comportement, on s'intéressera du point de vue neurologique à ce qui s'est passé la seconde d'avant ; on cherchera ce qui s'est passé dans l'environnement ou les stimuli perceptifs de quelques secondes à quelques minutes avant ; on étudiera les taux d'hormones de quelques heures à quelques jours avant ; on s'intéressera du point de vue génétique aux milliers d'années qui ont précédé le comportement.

Un comportement, dans cette manière d'envisager son explication, doit donc être considéré selon différentes strates. Cela n'est pas sans rapport avec la description que nous avons faite dans la section 1.1 du chapitre 4 de toutes les interactions que la psychiatrie était amenée à prendre en compte, ni sans rapport avec le projet RDoC présenté dans la section 2.2 du chapitre 4.

Le type de modélisation que nous envisageons ici est évidemment d'une très grande complexité, à la mesure des phénomènes à expliquer. Il ne préjuge pas du modèle d'explication le plus pertinent (réductionniste, émergentiste, holiste ou leur combinaison). L'ingénierie des connaissances a selon nous sa place dans un tel projet de modélisation, ainsi que les ontologies dans toute leur complexité, la voie ontologique pour la psychiatrie de précision s'inscrivant ici dans le chemin de la science.

Références

- ABDELGHANY A., DARWISH N. & HEFNI H. (2019). An Agile Methodology for Ontology Development. *International Journal of Intelligent Engineering and Systems*, **12**(2), 170–181. 26
- AIMÉ X., GEORGE S. & HORNING J. (2016). Vetivoc : A modular ontology for the fashion, textile and clothing domain. *Applied Ontology*, **11**(1), 1–28. 117
- ALEXANDER C., ISHIKAWA S., SILVERSTEIN M. & JACOBSON M. (1977). *A Pattern Language : Towns, Buildings, Construction*. Number 2 in Center for Environmental Structure Series. New York, NY : Oxford Univ. Press, 41. print edition. 27
- ALLEN J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, **26**(11), 832–843. 54
- AOUINA O., HILBEY J. & CHARLET J. (2023). Ontology-Based Semantic Annotation of French Psychiatric Clinical Documents. In M. HÄGGLUND, M. BLUSI, S. BONACINA, L. NILSSON, I. CORT MADSEN, S. PELAYO, A. MOEN, A. BENIS, L. LINDSKÖLD & P. GALLOS, Coordinateurs, *Studies in Health Technology and Informatics*. IOS Press. 125
- ARISTOTE (2002). *Catégories*. Number 415 in Collection des universités de France Série grecque. Paris : Belles Lettres, 2. tir edition. 45
- ARP R. & SMITH B. (2008). Function, Role, and Disposition in Basic Formal Ontology. *Nature Precedings*, p. 1–1. 54
- ARP R., SMITH B. & SPEAR A. D. (2015). *Building Ontologies with Basic Formal Ontology*. Cambridge, Massachusetts : Massachusetts Institute of Technology. 39
- AUSSENAC-GILLES N., DESPRES S. & SZULMAN S. (2008). The TERMINAE Method and Platform for Ontology Engineering from Texts. In P. BUITELAAR & P. CIMIANO, Coordinateurs, *Ontology Learning and Population : Bridging the Gap between Text and Knowledge*, volume 167 of *Frontiers in Artificial Intelligence and Applications*, p. 199–223. IOS Press. 26
- BAADER F., HORROCKS I., LUTZ C. & SATTLER U. (2017). *An Introduction to Description Logic*. Cambridge : Cambridge University Press. 21
- BACHIMONT B. (2000). Engagement sémantique et engagement ontologique : Conception et réalisation d'ontologies en ingénierie des connaissances. In J. CHARLET, M. ZACKLAD, G. KASSEL & D. BOURIGAULT, Coordinateurs, *Ingénierie Des Connaissances : Évolutions Récentes et Nouveaux Défis*. Paris : Eyrolles. 26, 37, 38
- BACHIMONT B. (2004). *Arts et Sciences Du Numerique : Ingenierie Des Connaissances et Critique de La Raison Computationnelle*. Habilitation à diriger des recherches, Université de Technologie de Compiègne. 65

- BACHIMONT B., GANDON F., POUPEAU G., VATANT B., TRONCY R., GANDON F., POUYLLAU S., MARTINEZ R., BATTISTI M. & ZACKLAD M. (2011). Enjeux et technologies : des données au sens. *Documentaliste-Sciences de l'Information*, **48**(4), 24–41. 32, 129
- BACHIMONT B., ISAAC A. & TRONCY R. (2002). Semantic Commitment for Designing Ontologies : A Proposal. In A. GÓMEZ-PÉREZ & V. R. BENJAMINS, Coordinateurs, *Knowledge Engineering and Knowledge Management : Ontologies and the Semantic Web*, Lecture Notes in Computer Science, p. 114–121, Berlin, Heidelberg : Springer. 26
- BANDROWSKI A., BRINKMAN R., BROCHHAUSEN M., BRUSH M. H., BUG B., CHIBUCOS M. C., CLANCY K., COURTOT M., DEROM D., DUMONTIER M., FAN L., FOSTEL J., FRAGOSO G., GIBSON F., GONZALEZ-BELTRAN A., HAENDEL M. A., HE Y., HEISKANEN M., HERNANDEZ-BOUSSARD T., JENSEN M., LIN Y., LISTER A. L., LORD P., MALONE J., MANDUCHI E., MCGEE M., MORRISON N., OVERTON J. A., PARKINSON H., PETERS B., ROCCA-SERRA P., RUTTENBERG A., SANSONE S.-A., SCHEUERMANN R. H., SCHÖBER D., SMITH B., SOLDATOVA L. N., STOECKERT C. J., TAYLOR C. F., TORNIAI C., TURNER J. A., VITA R., WHETZEL P. L. & ZHENG J. (2016). The Ontology for Biomedical Investigations. *PLoS ONE*, **11**(4), e0154556. 63
- BARTON A. & ROSIER A. (2016). Ontologies appliquées biomédicales et ontologie philosophique : Un développement complémentaire. *Lato Sensu, Revue de la Société de philosophie des sciences*, **3**, 1–8. 36
- BERGSON H. (2017). *L'énergie spirituelle*. Quadrige. Paris : PUF, nouvelle éd. édition. 74
- BIN K. (2008). Vers une psychopathologie en première personne. *Laval théologique et philosophique*, **64**(2), 377–385. 74
- BIOMARKERS DEFINITIONS WORKING GROUP. (2001). Biomarkers and surrogate endpoints : Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, **69**(3), 89–95. 8
- BIZER C., LEHMANN J., KOBILAROV G., AUER S., BECKER C., CYGANIAK R. & HELLMANN S. (2009). DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantics*, **7**(3), 154–165. 129
- BLEI D. M., NG A. Y. & JORDAN M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, **3**(null), 993–1022. 11
- BODENREIDER O. (2008). Biomedical ontologies in action : Role in knowledge management, data integration and decision support. *Yearbook of Medical Informatics*, p. 67–79. 31
- BONA J., KOHN G. & RUTTENBERG A. (2015). Ontology-driven patient history questionnaires. In *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO 2015)*, volume 1515, Lisbon, Portugal. 87
- BORGO S., GALTON A. & KUTZ O. (2022). Foundational ontologies in action : Understanding foundational ontology through examples. *Applied Ontology*, **17**(1), 1–16. 44
- BORSBOOM D. (2017). A network theory of mental disorders. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, **16**(1), 5–13. 84
- BORSBOOM D. & CRAMER A. O. J. (2013). Network analysis : An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, **9**, 91–121. 84
- BORST W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, University of Twente, Netherlands. 20
- BRINGAY S., BARRY C. & CHARLET J. (2004). Les documents et les annotations du dossier patient hospitalier. *Revue I3-Information Interaction Intelligence*, **4**(1). 124
- BRUGHA T., BEBBINGTON P., TENNANT C. & HURRY J. (1985). The List of Threatening Experiences : A subset of 12 life event categories with considerable long-term contextual threat. *Psychological Medicine*, **15**(1), 189–194. 87
- BUREK P., SCHERF N. & HERRE H. (2019). Ontology patterns for the representation of quality changes of cells in time. *Journal of Biomedical Semantics*, **10**(1), 16. 43
- CAPELLE H., MATHIAS F., JEAN C., AMIRAT-COMBRALIER V. & BERTAULT-PERES P. (2016). Analyse d'ordonnance en psychiatrie adulte : Interventions pharmaceutiques et mise en place d'actions correctives dans les unités de soins. In *Congrès de La Société Française de Pharmacie Clinique (SFPC)*, volume XVI, Montpellier, France. 94

- CARDOSO S. (2019). *Apports de la modélisation ontologique pour l'analyse des ruptures de parcours de soins dans la Sclérose Latérale Amyotrophique*. PhD thesis, Sorbonne Université. 68
- CARDOSO S., AIME X., MEININGER V., GRABLI D., COHEN K. B. & CHARLET J. (2018). De l'intérêt des ontologies modulaires. Application à la modélisation de la prise en charge de la SLA. In S. RANWEZ, Coordinateur, *29es Journées Francophones d'Ingénierie Des Connaissances, IC 2018*, 29es Journées Francophones d'Ingénierie Des Connaissances, IC 2018, p. 121–128, Nancy, France : AFIA. 117
- CEUSTERS W. & SMITH B. (2010). Foundations for a realist ontology of mental disease. *Journal of Biomedical Semantics*, **1**(1), 10. 77
- CHANDAK P., HUANG K. & ZITNIK M. (2023). Building a knowledge graph to enable precision medicine. *Scientific Data*, **10**(1), 67. 130
- CHARLET J., BACHIMONT B., MAZUEL L., DHOMBRES F., JAULENT M. & BOUAUD J. (2012). OntoMénélas. Motivations et retours d'expérience sur l'élaboration d'une ontologie noyau de la médecine. *Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques*, **31**, 125–147. 41
- CHAUMETTE B. (2016). *Identification de Facteurs Biologiques de La Transition Psychotique*. Thèse, Sorbonne Paris Cité. 102
- CIOMPI L. (1984). Model concepts of the interaction of biological and psychosocial factors in schizophrenia. *Fortschritte Der Neurologie-Psychiatrie*, **52**(6), 200–206. 17
- COLLÈGE NATIONAL DES UNIVERSITAIRES EN PSYCHIATRIE, ASSOCIATION POUR L'ENSEIGNEMENT DE LA SÉMIOLOGIE PSYCHIATRIQUE & COLLÈGE UNIVERSITAIRE NATIONAL DES ENSEIGNANTS EN ADDICTOLOGIE, Coordinateurs (2021). *Référentiel de psychiatrie et addictologie : psychiatrie de l'adulte, psychiatrie de l'enfant et de l'adolescent, addictologie*. L'officiel ECN. Tours : Presses universitaires François-Rabelais, 3e éd édition. 79, 81
- COLLINS F. S., MORGAN M. & PATRINOS A. (2003). The Human Genome Project : Lessons from large-scale biology. *Science (New York, N.Y.)*, **300**(5617), 286–290. 7
- COLLINS F. S. & VARMUS H. (2015). A new initiative on precision medicine. *The New England Journal of Medicine*, **372**(9), 793–795. 9
- COOPER R. (2013). Natural Kinds. In K. FULFORD, M. DAVIES, R. G. GIPPS, G. GRAHAM, J. Z. SADLER, G. STANGHELLINI & T. THORNTON, Coordinateurs, *Oxford Handbook of Philosophy and Psychiatry*, volume 1. Oxford University Press. 75
- CORCHO O., FERNÁNDEZ-LÓPEZ M. & GÓMEZ-PÉREZ A. (2007). Ontological Engineering : What are Ontologies and How Can We Build Them? In *Semantic Web Services : Theory, Tools and Applications*, p. 44–70. IGI Global. 24, 25
- CORCORAN C. M., CARRILLO F., FERNÁNDEZ-SLEZAK D., BEDI G., KLIM C., JAVITT D. C., BEARDEN C. E. & CECCHI G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, **17**(1), 67–75. 103
- COX S. J. & LITTLE C. (2022). Time Ontology in OWL. <https://www.w3.org/TR/owl-time/>. 70
- CRISP T. M. & SMITH D. P. (2005). 'Wholly Present' Defined. *Philosophy and Phenomenological Research*, **71**(2), 318–344. 43
- CUTHBERT B. N. & INSEL T. R. (2013). Toward the future of psychiatric diagnosis : The seven pillars of RDoC. *BMC Medicine*, **11**, 126. 83
- DAHLBERG I. (2008). The Information Coding Classification (ICC) : A Modern, Theory-Based Fully-Faceted, Universal System of Knowledge Fields. *Axiomathes*, **18**(2), 161–176. 56
- DANET F. (2013). La psychiatrie, une spécialité marginale dans la modernité. *L'information psychiatrique*, **Volume 89**(1), 33–41. 74
- D'AQUIN M., SCHLICHT A., STUCKENSCHMIDT H. & SABOU M. (2009). Criteria and Evaluation for Ontology Modularization Techniques. In H. STUCKENSCHMIDT, C. PARENT & S. SPACCAPIETRA, Coordinateurs, *Modular Ontologies : Concepts, Theories and Techniques for Knowledge Modularization*, Lecture Notes in Computer Science, p. 67–89. Berlin, Heidelberg : Springer. 119

- DE NICOLA A., MISSIKOFF M. & NAVIGLI R. (2005). A Proposal for a Unified Process for Ontology Building : UPON. In K. V. ANDERSEN, J. DEBENHAM & R. WAGNER, Coordinateurs, *Database and Expert Systems Applications, Lecture Notes in Computer Science*, p. 655–664, Berlin, Heidelberg : Springer. 25
- DECLERCK G., BANEYX A., AIMÉ X. & CHARLET J. (2012). A quoi servent les ontologies fondationnelles? In *23èmes Journées Francophones d'Ingénierie Des Connaissances (IC 2012)*, p. pp. 67–82, Paris, France. 37, 41
- DECLERCK G. & CHARLET J. (2014). Pourquoi notre sémantique naïve n'est pas formalisable et pourquoi c'est (presque) sans conséquence sur l'ingénierie ontologique. *Intellectica. Revue de l'Association pour la Recherche Cognitive*, **61**(1), 143–188. 65
- DEMAZEUX S. (2008). Les catégories psychiatriques sont-elles dépassées? *Philonsorbonne*, **2**, 67–88. 75
- DHOMBRES F., JOUANNIC J.-M., JAULENT M.-C. & CHARLET J. (2010). Choix méthodologiques pour la construction d'une ontologie de domaine en médecine périnatale. In *21èmes Journées Francophones d'Ingénierie des Connaissances*, p. 171 : Ecole des Mines d'Alès. 41
- DING H. & ZHANG Y. (2023). Speech Prosody in Mental Disorders. *Annual Review of Linguistics*, **9**(1), 335–355. 103
- DUMAS G., CERMOLACCE M., BATAIL J. M., QUILES C. & MICOULAUD-FRANCHI J.-A. (2019). Vers une approche physiologique de la sémiologie en psychiatrie. Partie 2 : perspectives offertes par la biologie systémique. *Annales Médico-psychologiques, revue psychiatrique*, **177**(3), 289–294. 76
- ECKER C., MARQUAND A., MOURÃO-MIRANDA J., JOHNSTON P., DALY E. M., BRAMMER M. J., MALTEZOS S., MURPHY C. M., ROBERTSON D., WILLIAMS S. C. & MURPHY D. G. M. (2010). Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, **30**(32), 10612–10623. 101
- EHLINGER L. & WÖSS W. (2016). Towards a Definition of Knowledge Graphs. In *Joint Proceedings of the Posters and Demos Track of 12th International Conference on Semantic Systems (SEMANTiCS2016) and 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS16)*, volume 1695, Leipzig, Germany. 128
- ENGEL G. L. (1980). The clinical application of the biopsychosocial model. *The American Journal of Psychiatry*, **137**(5), 535–544. 74
- ERXLBEN F., GÜNTHER M., KRÖTZSCH M., MENDEZ J. & VRANDEČIĆ D. (2014). Introducing Wikidata to the Linked Data Web. In P. MIKA, T. TUDORACHE, A. BERNSTEIN, C. WELTY, C. KNOBLOCK, D. VRANDEČIĆ, P. GROTH, N. NOY, K. JANOWICZ & C. GOBLE, Coordinateurs, *The Semantic Web – ISWC 2014, Lecture Notes in Computer Science*, p. 50–65, Cham : Springer International Publishing. 129
- FABRY P., BARTON A. & ETHIER J.-F. (2020). QUESTO – An Ontology for Questionnaires. In *11th International Conference on Biomedical Ontologies (ICBO 2020)*, volume 2807 of *CEUR Workshop Proceedings*, p. B.1–12, Bolzano, Italy : CEUR-WS.org. 87
- FALISSARD B. (2008). *Mesurer la subjectivité en santé : perspective méthodologique et statistique*. Abrégés. Issy-les-Moulineaux : Elsevier-Masson, 2e éd edition. 90
- FERNÁNDEZ-LÓPEZ M., GÓMEZ-PÉREZ A. & JURISTO N. (1997). METHONTOLOGY : From Ontological Art Towards Ontological Engineering. In *AAAI Conference on Artificial Intelligence*. 24
- FRAJERMAN A., SCORIELS L., KEBIR O. & CHAUMETTE B. (2021). Shared Biological Pathways between Antipsychotics and Omega-3 Fatty Acids : A Key Feature for Schizophrenia Preventive Treatment? *International Journal of Molecular Sciences*, **22**(13), 6881. 102
- GALTON A. & MIZOGUCHI R. (2009). The water falls but the waterfall does not fall : New perspectives on objects, processes and events. *Applied Ontology*. 44, 45
- GANDON F., BOYER R. & MONNIN A. (2016). DBpédia.fr : retour sur la publication de données de la culture française. *I2D - Information, données & documents*, **53**(2), 50. 129
- GANGEMI A., CATENACCI C., CIARAMITA M. & LEHMANN J. (2005). *Ontology Evaluation and Validation : An Integrated Formal Model for the Quality Diagnostic Task*. Rapport interne, Laboratory of Applied Ontologies – CNR, Rome, Italy. 28, 46, 48

- GANGEMI A., CATENACCI C., CIARAMITA M. & LEHMANN J. (2006). Modelling Ontology Evaluation and Validation. In Y. SURE & J. DOMINGUE, Coordinateurs, *The Semantic Web : Research and Applications*, Lecture Notes in Computer Science, p. 140–154, Berlin, Heidelberg : Springer. 30
- GANGEMI A., GUARINO N., MASOLO C., OLTRAMARI A. & SCHNEIDER L. (2002). Sweetening ontologies with DOLCE. In *Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management*, volume 2473, p. 166–181. 44, 55
- GANGEMI A. & PRESUTTI V. (2009). Ontology Design Patterns. In S. STAAB & R. STUDER, Coordinateurs, *Handbook on Ontologies*, International Handbooks on Information Systems, p. 221–243. Berlin, Heidelberg : Springer. 105
- GÄRDENFORS P. (2000). *Conceptual Spaces : The Geometry of Thought*. The MIT Press. 55
- GAULD C. & MICOULAUD-FRANCHI J. A. (2021). Analyse en réseau par fouille de données textuelles systématique du concept de psychiatrie personnalisée et de précision. *L'Encéphale*, **47**(4), 341–347. 13
- GAVRILOVA T., GOROVY V. & BOLOTNIKOVA E. (2010). Cognitive ergonomics of teaching ontologies. In *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, ECCE '10, p. 271–278, New York, NY, USA : Association for Computing Machinery. 47, 48
- GERBAUD T. (2023). L'intelligence artificielle n'existe (toujours) pas. 13
- GIBAUD B., KASSEL G., DOJAT M., BATRANCOURT B., MICHEL F., GAINARD A. & MONTAGNAT J. (2011). NeuroLOG : Sharing neuroimaging data using an ontology-based federated approach. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, **2011**, 472–480. 114
- GINZBURG C. (1980). Signes, traces, pistes : Racines d'un paradigme de l'indice. *Le Débat*, **6**(6), 3. 79
- GÓMEZ-PÉREZ A. (2004). Ontology Evaluation. In S. STAAB & R. STUDER, Coordinateurs, *Handbook on Ontologies*, International Handbooks on Information Systems, p. 251–273. Berlin, Heidelberg : Springer. 28
- GONÇALVES R. S., TU S. W., NYULAS C. I., TIERNEY M. J. & MUSEN M. A. (2017). An ontology-driven tool for structured data acquisition using Web forms. *Journal of Biomedical Semantics*, **8**(1), 26. 87
- GOTTESMAN I. I. & SHIELDS J. (1966). Schizophrenia in twins : 16 years' consecutive admissions to a psychiatric clinic. *The British Journal of Psychiatry : The Journal of Mental Science*, **112**(489), 809–818. 16
- GOURION D., GOUREVITCH R., LE PROVOST J.-B., OLIÉ J.-P., LÔO H. & KREBS M.-O. (2004). L'hypothèse neurodéveloppementale dans la schizophrénie. *L'Encéphale*, **30**(2), 109–118. 17
- GRENON P. & SMITH B. (2004). SNAP and SPAN : Towards Dynamic Spatial Ontology. *Spatial Cognition & Computation*, **4**(1), 69–104. 44, 50
- GROSJEAN J., LETORD C., ZANA I., ADVENIER-IAKOVLEV E., DUCLOS C., KREBS MO., CHARLET J. & DARMONI S. (2021). *Off-Label Drug Database*. Preprint, Health Informatics. 93
- GRUBER T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, **5**(2), 199–220. 20
- GRUBER T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, **43**(5), 907–928. 28
- GRÜNINGER M. & FOX M. S. (1995a). Methodology for the Design and Evaluation of Ontologies. In *IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal. 28
- GRÜNINGER M. & FOX M. S. (1995b). The Role of Competency Questions in Enterprise Engineering. In A. ROLSTADÅS, Coordinateur, *Benchmarking — Theory and Practice*, IFIP Advances in Information and Communication Technology, p. 22–31. Boston, MA : Springer US. 65
- N. GUARINO, Coordinateur (1998). *Formal Ontology in Information Systems : Proceedings of the First International Conference (FOIS '98), June 6 - 8, Trento, Italy*. Number 46 in Frontiers in Artificial Intelligence and Applications. Amsterdam : IOS-Press [u.a.]. 32
- GUARINO N. & GIARETTA P. (1995). Ontologies and Knowledge Bases : Towards a Terminological Clarification. In N. J. I. MARS, Coordinateur, *Towards Very Large Knowledge Bases : Knowledge Building and Knowledge Sharing*, p. 25–32 : IOS Press. 36

- GUARINO N., OBERLE D. & STAAB S. (2009). What Is an Ontology? In S. STAAB & R. STUDER, Coordinateurs, *Handbook on Ontologies*, International Handbooks on Information Systems, p. 1–17. Berlin, Heidelberg : Springer. 20
- GUARINO N. & WELTY C. (2000). A Formal Ontology of Properties. In G. GOOS, J. HARTMANIS, J. VAN LEEUWEN, R. DIENG & O. CORBY, Coordinateurs, *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*, volume 1937, p. 97–112. Berlin, Heidelberg : Springer Berlin Heidelberg. 45, 57
- GUARINO N. & WELTY C. (2002). Evaluating ontological decisions with OntoClean. *Communications of the ACM*, **45**(2), 61–65. 29
- GUARINO N. & WELTY C. A. (2004). An Overview of OntoClean. In S. STAAB & R. STUDER, Coordinateurs, *Handbook on Ontologies*, p. 151–171. Berlin, Heidelberg : Springer Berlin Heidelberg. 28, 52
- GUEDJ-BOURDIAU M.-J. (2018). *Manuel pratique des urgences psychiatriques : les principales situations cliniques en fiches pratiques*. Les précis. Paris : Lavoisier-Médecine sciences. 96
- HACKING I. (1998). *Rewriting the Soul : Multiple Personality and the Sciences of Memory*. Princeton Paperbacks. Princeton, NJ : Princeton Univ. Press, 2. print., and 1. paperback print., with corr edition. 76
- HACKING I. (2002). *L'émergence de la probabilité*. Liber. Paris : Seuil. 79
- HALLER A. & POLLERES A. (2020). Are we better off with just one ontology on the Web? *Semantic Web*, **11**(1), 87–99. 42
- HALLER A., POLLERES A., DOBRIY D., FERRANTI N. & RODRÍGUEZ MÉNDEZ S. J. (2022). An Analysis of Links in Wikidata. In P. GROTH, M.-E. VIDAL, F. SUCHANEK, P. SZEKLEY, P. KAPANIPATHI, C. PESQUITA, H. SKAF-MOLLI & M. TAMPER, Coordinateurs, *The Semantic Web*, Lecture Notes in Computer Science, p. 21–38, Cham : Springer International Publishing. 130
- HARROW I., BALAKRISHNAN R., JIMENEZ-RUIZ E., JUPP S., LOMAX J., REED J., ROMACKER M., SINGER C., SPLENDIANI A., WILSON J. & WOOLLARD P. (2019). Ontology mapping for semantically enabled applications. *Drug Discovery Today*, **24**(10), 2068–2075. 27
- HASTINGS J. (2020). *Mental Health Ontologies*. University of Exeter Press. 73, 74
- HASTINGS J., CEUSTERS W., JENSEN M., MULLIGAN K. & SMITH B. (2012). Representing mental functioning : Ontologies for mental health and disease. In *Towards an Ontology of Mental Functioning (ICBO Workshop), Proceedings of the Third International Conference on Biomedical Ontology*. 77
- HASTINGS J., CEUSTERS W., SMITH B. & MULLIGAN K. (2011). The Emotion Ontology : Enabling Interdisciplinary Research in the Affective Sciences. In M. BEIGL, H. CHRISTIANSEN, T. R. ROTH-BERGHOFER, A. KOFOD-PETERSEN, K. R. COVENTRY & H. R. SCHMIDTKE, Coordinateurs, *Modeling and Using Context*, volume 6967, p. 119–123. Berlin, Heidelberg : Springer Berlin Heidelberg. 77
- HASTINGS J., COX S., WEST R. & NOTLEY C. (2020). Addiction Ontology : Applying Basic Formal Ontology in the Addiction domain. *Qeios*. 77
- HASTINGS J. & SCHULZ S. (2012). Ontologies for human behavior analysis and their application to clinical data. *International Review of Neurobiology*, **103**, 89–107. 77
- HAWLEY K. (2023). Temporal parts. In E. N. ZALTA & U. NODELMAN, Coordinateurs, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2023 edition. 43
- HERNÁN M. A., HSU J. & HEALY B. (2019). A Second Chance to Get Causal Inference Right : A Classification of Data Science Tasks. *CHANCE*, **32**(1), 42–49. 137
- HERRE H., HELLER B., BUREK P., HOEHNDORF R., LOEBE F. & MICHALEK H. (2006). *General Formal Ontology (GFO) - Part I : Basic Principles*. Rapport interne, Institute of Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig. 39
- A. J. G. HEY, Coordinateur (2009). *The Fourth Paradigm : Data-Intensive Scientific Discovery*. Redmond, Washington : Microsoft Research. 64
- HILBEY J., AIMÉ X. & CHARLET J. (2022). Représentation des connaissances médicales temporelles au moyen d'ontologies. In *33es Journées Francophones d'Ingénierie des Connaissances, IC 2022*, p.26. 104

- P. HITZLER, A. GANGEMI, K. JANOWICZ, A. KRISNADHI & V. PRESUTTI, Coordinateurs (2016). *Ontology Engineering with Ontology Design Patterns : Foundations and Applications*. Number 25 in Studies on the Semantic Web. Amsterdam Berlin : IOS Press. 105
- HITZLER P., HALLER A., JANOWICZ K., COX S. J., LEFRANÇOIS M., TAYLOR K., LE PHUOC D., LIEBERMAN J., GARCÍA-CASTRO R., ATKINSON R. & STADLER C. (2019). The modular SSN ontology : A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. *Semantic Web*, **10**(1), 9–32. 33
- HOEHNENDORF R., SCHOFIELD P. N. & GKOUTOS G. V. (2015). The role of ontologies in biological and biomedical research : A functional perspective. *Briefings in Bioinformatics*, **16**(6), 1069–1080. 23
- HOLMES T. H. & RAHE R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, **11**(2), 213–218. 87
- HORROCKS I. (2013). What Are Ontologies Good For? In B.-O. KÜPPERS, U. HAHN & S. ARTMANN, Coordinateurs, *Evolution of Semantic Systems*, p. 175–188. Berlin, Heidelberg : Springer Berlin Heidelberg. 36
- HORROCKS I., KUTZ O. & SATTLER U. (2006). The Even More Irresistible SROIQ. In *Proceedings of the Tenth International Conference (KR2006)*. 21
- HUCKLENBROICH P. (2014). "Disease entity" as the key theoretical concept of medicine. *The Journal of Medicine and Philosophy*, **39**(6), 609–633. 75
- JACKSON R., MATENTZOGLU N., OVERTON J. A., VITA R., BALHOFF J. P., BUTTIGIEG P. L., CARBON S., COURTOT M., DIEHL A. D., DOOLEY D. M., DUNCAN W. D., HARRIS N. L., HAENDEL M. A., LEWIS S. E., NATALE D. A., OSUMI-SUTHERLAND D., RUTTENBERG A., SCHRIML L. M., SMITH B., STOECKERT JR. C. J., VASILEVSKY N. A., WALLS R. L., ZHENG J., MUNGALL C. J. & PETERS B. (2021). OBO Foundry in 2021 : Operationalizing open data principles to evaluate ontologies. *Database : The Journal of Biological Databases and Curation*, **2021**, baab069. 42
- JAMOUS H., COMMAILLE J. & PONS-VIGNON B. (1969). *Sociologie de la décision*. C.N.R.S. Editions. 74
- JANSEN L. (2008a). Chapter 7 : Classifications. In *Applied Ontology*, p. 159–172. De Gruyter. 41
- JANSEN L. (2008b). Chapter 8 : Categories : The Top-Level Ontology. In *Applied Ontology*, p. 173–196. De Gruyter. 40
- JASKOLLA L. J. (2011). On Storms in Teacups : Limitations of 3D-4D-Equivalence. *KRITERION – Journal of Philosophy*, **25**(1), 31–39. 44
- KAMBEITZ-ILANKOVIC L., KOUTSOULERIS N. & UPTHEGROVE R. (2022). The potential of precision psychiatry : What is in reach? *The British Journal of Psychiatry : The Journal of Mental Science*, **220**(4), 175–178. 15
- KEET C. M. (2011). The Use of Foundational Ontologies in Ontology Development : An Empirical Assessment. In G. ANTONIOU, M. GROBELNIK, E. SIMPERL, B. PARSIA, D. PLEXOUSAKIS, P. DE LEENHEER & J. PAN, Coordinateurs, *The Semantic Web : Research and Applications*, volume 6643, p. 321–335. Berlin, Heidelberg : Springer Berlin Heidelberg. 42, 46
- KEET C. M. (2013). Open World Assumption. In W. DUBITZKY, O. WOLKENHAUER, K.-H. CHO & H. YOKOTA, Coordinateurs, *Encyclopedia of Systems Biology*, p. 1567–1567. New York, NY : Springer. 23
- KEET C. M. (2018). *An Introduction to Ontology Engineering*. London : College Publications. 25, 26, 27
- KHAN Z. C. & KEET M. (2015). Foundational Ontology Mediation in ROMULUS. In A. FRED, J. L. DIETZ, K. LIU & J. FILIPE, Coordinateurs, *Knowledge Discovery, Knowledge Engineering and Knowledge Management, Communications in Computer and Information Science*, p. 132–152, Berlin, Heidelberg : Springer. 40
- KHOURY M. J., GWINN M., BURKE W., BOWEN S. & ZIMMERN R. (2007). Will genomics widen or help heal the schism between medicine and public health? *American Journal of Preventive Medicine*, **33**(4), 310–317. 9
- KIRYAKOV A. (2006). Ontologies for Knowledge Management. In J. DAVIES, R. STUDER & P. WARREN, Coordinateurs, *Semantic Web Technologies : Trends and Research in Ontology-Based Systems*, p. 115–138. Chichester, England; Hoboken, NJ : John Wiley & Sons. 39

- KLÖPPEL S., STONNINGTON C. M., BARNES J., CHEN F., CHU C., GOOD C. D., MADER I., MITCHELL L. A., PATEL A. C., ROBERTS C. C., FOX N. C., JACK C. R., ASHBURNER J. & FRACKOWIAK R. S. J. (2008). Accuracy of dementia diagnosis : A direct comparison between radiologists and a computerized method. *Brain : A Journal of Neurology*, **131**(Pt 11), 2969–2974. 101
- KÖHLER S., GARGANO M., MATENTZOGLU N., CARMODY L. C., LEWIS-SMITH D., VASILEVSKY N. A., DANIS D., BALAGURA G., BAYNAM G., BROWER A. M., CALLAHAN T. J., CHUTE C. G., EST J. L., GALER P. D., GANESAN S., GRIESE M., HAIMEL M., PAZMANDI J., HANAUER M., HARRIS N. L., HARTNETT M. J., HASTREITER M., HAUCK F., HE Y., JESKE T., KEARNEY H., KINDLE G., KLEIN C., KNOFLACH K., KRAUSE R., LAGORCE D., MCMURRY J. A., MILLER J. A., MUNOZ-TORRES M. C., PETERS R. L., RAPP C. K., RATH A. M., RIND S. A., ROSENBERG A. Z., SEGAL M. M., SEIDEL M. G., SMEDLEY D., TALMY T., THOMAS Y., WIAFE S. A., XIAN J., YÜKSEL Z., HELBIG I., MUNGALL C. J., HAENDEL M. A. & ROBINSON P. N. (2020). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, **49**(D1), D1207–D1217. 78
- KONDYLAKIS H., NIKOLAOS A., DIMITRA P., ANASTASIOS K., EMMANOUEL K., KYRIAKOS K., IRAKLIS S., STYLIANOS K. & PAPADAKIS N. (2021). Delta : A Modular Ontology Evaluation System. *Information*, **12**(8), 301. 119
- KOTOV R., KRUEGER R. F. & WATSON D. (2018). A paradigm shift in psychiatric classification : The Hierarchical Taxonomy Of Psychopathology (HiTOP). *World psychiatry : official journal of the World Psychiatric Association (WPA)*, **17**(1), 24–25. 83
- KREBS M.-O. (2015). *Signes précoces des schizophrénies*. Psychothérapies. Dunod. 16
- KREBS M. O., GUT-FAYAND A., BOURDEL M., DISCHAMP J. & OLIÉ J. (2000). Validation and factorial structure of a standardized neurological examination assessing neurological soft signs in schizophrenia. *Schizophrenia Research*, **45**(3), 245–260. 103
- KUHN T. S., MEYER L. & LUMINET J.-P. (2018). *La structure des révolutions scientifiques*. Champs. Paris : Flammarion. 41
- LARSEN R. R. & HASTINGS J. (2018). From Affective Science to Psychiatric Disorder : Ontology as a Semantic Bridge. *Frontiers in Psychiatry*, **9**, 487. 77
- LARSEN R. R., MASCHIÃO L. F., PIEDADE V. L., MESSAS G. & HASTINGS J. (2022). More phenomenology in psychiatry? Applied ontology as a method towards integration. *The Lancet Psychiatry*, **9**(9), 751–758. 77
- LE BOTERFF Q., RABAH A., CARMENT L., BENDJEMAA N., TÉRÉMETZ M., ALOUIT A., LEVY A., TANGUY G., MORIN V., AMADO I., CUENCA M., TURC G., MAIER M. A., KREBS M.-O. & LINDBERG P. G. (2023). A tablet-based quantitative assessment of manual dexterity for detection of early psychosis. *Frontiers in Psychiatry*, **14**. 103
- LE GLAZ A., HARALAMBOUS Y., KIM-DUFOR D.-H., LENCA P., BILLOT R., RYAN T. C., MARSH J., DEVYLDER J., WALTER M., BERROUIGUET S. & LEMEY C. (2021). Machine Learning and Natural Language Processing in Mental Health : Systematic Review. *Journal of Medical Internet Research*, **23**(5), e15708. 103
- LENAT D. B. (1995). CYC : A large-scale investment in knowledge infrastructure. *Communications of the ACM*, **38**(11), 33–38. 38
- LINDSAY P. H. & NORMAN D. A. (1977). *Human Information Processing : An Introduction to Psychology*. New York : Academic Press, 2d ed edition. 61
- LOSSIO-VENTURA J. A., JONQUET C., ROCHE M. & TEISSEIRE M. (2014). BIOTEX : A system for Biomedical Terminology Extraction, Ranking, and Validation. In *ISWC : International Semantic Web Conference*, volume CEUR Workshop Proceedings, p. 157. 80
- LUCARINI V., GRICE M., CANGEMI F., ZIMMERMANN J. T., MARCHESI C., VOGLEY K. & TONNA M. (2020). Speech Prosody as a Bridge Between Psychopathology and Linguistics : The Case of the Schizophrenia Spectrum. *Frontiers in Psychiatry*, **11**, 531863. 102
- MAILLARD A. & JEANTIN L. (2019). *Neurologie, psychiatrie et addictologie*. Number 03 in Majbook. Paris : Med-Line éditions. 96
- MARTINELLI C. & SHERGILL S. S. (2015). Everything you wanted to know about neuroimaging and psychiatry, but were afraid to ask. *BJPsych Advances*, **21**(4), 251–260. 101

- MASCARDI V., CORDI V. & ROSSO P. (2007). A Comparison of Upper Ontologies. In *WOA 2007 - 8th AI*IA/TABOO Joint Workshop "From Objects to Agents" : Agents and Industry : Technological Applications of Software Agents*, p. 55–64, Genova, Italy. 39
- MASOLO C., BORGIO S., GANGEMI A., GUARINO N., OLTRAMARI A. & SCHNEIDER L. (2002). *The WonderWeb Library of Foundational Ontologies*. Rapport interne, ISTC-CNR. 36, 39
- MCCALL S. & LOWE E. J. (2006). The 3D/4D Controversy : A Storm in a Teacup. *Nous*, **40**(3), 570–578. 44, 45
- MEEHL P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, **17**(12), 827–838. 16
- MICHIE S., HASTINGS J., JOHNSTON M., HANKONEN N., WRIGHT A. J. & WEST R. (2023). Developing and using ontologies in behavioural science : Addressing issues raised. *Wellcome Open Research*, **7**, 222. 76
- MICHIE S., WEST R., FINNERTY A. N., NORRIS E., WRIGHT A. J., MARQUES M. M., JOHNSTON M., KELLY M. P., THOMAS J. & HASTINGS J. (2020). Representation of behaviour change interventions and their evaluation : Development of the Upper Level of the Behaviour Change Intervention Ontology. *Wellcome Open Research*, **5**, 123. 77
- MICOULAUD-FRANCHI J.-A., GEOFFROY P.-A., AMAD A. & QUILES C. (2015). Le jardinier et le botaniste. Proposition d'une organisation minimale de la sémiologie psychiatrique pour l'étudiant en médecine. *Annales Médico-psychologiques, revue psychiatrique*, **173**(5), 460–469. 79, 84
- MICOULAUD-FRANCHI J.-A., QUILES C., BATAIL J. M., DAUDET C., CERMOLACCE M. & DUMAS G. (2019). Vers une approche physiologique de la sémiologie en psychiatrie. Partie 1 : approches RDC, DSM, RDoC et HiTOP. *Annales Médico-psychologiques, revue psychiatrique*, **177**(3), 282–288. 76
- MIZOGUCHI R. & TOYOSHIMA F. (2017). YAMATO : Yet Another More Advanced Top-Level Ontology with Analysis of Five Examples of Change. In *JOWO*. 39, 56
- MONNIN P., LEGRAND J., HUSSON G., RINGOT P., TCHECHMEDJIEV A., JONQUET C., NAPOLI A. & COULET A. (2019). PGxO and PGxLOD : A reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison. *BMC Bioinformatics*, **20**(4), 139. 130
- MORRIS J. H., SOMAN K., AKBAS R. E., ZHOU X., SMITH B., MENG E. C., HUANG C. C., CERONO G., SCHENK G., RIZK-JACKSON A., HARROUD A., SANDERS L., COSTES S. V., BHARAT K., CHAKRABORTY A., PICO A. R., MARDIROSSIAN T., KEISER M., TANG A., HARDI J., SHI Y., MUSEN M., ISRANI S., HUANG S., ROSE P. W., NELSON C. A. & BARANZINI S. E. (2023). The scalable precision medicine open knowledge engine (SPOKE) : A massive knowledge graph of biomedical information. *Bioinformatics*, **39**(2), btad080. 130
- MOTIK B., PATEL-SCHNEIDER P. F. & CUENCA GRAU B. (2012). *OWL 2 Web Ontology Language Direct Semantics (Second Edition)*. <https://www.w3.org/TR/owl2-direct-semantics/>. 22
- NAGEL T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, **83**(4), 435. 74
- NATIONAL RESEARCH COUNCIL (2011). *Toward Precision Medicine : Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC : The National Academies Press. 6
- NILES I. & PEASE A. (2001). Towards a standard upper ontology. In *Proceedings of the International Conference on Formal Ontology in Information Systems - Volume 2001*, FOIS '01, p. 2–9, New York, NY, USA : Association for Computing Machinery. 39
- NOY N., GAO Y., JAIN A., NARAYANAN A., PATTERSON A. & TAYLOR J. (2019). Industry-scale knowledge graphs : Lessons and challenges. *Communications of the ACM*, **62**(8), 36–43. 129
- OBRST L., CEUSTERS W., MANI I., RAY S. & SMITH B. (2007). The Evaluation of Ontologies. In C. J. O. BAKER & K.-H. CHEUNG, Coordinateurs, *Semantic Web : Revolutionizing Knowledge Discovery in the Life Sciences*, p. 139–158. Boston, MA : Springer US. 28, 30
- OGREN P. V. (2006). Knowtator : A protégé plug-in for annotated corpus construction. In *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Companion Volume : Demonstrations -*, p. 273–275, New York, New York : Association for Computational Linguistics. 127
- OSTROM V. & OSTROM E. (1977). *Public Goods and Public Choices*. Workshop in Political Theory and Analysis. Indiana University, Workshop in Political Theory and Policy Analysis. 92

- PAYKEL E. S., PRUSOFF B. A. & UHLENHUTH E. H. (1971). Scaling of life events. *Archives of General Psychiatry*, **25**(4), 340–347. 87
- PELLISSIER TANON T., WEIKUM G. & SUCHANEK F. (2020). YAGO 4 : A Reason-able Knowledge Base. In A. HARTH, S. KIRRANE, A.-C. NGONGA NGOMO, H. PAULHEIM, A. RULA, A. L. GENTILE, P. HAASE & M. COCHEZ, Coordinateurs, *The Semantic Web*, Lecture Notes in Computer Science, p. 583–596, Cham : Springer International Publishing. 129
- PERONI S. (2017). A Simplified Agile Methodology for Ontology Development. In M. DRAGONI, M. POVEDA-VILLALÓN & E. JIMENEZ-RUIZ, Coordinateurs, *OWL : Experiences and Directions – Reasoner Evaluation*, Lecture Notes in Computer Science, p. 55–69, Cham : Springer International Publishing. 26
- POKORSKA-BOCCI A., STEWART A., SAGOO G. S., HALL A., KROESE M. & BURTON H. (2014). ‘Personalized medicine’ : What’s in a name? *Personalized Medicine*, **11**(2), 197–210. 6, 7
- POLDRACK R. A., KITTUR A., KALAR D., MILLER E., SEPPA C., GIL Y., PARKER D. S., SABB F. W. & BILDER R. M. (2011). The cognitive atlas : Toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, **5**, 17. 78
- POVEDA-VILLALÓN M., GÓMEZ-PÉREZ A. & SUÁREZ-FIGUEROA M. C. (2014). OOPS! (OntOlogy Pitfall Scanner!) : An On-line Tool for Ontology Evaluation. *International Journal on Semantic Web and Information Systems*, **10**(2), 7–34. 29
- QUINE W. V. (1948). On What There Is. *The Review of Metaphysics*, **2**(5), 21–38. 20
- RABAH A., LE BOTERFF Q., CARMENT L., BENDJEMAA N., TÉRÉMETZ M., DUPIN L., CUENCA M., MAS J.-L., KREBS M.-O., MAIER M. A. & LINDBERG P. G. (2022). A novel tablet-based application for assessment of manual dexterity and its components : A reliability and validity study in healthy subjects. *Journal of Neuroengineering and Rehabilitation*, **19**(1), 35. 103
- RABOUDI-SOUILEM A. (2021). *Gestion Du Cycle de Vie Des Études Multimodales de Recherche Biomédicale Préclinique Pour Le Partage et La Réutilisation Des Données Hétérogènes*. Thèse de doctorat, Compiègne. 19, 43, 121, 122, 188
- RICHARD M. (2017). *Apport de la modélisation ontologique pour le partage des connaissances en psychiatrie*. Thèse, Université Pierre et Marie Curie - Paris VI. 72, 77, 79
- ROBINSON P. N. (2012). Deep phenotyping for precision medicine. *Human Mutation*, **33**(5), 777–780. 78
- ROBINSON P. N., KÖHLER S., BAUER S., SEELow D., HORN D. & MUNDLOS S. (2008). The Human Phenotype Ontology : A Tool for Annotating and Analyzing Human Hereditary Disease. *American Journal of Human Genetics*, **83**(5), 610–615. 78
- RUDNICKI R., SMITH B., MALYUTA T. & MANDRICK W. (2016). *Best Practices of Ontology Development*. White paper, National Institute of Standards and Technology. 62
- RUGGERO C. J., KOTOV R., HOPWOOD C. J., FIRST M., CLARK L. A., SKODOL A. E., MULLINS-SWEATT S. N., PATRICK C. J., BACH B., CICERO D. C., DOCHERTY A., SIMMS L. J., BAGBY R. M., KRUEGER R. F., CALLAHAN J., CHMIELEWSKI M., CONWAY C. C., DE CLERCQ B. J., DORNBACH-BENDER A., EATON N. R., FORBES M. K., FORBUSH K. T., HALTIGAN J. D., MILLER J. D., MOREY L. C., PATALAY P., REGIER D. A., REININGHAUS U., SHACKMAN A. J., WASZCZUK M. A., WATSON D., WRIGHT A. G. & ZIMMERMANN J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into Clinical Practice. *Journal of consulting and clinical psychology*, **87**(12), 1069–1084. 84
- SANTOR D. A., GREGUS M. & WELCH A. (2006). Eight Decades of Measurement in Depression. *Measurement : Interdisciplinary Research & Perspective*, **4**(3), 135–155. 85
- SAPOLSKY R. M. (2017). *Behave : The Biology of Humans at Our Best and Worst*. London : Bodley Head. 138
- SAWYER S. M., AZZOPARDI P. S., WICKREMARATHNE D. & PATTON G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health*, **2**(3), 223–228. 45
- SCHEUERMANN R. H., CEUSTERS W. & SMITH B. (2009). Toward an Ontological Treatment of Disease and Diagnosis. *Summit on Translational Bioinformatics*, **2009**, 116–120. 62, 67
- SCHULZ S., BOEKER M. & MARTINEZ-COSTA C. (2017). The BioTop Family of Upper Level Ontological Resources for Biomedicine. *Studies in Health Technology and Informatics*, **235**, 441–445. 64

- SHAH N. H. & TENENBAUM J. D. (2012). The coming age of data-driven medicine : Translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association : JAMIA*, **19**(e1), e2–e4. 64
- SINGHAL A. (2012). Introducing the Knowledge Graph : Things, not strings. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>. 32, 128, 129
- SLATER L. T., GKOUTOS G. V. & HOEHDORF R. (2020). Towards semantic interoperability : Finding and repairing hidden contradictions in biomedical ontologies. *BMC Medical Informatics and Decision Making*, **20**(10), 311. 42
- SMITH B. (1998). Basic concepts of formal ontology. In N. GUARINO, Coordinateur, *Formal Ontology in Information Systems*, p. 19–28. IOS Press. 39
- SMITH B. & CEUSTERS W. (2015). Aboutness : Towards Foundations for the Information Artifact Ontology. In *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO)*, volume CEUR vol. 1515, p. 1–5. 63
- SMITH K. A., BLEASE C., FAURHOLT-JEPSEN M., FIRTH J., DAELE T. V., MORENO C., CARLBRING P., EBNER-PRIEMER U. W., KOUTSOULERIS N., RIPER H., MOUCHABAC S., TOROUS J. & CIPRIANI A. (2023). Digital mental health : Challenges and next steps. *BMJ Ment Health*, **26**(1). 76
- SOLDATOVA L. N. & KING R. D. (2006). An ontology of scientific experiments. *Journal of The Royal Society Interface*, **3**(11), 795–803. 105
- STUDER R., BENJAMINS V. R. & FENSEL D. (1998). Knowledge engineering : Principles and methods. *Data & Knowledge Engineering*, **25**(1), 161–197. 20
- SUÁREZ-FIGUEROA M. C., GÓMEZ-PÉREZ A. & FERNÁNDEZ-LÓPEZ M. (2012). The NeOn Methodology for Ontology Engineering. In M. C. SUÁREZ-FIGUEROA, A. GÓMEZ-PÉREZ, E. MOTTA & A. GANGEMI, Coordinateurs, *Ontology Engineering in a Networked World*, p. 9–34. Berlin, Heidelberg : Springer Berlin Heidelberg. 25
- TABB K. & LEMOINE M. (2021). The prospects of precision psychiatry. *Theoretical Medicine and Bioethics*, **42**(5-6), 193–210. 74, 75
- TIMÓN-REINA S., RINCÓN M. & MARTÍNEZ-TOMÁS R. (2021). An overview of graph databases and their applications in the biomedical domain. *Database*, **2021**, baab026. 23
- TONELLI M. R. & SHIRTS B. H. (2017). Knowledge for Precision Medicine : Mechanistic Reasoning and Methodological Pluralism. *JAMA*, **318**(17), 1649–1650. 9
- TORKAMANI A., ANDERSEN K. G., STEINHUBL S. R. & TOPOL E. J. (2017). High Definition Medicine. *Cell*, **170**(5), 828–843. 64
- TUKEY J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science. Reading, Mass : Addison-Wesley Pub. Co. 131
- TUKEY J. W. (1980). We Need Both Exploratory and Confirmatory. *The American Statistician*, **34**(1), 23. 131
- TURNER J. A. & LAIRD A. R. (2012). The cognitive paradigm ontology : Design and application. *Neuroinformatics*, **10**(1), 57–66. 106
- USCHOLD M. & GRÜNINGER M. (1996). Ontologies : Principles, methods and applications. *The Knowledge Engineering Review*, **11**(2), 93–136. 61, 62
- VARZI A. C. (1994). On the Boundary Between Mereology and Topology. In R. CASATI, B. SMITH & G. WHITE, Coordinateurs, *Philosophy and the Cognitive Sciences*, p. pp. 423–442, Vienna : Hölder-Pichler-Tempsky. 53
- VERBOON A. R. (2014). The Medieval Tree of Porphyry : An Organic Structure of Logic. In P. SALONIUS & A. WORM, Coordinateurs, *The Tree*, volume 20, p. 95–116. Turnhout : Brepols Publishers. 116
- VERDONCK M., GAILLY F., PERGL R., GUIZZARDI G., MARTINS B. & PASTOR O. (2019). Comparing traditional conceptual modeling with ontology-driven conceptual modeling : An empirical study. *Information Systems*, **81**, 92–103. 42

- VRANDEČIĆ D. (2009). Ontology Evaluation. In S. STAAB & R. STUDER, Coordinateurs, *Handbook on Ontologies*, International Handbooks on Information Systems, p. 293–313. Berlin, Heidelberg : Springer. 28, 30
- WEIBEL L. (2012). *Evaluation de la prescription médicamenteuse inappropriée en psychiatrie de l'âge avancé*. Maîtrise universitaire d'études avancées en pharmacie hospitalière, Université de Genève, Genève, Suisse. 94
- WHITE B. Y. (1993). Intermediate Causal Models : A Missing Link for Successful Science Education? In *Advances in Instructional Psychology*. Routledge. 61
- WILKINSON M. D., DUMONTIER M., AALBERSBERG I. J., APPLETON G., AXTON M., BAAK A., BLOMBERG N., BOITEN J.-W., DA SILVA SANTOS L. B., BOURNE P. E., BOUWMAN J., BROOKES A. J., CLARK T., CROSAS M., DILLO I., DUMON O., EDMUNDS S., EVELO C. T., FINKERS R., GONZALEZ-BELTRAN A., GRAY A. J. G., GROTH P., GOBLE C., GRETHE J. S., HERINGA J., 'T HOEN P. A. C., HOOFT R., KUHN T., KOK R., KOK J., LUSHER S. J., MARTONE M. E., MONS A., PACKER A. L., PERSSON B., ROCCA-SERRA P., ROOS M., VAN SCHAIK R., SANSONE S.-A., SCHULTES E., SENGSTAG T., SLATER T., STRAWN G., SWERTZ M. A., THOMPSON M., VAN DER LEI J., VAN MULLIGEN E., VELTEROP J., WAAGMEESTER A., WITTENBURG P., WOLSTENCROFT K., ZHAO J. & MONS B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, **3**(1), 160018. 65
- YACOUBI N., CHARRAD M., VIDAL M.-E., BEN AHMED M. & AMDOUNI S. (2011). A Semantic Framework for Web service Annotation, Matching and Classification in Bioinformatics. *Journal I3 Interaction Intelligence Information*, **11**, 9–38. 126
- ZACHAR P. (2006). Les troubles psychiatriques et le modèle des espèces pratiques. *Philosophiques*, **33**(1), 81–97. 75
- ZACKLAD M. (2010). Évaluation des systèmes d'organisation des connaissances. *Les Cahiers du numérique*, **6**(3), 133–166. 21
- ZEMACH E. (1970). Four ontologies. *The Journal of Philosophy*, **LXVII**(8), 231–247. 50
- ZHANG S. & ELHADAD N. (2013). Unsupervised biomedical named entity recognition : Experiments with clinical and biological texts. *Journal of Biomedical Informatics*, **46**(6), 1088–1098. 125
- ZOHAR J., STAHL S., MOLLER H.-J., BLIER P., KUPFER D., YAMAWAKI S., UCHIDA H., SPEDDING M., GOODWIN G. M. & NUTT D. (2015). A review of the current nomenclature for psychotropic agents and an introduction to the Neuroscience-based Nomenclature. *European Neuropsychopharmacology*, **25**(12), 2318–2325. 95
- ZUBIN J. & SPRING B. (1977). Vulnerability—a new view of schizophrenia. *Journal of Abnormal Psychology*, **86**(2), 103–126. 16

A Requête sur la psychiatrie de précision dans PubMed

```
("precision psychiatry"[tiab]
OR
"personalized psychiatry"[tiab]
OR
((Psychiatry[majr]
OR psychiatry[tiab]
OR "Mental Disorders"[majr]
OR "mental disorder"[tiab]
OR "mental disorders"[tiab]))
AND
("Precision Medicine"[majr]
OR "Precision Medicine"[tiab]
OR "personalised medicine"[tiab]
OR "personalized medicine"[tiab]))))
NOT "Forensic Psychiatry"[MH]
```

```
AND ("journal article"[pt]
NOT Bibliography[pt]
NOT Comment[pt]
NOT Editorial[pt]
NOT Letter[pt]
NOT News[pt]
NOT Case Reports[pt]
NOT Published Erratum[pt]
NOT Historical Article[pt])
```

```
AND "english"[la]
```

```
AND hasabstract
```

B Petite Ontologie Fondationnelle – Mode d’emploi

Cette annexe présente les classes (figure B.1) et les relations (figure B.2) de la Petite Ontologie Fondationnelle (ontoPOF). Elle se veut un mode d’emploi à l’attention de toute personne souhaitant réutiliser cette ontologie fondationnelle hors du contexte dans lequel elle a été créée.

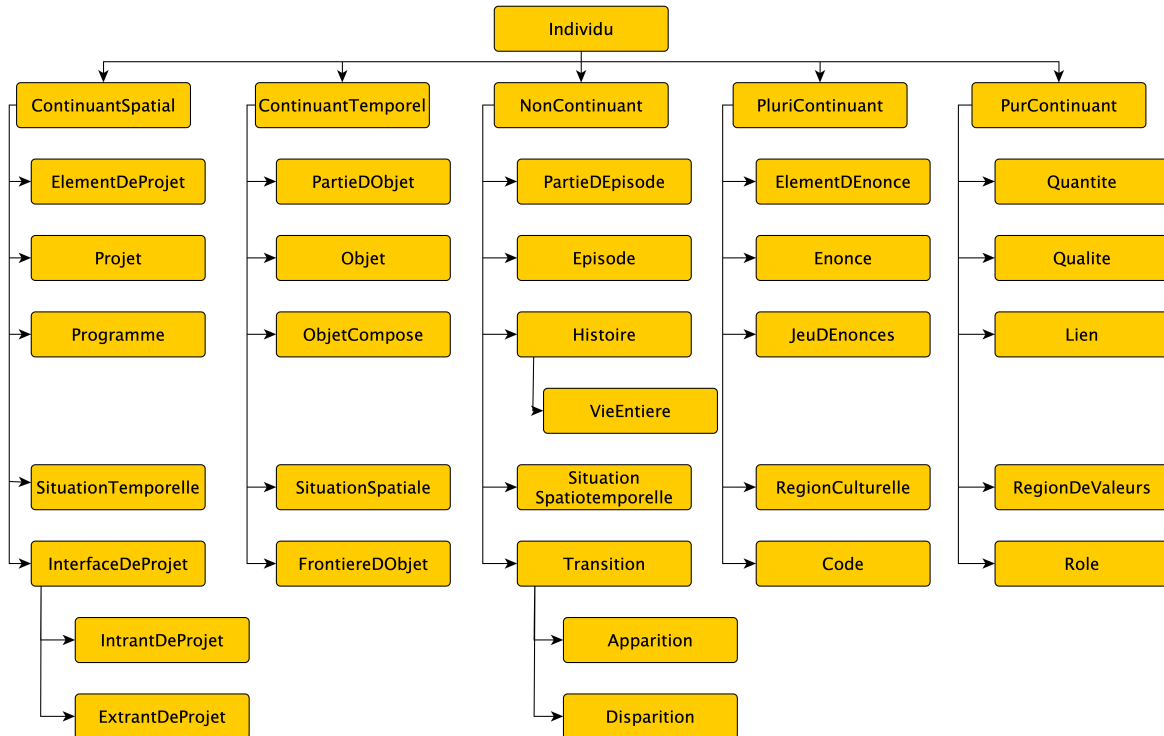


FIGURE B.1 – Les classes de la Petite Ontologie Fondationnelle.

B.1 Principes généraux

Le premier principe à l’œuvre dans ontoPOF est de ne prendre en considération que des entités qui sont individuées. C’est ce que marque le fait d’ajouter, sous la racine owl : Thing, la classe Individu. Nous verrons plus bas les différentes formes d’individuation qui sont considérées, mais marquons ici la conséquence de ce choix initial : nous ne considérons pas de classe représentant une substance, par exemple l’eau ou le paracétamol, mais seulement des classes représentant des portions de substance individuées (l’eau contenue dans ce verre, ce comprimé de paracétamol) ; de même, si nous parlons de propriétés d’un individu, nous considérons la couleur individuelle de ce ballon ou la longueur individuelle de ce crayon.

Deux autres principes guident l’organisation d’ontoPOF :

1. un principe d’individuation au regard de l’espace et du temps, qui prend en compte la limitation ou non dans l’espace et dans le temps des entités (sans considération de leur caractère matériel ou immatériel), et aboutit à distinguer intuitivement :
 - un monde des événements,

- un monde des objets,
 - un monde des projets,
 - un monde des propriétés,
 - un monde des informations.
2. un principe méréotopologique, s’attachant à la fois aux relations parties-tout (méréologie) et à la différence entre continuité et contiguïté ou à la position relative d’entités du même type (topologie), qui aboutit à différencier :
- un tout,
 - une partie d’un tout,
 - l’interface d’un tout avec un autre tout du même type ou avec son environnement,
 - une composition d’individus pouvant chacun être considéré comme un tout,
 - une situation dans un espace de référence propre au type d’entité considéré.

Les entités distinguées selon le principe d’individuation au regard de l’espace et du temps ne se situent pas toutes au même niveau d’abstraction :

- les événements sont les seules entités entièrement concrètes;
- les objets sont des abstractions obtenues à partir de ce que le langage courant désigne ainsi, par sélection de ce qui confère aux objets leur identité à travers le temps;
- les projets peuvent connaître différentes réalisations dans différents événements et ils sont, à ces événements, dans la relation d’un modèle logique à un modèle physique (pour reprendre le vocabulaire de la méthode MERISE);
- les propriétés sont des abstractions par sélection – en isolant un trait – pouvant être opérées sur des individus de tout type;
- les objets informationnels peuvent être considérés à différents niveaux d’abstraction selon les cas d’usage, mais ils sont ici considérés à tout le moins comme des abstractions par généralisation à partir d’objets concrets ayant un même contenu, un même aspect et les mêmes circonstances de production.

Ces deux principes sont croisés, le principe méréotopologique étant appliqué sur chacun des grands types ontologiques distingués par le principe d’individuation au regard de l’espace et du temps. Le noyau d’ontoPOF consiste donc en une matrice de vingt-cinq classes. Si l’application du principe méréotopologique aux objets est largement consensuelle, son application à des entités ayant des parties temporelles - ou à des entités n’ayant ni partie spatiale ni partie temporelle - implique quelques aménagements. Nous détaillons ci-dessous ce noyau et ces cas particuliers.

B.2 Les classes

B.2.1 Les non continnants

Un non continuant est une entité limitée dans l’espace et dans le temps, qui présente à ce titre des parties spatiales et des parties temporelles et peut donc être divisé selon des parties temporelles (tranches temporelles), selon des parties spatiales (tranches spatiales), voire selon des parties spatiotemporelles (morceaux spatiotemporels). Intuitivement, un non continuant correspond, s’il est matériel, à quelque chose qui « se passe », qui « a lieu », qui se déploie dans l’espace et dans le temps avant de s’interrompre, et s’il est immatériel à un espace-temps. La question de l’individuation se pose ici de façon particulièrement aiguë, à la fois pour déterminer les limites d’un individu non continuant et pour décider d’un « individu de base ». Nous empruntons, pour surmonter cette difficulté, au vocabulaire du récit.

L'individu de base est l'**épisode**, entendu comme ce qui présente à la fois une continuité spatiotemporelle et une unité narrative.

La **partie d'épisode** peut être un événement soudain, une action, un processus, qui ne possède pas par elle-même l'unité narrative pour constituer un épisode mais en forme seulement une partie spatiotemporelle.

L'interface d'un épisode avec un autre épisode ou au sein d'un « vide » narratif, appelée génériquement **transition**, recouvre en raison de l'irréversibilité du temps à la fois l'instant et le lieu de début d'un épisode, mais aussi l'instant et le lieu de fin d'un épisode. C'est pourquoi cette classe générique est moins susceptible d'être utilisée directement que ses sous-classes **apparition** et **disparition**.

Afin de prévenir la nécessité de définir autant de classes *ad hoc* que de niveaux de granularité de non continuants, une **histoire** est une composition à tout le moins d'épisodes, mais éventuellement d'histoires. Une histoire rassemble donc des épisodes, ou des compositions d'épisodes, qui présentent un lien narratif et un lien spatiotemporel, mais avec une exigence moindre que dans le cas d'un épisode : une histoire peut rassembler des épisodes n'étant pas spatiotemporellement connexes. Par son caractère générique et essentiel dans la modélisation des objets, la sous-classe **vie entière** est une histoire spécifique, qui représente l'intégralité de l'existence d'un continuant temporel, le « ver spatiotemporel » de son existence.

Afin de situer un non continuant dans un espace de référence, on le lie à un type particulier de non continuant, immatériel : une **situation spatiotemporelle**. Le non continuant entendu comme « ce qui se passe » est donc distingué de l'espace-temps au cours duquel il a lieu. Afin de compléter cette description, une situation spatiotemporelle est elle-même définie par son extension spatiale et son extension temporelle, qui sont respectivement une situation spatiale et une situation temporelle.

B.2.2 Les continuants temporels

Un continuant temporel est une entité limitée dans l'espace et continue dans le temps, qui présente à ce titre des parties spatiales mais pas de partie temporelle, et peut être divisé selon des parties spatiales. Intuitivement, un continuant temporel est ce qui occupe un certain espace, et ce de façon continue dans le temps. Classer comme continuants temporels les objets et les êtres vivants qui nous entourent permet de rendre compte de leur identité à travers le temps (marquée par cette occupation continuée d'espace dans le temps) et de leur attribuer les propriétés rigides qui les caractérisent. Cela n'exonère pas toutefois d'associer aux continuants temporels ainsi constitués des non continuants correspondant à leurs vies entières, ce qui permet de rendre compte des changements qui affectent ces mêmes objets et êtres vivants au cours de leur existence (au premier rang desquels leur apparition et leur disparition). La question de la granularité telle qu'elle apparaît dans une « pyramide de la complexité », des particules élémentaires aux organismes vivants voire aux écosystèmes, se pose ici et le choix de l'individu de base peut créer des difficultés dans l'articulation d'ontologies de domaines, un objet pouvant apparaître ici comme un individu de base et là comme une composition.

L'individu de base est l'**objet** mais comme nous venons de le remarquer, ontoPOF ne décide pas de la granularité à laquelle se situer. L'utilisation d'ontoPOF pour modéliser un domaine de connaissances implique ce choix initial, et de s'y tenir dans le cas de l'articulation d'ontologies de « cosmicités » différentes (microcosme, mésocosme, macrocosme).

Une **partie d'objet** est une partie spatiale d'un continuant temporel (considéré comme individu de base), qui peut être définie arbitrairement ou à partir de considérations cognitives ou fonc-

tionnelles, qui est ou a originellement été attachée à l’objet dont elle est une partie.

L’interface d’un continuant temporel avec un autre continuant temporel ou avec son environnement est une **frontière d’objet**. C’est à travers cette frontière qu’ont lieu les échanges avec l’extérieur du continuant temporel, et c’est sa présence qui permet de différencier les situations de continuité entre deux parties d’un même continuant temporel des situations de contiguïté de deux continuants temporels.

Un **objet composé** est un continuant temporel individué pour des raisons cognitives ou fonctionnelles, dont les composants ne sont pas des parties en tant qu’ils sont eux-mêmes des continuants temporels spatialement non connexes. Un agrégat, un groupe humain, un système, une machine, peuvent être considérés selon le contexte comme des objets composés.

Une **situation spatiale** est un continuant temporel immatériel qui permet de situer un continuant temporel matériel, de le localiser, que ce soit dans un système de référence (par exemple de coordonnées géographiques), dans un système topologique, voire dans un graphe.

B.2.3 Les continuants spatiaux

Un continuant spatial est une entité limitée dans le temps et continue dans l’espace, qui présente à ce titre des parties temporelles mais pas de partie spatiale, et peut être divisé selon des parties temporelles. Intuitivement, ce qui est individué selon un critère spécifiquement temporel relève de l’intentionnalité, d’une succession d’étapes projetées appelée à se réaliser dans un ou plusieurs non continuants. Un continuant spatial est conçu par un individu ou un groupe d’individus humains.

L’individu de base est le **projet**, dont l’unité est assurée par l’objectif poursuivi, cet objectif devant être « minimal », c’est-à-dire ne pas pouvoir être divisé en sous-objectifs susceptibles de faire eux-mêmes l’objet d’une intention, mais seulement en étapes menant à cet objectif.

Un **élément de projet** est une étape d’un continuant spatial n’obéissant pas à une intentionnalité précise, et devant à ce titre être réinscrite dans une succession pour prendre son sens.

L’interface d’un continuant spatial appelée génériquement une **interface de projet**, doit comme pour les non continuants prendre en compte la logique temporelle dans laquelle s’inscrit un continuant spatial. A l’interface du projet, on trouve en amont un ou plusieurs exemplaires d’**intran** de projet, qui constituent les ressources pour faire aboutir le projet, envisagées du même point de vue abstrait que le projet lui-même par rapport à sa réalisation; en aval du projet se trouvent un ou plusieurs exemplaires d’**extrant de projet**, qui constituent les résultats attendus du projet. De tels résultats peuvent être de natures différentes : produit, état, impact, bénéfice, etc. En ce sens, il est tout à fait possible que l’extrant d’un continuant spatial soit l’intran d’un autre.

Un **programme** est un ensemble de projets coordonnés présentant une unité intentionnelle. Un projet d’essai clinique multicentrique, un programme de construction d’un lotissement, un progiciel de gestion intégrée, peuvent constituer des exemples de programmes.

Une **situation temporelle** est un continuant spatial immatériel qui permet de situer un continuant spatial matériel, que ce soit dans un système de référence chronologique, ou relativement à d’autres continuants spatiaux (par exemple en suivant l’algèbre d’Allen), voire en termes de complexité algorithmique.

B.2.4 Les purs continuants

Un pur continuant est une entité qui n’est limitée ni dans l’espace, ni dans le temps, et à ce titre ne présente ni partie spatiale, ni partie temporelle. Intuitivement, sont individuées de cette

manière, par le sens commun, les propriétés attribuées à un individu. Une propriété peut concerner une partie d'un individu, ou un individu pendant un certain temps, mais l'attribution d'une partie de propriété n'a aucun sens. L'application du principe méréotopologique pour différencier des sous-classes dans la classe des purs continuants n'a dès lors à *la rigueur* aucune validité, puisqu'ils n'ont pas de partie. On peut toutefois distinguer les propriétés selon leur correspondance privilégiée (mais non stricte) à l'une ou l'autre des sous-classes distinguées par la méréotopologie pour les autres types, selon qu'elles seront de façon privilégiée attribuées à un tout, une partie, une interface, une composition. La question de la situation d'une propriété dans un espace de référence qui lui est propre est résolue par le fait qu'une propriété reçoit des valeurs. Remarquons enfin que les propriétés permettent de qualifier, de caractériser, et donc de discriminer. Elles sont appelées à recevoir des valeurs et permettent de modéliser les variables.

Une **qualité** est un pur continuant qui peut être préférentiellement associé à un « tout », au sens où on recourra fréquemment à une qualité pour distinguer deux individus de base d'un même type ontologique.

Une **quantité** est associée à une partie, au sens où deux parties d'un même tout pourront être comparées ou discriminées au moyen d'une quantité.

Pour qualifier une interface, on précisera son **rôle**. Rôle est à entendre ici en un sens large, comprenant aussi bien un rôle social qu'une fonction ou une propriété dispositionnelle.

Pour qualifier une composition, on précisera le **lien** à l'œuvre dans cette composition, par exemple la simple agrégation, la composition au sens spatial et presque pictural du terme, la structure (y compris celle d'une structure de données), voire la correspondance (pour un mapping de termes par exemple).

Une instance de la classe des purs continuants sera située par la place qu'elle occupe dans une **région de valeurs** spécifique.

B.2.5 Les pluricontinuants

Les possibilités qu'offraient le croisement de la continuité ou non dans l'espace avec la continuité ou non dans le temps sont épuisées par les quatre types d'entités précédents. Toutefois, pour une entité susceptible d'occuper plusieurs espaces-temps, les extensions temporelles de ceux-ci étant éventuellement simultanées et leurs extensions spatiales non-connexes, nous utiliserons le terme de pluricontinuant. Intuitivement, cette « ubiquité » n'est pas facile à appréhender, mais caractérise aussi bien les objets informationnels que les œuvres d'art, et de manière générale ce qui peut être communiqué ou répliqué - et ainsi occurrer dans un nouvel espace-temps sans cesser d'occurrer dans les espaces-temps où il occurrirait précédemment. Comme nous l'avons signalé plus haut, la question de l'individuation se pose en termes de niveau d'abstraction auquel on se situe et doit faire l'objet d'un choix de modélisation dépendant du cas d'usage. D'autre part, un problème de nommage se pose : les différents domaines artistiques, la théorie de l'information, ont chacun développé indépendamment leur propre vocabulaire. Nous empruntons, pour surmonter cette difficulté, au vocabulaire de la linguistique, la notion de langage étant ici la plus générique.

L'entité minimale de contenu informationnel est un **énoncé**. On suppose ici que la question de savoir ce qui est effectivement porteur d'une signification, qui peut faire l'objet d'une négociation entre l'émetteur et le récepteur d'une information, est résolue. Comme nous l'avons remarqué, un énoncé est susceptible de désigner une grande variété d'entités, qu'il s'agisse d'une donnée, d'un énoncé assertif, interrogatif, impératif ou performatif dans une langue naturelle, d'une image, d'une mélodie, etc.

Un **élément d’énoncé** est un élément d’un pluricontinuant qui n’est pas par lui-même significatif mais peut entrer légitimement dans la construction d’un énoncé. C’est le cas par exemple d’un identifiant mentionné seul ou d’un terme d’une terminologie détaché de sa définition.

Ce qui constitue l’interface d’un pluricontinuant est ce qui va permettre sa manipulation, qui va permettre au récepteur sa compréhension, qui va éventuellement permettre sa traduction. Nous utilisons pour désigner cette interface le terme de **code**, adapté à un grand nombre de contextes et suffisamment explicite.

Un **jeu d’énoncés** est un ensemble d’énoncés présentant une unité informationnelle, que ce soit par son contenu informationnel ou par ses circonstances de production. Le terme de « jeu » est choisi pour sa neutralité par rapport à la structure que présente cet ensemble; il est calqué sur l’expression « jeu de données » et peut désigner tous les niveaux de composition, ses composants étant soit des énoncés, soit des jeux d’énoncés.

Ce qui permet de situer un pluricontinuant dans un espace de référence est une **région culturelle**, c’est-à-dire un domaine de la connaissance auquel est susceptible d’appartenir ce pluricontinuant, ces domaines étant organisés selon une classification au sens des sciences de l’information, dont les classifications bibliographiques sont un exemple.

B.3 Les relations

Les relations sont organisées selon une arborescence qui s’appuie sur les principes généraux exposés plus haut. Une relation peut concerner des instances appartenant à des grands types ontologiques distincts, auquel cas elle est dite « hétérotypique ». Elle peut également concerner des instances appartenant au même type ontologique, et est alors désignée comme « homotypique ». Dans ce dernier cas, elle peut soit concerner des instances qui sont entre elles dans une relation méréotopologique, soit concerner des instances qui bien que du même type ontologique, ne sont pas dans une relation méréotopologique.

B.3.1 Les relations homotypiques

Les relations méréotopologiques sont les relations établies entre les classes distinguées selon la méréotopologie.

- Les relations `ontoPOF:aPourPartie` et `ontoPOF:estPartieDe` (son inverse), dans la mesure où elles concernent des individus qui ont été individués selon l’espace et le temps, sont déclinées pour :
 - les parties spatiotemporelles,
 - `ontoPOF:aPourPartieSpatiotemporelle`
 - `ontoPOF:estPartieSpatiotemporelleDe`
 - les parties spatiales,
 - `ontoPOF:aPourPartieSpatiale`
 - `ontoPOF:estPartieSpatialeDe`
 - les parties temporelles.
 - `ontoPOF:aPourPartieTemporelle`
 - `ontoPOF:estPartieTemporelleDe`
- Les relations `ontoPOF:aPourInterface` et `ontoPOF:estInterfaceDe` (son inverse) sont déclinées selon les entités considérées :
 - les non continnants (directement vers les interfaces temporellement différenciées),

- ontoPOF:aPourApparition
 - ontoPOF:estApparitionDe
 - et
 - ontoPOF:aPourDisparition
 - ontoPOF:estDisparitionDe
- les continuants temporels,
 - ontoPOF:aPourFrontiere
 - ontoPOF:estFrontiereDe
- les continuants spatiaux (directement vers les interfaces temporellement différenciées),
 - ontoPOF:aPourIntrant
 - ontoPOF:estIntrantDe
 - et
 - ontoPOF:aPourExtrant
 - ontoPOF:estExtrantDe
- les purs continuants,
 - ontoPOF:aPourRole
 - ontoPOF:estRoleDe
- les pluricontinuants.
 - ontoPOF:aPourCode
 - ontoPOF:estCodeDe
- Les relations ontoPOF:aPourComposant et ontoPOF:estComposanteDe (son inverse) sont déclinées pour :
 - les non continuants,
 - ontoPOF:aPourComposantEpisode
 - ontoPOF:estComposantEpisodeDe
 - les continuants temporels,
 - ontoPOF:aPourComposantObjet
 - ontoPOF:estComposantObjetDe
 - les continuants spatiaux,
 - ontoPOF:aPourComposantProjet
 - ontoPOF:estComposantProjetDe
 - les pluricontinuants.
 - ontoPOF:aPourComposantEnonce
 - ontoPOF:estComposantEnonceDe

Elles ne concernent pas les purs continuants. Il faut noter d'autre part qu'elles ne distinguent pas différents niveaux de composition et sont nommées d'après le niveau le plus simple auquel elles sont susceptibles d'intervenir (relation entre un composé et les « individus de base » qui le composent) mais restent valables pour une composition de compositions.

- Les relations ontoPOF:aPourSituation et ontoPOF:estSituationDe (son inverse) sont déclinées pour tous les grands types d'entités :
 - les non continuants,
 - ontoPOF:aPourSituationSpatiotemporelle
 - ontoPOF:estSituationSpatiotemporelleDe

- les continuants temporels,
 - ontoPOF:aPourSituationSpatiale
 - ontoPOF:estSituationSpatialeDe
- les continuants spatiaux,
 - ontoPOF:aPourSituationTemporelle
 - ontoPOF:estSituationTemporelleDe
- les purs continuants,
 - ontoPOF:aPourRegionDeValeurs
 - ontoPOF:estRegionDeValeursDe
- les pluricontinuants,
 - ontoPOF:aPourRegionCulturelle
 - ontoPOF:estRegionCulturelleDe

Les relations homotypiques non méréotopologiques interviennent, dans ontoPOF, parce que la séparation entre relations homotypiques et relations hétérotypiques crée un cas particulier pour les classes d’individus susceptibles d’être en relation avec toutes les autres classes d’individus de l’ontologie, à savoir les propriétés et les objets informationnels. Sont donc spécifiées des relations pour :

- les propriétés qui qualifient / sont qualifiées par des propriétés,
 - ontoPOF:estMetaproprieteDe
 - ontoPOF:aMetapropriete
- les informations qui portent sur / sont sujets d’autres informations.
 - ontoPOF:estMetainformationSur
 - ontoPOF:aMetainformation

Il faut toutefois noter que des relations homotypiques non méréotopologiques sont susceptibles d’apparaître plus fréquemment dans des ontologies de plus bas niveau pour exprimer par exemple les liens de parentalité reliant des personnes humaines ou des liens de causalité entre événements.

B.3.2 Les relations hétérotypiques

Les relations hétérotypiques sont le plus souvent bien définies par, et même conçues pour relier, leur domaine et leur codomaine. Ainsi, sont prévues des relations entre :

- un non continuant
 - et
 - un projet, un programme, un élément de projet ou une interface de projet (pour rendre compte de la **concrétisation d’un projet** dans un événement),
 - ontoPOF:realise
 - ontoPOF:estRealisePar
- un épisode, une histoire, une partie d’épisode ou une transition
 - et
 - n’importe quel individu (pour rendre compte de la **participation à un événement**),
 - ontoPOF:aPourParticipant
 - ontoPOF:participeA

- une situation spatiotemporelle
et
un continuant temporel
(pour rendre compte du **lieu d'un événement**),
 - ontoPOF:aExtensionSpatiale
 - ontoPOF:estExtensionSpatialeDe
 - une situation spatiotemporelle
et
un continuant spatial
(pour rendre compte du **moment d'un événement**),
 - ontoPOF:aExtensionTemporelle
 - ontoPOF:estExtensionTemporelleDe
 - un continuant temporel
et
un non continuant
(pour articuler l'aspect tridimensionnel et l'aspect quadridimensionnel d'un **individu matériel**),
 - ontoPOF:aPourVieEntiere
 - ontoPOF:estVieEntiereDe
 - un objet ou un objet composé
et
un continuant spatial
(pour rendre compte de l'autorité de la **conception d'un projet**),
 - ontoPOF:concoit
 - ontoPOF:estConcuPar
 - un pur continuant
et
n'importe quel individu n'étant pas un pur continuant
(pour rendre compte de l'**attribution de propriétés** à une entité),
 - ontoPOF:qualifie
 - ontoPOF:estQualifiePar
 - un pluricontinuant
et
n'importe quel individu n'étant pas un pluricontinuant
(pour rendre compte de l'**apport d'information** sur une entité),
 - ontoPOF:porteSur
 - ontoPOF:estSujetDe
- et plus spécifiquement (pour rendre compte de la **dénotation**)
- ontoPOF:denote
 - ontoPOF:estDenotePar

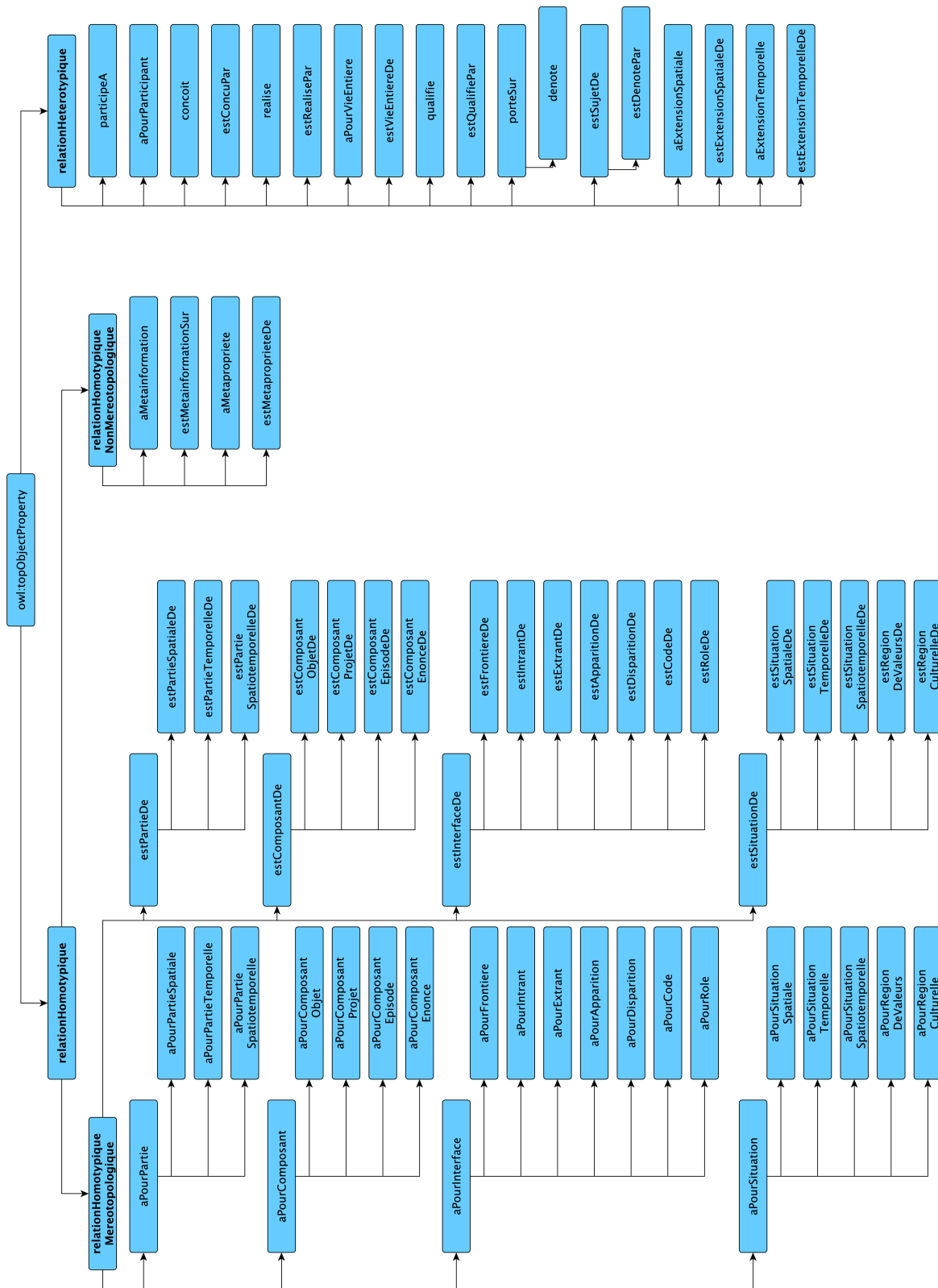


FIGURE B.2 – Les propriétés d’objet de la Petite Ontologie Fondationnelle.

C Histoire externe, histoire interne

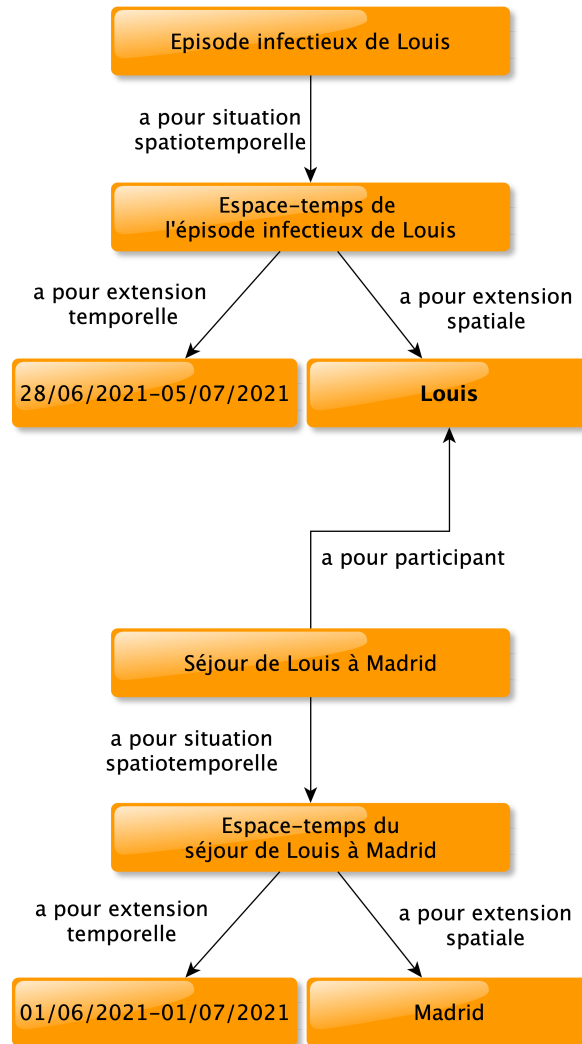


FIGURE C.1 – Histoire externe, histoire interne.

La situation que nous modélisons est la suivante : « Louis a contracté une infection fin juin pendant son séjour à Madrid ». Dans la phrase, on voit que tout est centré sur Louis. Mais du point de vue des événements impliqués, on est face à deux événements, l'un auquel Louis participe (son séjour à Madrid) et qui fait donc partie de son « histoire externe », l'autre dont Louis est le siège (son infection) et qui fait donc partie de son « histoire interne ».

D Propriétés d'objet des relations de proximité entre personnes

estUnProcheDe

aLienAmicalAvec

- aPourAmi_eDeLaFamille
- estAmi_eDe
- estAmi_eDeLaFamilleDe

aLienAmoureuxAvec

- estFiance_eDe
- estPetit_eAmi_eDe

aLienFamilialAvec

- estAdelpheDe
 - estFrereDe
 - estFrereAdoptifDe
 - estFrereBiologiqueDe
 - estDemiFrereDe
 - estFrereJumeauDe
 - estFauxJumeauDe
 - estVraiJumeauDe
- estSoeurDe
 - estSoeurAdoptiveDe
 - estSoeurBiologiqueDe
 - estDemiSoeurDe
 - estSoeurJumelleDe
 - estFausseJumelleDe
 - estVraieJumelleDe

- estConjoint_eDe
 - estConcubin_eDe
 - estEpoux_seDe
 - estPacsé_eA

- estCousin_eDe
 - estCousinDe
 - estCousinGermainDe
 - estCousinMaternelDe
 - estCousinPaternelDe
 - estCousinIssuDeGermainDe
 - estCousineDe
 - estCousineGermaineDe
 - estCousineMaternelleDe
 - estCousinePaternelleDe
 - estCousineIssueDeGermainDe

- estEnfantDe
 - estFilleDe
 - estFilleAdoptiveDe
 - estFilleBiologiqueDe
 - estFilsDe
 - estFilsAdoptifDe

- estFilsBiologiqueDe
- estGrandParentDe
 - estGrandMereDe
 - estGrandMereMaternelleDe
 - estGrandMerePaternelleDe
 - estGrandPereDe
 - estGrandPereMaternelDe
 - estGrandPerePaternelDe
- estNeveu_nieceDe
 - estNeveuDe
 - estNieceDe
- estOncle_tanteDe
 - estOncleDe
 - estOncleMaternelDe
 - estOnclePaternelDe
 - estTanteDe
 - estTanteMaternelleDe
 - estTantePaternelleDe
- estParentDe
 - estMereDe
 - estMereAdoptiveDe
 - estMereBiologiqueDe
 - estPereDe
 - estPereAdoptifDe
 - estPereBiologiqueDe
- estPetitEnfantDe
 - estPetitFilsDe
 - estPetiteFilleDe
- estRelationFamilialeParAllianceDe
 - estBeauParentDe
 - estBeauPereDe
 - estBelleMereDe
 - estBelAdelpheDe
 - estBeauFrereDe
 - estBelleSoeurDe
 - estBelEnfantDe
 - estBeauFilsDe
 - estBelleFilleDe
- aLienJuridiqueAvec**
 - aPourCurateur
 - aPourTuteur
 - estCurateurDe
 - estTuteurDe

E Arborescence de signes cliniques en psychiatrie

signe clinique

signe clinique général

Signe clinique général d'intérêt psychiatrique

Amaigrissement

Asthénie

Cyanose

Ictère

Œdèmes

Pâleur

Pli cutané

Prise de poids

Signe clinique neurovégétatif d'intérêt psychiatrique

Hyperréactivité neurovégétative

Signe sympathique d'anxiété

Signe clinique neuropsychiatrique

Évaluation du fonctionnement

Signe relatif à l'activité professionnelle

Signe relatif à l'autonomie

Signe relatif aux finances

Signe relatif aux relations interpersonnelles

Évaluation du jugement et de la conscience du trouble

Évaluation de l'attitude face aux soins

Adhésion aux soins

Réticence aux soins

Évaluation de l'insight

Insight altéré

Insight normal

Évaluation du jugement

Jugement altéré

Jugement normal

Évaluation du risque suicidaire

Présentation et description générale du patient

Comportement auto-agressif

Comportement pour diminuer l'anxiété

Comportement d'évitement

Compulsion

Signe relatif à la tenue vestimentaire

Tenue bizarre

Tenue correcte

Tenue négligée

Signe relatif à l'hygiène

Bonne hygiène

Incurie

Signe relatif au comportement interpersonnel

Comportement interpersonnel coopérant

- Comportement interpersonnel désinhibé
- Comportement interpersonnel diminué
- Comportement interpersonnel inadapté
- Signe relatif au contact
 - Contact anormal
 - Contact correct
 - Contact familier
 - Contact médiocre
 - Contact réticent
- Signe relatif à la motivation
 - Motivation augmentée
 - Motivation diminuée
 - Motivation normale
- Signe relatif à la motricité
 - Signe relatif à la mimique faciale
 - Mimique faciale altérée
 - Mimique faciale augmentée
 - Mimique faciale diminuée
 - Mimique faciale normale
 - Signe relatif à l'activité psychomotrice globale
 - Activité psychomotrice globale altérée
 - Activité psychomotrice globale augmentée
 - Activité psychomotrice globale diminuée
 - Activité psychomotrice globale normale
- Signe relatif à la perception
 - Perception altérée
 - Hallucinations
 - Hallucinations acoustico-verbales
 - Hallucinations auditives
 - Hallucinations cénesthésiques
 - Hallucinations gustatives
 - Hallucinations intrapsychiques
 - Hallucinations olfactives
 - Hallucinations tactiles
 - Hallucinations visuelles
 - Illusions
 - Perception dissociée
 - Perception augmentée
 - Perception diminuée
 - Perception normale
- Signe relatif à l'humeur
 - Humeur augmentée
 - Humeur diminuée
 - Humeur labile
 - Humeur normale
- Signe relatif aux émotions
 - Anhédonie

- Anxiété
- Émotions augmentées
- Émotions diminuées
- Émotions inadaptées
- Émotions labiles
- Émotions normales
- Hyperhédonie
- Signe relatif à l'attention ou à la mémoire
 - Signe relatif à la mémoire
 - Mémoire altérée
 - Mémoire augmentée
 - Mémoire diminuée
 - Mémoire normale
 - Signe relatif à la vigilance et au niveau de conscience
 - Vigilance et niveau de conscience augmentés
 - Vigilance et niveau de conscience diminués
 - Vigilance et niveau de conscience normaux
 - Signe relatif à l'attention
 - Attention augmentée
 - Attention diminuée
 - Attention normale
 - Signe relatif à l'orientation temporo-spatiale
 - Désorientation temporo-spatiale
 - Orientation temporo-spatiale normale
- Signe relatif au langage
 - Signe relatif à la prosodie
 - Prosodie altérée
 - Prosodie augmentée
 - Prosodie diminuée
 - Prosodie normale
 - Signe relatif à la signification du langage
 - Signification du langage altérée
 - Signification du langage normale
 - Signe relatif à la syntaxe
 - Syntaxe altérée
 - Syntaxe normale
 - Signe relatif à l'intensité sonore de la voix
 - Intensité sonore de la voix augmentée
 - Intensité sonore de la voix diminuée
 - Intensité sonore de la voix normale
 - Signe relatif au rythme ou à la quantité de langage
 - Rythme et quantité de langage normaux
 - Rythme ou quantité de langage altérés
 - Rythme ou quantité de langage augmentés
 - Rythme ou quantité de langage diminués
- Signe relatif aux fonctions physiologiques
 - Signe relatif à la sexualité

- Augmentation du désir ou de l'excitation
- Baisse du désir ou de l'excitation
- Sexualité normale
- Signe relatif à l'alimentation
 - Alimentation normale
 - Anorexie
 - Hyperphagie
- Signe relatif au sommeil
 - Cauchemar
 - Hypersomnolence
 - Insomnie
 - Parasomnie
 - Sommeil normal
- Signe relatif aux pensées
 - Signe relatif au contenu de la pensée
 - Idées de mort ou de suicide
 - Idées délirantes
 - Adhésion au délire
 - Évolution du délire
 - Mécanisme du délire
 - Retentissement émotionnel et comportemental du délire
 - Systématisation du délire
 - Thème du délire
 - Idées liées à l'humeur dépressive
 - Idées liées à l'humeur exaltée
 - Idées liées à une anxiété anticipatoire
 - Idées obsessionnelles
 - Idées phobiques
 - Soucis et inquiétudes excessifs
 - Reviviscences, remémorations, flashbacks
 - Signe relatif au cours de la pensée
 - Cours de la pensée illogique et incohérent
 - Cours de la pensée normal
 - Signe relatif au rythme de la pensée
 - Rythme de la pensée augmenté
 - Rythme de la pensée diminué
 - Rythme de la pensée normal

F Arborescence d'événements vécus

événement vécu

événement vécu relatif à l'amour avec cohabitation

- dispute sérieuse avec un fiancé
- fiançailles
- rupture d'une relation stable (mais pas de vie commune)
- rupture de fiançailles

événement vécu d'ordre conjugal

- amélioration marquée de la relation avec le partenaire
- aventure extra-conjugale du partenaire
- début d'une relation extra-conjugale
- dispute sérieuse avec le conjoint
- divorce
- mariage
- changement important de fréquence des disputes conjugales
- réconciliation conjugale
- séparation conjugale
 - séparation avec votre partenaire (vie commune) à cause de problèmes
 - séparation conjugale pendant un mois, non dû à une dispute
 - séparation conjugale liée à une dispute

changement de résidence

- déménagement dans un autre pays
- déménagement dans une autre ville
- déménagement dans la même ville

événement vécu de deuil

- un autre parent ou un ami proche est décédé
 - décès d'un proche
 - décès d'un ami proche
- un de vos parents, votre conjoint (partenaire) ou un enfant est décédé
 - décès du conjoint
 - décès d'un enfant (légitime, adopté)
 - décès d'un membre de la famille proche
- vous avez perdu ou on vous a volé quelque chose de valeur

événement vécu relatif à l'enseignement reçu

- arrêt d'un enseignement à temps plein
- changement d'école
- début d'un enseignement partiel ou à temps plein
- début ou fin de la scolarité formelle
- important échec scolaire
- préparation au passage d'un examen important

événement vécu d'ordre financier

- difficultés financières modérées
- contraction d'un prêt à la consommation
- contraction d'un prêt hypothécaire important
- changement important de la situation financière
 - amélioration substantielle sur le plan financier

crise financière majeure
graves difficultés financières (vous-même ou votre partenaire)
graves problèmes de logement, incluant le fait de se retrouver sans abri
saisie liée à une hypothèque ou à un emprunt

problèmes avec la police ou la justice

délit plus important avec passage devant le tribunal
délit mineur sans passage devant le tribunal
peine de prison
problèmes juridiques complexes et inhabituels
problèmes juridiques chez un membre de la famille proche

événement vécu personnel

succès personnel exceptionnel
changement important des conditions de vie
révision des habitudes de vie
témoin d'une agression grave ou autre événement traumatique
victime de violence (y compris au foyer), vol ou cambriolage

événement vécu relatif à la santé

changement important d'habitudes de sommeil
difficultés sexuelles
fausse-couche, avortement, enfant mort-né
grossesse
 grossesse désirée
 grossesse indésirée
maladie ou accident (blessure) sévère vous concernant
maladie ou accident (blessure) sévère concernant l'un de vos proches
ménopause
changement important d'habitudes alimentaires
changement important d'état de santé ou de comportement d'un parent
naissance d'un enfant en vie

événement vécu d'ordre sociofamilial

amélioration sensible avec un membre de la famille ou un ami proche
nouveau membre dans la famille
 adoption ou prise en charge à long terme d'un enfant
 naissance d'un enfant
changement important dans les activités religieuses
changement important dans les activités sociales
changement important dans le type ou la quantité de loisirs
changement important du nombre de réunions de famille
enfant quittant le foyer
 enfant quittant le domicile pour d'autres raisons
 fiançailles d'un enfant
 mariage d'un enfant avec consentement
 mariage d'un enfant sans consentement
dispute sérieuse avec un membre de la famille résident
dispute sérieuse avec un non-résident
jour férié important
nouvelle personne arrivant dans le ménage

nouvelles choquantes ou importantes et inattendues concernant votre conjoint
(partenaire) ou vos enfants

- problème grave avec un ami proche, un voisin ou un apparenté
- problèmes avec la belle-famille
- problèmes conjugaux chez des membres de la famille proche
- problèmes graves et continus avec votre partenaire ou vos enfants
- séparation d'une personne importante
- vacances

événement vécu relatif au travail exercé

- changement important des horaires ou des conditions de travail
 - changement dans les conditions de travail
 - changement dans les heures de travail
- changement important des responsabilités au travail
 - promotion
 - régression professionnelle
- changement de type de travail
- conjoint commençant ou cessant de travailler à l'extérieur du foyer
- faillite
- passage à la retraite
- problème grave au travail
 - désaccords au travail avec sanctions ou menaces
 - problèmes avec le patron
- réajustement important de l'activité professionnelle
- licenciement
- perte d'emploi ou recherche d'emploi infructueuse

G Abréviations utilisées en psychiatrie

AA	Alcooliques Anonymes OU Air ambient
AAH	Allocation aux Adultes Handicapés
AAN	Anticorps anti-nucléaires
ACR	Arrêt cardiorespiratoire
ACT	Acceptance and Commitment Therapy
ACTH	Corticotrophine
AD	Antidépresseur
ADH	Anti-diuretic hormone
ADHD	Attention Deficit Hyperactivity Disorder
ADK	Adénocarcinome
ADP	Adénopathies OU Analyse de la pratique professionnelle
ADT	Antidépresseur tricyclique
AE	Auto-Entreprise
AED	Aide Educative à Domicile
AEMO	Action Educative en Milieu Ouvert
AES	Allocation d'Education Spéciale
AFP	Association Française de psychiatrie
AFPA	Association pour la Formation Professionnelle des Adultes
AFT	Accueil Familial Thérapeutique
AFTCC	Association Française de Thérapie Comportementale et Cognitive
AFTOC	Association Française de malades souffrant de troubles obsessionnels compulsifs
AG	Anesthésie générale
AINS	Anti-inflammatoire non stéroïdien
AIT	Accident ischémique transitoire
AL	Anesthésie locale
ALD	Affection Longue Durée
AME	Aide Médicale de l'Etat
AMM	Autorisation de Mise sur le Marché
AMP	Aide Médico-Psychologique OU Antenne Médico-Psychologique
AMS	Atrophie multi-systématisée
ANPAA	Association Nationale de Prévention en Alcoologie et Addictologie
AOMI	Artérite oblitérante des membres inférieurs
AP	Antipsychotique OU Atelier Protégé OU Attaque de Panique
APA	Activité Physique Adaptée OU Allocation personnalisée d'autonomie OU Association de psychiatrie américaine
APAP	Antipsychotique à Action Prolongée
APL	Aide Personnalisée au Logement
ARH	Agence Régionale de l'Hospitalisation
ARM	Angiographie par résonance magnétique
AS	Aide Soignant OU Assistant(e) Social(e)
ASE	Aide Sociale à l'Enfance OU Assistant(e) Socio-Educatif(ve)
ASF	Allocation de Soutien Familial
ASH	Agent des Services Hospitaliers
ASP	Admission en soins psychiatriques

ASPA	Allocation de Solidarité aux Personnes Agées
ASPD	Admission en soins psychiatriques sur demande d'un tiers
ASPPI	Admission en soins psychiatriques en cas de péril imminent
ASS	Assistant(e) de Service Social
AT	Atelier Thérapeutique OU Arrêt de Travail
ATB	Antibiotique OU Antibiothérapie
ATCD	Antécédents
ATD	Anti-Dépresseur
AV	Acuité visuelle
AVC	Accident vasculaire cérébral
AVF	Algie vasculaire de la face
AVK	Anti-vitamine K
AVP	Accident de la voie publique
AVQ	Activité de la Vie Quotidienne
AVS	Auxiliaire de Vie Sociale OU Auxiliaire de vie scolaire
BAT	Biopsie de l'artère temporale
BAV	Bloc atrioventriculaire ou Baisse de l'acuité visuelle
BDA	Bouffée Délirante Aiguë
BGN	Bacille Gram Négatif
BHA	Bruit Hydroaérique
BK	Bacille de Koch, par extension tuberculose
BLIPS	Brief Limited Intermittent Psychotic Symptoms
BMI	Indice de masse corporelle
BPCO	Bronchopneumopathie Chronique Obstructive
BPRS	Brief Psychiatric Rating Scale
BU	Bandelette urinaire
BZD	Benzodiazépines
CA	Conseil d'Administration
CAC	Centre d'Accueil et de Crise
CADA	Commission d'Accès aux Documents Administratifs
CAE	Conduit auditif externe
CAF	Caisse d'Allocations Familiales
CAI	Conduit auditif interne
CAJ	Centre d'Accueil de Jour
CAMSP	Centre d'Action Médico-Sociale Précoce
CAO	Commission d'appel d'offres
CASA	Centre d'Accueil et de Soins pour Adolescents
CAT	Centre d'Aide par le Travail OU Conduite à tenir
CATEB	Centre d'Accueil et de Thérapies Brèves
CATTP	Centre d'Accueil Thérapeutique à Temps Partiel
CCA	Centre de Consultations pour Adolescents
CCAS	Centre Communal d'Action Sociale
CCI	Commission de Conciliation et d'Indemnisation des accidents médicaux
CCR	Cancer colorectal
CD	Conseil Départemental
CDA	Consommation déclarée d'alcool
CDJ	Centre De Jour

CDT	Carbohydate Deficient Transferrin
CE	Corps étranger
CESF	Conseiller(e) en économie sociale et familiale
CG	Conseil Général
CH	Centre Hospitalier
CHC	Carcinome hépatocellulaire
CHG	Centre Hospitalier Général
CHR	Centre Hospitalier Régional
CHRS	Centre d'Hébergement et de Réinsertion Sociale
CHS	Centre Hospitalier Spécialisé
CHT	Communauté Hospitalière de Territoire
CHU	Centre Hospitalier Universitaire
CI	Cadre Infirmier OU Contre-indication OU Chargé(e) d'Insertion
CIM	Classification Internationale des Maladies
CIVD	Coagulation intravasculaire disséminée
CJAAD	Centre Jeunes Adultes et Adolescents
CK	Créatine Kinase
CLIC	Centre Local d'Information et de Coordination
CMH	Coordination Médicale Hospitalière
CMI	Carte Mobilité Inclusion
CMM	Commission des Maladies Mentales
CMP	Centre Médico-Psychologique
CMPP	Centre Médico-Psycho-Pédagogique
CMU	Couverture Maladie Universelle
CMV	Cytomégalovirus
CNAM	Caisse Nationale d'Assurance Maladie
CNAV	Caisse Nationale d'Assurance Vieillesse
CNRS	Centre Nationale de Recherche Scientifique
CNS	Conférence Nationale de Santé
CNV	Communication Non Violente
CO	Monoxyde de carbone
COTOREP	Commission Technique d'Orientation et de Reclassement Professionnel
CP	Comprimé
CPAM	Caisse Primaire d'Assurance Maladie
CPK	Créatine phospho-kinase
CPOA	Centre Psychiatrique d'Orientation et d'Accueil
CPR	Complément de Ressources
CPT	Communauté Psychiatrique de Territoire
CRA	Centre Ressources Autisme
CRC	Centre de Recherche Clinique
CRH	Compte-Rendu d'Hospitalisation
CRIP	Cellule de Recueil des Informations Préoccupantes
CRP	Cognitive Remediation Therapy OU protéine C-réactive
CSAPA	Centre de Soins d'Accompagnement et de Prévention en Addictologie
CSST	Centre Spécialisé de Soins aux Toxicomanes
CTCG	Crise tonico-clonique généralisée
CTR	Centre Thérapeutique Résidentiel

CV	Cordes vocales, cardiovasculaire, champ visuel
DA	Directeur Adjoint OU Dopamine
DAEU	Diplôme d'Accès aux Etudes Universitaires
DALO	Droit Au Logement Opposable
DCI	Dénomination commune internationale
DDASS	Direction Départementale des Affaires Sanitaires et Sociales
DDB	Dilatation des bronches
DI	Déficiência Intellectuelle
DIM	Département d'Information Médicale
DIS	Drogue info service
DLCO	Diffusion Libre du CO
DMLA	Dégénérescence maculaire liée à l'âge
DMS	Durée Moyenne de Séjour
DNA	dotation non affectée
DNID	Diabète non insulino-dépendant
DR	Décollement de rétine
DRASS	Direction Régionale des Affaires Sanitaires et Sociales
DRH	Directeur des Ressources Humaines OU Direction des Ressources Humaines
DSM	Diagnostic and Statistical Manual of mental disorders OU Directeur des secours médicaux
DT	Douleur thoracique OU Delirium tremens
DU	Direction des Usagers
DUP	Duration of Untreated Psychosis
EAL	Examen d'une anomalie lipidique
EBV	Epstein-Barr virus
ECBU	Examen cyto-bactériologique des urines
ECC	Évaluation cognitivo-comportementale
ECG	Électrocardiogramme
ECT	Electroconvulsivothérapie
EDC	Épisode dépressif caractérisé
EDM	Épisode dépressif majeur
EDTSA	Écho-doppler des troncs supra-aortiques
EEG	Électroencéphalogramme
EER	Épuration extra-rénale
EG	État général
EHPAD	Établissement d'Hébergement pour Personnes Âgées Dépendantes
EHS	Entraînement aux habiletés sociales
EI	Effet indésirable OU Entreprise d'Insertion
EMC	Entraînement des habiletés méta-cognitives
EMDR	Eye Movement Desensitization and Reprocessing
EMG	Électromyogramme
EMPP	Équipe Mobile Psychiatrie Précarité
ENMG	Électroneuromyogramme
EP	Embolie pulmonaire
EPP	Évaluation des Pratiques Professionnelles OU Électrophorèse des protéines plasmatiques
EPS	Établissement Public de Santé

EPSM	Établissement Public de Santé Mentale
ES	Éducateur/trice Spécialisé(e) OU Effet secondaire
ESA	État de stress aigu
ESAT	Établissement et Service d'Aide par le Travail
ESM	Établissement de Santé Mentale
ESP	Équipe de Soins Primaires
ESPT	État de Stress Post Traumatique
ESS	Équipe de Suivi de Scolarisation
ETO	Échographie trans-œsophagienne
ETP	Équivalent Temps Plein
ETT	Échographie trans-thoracique
FAM	Foyer d'Accueil Médicalisé
FAN	Facteur anti-nucléaire
FC	Fréquence cardiaque
FCS	Fausse couche spontanée
FFI	Faisant Fonction d'Interne
FH	Foyer d'Hébergement
FL	Foyer Logement
FO	Fond d'oeil OU Foyer occupationnel
FP	Faux positifs
FR	Fréquence respiratoire
FV	Foyer de Vie
GCS	Groupement de Coopération Sanitaire
GDS	Gaz du sang
GEM	Groupe d'entraide Mutuelle OU Glomérulonéphrite extra-membraneuse
GGT	Gamma-glutamyl-transférase
GHB	Gamma-hydroxybutyrate
GIR	Groupe Iso Ressource
GMP	GIR Moyen Pondéré
GPA	Granulomatose avec polyangéite
GPAO	Glaucome primitif à angle ouvert
GR	Globule rouge
HAD	Hospitalisation à domicile
HAS	Haute Autorité de Santé
HAV	Hallucination acoustico-verbale
HBP	Hypertrophie bénigne de prostate
HBPM	Héparine de bas poids moléculaire
HBV	Virus de l'hépatite B
HC	Hospitalisation à temps Complet
HCV	Virus de l'hépatite C
HDJ	Hôpital De Jour
HDM	Histoire De la Maladie
HDT	Hospitalisation à la Demande d'un Tiers
HG	Hôpital général
HIA	Hémorragie intra-alvéolaire
HIV	Hémorragie intravitréenne OU Virus de l'immunodéficience humaine
HJ	Hôpital de Jour

HL	Hospitalisation libre
HLM	Habitation à Loyer Modéré
HMG	Hépatomégalie
HNF	Héparine non fractionnée
HO	Hospitalisation d'office
HP	Hôpital psychiatrique
HPN	Hémoglobinurie paroxystique nocturne
HPV	Human Papilloma Virus
HRP	Hématome rétro-placentaire
HSA	Hémorragie sous-arachnoïdienne
HSD	Hématome sous-dural
HSM	Hépatosplénomégalie
HSMG	Hépatosplénomégalie
HSV	Herpes Simplex Virus
HTA	Hypertension artérielle
HTAP	Hypertension artérielle pulmonaire
HTIC	Hypertension intracrânienne
HTP	Hypertension pulmonaire
HU	Hauteur utérine
IAD	Infirmière à domicile
IADL	Instrumental Activity of Daily Living
IAE	Ivresse éthylique aiguë
IC	Insuffisance cardiaque
ICG	Insuffisance cardiaque gauche
IDE	Infirmier Diplômé d'État
IDM	Infarctus du myocarde
IDR	Intradermoréaction
IDS	Idées suicidaires
IFSI	Institut de Formation en Soins Infirmiers
IG	Infirmière générale
IJ	Indemnités journalières
IM	Intramusculaire
IMAO	Inhibiteur de la monoamine oxydase
IMC	Indice de Masse Corporelle
IME	Institut médico-éducatif
IMG	Interruption médicale de grossesse
IMP	Institut médico-pédagogique
IMPRO	Institut médico-professionnel
IMV	Intoxication médicamenteuse volontaire
INED	Institut National d'Etudes Démographiques
INF	Interféron
INR	Injection Neuroleptique Retard OU International Normalized Ratio
INSEE	Institut National de la Statistique et des Études Économiques
INVS	Institut National de Veille Sanitaire
IOT	Intubation oro-trachéale
IP	Incapacité Permanente
IPP	Incapacité Partielle Permanente

IPPP	Infirmierie Psychiatrique de la Préfecture de Police
IR	Insuffisance rénale OU Injection Retard
IRA	Insuffisance rénale aiguë
IRC	Insuffisance rénale chronique
IRD	Institut de Recherche et Développement
IRM	Imagerie par Résonance Magnétique
IRP	Institut de Rééducation Pédagogique
IRS	Inhibiteur de la recapture de la sérotonine
IRSNA	Inhibiteurs de la recapture de la sérotonine et de la noradrénaline
ISRS	Inhibiteur sélectif de la recapture de la sérotonine
IST	Infection sexuellement transmissible
IT	Injonction Thérapeutique
ITEP	Institut Thérapeutique Educatif et Pédagogique
ITT	Incapacité Temporaire de Travail
IV	Intraveineux
IVG	Interruption volontaire de grossesse
JAF	Juge aux Affaires Familiales
JLD	Juge Des Libertés et de la Détention
KT	Cathéter OU cathéterisme
LAF	Lampe à fente
LBA	Liquide broncho-alvéolaire
LCR	Liquide céphalo-rachidien
LCS	Liquide cérébrospinal
LED	Lupus érythémateux disséminé
LP	Libération prolongée
LSD	Diéthyllysergamide
MADRS	Échelle de dépression de Montgomery et Asberg
MAP	Menace d'accouchement prématuré
MAPA	Maison d'Accueil pour Personnes Agées
MAS	Maison d'Accueil Spécialisée
MASP	Mesure d'Accompagnement Social Personnalisé
MAV	Malformation artério-veineuse
MBCT	Mindfulness-Based Cognitive Therapy
MCI	Mise en chambre d'isolement OU Mild Cognitive Impairment
MDA	Maison Des Adolescents
MDMA	Méthylène-dioxy-3,4-méthamphétamine
MDPH	Maison Départementale des Personnes Handicapées
MG	Médecin généraliste
MHD	Mesures hygiéno-diététiques
MI	Membre inférieur
MID	Membre inférieur droit
MIG	Membre inférieur gauche
MINI	Mini-International Neuropsychiatric Interview
ML	Mission locale
MMS	Mini-Mental State
MMSE	Mini-Mental State Examination
MS	Membre supérieur

MSA	Mutualité Sociale Agricole
MSD	Membre supérieur droit
MSG	Membre supérieur gauche
MT	Médecin traitant
MTX	Méthotrexate
NAP	Neuroleptique à action prolongée
NEAR	Neuropsychological and Educational Approach of Remediation
NFS	Numération de la formule sanguine
NGC	Noyaux gris centraux
NL	Neuroleptique
NORB	Neuropathie optique rétrobulbaire
NRS	Nourrisson
OACR	Occlusion de l'artère centrale de la rétine
OAP	Cedème aigu du poumon
OCT	Tomographie par cohérence optique
OD	Oreillette droite
OG	Oreillette gauche
OGE	Organes génitaux externes
OH	Alcoolisé OU Alcool OU Alcoolisme
OMA	Otite moyenne aiguë
OMI	Cedèmes des membres inférieurs
OMS	Organisation Mondiale de la Santé
OPP	Ordonnance de placement provisoire
PA	Pression artérielle
PAN	Péri-artérite noueuse
PANSS	Positive And Negative Syndrome Scale
PAS	Prestation d'Appui Spécifique
PASS	Point d'Accueil Santé Solidarité
PC	Perte de connaissance OU Périmètre crânien OU Centre de postcure
PCH	Prestation de Compensation du Handicap
PCR	Polymerase Chain Reaction
PDS	Programme de soins psychiatriques sans consentement
PEA	Potentiels évoqués auditifs
PEC	Prise en charge
PEP	Premier épisode psychotique
PH	Praticien hospitalier
PHC	Psychose hallucinatoire chronique
PHRC	Programme Hospitalier de Recherche Clinique
PJJ	Protection Judiciaire de la Jeunesse
PL	Ponction lombaire
PMA	Poste médical avancé
PMD	Psychose maniaco-dépressive
PMI	Protection maternelle et infantile
PNA	Pyélonéphrite aiguë
PO	Per os
PP	Post-partum OU Prévention du pellagre
PPC	Plan Personnalisé de Compensation

PPR	Photocoagulation pan-rétinienne
PR	Polyarthrite rhumatoïde
PSP	Paralysie supra-nucléaire progressive
PTA	Plateforme Territoriale d'Appui
PTSD	Trouble de stress post-traumatique
PUMA	Protection Universelle Maladie
QI	Quotient intellectuel
QIG	Quotient intellectuel général
QIT	Quotient intellectuel total
QSP	Quantité Suffisante Pour
RAA	Rhumatisme articulaire aigu
RAD	Retour à domicile
RAS	Rien à signaler
RCH	Rectocolite hémorragique
RCIU	Retard de croissance intra-utérin
RD	Rétinopathie diabétique
RDR	Réduction des risques liés à l'usage de drogues
RDV	Rendez-vous
RGO	Reflux gastro-œsophagien
RHJ	Reflux hépato-jugulaire
RM	Rétrécissement mitral
RMI	Revenu Minimum d'Insertion
ROT	Réflexes ostéotendineux
RPCA	Résistance à la protéine C activée
RPM	Rupture prématurée des membranes
RQTH	Reconnaissance de la qualité de travailleur handicapé
RSA	Résumé de sortie anonyme OU Revenu de solidarité active
RSR	Responsable scientifique régional
SA	Semaines d'aménorrhée
SAD	Service à domicile
SAMU	Service d'Aide Médicale d'Urgence
SAOS	Syndrome d'apnée obstructive du sommeil
SAPAD	Service d'Assistance Pédagogique A Domicile
SAPL	Syndrome des antiphospholipides
SAU	Service d'accueil des urgences
SAVS	Service d'Accompagnement à la Vie Sociale
SB	Substance blanche (cérébrale)
SC	Sous-cutané OU Schizophrénie
SCAM	Sortie contre avis médical
SCZ	Schizophrénie
SDF	Sans domicile fixe
SDRE	Soins à la Demande d'un Représentant de l'Etat
SDT	Soins à la Demande d'un Tiers
SDTU	Soins à la Demande d'un Tiers en Urgence
SEP	Sclérose en plaques
SG	Substance grise (cérébrale)
SIAO	Service Intégré d'Accueil et d'Orientation

SIDA	Syndrome d'immunodéficience acquise
SL	Soins libres
SMPR	Service Médico-Psychologique Régional
SMS	Structures Sociales et Médico-Sociales
SN	Syndrome néphrotique
SNC	Système nerveux central
SNG	Sonde naso-gastrique
SOFAS	Social and Occupational Functioning Assessment Scale
SPA	Symptômes psychotiques atténués
SPDT	Admission en soins psychiatriques sur demande d'un tiers
SPI	Soins en cas de péril imminent
SPIP	Service Pénitentiaire d'Insertion et de Probation
SPL	Soins psychiatriques libres
SPPI	Soins Psychiatrique en cas de Péril Imminent
SSIAD	Service de Soins Infirmiers A Domicile
SSR	Soins de suite et de réadaptation
TA	Tentative d'autolyse OU Tension artérielle OU Troubles de l'adaptation
TAG	Trouble anxieux généralisé
TB	Trouble bipolaire
TBL	Trouble
TBP	Trouble bipolaire
TC	Tronc cérébral OU Trauma crânien
TCA	Temps de céphaline activée OU Trouble du comportement alimentaire
TCC	Thérapie Comportementale et Cognitive
TCK	Temps de Céphaline Kaolin
TDA	Trouble déficitaire de l'attention sans hyperactivité
TDAH	Trouble déficitaire de l'attention avec hyperactivité
TDM	Tomodensitométrie
TED	Troubles envahissants du développement
TFI	Troubles fonctionnels intestinaux
TGI	Tribunal de Grande Instance
THC	Tétrahydrocannabinol
THM	Traitement hormonal de la ménopause
TI	Tribunal d'Instance
TISF	Technicien d'Intervention Sociale et Familiale
TJ	Turgescence jugulaire
TM	Ticket modérateur
TMS	Stimulation Magnétique Transcrânienne
TO	Tonus oculaire
TOC	Trouble obsessionnel compulsif
TOP	Trouble oppositionnel avec provocation
TP	Trouble de personnalité OU Taux de prothrombine OU Trouble panique
TPHA	Treponema Pallidum Haemagglutination Assay
TR	Toucher rectal
TS	Tentative de suicide
TSA	Troncs supra-aortiques OU Troubles du spectre autistique
TSH	Thyréostimuline

TSO	Traitement de substitution aux opiacés
TV	Tachycardie ventriculaire
TVC	Thrombose veineuse cérébrale
UCSA	Unité de Consultations et de Soins Ambulatoires
UDAF	Union Départementale des Associations Familiales
UEROS	Unité d'Evaluation, de Réentraînement et d'Orientation Sociale et/ou professionnelle
UF	Unité fonctionnelle
UGD	Ulcère gastro-duodéal
UHC	Unité d'hospitalisation complète
UHCD	Unité d'hospitalisation de courte durée
UHR	Ultra Haut Risque
UHSA	Unité Hospitalière Spécialement Aménagée
ULIS	Unité Localisée pour l'Inclusion Scolaire
UMD	Unité pour Malades Difficiles
UPAD	Unité pour Personnes Agées Désorientées
USLD	Unité de Soins de Longue Durée
VAD	Visite à domicile
VADS	Voies aérodigestives supérieures
VAE	Validation des acquis de l'expérience
VB	Voies biliaires
VDRL	Veneral Disease Research Laboratory
VGM	Volume globulaire moyen
VHA	Virus de l'hépatite A
VHB	Virus de l'hépatite B
VHC	Virus de l'hépatite C
VHD	Virus de l'hépatite D
VHE	Virus de l'hépatite E
VIH	Virus de l'immunodéficience humaine
VSL	Véhicule sanitaire léger
XT	Ecstasy

H Sources des données utilisées par le projet PsyCARE

H.1 Études antérieures

Nom de l'étude	Description	Promoteur	Statut	Types de données
PSYDEV	étude familiale clinique et génétique des aspects développementaux dans les maladies psychiatriques	Inserm	collection à renouveler	clinique omiques cognition neurologie
ICAAR	influence du cannabis sur le devenir psychopathologique des patients à risque (Ultra High Risk, FEP, Help seeker controls)	GHU Paris	terminée	clinique neurologie cognition imagerie omiques
AUSZ	Continuum autisme et schizophrénie : recherche de marqueurs phénotypiques cliniques, cognitifs, biologiques et en imagerie	GHU Paris	terminée	clinique cognition neurologie imagerie omiques
GENTOC	étude des facteurs de susceptibilité génétiques dans le Trouble obsessionnel compulsif	Inserm	terminée	clinique génétiques
GENSZ	étude familiale et génétique dans la schizophrénie	Inserm	terminée	clinique cognition génétique
CIMOCS	inhibition corticale et contrôle moteur dans la schizophrénie	GHU Paris	terminée	clinique cognition neurologie / neurophysiologie imagerie omiques
MOTOS	contribution des réseaux cortico-cérébelleux dans les signes neurologiques mineurs	GHU Paris	terminée récemment	clinique cognition neurologie / neurophysiologie
PRIMEPI	étude des facteurs prédictifs d'évolution lors d'un premier épisode psychotique	GHU Paris	terminée récemment	clinique biologie imagerie
STAR-T	marqueurs de stress et efficacité des thérapies comportementales et cognitives chez les sujets jeunes présentant un état mental à risque	GHU Paris	terminée récemment	clinique cognition neurologie imagerie omiques

H.2 Études associées

Nom de l'étude	Description	Promoteur	Statut	Types de données
SCOPE	étude évaluant l'efficacité et l'efficience d'une application de santé en ligne fournissant un entraînement cognitif personnalisé	GHU Paris	à venir	clinique neurologie cognition imagerie omiques enregistrement
SCOPE PILOT	<i>idem</i>	GHU Paris	à venir	clinique biologie imagerie
ÉTUDE DU WP8.1	formulaire en ligne sur les attentes des patients, des cliniciens et des soignants	APHP / Université Paris Cité	à venir	formulaire en ligne
ÉTUDE DU WP6.2	étude de l'efficacité d'une application mobile pour réduire le désengagement dans les soins	GHU Paris	à venir	clinique
DEMETER	détermination des Microétats EEG associés à la Transition psychotique dans les États à Risque	GHU Paris	en cours	clinique biologie imagerie neurophysiologie
PSYMAC	imagerie cérébrale et prédiction de la psychose	CHU Lille	en cours	biologie imagerie neurophysiologie
PREDIPSY	Faciliter les liens entre la psychiatrie et la radiologie	CHU Lille	en cours	clinique neurologie cognition imagerie
EDIPAJ	évaluation des dispositifs de détection et d'intervention pour les adolescents et adultes jeunes	GHU Paris	en cours	cognition imagerie omiques biologie
LONGIPEP	constitution d'une patienthèque chez des patients présentant une psychose débutante un état mental à risque	CHRU Brest	en cours	clinique cognition imagerie omiques
ASESID	analyse sémantique et syntaxique informatisée du discours	CHRU Brest	en cours	clinique cognition enregistrement

H.3 Étude prospective

Nom de l'étude	Description	Promoteur	Statut	Types de données
SCOPE	efficacité d'une prise en charge personnalisée composite sur les résultats fonctionnels dans la psychose précoce	GHU Paris	à venir	clinique cognition neurologie imagerie enregistrement omiques

I Alignements entre classes d'ontoDOME et classes de bareBMSLM

Classe de bareBMSLM	Relation de mapping	Classe d'ontoDOME
acquisition de données	owl:equivalentClass	CollecteDeDonnees
administration d'agent pharmaceutique	owl:equivalentClass	AdministrationDeMedicament
appareil	owl:equivalentClass	Equipement
chaîne de traitement de données	owl:equivalentClass	ChaineDeTraitement
contenu publié	owl:equivalentClass	PublicationScientifique
échantillon biologique	owl:equivalentClass	EchantillonBiologique
étude de recherche biomédicale	owl:equivalentClass	RechercheBiomedicale
examen	owl:equivalentClass	PassationExamenMedical
groupe de sujets dans l'étude	owl:equivalentClass	GroupeHumain
intervention biomédicale	owl:equivalentClass	InterventionMedicale
organisme étudié	rdfs:subClassOf	OrganismeVivant
outil logiciel	owl:equivalentClass	Logiciel
produit agent	owl:equivalentClass	PortionDeSubstanceChimique
protocole d'acquisition de données	owl:equivalentClass	ProtocoleDeCollecteDeDonnees
protocole d'examen	owl:equivalentClass	ProtocoleDEXamenMedical
protocole d'intervention biomédicale	owl:equivalentClass	ProjetDInterventionMedicale
protocole d'unité de traitement de données	owl:equivalentClass	ProtocoleUniteDeChaineDeTraitement
protocole de chaîne de traitement de données	owl:equivalentClass	ProtocoleChaineDeTraitement
résultats d'acquisition de données	owl:equivalentClass	ResultatsCollecteDeDonnees
résultats d'examen	owl:equivalentClass	ResultatsDEXamen
spécification d'unité de données	owl:equivalentClass	ResultatAttendu
unité de chaîne de traitement de données	owl:equivalentClass	UniteDeChaineDeTraitement
unité de données brutes résultant	owl:equivalentClass	ResultatsEtapeCollecteDeDonnees
unité de données dérivées résultant	owl:equivalentClass	ResultatsUniteDeChaineDeTraitement

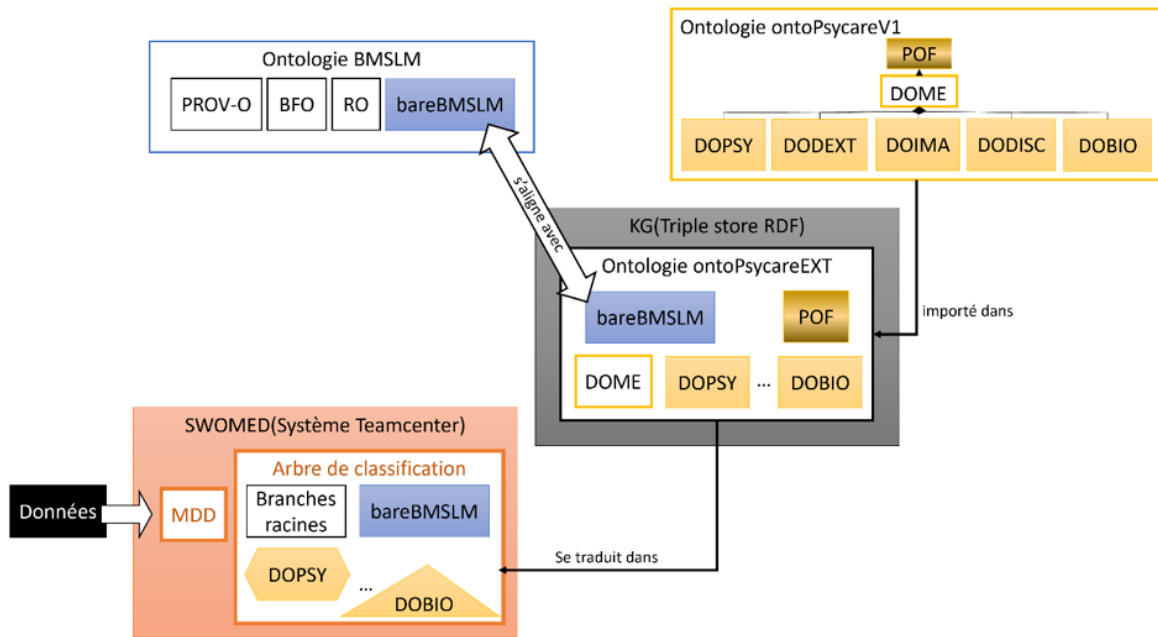


FIGURE I.1 – Regroupement BMS-LM – ontoPSYCARE (doc. Fealinx).

Table des figures

1	Schéma d'organisation du mémoire.	3
1.1	Tendances d'emploi des termes exprimant la médecine de précision.	6
1.2	Psychiatrie de précision / personnalisée : nombre d'articles par année de publication.	10
1.3	Nombre optimal de sujets pour la LDA.	11
1.4	Sujet n°5 de la LDA.	12
1.5	Lots de travaux du RHU PsyCARE (doc. PsyCARE).	19
2.1	Années de lancement initial des principales ontologies de haut niveau.	38
2.2	L'organisation des relations de la Petite Ontologie Fondationnelle.	58
4.1	Séquence canonique des sections dans un compte rendu d'hospitalisation en psychiatrie.	80
4.2	Exemple d'examen à l'entrée.	80
4.3	TIPSY : page d'un domaine d'intérêt psychiatrique.	91
4.4	TIPSY : page d'un outil d'évaluation.	92
4.5	Les trois branches de l'ontologie ontoMEDPSY.	94
5.1	Les différents niveaux de granularité des événements d'acquisition de données.	105
5.2	Un patron de conception d'ontologie pour les paradigmes expérimentaux.	106
5.3	Application aux données de neuroimagerie.	108
5.4	Application aux données de biologie médicale.	109
5.5	Application aux données d'analyse de discours.	112
5.6	Application aux données de dextérité.	113
6.1	Flux de données dans le projet PsyCARE (doc. DMP PsyCare4.0).	120
6.2	SWOMed : cycle de vie des données (doc. Raboudi-Souilem (2021)).	121
6.3	Schéma de l'infrastructure de la plateforme PsyCare4.0 (doc. Fealinx).	122
6.4	Concepts de l'ontologie BMS-LM (doc. Raboudi-Souilem (2021)).	122
6.5	Intégration des données dans SWOMed (doc. PsyCare4.0).	124
6.6	Processus d'annotation (doc. O. Aouina).	126
6.7	Annotation avec Knowtator (doc. O. Aouina).	128
B.1	Les classes de la Petite Ontologie Fondationnelle.	152

B.2	Les propriétés d'objet de la Petite Ontologie Fondationnelle.	161
C.1	Histoire externe, histoire interne.	162
I.1	Regroupement BMS-LM – ontoPSYCARE (doc. Fealinx).	187

Table des tableaux

1.1	Nombre de mentions des termes de l'intelligence artificielle.	15
2.1	Métriques d'ergonomie cognitive pour BFO-2020, DOLCE-Lite et ontoPOF	49
2.2	Les principales classes de la Petite Ontologie Fondationnelle sous forme de matrice.	56
5.1	Éléments communs selon les différentes sources de données de PsyCARE.	104

Résumé : La médecine de précision est une approche médicale récente prenant en compte les caractéristiques génomiques, biologiques, environnementales et sociales des patients pour proposer une meilleure évaluation des risques de maladies, un diagnostic plus fin, une meilleure capacité pronostique, une plus grande personnalisation des traitements. Elle se caractérise donc par le recours à une très large quantité de données, de sources et de natures diverses. L'exploitation de ces données soulève en premier lieu des enjeux d'intégration et d'interopérabilité. Leur réutilisabilité ultérieure suppose la compréhension de leur sémantique et du contexte dans lequel elles ont été produites. Sous ces deux aspects, l'utilisation d'ontologies informatiques constitue une solution viable et éprouvée. Un troisième aspect concerne l'analyse de ces données, pour laquelle l'organisation de ces dernières en un graphe de connaissances ouvre des perspectives propres. C'est en ce sens que nous considérons qu'il existe une « voie ontologique » pour la médecine de précision.

Dans le cadre du projet de recherche hospitalo-universitaire PsyCARE, qui témoigne de l'émergence en psychiatrie de cette approche médicale, nous proposons une ontologie modulaire. Le module de haut niveau accorde une importance majeure à la représentation du changement et de la temporalité. Le module de niveau intermédiaire est une ontologie noyau des données médicales. Le module dédié à la clinique psychiatrique vise également à l'annotation sémantique de documents de psychiatrie. Plusieurs ontologies de domaine correspondant aux disciplines convoquées (imagerie, biologie, analyse du discours, évaluation de la psychomotricité) sont développées à partir d'un patron de conception ontologique élaboré à cette occasion.

Mots-clés : Ingénierie des connaissances, informatique médicale, ontologies, traitement du langage naturel, systèmes d'aide à la décision.

Precision medicine : the ontological way

Application in psychiatric field

Abstract : Precision medicine is a recent medical approach that takes into account patients' genomic, biological, environmental and social features to provide better disease risk assessment, more accurate diagnosis, better prognostic capacity and more personalized treatment. It is therefore characterized by the use of vast quantities of heterogeneous data. The use of such data raises questions of integration and interoperability. Their subsequent re-use presupposes an understanding of their semantics and the context in which they were produced. In both respects, the use of computerized ontologies is a viable and proven solution. A third aspect concerns data analysis, for which the organization of data in a knowledge graph opens up new perspectives. In this sense, we see an "ontological way" for precision medicine.

As part of the PsyCARE university-hospital research project, which bears witness to the emergence of this medical approach in psychiatry, we are proposing a modular ontology. The top-level module focuses on the representation of change and temporality. The mid-level module is a core ontology for medical data. The module dedicated to the psychiatric clinic also aims at the semantic annotation of psychiatric documents. Several domain ontologies corresponding to the disciplines concerned (brain imaging, biology, speech analysis, psychomotricity assessment) are being developed on the basis of an Ontology Design Pattern developed for this purpose.

Keywords : Knowledge Engineering, Health Informatics, Ontologies, Natural Language Processing, Clinical Decision Support System.