



Document ID. CrEDIBLE-13-1-v1
Décembre 2013

Sémantique des données de l'observation : une approche ontologique

B. Gibaud¹, G. Kassel²

¹ LTSI, U 1099, INSERM, Université de Rennes 1, Rennes, France

² MIS, EA 4290, Université de Picardie Jules Verne, Amiens, France

Résumé

Nous proposons dans ce document un cadre ontologique général, nommé « DataTop », pour rendre compte de la sémantique de données produites et exploitées lors d'études scientifiques.

Ce cadre consiste en un modèle explicite de connaissances définissant les principales catégories d'entités intervenant dans des situations d'études scientifiques : *données, entités observées, processus d'observation, instruments de mesure*. Il est générique dans la mesure où il reste indépendant de tout domaine scientifique.

DataTop s'appuie sur une analyse ontologique de théories empiriques de la mesure et tout particulièrement sur des travaux menés au LOA (ISTC-CNR, Trento, Italie) portant sur l'ontologie des objets (observés) et de leurs propriétés (qualités). DataTop étend le noyau ontologique DOLCE-CORE en cours de définition au LOA au moyen de modules élaborés par les auteurs dans le cadre du projet NeuroLOG (2007-2010).

L'ontologie est présentée dans ce document de façon essentiellement discursive. Des extensions de DataTop, sa formalisation et son évaluation dans le domaine des neurosciences font partie de travaux en cours ou prévus à court terme discutés en conclusion.

Table des matières

1	INTRODUCTION.....	3
2	FONDEMENTS ONTOLOGIQUES DE L'OBSERVATION.....	5
2.1	LE PROCESSUS D'OBSERVATION.....	5
2.2	L'ENTITE OBSERVEE.....	7
2.3	LES DONNEES D'OBSERVATION.....	9
2.4	BILAN SUR LES DONNEES ET LEUR SEMANTIQUE.....	11
3	MODELE ONTOLOGIQUE DE DOLCE-CORE.....	12
3.1	PRINCIPALES CATEGORIES.....	12
3.2	COMPLEMENTS SUR LES QUALITIES.....	13
3.3	COMPLEMENT SUR LES REGIONS.....	15
3.4	BILAN SUR LE MODELE ONTOLOGIQUE DE DOLCE-CORE.....	17
4	EXTENSIONS APPORTEES AU NOYAU ONTOLOGIQUE DOLCE-CORE.....	17
4.1	COLLECTIONS.....	18
4.2	VALEURS ET UNITES DE MESURE.....	19
4.3	CONCEPTUALISATION DE L'ACTION D'OBSERVATION.....	22
4.4	CONCEPTUALISATION DES ENTITES JOUANT LE ROLE D'INSTRUMENT.....	24
4.5	CONCEPTUALISATION DES DONNEES ET D'ENTITES « PORTEUSES » D'INFORMATIONS.....	25
5	BREVE DISCUSSION.....	27
6	CONCLUSIONS ET PERSPECTIVES.....	29
7	RÉFÉRENCES.....	29

1 Introduction

« ... *The three most important research problems in Databases used to be 'Performance', 'Performance', 'Performance'; in years to come the three most important and challenging problems will be 'Semantics', 'Semantics', and 'Semantics'...* »¹

La sémantique des données est un problème reconnu dans le domaine des Bases de Données depuis les premiers jours du Modèle Relationnel. Il reste aujourd'hui inscrit en bonne place dans l'agenda des recherches en Bases de Données, pour une raison simple : les grandes masses de données produites et stockées nécessitent de disposer de capacités d'exploitation performantes. Ce constat, qui concerne notamment tous les domaines scientifiques, est à l'origine du défi MASTODONS du CNRS. L'exploitation des données soulève de nombreux problèmes, dont celui de leur sémantique, lié à l'exploitation de leur contenu. Dans ce document, nous explorons une voie analogue à celle suivie dans les travaux sur le Web Sémantique, consistant à doter les données d'une sémantique déclarative prenant la forme d'un modèle de connaissances. En préambule, afin de préciser le propos et la contribution de ce document, nous commençons par indiquer en quel sens nous entendons les termes « donnée », plus précisément « donnée de l'observation », et « sémantique de donnée ».

Le terme « donnée » signifie, dans une acception large, « ce qui est connu et admis, et qui sert de base, à un raisonnement, à un examen ou à une recherche » (une telle acception est attestée, formulée pratiquement à l'identique, dans des dictionnaires comme Le Trésor de la langue française, le Larousse ou le Littré). Littéralement, une donnée est donc tout ce qui peut être saisi et pris en considération par un être humain dans un raisonnement. Dans ce sens, les données recouvrent à la fois des informations singulières portant sur des situations particulières du monde, comme dans l'expression « les données d'un problème », et des connaissances générales attestées dans un domaine, comme dans l'expression « les données de la biologie ». Toujours dans ce sens large, les données peuvent être extérieures à un être humain et saisies par lui au moyen de ses sens, ou être déjà acquises et conceptualisées pour être utilisées mentalement.

Dans ce document, pour répondre aux attendus du programme MASTODONS d'aborder la question de la gestion de *données scientifiques*, nous nous focalisons sur une classe particulière de données, que nous nommons « données de l'observation ». Nous nous intéressons ainsi aux données / informations singulières qui sont exploitées lors d'observations ou de mesures réalisées dans le cadre d'études scientifiques. Avant toute chose, notons à propos des termes « observation » et « mesure » que pour nous conformer à leur usage courant, nous considérons le second comme étant un hyponyme du premier : une mesure est ainsi une observation conduisant à une quantification. Fondamentalement, ce contexte de l'observation met en scène un *phénomène* observé, autrement dit des entités observées (ex : on mesure la température d'un patient, l'âge d'un fossile, l'impact de l'état de dépression d'un patient sur son entourage, la proportion de clients d'une enseigne, dans une certaine classe d'âge, achetant un certain produit). Les « données de l'observation » sont dès lors toute information en lien avec le phénomène observé, qu'il s'agisse d'informations prises en compte au cours de l'observation, établies à cette occasion ou correspondant au résultat de l'observation (Bogen and Woodward, 1988). Ajoutons toutefois, pour préciser le sens dans lequel

¹ Déclaration de Stefano Ceri lors de la table ronde « *Data Semantics Can't Fail This Time!* » organisée dans le cadre de la conférence CAISE'98 (Pise, 11 juin 1998).

nous entendons le terme « donnée » et distinguer la *donnée* de l'*information*, que nous identifions les données à des enregistrements publics, accessibles à plusieurs interprétants. Ces enregistrements, jouissant d'une certaine stabilité dans le temps, demeurent exploitables pour des traitements, notamment leur interprétation en des informations. Cette disponibilité pour des interprétations ultérieures explique qu'ils soient désignés par le terme « donnée », même lorsqu'ils correspondent au résultat d'une observation.

Plus spécifiquement, dans le cadre du projet CrEDIBLE², les données considérées sont des données images (des intensités de signaux) et des données provenant de tests et de questionnaires et rendant compte de l'état neurologique de sujets. Ces données se présentent sous la forme de collections de nombres, de valeurs scalaires et d'items d'échelles de mesure et sont enregistrées dans des fichiers et/ou des Bases de Données. Nous en venons à évoquer la *sémantique* de ces données.

Associer une « sémantique » à une donnée relève, d'un point de vue général, d'une interprétation consistant à établir un lien entre la donnée et d'autres entités (Sheth, 1996). Pour une donnée de l'observation, cela consiste plus précisément à mettre la donnée en relation avec d'autres entités participant, soit à un phénomène observé (ex : une valeur scalaire est identifiée à la température d'un sujet), soit à un processus de l'observation (ex : cette même valeur est reliée à une action d'observation, son observateur, l'instrument de mesure utilisé et l'instant de la mesure). Par ailleurs, comme pour toute donnée, des propriétés peuvent être attachées à la donnée d'observation pour fournir des indications sur sa localisation, son format, la date de sa mise à jour, son accessibilité et des contraintes d'usage. Cette association peut correspondre à une expérience de pensée (une interprétation d'un agent cognitif) ou bien à l'établissement d'un lien, dans un système d'information, avec une autre donnée. Cette autre donnée, appelée ordinairement « métadonnée » ou « métadonnée sémantique », a pour rôle de conférer un sens à une donnée en augmentant son intelligibilité.

Dans le cadre du projet CrEDIBLE, nous décidons de fonder la sémantique de nos données sur une ontologie explicite, poursuivant ainsi des travaux menés par les auteurs du document dans le cadre des projets NeuroBase (Temal *et al.*, 2006), NeuroLOG³ (Gibaud *et al.*, 2011) et Green Computing Observatory (Germain-Renaud *et al.*, 2012). L'approche nous conduit dès lors à associer aux données deux composants : (i) des métadonnées formelles (des représentations de connaissances) exprimées au moyen de primitives conceptuelles – des concepts et relations – définies dans (ii) une ontologie. Comme justifié dans (Gibaud *et al.*, 2011), dans une perspective de fédération de sources de données hétérogènes, nous souhaitons en effet disposer de métadonnées riches sur le plan sémantique et reposant sur des catégories conceptuelles définies de manière rigoureuse.

Ces précisions terminologiques ayant été apportées et la démarche visée dans le projet CrEDIBLE (en tout cas en matière de sémantique de données) ayant été indiquée, nous en venons à préciser le contenu du présent document.

L'objet du document est de proposer un cadre ontologique général, nommé 'DataTop' (nom abrégé pour 'Scientific Data Semantics Top-level Ontology'), permettant de rendre compte de la sémantique des données de l'observation. Un tel cadre définit la nature des entités participant à des situations d'observation – en particulier les données, mais aussi les entités observées, les processus d'observation et les instruments de mesure – et les liens qu'entretiennent ces entités, en faisant abstraction d'un domaine scientifique particulier. La contribution dépasse donc le contexte du projet CrEDIBLE, ancré dans le domaine des neurosciences. Potentiellement, l'ontologie est à même d'être utilisée dans tout projet du programme

² <http://credible.i3s.unice.fr/>.

³ <http://neurolog.i3s.unice.fr/>.

MASTODONS, moyennant une spécialisation pour tenir compte des entités du domaine considéré.

Pour établir DataTop, nous avons suivi la démarche suivante. Jusqu'à présent, dans nos travaux antérieurs (projets NeuroLOG et GCO), nous avons fait le choix de DOLCE (Masolo *et al.*, 2003) comme ontologie fondatrice pour structurer nos ontologies d'application. Un travail en cours au LOA⁴ (ISTC-CNR, Trento, Italie) vise à redéfinir le noyau de DOLCE – la nouvelle version s'appelle DOLCE-CORE – pour répondre notamment à des desiderata relevant de l'ontologie de l'observation (Borgo & Masolo, 2009)(Masolo, 2010b). Pour le projet CrEDIBLE, nous faisons le choix de DOLCE-CORE comme ontologie fondatrice. Un tel noyau nécessite toutefois d'être étendu afin d'introduire des entités spécifiques de l'observation et de ses données (ex : des actions, des instruments de mesure, des inscriptions et leur support matériel). À cette fin, nous nous appuyons sur les travaux que nous avons réalisés dans le cadre du projet NeuroLOG, sur l'ontologie des images en neuroimagerie (Temal *et al.*, 2008) et l'ontologie d'instruments pour l'évaluation neurologique, comportementale et cognitive de sujets (Batrancourt *et al.*, 2010). Par ailleurs, nous prenons en compte des travaux récents consacrés à l'ontologie de l'observation et de la mesure (Probst, 2008)(Kuhn, 2009)(Masolo, 2010a)(Bottazzi *et al.*, 2012).

Le plan du document est le suivant. La section 2 analyse les fondements ontologiques de l'observation. La section 3 positionne DOLCE-CORE comme socle de DataTop. En section 4, des extensions sont apportées, de sorte à disposer d'un cadre ontologique complet. En section 5, nous discutons la portée de DataTop puis en conclusion nous envisageons plusieurs perspectives à ce travail.

2 Fondements ontologiques de l'observation

Dans cette section, nous examinons les fondements ontologiques de l'observation. L'objectif de cette étude est d'établir un cadre conceptuel qui nous serve de guide pour la définition de notre cadre ontologique, dans les sections suivantes. Pour mener cette étude, nous nous basons principalement sur des théories empiriques de l'observation⁵ ainsi que sur des travaux portant sur l'ontologie de l'observation. Nous en profitons pour nous situer vis-à-vis des ontologies de haut niveau BFO (Grenon & Smith, 2004) et DOLCE (Masolo *et al.*, 2003). Ce faisant, nous structurons notre examen autour de trois questions principales : en quoi consiste le processus d'observation ? Quelle est la nature de l'entité observée ? En quoi consistent les données de l'observation et comment caractériser leur sémantique ?

En guise de point de départ et pour servir de fil conducteur à notre discussion, nous adoptons comme référence la définition proposée par (van Fraassen, 2008) considérant l'observation comme (p. 179) "*a physical interaction, set up by agents, in a way that allows them to gather information. The outcome of a measurement provides a representation of the entity (object, event, process) measured, selectively, by displaying values of some physical parameters that – according to the theory governing this context – characterize that object*".

2.1 Le processus d'observation

Selon la définition de van Fraassen, l'observation consiste en un acte intentionnel visant à produire une représentation d'une entité, l'*entité observée*, selon un (ou plusieurs) paramètre(s). En résultat, l'entité observée se trouve positionnée dans un

⁴ <http://www.loa.istc.cnr.it/>.

⁵ Nous laissons de côté les théories formelles de la mesure focalisées sur l'étude des propriétés algébriques des espaces de valeurs (les échelles de mesures) pour privilégier des théories se focalisant sur la phénoménologie du processus de l'observation.

espace d'états possibles, dépendant du (ou des) paramètre(s) considéré(s), permettant de la classer et de la comparer à d'autres entités selon cette (ou ces) dimension(s).

Des exemples types d'observations sont la mesure de la température corporelle d'une personne ou la caractérisation de la forme d'une bactérie. Mais, comme le souligne la définition de van Fraassen, les observations ne se limitent pas à la mesure d'objets physiques comme un être humain ou une bactérie. Les entités observées peuvent être également des objets *temporels* comme la vie d'un être humain, un mal de tête, une division cellulaire, un match de football. Cette distinction (« physique » vs « temporel ») est reconnue au plus haut niveau des ontologies DOLCE et BFO sous la forme d'une partition entre « endurements » (ou « continuants ») et « perdurants » (ou « occurrents »)⁶. Des exemples d'observations d'objets temporels sont l'estimation de l'intensité d'un mal de tête, de la durée d'une division cellulaire ou l'évaluation de la qualité d'un match de football.

Cette première caractérisation du processus étant posée, des précisions restent à apporter quant à la nature ontologique du résultat. En particulier, deux questions se posent :

- Quelle relation le résultat entretient-il vis-à-vis de la langue ? Notamment, le résultat est-il une expression linguistique (ou une expression dans un quelconque langage de communication) ?
- Quelle relation le résultat entretient-il vis-à-vis des phénomènes observés ? Notamment, l'entité observée est-elle un objet du monde physique réel ?

Pour répondre à la première question, on notera, dans les définitions courantes du processus d'observation, une indétermination quant à la nature exacte de son résultat. Pour certains auteurs (ex : (Probst, 2008)(Kuhn, 2009)) le résultat consiste en une expression symbolique, autrement dit en un contenu conceptuel encodé au moyen de signes d'un langage de communication. Dans le cas d'une langue naturelle, le résultat (partiel ou global) peut ainsi être un mot (ex : « moyenne »), un syntagme (ex : « 38,4 °C ») ou une phrase (ex : « la température du patient est de 38,4 degrés Celsius »). L'expression langagière du résultat ne paraît toutefois pas indispensable à la caractérisation d'une action en tant qu'observation, comme en atteste la définition de van Fraassen. Nous admettons donc que des observations peuvent se conclure sur un résultat conceptuel « interne » à l'observateur. Le cadre ontologique que nous proposons dans les sections suivantes tient compte de ces deux possibilités. Cette question n'est toutefois pas anodine. En effet, compte tenu des limitations expressives d'une langue et des compétences langagières d'un observateur, la mise en langue d'un contenu conceptuel peut introduire des imprécisions et des ambiguïtés qui handicapent l'interprétation ultérieure de ce contenu.

Pour répondre à la seconde question, nous allons procéder par étapes. Dans un premier temps, nous nous focalisons sur un processus constitutif (en général) de l'observation, la *perception*, pour distinguer le phénomène observé, les sensations acquises et, finalement, l'entité observée.

Comme le montre l'étude approfondie de (Bogen, 2010), le terme « observation » désigne des processus très variés. Certaines observations relèvent d'expériences de

⁶ D'après les théories métaphysiques sur lesquelles se fondent DOLCE et BFO, ces entités se distinguent suivant le comportement dans le temps de leurs parties : un endurement est pleinement présent (toutes ses parties sont présentes) à tout instant de son existence, tandis qu'un perdurant n'est que partiellement présent (seules certaines de ses parties sont présentes) à tout instant de son existence. Bien que ces théories soient couramment adoptées pour structurer le haut niveau d'ontologies fondatrices, elles ne sont pas exemptes de questionnements, par exemple concernant la nature des *processus* par opposition aux *événements*. Le lecteur trouvera dans (Galton & Mizoguchi, 2009) un exposé récent de ces questions.

pensée pure, de jugements effectués à partir d'informations dont dispose déjà l'observateur. D'autres observations, au contraire, procèdent d'une acquisition d'informations extérieures et mobilisent le système perceptuel de l'observateur. Ces observations se distinguent alors suivant que le système perceptif de l'observateur interagit directement avec le phénomène étudié (le phénomène est regardé, écouté, touché, goûté ou senti)⁷ ou bien indirectement, par l'intermédiaire d'un instrument de mesure⁸ (nous laissons de côté momentanément les observations procédant de l'interprétation de données, que nous considérons en section 2.3).

Peu important, pour la question que nous nous posons, les détails de ces différents processus. Nous retenons par contre que trois entités sont à distinguer : le stimulus provenant du monde extérieur et entrant en interaction avec le système perceptuel de l'observateur⁹ ; la perception, ou sensation, acquise ; la conception qui en est élaborée. Une telle description est compatible avec les théories courantes de la perception, notamment pour ce qui concerne la distinction des mécanismes perceptuel et conceptuel (voir, par exemple, (Pylyshyn, 2007)). On la retrouve adoptée par Kuhn (2009), dans son ontologie de l'observation, pour qui l'identification de l'entité observée relève du niveau conceptuel (p.32) : *“Neither the choice of the observed entity, nor the quality assigned to it, nor its link to a stimulus, nor the value assigned to the quality are mind-independent. All of them involve human conceptualizations...”*. La conséquence importante que nous tirons de cette description est que, ce que nous nommons *entité observée*, ne relève pas du phénomène observé ni même de la perception immédiate mais correspond à un construit conceptuel élaboré par l'observateur à partir de stimuli qu'il perçoit, un point de vue que l'on trouve également récemment résumé par (Bottazzi *et al.*, 2012, p.61) : *“Measurement apparatuses don't have the 'innate' ability of carving-up the world into objects (and properties). Outputs need to be explicitly selected and manipulated to identify objects”*. Dans la section suivante, nous tirons des conséquences de cette conception de l'entité observée.

2.2 L'entité observée

Comme caractérisation de l'entité observée, nous retenons donc que celle-ci est le fruit d'une conception de l'observateur, ce qui indique sa nature – a priori – subjective. L'identification de l'entité observée résulte d'un processus cognitif s'appuyant notamment sur des sensations acquises par le système perceptuel de l'observateur augmenté, suivant les cas, d'appareils de mesure. Dans cette section, nous tirons plusieurs enseignements de cette conception.

En premier lieu, sur un plan ontologique, celle-ci revient à distinguer, d'un côté, les phénomènes réels avec lesquels nos capteurs, y compris nos systèmes perceptuels, interagissent directement et, d'un autre côté, des entités cognitivement construites ou socialement construites, entretenant un certain lien avec le monde réel. En termes de posture ontologique, ceci revient à adopter une position intermédiaire entre *réalisme* et *constructivisme* : certes, des entités existent dans le monde indépendamment d'agents cognitifs reconnaissant leur existence (réalisme) mais les

⁷ Ces observations directes peuvent être facilitées par des instruments permettant de mieux percevoir le phénomène. Par exemple, pour la vue, des lunettes, un microscope ou un télescope permettent respectivement de voir plus clairement et de voir des choses trop petites ou trop éloignées.

⁸ Le principe d'un instrument de mesure (par exemple, d'un thermomètre à mercure) est d'exploiter une corrélation entre une qualité d'une entité relevant du phénomène étudié (ex : la température corporelle d'un être humain) et la qualité d'une autre entité – directement perçue (ex : la hauteur d'un volume de mercure dans un tube gradué).

⁹ On notera que si des appareils d'aide à la perception ou de mesure sont utilisés, les stimuli produits par ces appareils peuvent renvoyer une vue déformée, voire erronée, du phénomène étudié.

entités que nous manipulons, notamment pour communiquer des résultats d'observation, sont des entités construites (constructivisme) tenant lieu de notre connaissance des entités du monde réel.

Nous avons rappelé, dans la section précédente, que les ontologies de haut niveau courantes distinguent deux catégories principales d'objets, à savoir les *objets physiques* et les *objets temporels*. Une telle distinction relève d'une caractérisation du mode de persistance des objets (et du comportement de leurs parties) plutôt que d'un mode d'existence. Concernant le mode d'existence des objets, des ontologies comme BFO et DOLCE affinent la précédente distinction en considérant, d'une part, des objets relevant d'une réalité physique et, d'autre part, des objets relevant d'une réalité cognitive ou sociale¹⁰. Un point de vue plus « extrême », défendu par Bottazzi *et al.* (2012), consiste à considérer que toutes les entités dont nos catégories rendent compte, et qui sont donc potentiellement des entités observées, possèdent une part cognitive ou sociale. Ce point de vue est en accord avec la théorie psychologique des *affordances* de Gibson (1979) selon laquelle les objets sont perçus par les observateurs en fonction d'actions qu'ils permettent. Selon cette théorie, une marche d'escalier sera conçue par des personnes valides comme un moyen de monter pour atteindre un lieu plus élevé. Au contraire, pour des personnes handicapées, elle sera conçue comme un obstacle. Dans le contexte des études scientifiques, très tôt le poids des théories scientifiques dans l'interprétation des phénomènes perçus et des résultats d'expériences a été mis en évidence¹¹. Ceci nous conduit à interroger le caractère subjectif des résultats de l'observation.

Si les entités observées comportent une part de dépendance vis-à-vis d'agents, qu'en est-il des propriétés que nous leur attribuons et, tout particulièrement, des qualités ou dimensions selon lesquelles nous les observons et comparons ? La cohérence avec les entités observées voudrait qu'on leur accorde le même statut. Tel est le point de vue défendu par Masolo (2010) qui souligne le rôle central joué par les appareils et procédures de mesure dans la définition des qualités des objets et le positionnement des objets selon ces qualités, un rôle que résume le slogan : « les qualités des objets sont ce que mesurent nos instruments ». Selon ce point de vue, les résultats de l'observation acquièrent une dimension *inter-subjective* du fait qu'ils reposent à la fois sur des instruments calibrés fondés sur des théories reconnues et sur des procédures de mesures standardisées, à l'image des tests et questionnaires standards utilisés en neurosciences (Batrancourt *et al.*, 2010). Ainsi, le caractère *social* des attributs (au sens de leur dépendance vis-à-vis de communautés d'observateurs) concourt à l'objectivité (ou plutôt *inter-subjectivité*) des résultats de mesures et à leur partage.

Une dernière raison de cette inter-subjectivité, que nous soulignons pour clore cette section, tient à la part d'arbitraire intervenant dans l'identification des entités mesurées. Ainsi, pour reprendre un exemple donné par (Kuhn, 2009), lorsqu'une température mesurée par un thermomètre dans une pièce est communiquée comme étant LA température de LA pièce, celle-ci correspond en fait à la température d'une

¹⁰ Dans BFO et DOLCE, les entités dépendant d'agents cognitifs sont distinguées des autres au moyen de catégories explicites les rassemblant. Pour BFO, il s'agit des *Fiat objects* (Smith, 2001), lesquels peuvent être physiques (ex : la frontière d'un pays, le centre de gravité de la Terre) ou temporels (ex : une partie particulière d'un processus comme la naissance ou la mort d'un être humain). Pour DOLCE, ce sont des *Non-physical objects* (Masolo *et al.*, 2004), parmi lesquels sont distingués les *Mental objects* dépendant d'un agent (ex : un procédé mnémotechnique propre à une personne) et les *Social objects* dépendant d'une pluralité d'agents (ex : une société commerciale, les règles d'un jeu de société). À noter que les exemples sont donnés ici à titre d'illustration de ces distinctions mais qu'ils ne constituent pas une proposition d'une ontologie particulière du monde.

¹¹ Bogen (2010) en donne plusieurs exemples dans la section 4 : *How observational evidence might be theory laden*.

masse d'air environnant le thermomètre. Une masse d'air différente située dans une autre partie de la pièce, par exemple à une autre hauteur du sol, aurait pu donner un autre résultat. On voit là que l'identification de l'entité mesurée comporte une part de convention (la plupart du temps sociale). À noter que, dans certaines situations, cette identification relève d'opérations complexes. Ainsi, dans le cas de données complexes comme une image (par exemple une image médicale) il est naturel de faire intervenir le sujet qui fait l'objet de la procédure, et qui est placé dans le champ de vue de l'appareil, selon une position et pendant une durée bien définies, adaptées aux objectifs de la procédure. À partir de l'image acquise, d'autres observations et mesures peuvent éventuellement être faites, en focalisant l'attention sur des régions particulières de l'image obtenue.

Cette dernière remarque nous amène à évoquer les données et leur rôle dans la construction des résultats de l'observation.

2.3 Les données d'observation

Comme nous l'avons indiqué en introduction, l'usage courant du terme « donnée » lui confère une acception large désignant toute entité porteuse de sens pour un être humain. Le contexte des études scientifiques dans lequel nous nous situons nous conduit toutefois, à l'instar de (Woodward, 2011), à nous focaliser sur une classe particulière de données correspondant à des *enregistrements physiques* porteurs de sens (p. 166) : « *Data are public records produced by measurement and experiment that serve as evidence for the existence or features of phenomena* ».

Comme le laisse entendre cette définition, lors d'études scientifiques, ces données interviennent dans des raisonnements variés allant de l'observation à des expérimentations complexes visant à établir de nouvelles connaissances sur des phénomènes singuliers ou généraux, par exemple : la prise de température d'un patient, tenant compte de l'enregistrement d'une valeur affichée sur un thermomètre ; la datation d'un fossile à partir d'une évaluation de la quantité de carbone-14 qu'il contient ; le calcul par Eddington en 1919 de la déviation de la lumière des étoiles par le soleil à partir de plaques photographiques (prises lors d'une éclipse totale du soleil) reflétant la position d'étoiles¹². Notre propos dans cette section n'est pas d'étudier la contribution de ces données à la construction ou l'évaluation de théories scientifiques, mais plutôt de caractériser leur nature ontologique en précisant notamment leur disposition à porter de l'information. Ainsi, nous nous attachons à rendre compte à la fois de leur matérialité (les données sont des enregistrements physiques matériels), leur artefactualité (elles sont, pour la plupart, intentionnellement produites) et leur fonctionnement sémiotique (elles sont porteuses de sens pour un être humain).

Pour suivre leur cycle de vie, précisons tout d'abord que ces données sont *acquises*, ce processus consistant en l'*inscription sur un support d'une forme matérielle*. Si on ne peut exclure des acquisitions fortuites (accidentelles) de données, on peut noter que, dans la très grande majorité des cas, les acquisitions sont réalisées intentionnellement et dans un but précis (elles répondent aux besoins d'une observation courante ou prévue). Afin de pérenniser la donnée acquise, un support est nécessaire. À nouveau, si divers objets physiques peuvent accidentellement

¹² Ces deux derniers exemples sont cités par Woodward (2011) pour mettre en évidence le rôle joué lors de raisonnements par des hypothèses et des théories, en complément des données, pour aboutir à des conclusions. Dans le cas de la datation du fossile, des hypothèses portant sur le cycle de vie du carbone-14 ainsi que l'exposition du fossile aux conditions atmosphériques doivent être prises en compte. Dans le cas de l'expérience d'Eddington, la théorie de la Relativité Générale portant sur les champs gravitationnels a servi de cadre pour définir les conditions de l'expérience et calculer la valeur de la déviation des rayons lumineux des étoiles.

servir de support (comme une table couverte de graffitis), on peut noter qu'il existe des objets intentionnellement produits pour réaliser cette fonction. C'est notamment le cas des systèmes d'information informatisés, qui nous concernent au premier chef dans ce document. Dans le cas des documents numériques et bases de données, il correspond en réalité à une donnée, non pas une mais au moins deux inscriptions. En effet, une inscription sur un support d'enregistrement (ex : une zone de mémoire interne d'ordinateur) non accessible à un être humain est transformée en une autre inscription sur un support de restitution (ex : le moniteur de l'ordinateur) pour être interprétée cette fois par un être humain. Comme nous ne souhaitons pas descendre à ce niveau de détail, nous considérons que notre *donnée* désigne ces différentes inscriptions réifiées. Par là même, nous nous contentons de considérer (comme nous l'avons fait jusqu'à présent) que l'interprétant est un être humain, laissant de côté les dispositifs matériels et logiciels conçus pour interpréter les inscriptions sur des supports d'enregistrement.

Comme le montrent les pratiques courantes de systèmes informatisés, les données peuvent être effacées, détruites, dupliquées, compressées, cryptées, etc. Certains de ces processus (intentionnels ou non) ne dépendent que de la matérialité des données. D'autres, comme la compression, tiennent compte du contenu des données : la compression doit se faire sans perte d'information. Ceci nous conduit à évoquer le contenu sémantique des données.

La caractérisation du contenu sémantique des données est une question centrale pour la philosophie de l'information (Floridi, 2011a). Selon un cadre historiquement proposé par (Carnap and Bar-Hillel, 1952) et aujourd'hui largement accepté¹³, une donnée « signifiante » constitue une ancre physique matérielle pour un contenu mis en forme et, qui plus est, bien formé. Cette définition suppose qu'un code ou langage de communication soit utilisé. Le terme « bien formé » signifie que la donnée, en tant que forme, est structurée selon des règles (une syntaxe) gouvernant ce code ou langage de communication. Le contenu sémantique, ou « signification », est dès lors contraint par la sémantique du code ou langage choisi. Dans le cas de nos données d'observation, le langage est habituellement une langue naturelle et la forme est celle d'une expression linguistique composée d'entités signifiantes. Lorsque ces entités sont des mots, les données ont pour contenu des nombres, des valeurs comportant une unité de mesure ou d'autres entités nommées. Dans le cas d'entités linguistiques plus complexes, les données ont des contenus eux-mêmes plus complexes, parmi lesquels deux catégories sont principalement distinguées : des contenus factuels (ex : « la température du patient est de 23°C ») et des instructions (ex : « pour prendre le pouls d'un patient, faire ... puis ... »).

Le contenu sémantique *littéral* de la donnée, tel que nous venons de le définir, n'épuise toutefois pas la sémantique de la donnée, ceci pour plusieurs raisons. D'une part, des données prises isolément peuvent ne pas avoir un contenu significatif vis-à-vis des entités observées. C'est notamment le cas des pixels composant une image d'IRM fonctionnelle. Ces pixels peuvent en théorie recevoir une interprétation¹⁴, cependant celle-ci n'a pas d'intérêt pratiquement (en revanche, une collection de pixels déterminée par segmentation pourra être interprétée comme représentant une structure anatomique). D'autre part, des données peuvent avoir un contenu codé, comme lorsque des échelons d'échelles de mesure sont codés au

¹³ Pour une présentation complète tout en restant synthétique de ce cadre, le lecteur peut se reporter à (Floridi, 2011b).

¹⁴ Comme l'indique (Bogen, 2010, Section 3), plusieurs interprétations sont du reste possibles : “*If fMRI images record observations, it's hard to say what was observed – neuronal activity, blood oxygen levels, proton precessions, radio signals, or something else. (if anything is observed, the radio signals that interact directly with the equipment seem to be better candidates than blood oxygen levels or neuronal activity)*”.

moyen d'un adjectif ou d'un nombre¹⁵. Dans ce cas, un lien est à établir avec un autre contenu plus expressif. Enfin, et pour mentionner une situation fréquente, des données peuvent avoir des contenus variables dépendant de choix de concepts, réalisés par l'observateur pour le décrire. Ainsi, une même solution pourra être décrite par un observateur selon sa couleur et par un autre observateur selon la valeur de son ph ou encore son niveau d'acidité. Dans ce cas, une bonne interprétation du contenu suppose de disposer de règles pour passer d'une description à une autre.

Au-delà donc du sens littéral d'une donnée, la contextualisation de la donnée et de son contenu concourent à la doter d'une sémantique *pragmatique* (Shet, 1996). Ceci revient à relier conceptuellement le contenu à d'autres entités, notamment l'entité observée, le processus d'observation, l'instrument de mesure utilisé, l'observateur, etc. Par ailleurs, des informations sur la donnée elle-même, sa provenance, sa fiabilité, etc., viennent compléter l'intelligibilité de la donnée et accroître ainsi son utilisabilité.

2.4 Bilan sur les données et leur sémantique

Au terme du parcours que nous venons de réaliser pour analyser la nature ontologique des données de l'observation (ou, plus généralement, « données scientifiques »), nous résumons ici les notions de *donnée* et de *sémantique de donnée* que nous adoptons.

- Les données sont des enregistrements publics possédant un contenu informationnel. Des données élémentaires peuvent être membres de collections (spatiales ou temporelles) de données ou être des composants de données structurées auxquelles correspondent des contenus informationnels complexes.
- Les données sont exploitées et produites à l'occasion d'observations, de mesures ou d'études scientifiques reposant sur des expérimentations complexes. Que ces actions soient passives ou actives (en termes d'interactions avec l'environnement), simples ou complexes (en termes de composition), les données possèdent un lien avec l'objectif poursuivi et ainsi avec une entité, ou classe d'entités, objet(s) de l'observation ou de l'étude.
- Les entités observées ou étudiées sont des entités conçues par un agent ou une communauté d'agents et relèvent de réalités diverses : physique, cognitive, sociale. Ces entités rendent compte d'aspects statique/structurel et dynamique/comportemental des phénomènes observés/étudiés.
- La sémantique des données consiste en les relations que les données entretiennent avec d'autres entités intervenant dans le cadre de l'observation/étude. Certaines de ces entités relèvent des phénomènes observés/étudiés et constituent le contenu informationnel des données. D'autres entités relèvent de l'observation/étude elle-même et viennent compléter la sémantique de ces données.

Dans la section suivante, nous partons de ce bilan pour proposer un cadre ontologique permettant de représenter explicitement les données et leur sémantique.

¹⁵ Par exemple, l'échelon codé par le nombre 4.0 de l'instrument neurologique EDSS (Kurtzke, 1983) et synthétisant la capacité de marche d'un patient atteint de sclérose en plaques, a pour signification : *"Fully ambulatory without aid, self-sufficient, up and about 12 hours a day despite relatively severe disability consisting of one FS grade 4, or combination of lesser grades exceeding limits of previous steps ; able to walk without aid or rest some 500 meters."*

3 Modèle ontologique de DOLCE-CORE

Comme nous l'indiquions en introduction, nous avons fait le choix de l'ontologie fondatrice DOLCE-CORE pour structurer notre cadre ontologique. Dans cette section, nous détaillons donc cette ontologie et ses engagements vis-à-vis de l'observation et de la mesure. DOLCE-CORE présente des différences substantielles par rapport à DOLCE, motivées notamment par des réflexions sur la nature des propriétés des objets, en lien avec des théories expérimentales de la mesure (Masolo, 2010a). Contrairement à DOLCE, DOLCE-CORE n'a pas été encore publiée dans son intégralité et des parties de cette ontologie font toujours (fin 2013) l'objet de travaux. Pour cette présentation, nous nous basons principalement sur les publications : (Borgo and Masolo, 2009)(Masolo, 2010a, 2010b).

3.1 Principales catégories

DOLCE-CORE distingue quatre principales catégories d'entités jouant des rôles spécifiques pour la représentation de phénomènes, observés ou conceptualisés (cf. Figure 1). Toutes ces entités sont *concrètes* dans le sens où elles possèdent une extension spatio-temporelle.

Les « *Objects* » et « *Events* » sont deux catégories disjointes d'entités observables ou, plus généralement, classables. Leur nature différente tient à la façon dont ces entités sont localisées dans l'espace-temps : les *Objects*, notamment les objets physiques, sont *principalement* reliés à l'espace tandis que les *Events* sont *principalement* reliés au temps¹⁶. Les *Objects* n'acquièrent de localisation temporelle que secondairement, par l'intermédiaire des *Events* auxquels ils participent. Réciproquement, les *Events* n'acquièrent de localisation spatiale que secondairement, par l'intermédiaire des *Objects* qui les engendrent.

Les « *Qualities* » sont les dimensions ou aspects selon lesquels les agents perçoivent, classent et comparent les *Objects* et *Events*. Les *Objects* et *Events* individuels possèdent des *Qualities* individuelles qui leur sont propres. Ces *Qualities* individuelles relèvent de types de *Qualities* correspondant à des dimensions communes à des classes d'entités. Il convient de noter (voir Figure 1) que les *Qualities* d'*Objects* sont distinctes des *Qualities* d'*Events* : les *Objects* physiques ont une masse, une forme, une texture, etc., tandis que les *Events* ont une durée, sont rapides ou lents, se chevauchent temporellement ou bien au contraire surviennent avant ou après d'autres *Events*¹⁷.

Le classement d'un *Object* ou *Event* selon une dimension (une *Quality*) consiste à localiser sa *Quality* individuelle dans un espace de positions propre au type de *Quality* (ex : « être d'une longueur de 2 mètres », « être d'une durée de 36 secondes »). Les auteurs de DOLCE-CORE s'appuient ici sur la notion d'espace conceptuel de Gärdenfors (2000). La catégorie « *Region* » (voir Figure 1) regroupe à la fois des positions atomiques, appelées (au pluriel) « *Qualia* », et des

¹⁶ En retenant cette distinction, les auteurs de DOLCE-CORE suivent la conception de Hacker (1982). On notera à ce propos une différence avec la distinction établie entre *Endurants* et *Perdurants* dans DOLCE, laquelle reposait sur le comportement des parties des entités. Par contre, contrairement à Hacker (1982), pour qui les objets sont assimilés à des objets physiques, la catégorie *Object* de DOLCE-CORE a une extension plus large car elle embrasse des entités cognitives et sociales relevant d'autres réalités que la réalité physique.

¹⁷ L'existence même de ces aspects distincts est un argument mis en avant pour considérer dans l'ontologie les deux catégories d'entités – *Objects* et *Events* – et ne pas chercher, dans une démarche réductionniste, à vouloir réduire l'une à l'autre.

« Spaces » correspondant à la structure recouvrant (méréologiquement) l'ensemble des *Qualia* pour un type de *Quality* donné.

Dans les sections suivantes, nous détaillons les catégories *Quality* et *Region* pour mettre en évidence la puissance d'expression apportée par DOLCE-CORE. Ce faisant, nous nous référons, à titre de comparaison et de positionnement, au *Système International (SI) des Unités* (NIST, 2008). Le SI repose sur une terminologie et des concepts (ex : *grandeur*, *unité de mesure*, *valeur*) définis dans le *Vocabulaire International de Métrologie (VIM)* (JCGM, 2008). Cette référence nous accompagnera dans les extensions de DOLCE-CORE présentées en Section 4.

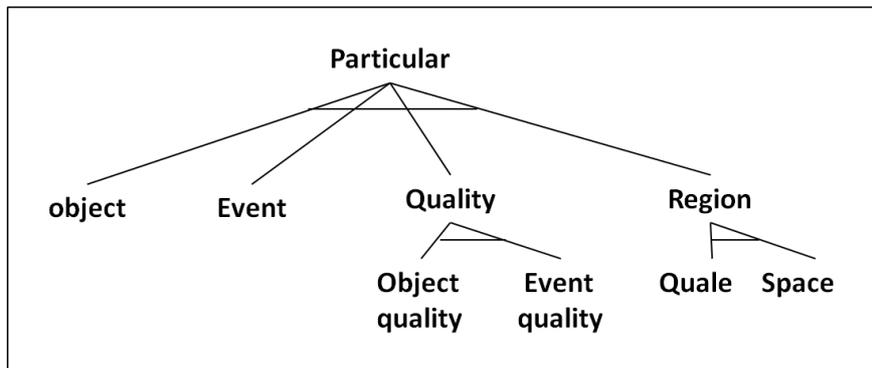


Figure 1. Principales catégories d'entités considérées par DOLCE-CORE. Un trait descendant représente un lien de subsomption (le concept placé plus haut subsume le concept placé plus bas) ; un trait horizontal indique que les concepts fils subsumés sont incompatibles (les classes d'entités réalisant ces concepts sont disjointes).

3.2 Compléments sur les *Qualities*

Comme le précisent Masolo et Borgo (2005), dans leur synthèse sur le traitement des qualités dans le domaine de l'ontologie formelle, les *Qualities* de DOLCE sont à rapprocher des *quantités* du SI. Dans leur portée, les quantités du SI rendent compte essentiellement de grandeurs quantitatives mais aussi qualitatives (ces dernières, comme le *sexe* d'une personne ou la *couleur* d'un objet, sont nommées « propriétés qualitatives »). Les types de *Qualities* de DOLCE (et DOLCE-CORE) correspondent aux quantités dans un *sens général* du SI tandis que les *Qualities* individuelles correspondent aux quantités dans un *sens particulier*. Les *Qualities* individuelles et les quantités dans un sens particulier sont *inhérentes* à un *hôte* (ou *porteur*) : elles lui sont propres (le poids de Paul et le poids de Marie sont deux individus différents) et elles ont la même extension temporelle que leur hôte (le poids de Paul existe tout au long de l'existence de Paul)¹⁸.

Pour poursuivre la comparaison, ajoutons que le SI considère que certaines quantités sont de même *nature* que d'autres quantités. Ainsi, par exemple, un *diamètre*, une *circonférence* ou une *longueur d'onde* sont de même nature qu'une *longueur*, tandis qu'une *chaleur*, une *énergie cinétique* ou une *énergie potentielle*

¹⁸ Sur ce point, DOLCE et le SI se démarquent de EngMath (Gruber and Olsen, 1994), une première ontologie à avoir conceptualisé le domaine des quantités et des unités de mesure pour des applications d'ingénierie exploitant des modélisations mathématiques. Si EngMath considère des types de quantités, en revanche les instances de ces quantités génériques ne dépendent pas d'entités hôtes et correspondent plutôt à la notion de *valeur* dans le SI. Ainsi, dans EngMath, les qualités individuelles ne sont pas représentées en tant que telles. La conséquence, sur le plan de l'expressivité de l'ontologie, est qu'il n'est pas possible de se référer à des qualités d'entités pour exprimer, par exemple, que : le *poids de Paul* a baissé durant une période donnée, la *température de Marie* a été prise à tel moment de la journée ou sa *température* est anormalement élevée depuis une date donnée.

peuvent être considérées comme des quantités de la nature *énergie*. Bien que DOLCE-CORE ne se prononce pas sur cette notion de proximité de nature, il paraît clair que la relation de spécialisation entre types de *Qualities* est un moyen adéquat d'en rendre compte. La structuration des types de *Qualities* n'est pas contrainte dans DOLCE-CORE, excepté la classification *Object vs Event quality*. Mais cette distinction n'empêche pas d'introduire (comme on le voit dans des ontologies courantes) d'autres distinctions comme *Physical vs Chirical vs Biological quality*.

Pour compléter l'analyse de la nature des *Qualities*, il est intéressant de faire état d'une classification proposée par Gärdenfors (2004), conduisant à considérer deux catégories principales de qualités suivant leur origine. D'un côté, des qualités comme la *couleur*, la *forme* d'un objet ou le *goût* d'un aliment sont des qualités « psychologiques » dérivant directement d'output des systèmes sensoriels d'êtres humains et d'animaux. Pour la plupart, ces qualités sont innées et se développent au gré des expériences sensorielles des individus. D'un autre côté, des qualités comme le *poids* d'un objet (au sens de la mécanique newtonienne, c'est-à-dire distinct de la *masse*), le *prix* d'un produit ou la *fonction* d'un artefact sont des qualités « théoriques » dépendant de facteurs culturels, par exemple de théories scientifiques. Ces dernières sont acquises et les dimensions qui leur sont associées (les *Regions*, au sens de DOLCE-CORE) sont façonnées dans leur structure par les besoins d'échanges d'information entre agents. Nul doute que ces distinctions sont importantes à prendre en considération pour structurer une taxinomie de *Qualities*. La notion de qualité « théorique » rejoint la visée de DOLCE-CORE de rendre compte d'autres réalités que la réalité physique, notamment la réalité sociale. Ainsi, la notion de *Quality* est-elle amenée à rendre compte, par exemple, de la *cohésion* d'une organisation, de la *clarté* d'un texte ou de la *difficulté* des règles d'un jeu.

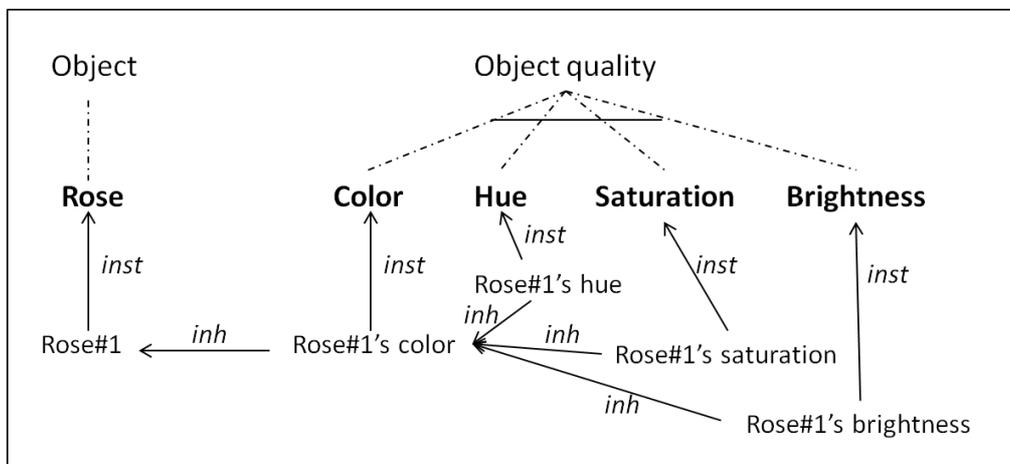


Figure 2. Modélisation des *Qualities* dans DOLCE-CORE. Un trait descendant en pointillés indique l'existence de concepts intermédiaires. La relation *inst(x,y)* tient pour « x est une instance de y » (ex : l'individu *Rose#1* est une instance du type *Rose*) ; la relation *inh(x,y)* tient pour « x est inhérente à y » (ex : la *Quality* individuelle *Rose#1's color* est inhérente à l'individu *Rose#1*).

Pour clore la présentation des *Qualities*, précisons qu'un aspect pris en compte dans DOLCE, mais qui est absent du SI, est le traitement de certains types de *Qualities* qualifiées de « multi-dimensionnelles » (ou « multi-aspectuelles »). Une *Quality* multidimensionnelle correspond à un aspect comme la *couleur* dépendant de plusieurs autres aspects : *teinte*, *saturation*, *intensité*. Dans la littérature, d'autres qualités sensorielles comme la *hauteur* d'un son ou le *goût* ont été caractérisées

comme multi-dimensionnelles¹⁹. Dans DOLCE-CORE (cf. Figure 2), les *Qualities* indépendantes de base constituant une *Quality* multi-dimensionnelle sont modélisées comme étant inhérentes à cette dernière²⁰.

3.3 Complément sur les *Regions*

Comme nous l'avons indiqué, les auteurs de DOLCE s'appuient sur une théorie cognitive, celle des espaces conceptuels de Gärdenfors (2000), pour expliquer comment des agents, au premier rang desquels les humains, classent les *Objects* et *Events* qu'ils perçoivent et conçoivent selon des *Qualities* jouant le rôle de dimensions de classement.

Les *Qualities* individuelles (qu'il s'agisse des *Object qualities* ou des *Event qualities*) occupent temporairement des positions dans un *Space* recouvrant toutes les positions pouvant être occupées par des *Qualities* individuelles du même type²¹. Ce mapping (ou localisation) temporaire revient à abstraire les *Qualities* individuelles de leur hôte et du temps, mais non de leur type. Ainsi (cf. Figure 3), une position 32 cm width est-elle différente de la position 32 cm height²². Une *Quale* correspond à une position occupée par une *Quality* à un instant de temps²³. Un (*Quality*-)*Space* correspond à la somme méréologique des *Qualia* pour un type de *Quality* donné. Entre les *Qualia* et les *Spaces*, les *Regions* composées (méréologiquement) correspondent aux positions occupées par une *Quality* individuelle pendant un intervalle de temps (ex : *Red*, comme la somme de *Scarlet* et de *Crimson* ; la *Region* ayant pour parties exactement les deux positions 32 cm width et 33 cm width).

La nouveauté introduite dans DOLCE-CORE (par rapport à DOLCE) est d'autoriser que plusieurs *Spaces* soient associés à un même type de *Quality*, rendant compte du fait que des théories et/ou des instruments différents peuvent être utilisés lors des mesures. Ainsi (cf. Figure 3), les positions *Red*₁ et *Red*₂ localisant la couleur d'une rose sont-elles considérées comme deux individus distincts faisant partie de deux espaces de couleurs différents (*Colored*₁ et *Colored*₂). De la même façon, pour mesurer la hauteur d'un cube, nous pourrions recourir à un espace de hauteur qualitatif comportant trois échelons : *small*, *medium* et *big*. On rappellera par ailleurs que les *Regions* sont considérées dans DOLCE-CORE comme des entités

¹⁹ Pour ces qualités, il n'existe toutefois pas toujours de consensus dans l'identification des qualités de base constituantes. Ainsi, pour le *goût*, des théories à 3, 4 voire plus de qualités de base, ont été proposées dans la littérature (voir à ce propos (Schiffman, 1982, chap. 9) pour un exposé de ces théories).

²⁰ Selon la théorie des espaces conceptuels de Gärdenfors (2000), la *couleur* est un ensemble *intégral* de dimensions : un objet possédant une localisation pour une des dimensions de l'ensemble (ex : la *teinte*), possède nécessairement une localisation pour les autres dimensions (ex : la *saturation* et l'*intensité*). En modélisant le lien entre les *Qualities* de base et la *Quality* multi-dimensionnelle comme une relation d'inhérence, DOLCE-CORE s'assure d'une dépendance mutuelle entre toutes ces *Qualities*.

²¹ Dans le cas des *Events*, il faut considérer que ces objets persistent en accumulant des parties (le lecteur peut penser ici à une conférence consistant en l'enchaînement d'événements divers). Techniquement, à un instant donné, la partie maximale de l'*Event* est prise en compte pour déterminer la position occupée par une de ses qualités (par exemple, pour évaluer la *durée* ou l'*intérêt* de la conférence).

²² On notera la différence entre la notion de *position* dans DOLCE-CORE et celle de *valeur* dans le SI : une même *valeur* correspond à des positions distinctes. Nous revenons sur ce point en Section 4.2.

²³ DOLCE-CORE considère des instants et des intervalles de temps qui sont des *Regions* d'un espace de temps.

concrètes (et non des entités abstraites, comme dans DOLCE) pour tenir compte du fait, qu'à l'instar des échelles de mesure, ces *Regions* sont créées, maintenues puis éventuellement abandonnées par des communautés d'agents.

Un rapprochement utile peut justement être établi entre la catégorie *Space* de DOLCE-CORE et la notion d'*échelle de mesure*. Ce rapprochement se justifie par le fait qu'une même fonction cognitive soit attribuée à ces entités dans la littérature. De façon abstraite, les échelles de mesure ont été définies comme un moyen permettant in fine d'affecter des nombres ou mots (nous parlerons de *valeur*) à des objets ou événements en guise de mesure, (ex : Stevens, 1946, p. 677) : « *We may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules* ». Toujours selon Stevens, ces « règles » ou « échelles de mesure », déterminent avant tout notre capacité à classer et comparer des entités selon différents aspects (Steven, 1946, p. 677) : « *In dealing with the aspects of objects we invoke empirical operations for determining equality (classifying), for rank-ordering, and for determining when differences and when ratios between the aspects of objects are equal* ». Comme on le voit, les définitions sont proches entre les *dimensions qualités* de Gärdenfors (2000, 2004) et les *échelles de mesure* (comme le sont du reste les exemples donnés à titre d'illustration des deux côtés). Un intérêt de ce rapprochement est de souligner que des échelles « simples » (en termes de structure) comme des échelles *nominales* ou *ordinales* peuvent faire office de *Space*. À ce propos, notons un exemple d'échelle de mesure nominale jouant un rôle important, celle de la normalité distinguant 2 positions : *normale* et *anormale*. De tels *Spaces*, appliqués à différents types *Qualities*, peuvent servir à classer les *Qualities* en normales et anormales (ex : une *Normal Temperature* est une *Temperature* ayant pour location temporaire (*loc*) la position *Normal*²⁴.

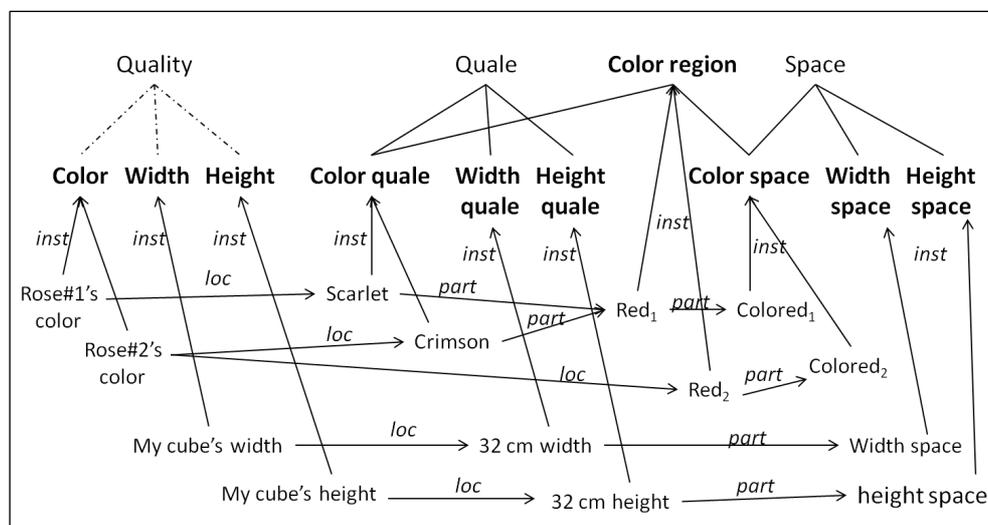


Figure 3. Modélisation des *Regions* dans DOLCE-CORE. La relation *loc(x,y)* tient pour « x (une *Quality*) a pour localisation y (une *Region*) »²⁵ (ex : *Rose#1's color* a pour localisation *Scarlet*); la relation *part(x,y)* tient pour « x est une partie de y » (ex : *Crimson* est une partie de *Red₁*).

²⁴ À titre d'illustration, mentionnons l'ontologie PATO (Phenotype And Trait Ontology) définissant une ontologie de qualités/traits pour caractériser les caractères phénotypiques d'espèces animales (Gkoutos *et al.*, 2004). Une large proportion des qualités de PATO sont classées sous les classes génériques : 'decreased object quality' et 'increased object quality'.

²⁵ Comme il a été précisé dans le texte, la relation de localisation de DOLCE-CORE comporte un argument complémentaire, le temps. Pour des raisons de clarté de la figure, nous ne faisons pas apparaître cet argument.

3.4 Bilan sur le modèle ontologique de DOLCE-CORE

Pour clore la présentation de DOLCE-CORE, nous formulons quelques remarques sur la généralité de ce modèle ontologique en même temps que son caractère partiel pour rendre compte de tous les aspects d'un processus d'observation.

En considérant la classification d'objets et d'événements quelconques selon une dimension, le modèle dépasse dans sa portée le contexte de l'observation (ou de la perception) pour s'appliquer à des *objets non physiques* conçus par un (ou plusieurs) agent(s). Dans DOLCE, les `Non-physical objects` sont définis comme des objets dépendant d'agents individuels (pour les `Mental objects`) ou de communautés d'agents (pour les `Social objects`). En tant qu'`Objects`, ils sont *principalement* localisés dans l'espace, mais cette localisation n'est, en quelque sorte, qu'*indirecte*. Un procédé mnémotechnique, un texte de loi, les règles d'un jeu de stratégie, une organisation commerciale, etc., en sont des exemples²⁶. Le modèle ontologique de DOLCE-CORE que nous venons de présenter couvre donc la classification de ces objets non physiques. Ainsi, par exemple dans une activité relevant de l'intelligence économique, le modèle permet de rendre compte du classement d'informations selon leur importance ou fiabilité. Il permet également de classer des ensembles (un autre type d'entités conçues) selon leur cardinalité. Cette généralité a du reste été soulignée par Gärdenfors (2000) comme un atout de sa théorie des espaces conceptuels.

Dans le même temps, le modèle n'introduit qu'une partie des entités se trouvant impliquées dans un processus d'observation. Nous avons déjà évoqué les *valeurs* de quantités physiques (ex : 15 grammes), dont la conceptualisation nécessite de faire un lien entre les `Qualia` de DOLCE-CORE, les nombres et les unités de mesure. Par ailleurs, une observation, en tant qu'action, fait intervenir un agent observateur et, suivant les cas, des instruments de mesure et des données. Qui plus est, comme indiqué en Section 2.1, le résultat peut consister en une expression langagière. De telles expressions prennent couramment la forme de chiffres (ex : « 15 »), éventuellement complétés d'un symbole d'unité de mesure (ex : « 15 cm »), ou d'adjectifs (ex : « petit »). Il nous faut introduire de telles expressions (notamment linguistiques) qui dénotent des *valeurs*, au sens du SI. Il nous reste également à introduire des supports matériels physiques pour rendre compte des enregistrements de ces expressions que constituent les données. Dans la section suivante, nous proposons des extensions de DOLCE-CORE pour conceptualiser ces différentes entités.

4 Extensions apportées au noyau ontologique DOLCE-CORE

Dans cette section, nous décrivons des extensions apportées à DOLCE-CORE pour obtenir un modèle rendant compte de l'ensemble des entités impliquées dans la sémantique des données.

Tout d'abord, nous complétons le modèle du classement des entités (`Objects` et `Events`) en introduisant les *valeurs* (au sens du SI) servant à communiquer le résultat de mesures (4.2). Auparavant, afin d'obtenir un modèle plus large, nous introduisons des *collections* (une forme d'ensembles) dont la fonction cognitive est de permettre de dénombrer, ce qui est une forme de mesure (4.1).

Par la suite, nous élargissons notre perspective en prenant en compte les actions d'observation et de mesure, leurs données et résultats. Nous introduisons ainsi (4.3) l'action d'observation, celle-ci faisant le lien avec l'ensemble des entités impliquées

²⁶ Ces différents objets ont des contreparties matérielles physiques (ex : des documents, des personnes, des représentations mentales) leur conférant une localisation spatiale.

lors d'une observation : l'agent de l'observation, l'entité observée, la (ou les) position(s) occupée(s) par différentes qualités de l'entité observée, le (ou les) instrument(s) utilisé(s), les données produites et/ou exploitées. Il nous reste alors à pouvoir décrire les instruments de mesure et les données. En section 4.4, nous introduisons un ensemble minimal de concepts, définis dans l'ontologie formelle d'artefacts de Kassel (2010), pour permettre de rendre compte de l'origine intentionnelle des instruments de mesure et de leur fonction. En section 4.5, nous réutilisons le module ontologique I&DA (Fortier and Kassel, 2004) dédié à la conceptualisation d'informations pour définir les *données* en tant que des enregistrements / inscriptions physiques matérialisant des expressions codées dans un langage de communication et possédant un contenu informationnel pour des agents (leur sémantique). L'entité *donnée* étant introduite, il est alors possible de leur attribuer des propriétés pour décrire par exemple leur provenance ou les classer selon leur fiabilité pour des agents, ces informations constituant autant de métadonnées.

4.1 Collections

Le terme « collection » est couramment utilisé de façon large pour désigner différentes notions qui ont été définies en Logique et en Ontologie Formelle : *ensemble, classe, collectif, entité plurielle, multitude*. Un exposé exhaustif de ces notions étant hors de portée pour ce document, nous nous contentons de présenter ici une notion de *collection* que nous avons adoptée dans des travaux précédents (Kassel *et al.*, 2012) et que nous retenons comme catégorie (nommée *Collection*) dans DataTop.

Cette catégorie *Collection* s'inspire très fortement de la notion d'*ensemble naturalisé (incarné)* proposée par Bottazzi *et al.* (2006)²⁷. Essentiellement, une *Collection* est une *unité* ou *ensemble* d'entités mentalement construite par un agent ou une communauté d'agents. Il s'agit donc d'un *Non-physical object* pouvant être, suivant les cas, un *Mental object* ou un *Social object*. L'unité / ensemble joue en quelque sorte le rôle d'un container contenant temporairement des entités. Ces dernières *appartiennent* dès lors temporairement à la *Collection* (relation : *isMemberOfAt*) dans le sens où elles sont conçues comme étant contenues dans la *Collection*.

Une propriété importante des *Collections* est que leurs membres peuvent varier au cours du temps. Une *Collection* est ainsi un ensemble intensionnel dont l'identité est définie au moyen d'une propriété, comme pour les exemples : *le fond documentaire de la Bibliothèque Nationale de France (BNF), ma collection de toupies, l'ensemble des membres participant au projet CrEDIBLE, la collection des briques constituant le mur de mon jardin*. La propriété jouant le rôle de critère d'identité fait systématiquement référence à une propriété s'appliquant à une entité (ex : « être un ouvrage géré par la BNF », « être une toupie que je collectionne ») dont la possession à un instant *t* par l'entité détermine son appartenance au même instant *t* à la *Collection*. Une *Quality* des *Collections* est leur *Cardinal*, c'est-à-dire le nombre de leurs membres (un entier pouvant être nul).

Pour compléter cette description des *Collections*, nous formulerons deux remarques. Nous noterons tout d'abord qu'en conséquence de la nature intensionnelle des *Collections*, deux *Collections* distinctes peuvent être extensionnellement équivalentes, en ayant les mêmes membres. Ceci se produit lorsque les propriétés jouant le rôle de critère d'identité sont extensionnellement

²⁷ Dans cette référence, les auteurs analysent finement les différences entre leur notion d'ensemble naturalisé et la notion d'ensemble mathématique.

équivalentes (ex : « être une brique utilisée pour bâtir le mur de mon jardin » et « être une brique achetée dans tel magasin à telle date »). Nous soulignerons ensuite le caractère *primitif* de la relation `isMemberOfAt` de `DataTop`. Ainsi, contrairement à un usage courant (voir, par exemple, (Wood and Galton, 2009)), nous ne définissons pas cette relation comme sous-relation de `isPartOfAt`. Ceci permet de considérer que des `Physical objects` comme des briques soient membres d'une `Collection` (assimilée à un `Non-physical Object`) sans violer la contrainte d'homogénéité de DOLCE selon laquelle un `Non-physical Object` n'a pour parties que des `Non-physical objects`.

4.2 Valeurs et unités de mesure

Comme nous l'avons vu, la conceptualisation des `Qualities` de DOLCE-CORE s'appuie sur la théorie proposée par Gärdenfors (2000) consistant en un modèle des représentations d'informations d'agents cognitifs. Selon cette théorie, les agents se représentent des entités suivant des dimensions/qualités dotées de structures géométriques. Selon l'interprétation qui est faite dans DOLCE-CORE de cette théorie – la distinction des catégories `Quality` et `Region` – les positions qu'occupent temporairement les entités selon ces dimensions permettent à des agents de les comparer suivant ces mêmes dimensions, par exemple de conceptualiser le fait que *Marie est plus petite en taille que Paul*. Par contre, la catégorie `Region` de DOLCE-CORE ne permet pas de rendre compte de la communication, notamment langagière, de résultats de mesure entre agents. Du côté du SI, en revanche, cette fonction de communication de résultats de mesures est clairement dévolue à l'entité *valeur*. La *valeur* correspond au corrélat conceptuel d'expressions comme « 324 mt » ou, tout simplement, « 324 ». Une valeur peut aussi être qualitative et s'exprimer par des mots comme « grand », « profond », ou le syntagme « très sévère ». Ainsi donc, les *valeurs* du SI paraissent complémentaires des catégories actuellement proposées par DOLCE-CORE. Pour continuer à souligner cette complémentarité, il suffit de constater que la plupart des instruments de mesure affichent directement des valeurs, laissant à l'humain le soin (suivant les cas) de positionner ces valeurs vis-à-vis d'autres valeurs se rapportant à la qualité mesurée pour un type d'entité mesurée. Inversement, il paraît clair que des tâches cognitives ne nécessitant pas une communication (ou tout simplement une conceptualisation) de valeur s'appuient uniquement sur une représentation d'information basée sur une localisation de qualité²⁸. Partant de ce constat de complémentarité, nous faisons le choix d'introduire la notion de *valeur* (catégorie `Value`) dans `DataTop`.

Une `Value` est une entité conceptuelle exprimée par un terme (il s'agit d'une entité nommée) et dont la fonction est de pouvoir représenter une `Quality` d'une entité en jouant le rôle de résultat d'une mesure de cette `Quality` individuelle. Une `Value` peut être (cf. Figure 4) qualitative (`Qualitative value`) ou quantitative en intégrant un nombre (`Quantitative value`) et, dans ce dernier cas, correspondre à un `Number`, une `Scalar value` ou une `Tensorial value`. Ces `Values` sont adossées respectivement aux `Regions` et `Qualities` au moyen de deux relations :

- Une `Region` a comme valeur (`rValue`) une `Value`. Par exemple (cf. Figure 4), la `Quality` : 324 mt height localisant la `Quality` individuelle Eiffel tower's height possède comme `Scalar value` : 324 mt. La particularité d'une `Value` par rapport à une `Region` est de s'abstraire du type de `Quality` :

²⁸ Ce point est du reste souligné par Gärdenfors (2004, p. 11) : « *The quality dimensions are taken to be independent of symbolic representations in the sense that we and other animal can represent the qualities of objects, for example when planning an action, without presuming an internal language or another symbolic system in which these qualities are expressed* ».

324 *mt* peut ainsi valuer indifféremment une profondeur, une longueur ou une largeur. On notera que toutes les *Regions* n'ont pas nécessairement de *Value* associée. Une telle valuation dépend de la nature de la *Quality* et de la structuration du *Space* qui lui correspond. Dans le cas d'échelles de mesure conventionnellement définies pour communiquer des informations, tous les échelons (correspondant à des *Qualia*) sont exprimables en langue et se voient donc associer une *Value*. En revanche, dans le cas de *Qualities* et de *Spaces* dérivés d'un système sensoriel d'un agent, comme la *couleur*, nous savons qu'il existe des positions perçues qui ne sont pas exprimables en langue.

- Une *Quality* individuelle a temporairement comme valeur (*qValue*) une *Value*. La relation *qValue* correspond à la composition des relations *loc* et *rValue* lorsqu'une *Quality* individuelle possède une localisation dans un *Space*. Par contre, comme nous l'avons vu plus haut, une *Value* peut être associée directement à une *Quality* individuelle en l'absence de toute localisation, ce qui justifie l'introduction de cette relation *qValue*.

Un cas particulier de mesure est le dénombrement de collections. De telles situations de mesure sont fréquentes notamment en psychométrie, lorsque l'administration d'un test conduit à faire réaliser des actions par le sujet (Batrancourt *et al.*, 2010). Dans ces cas, nous avons à considérer des collections de résultats (ex : un ensemble d'objets correctement identifiés par le sujet ou, au contraire, un ensemble d'actions mal réalisées). De même, des questionnaires peuvent porter sur des événements particuliers ayant émaillé la vie d'un patient sur une période de temps donnée (ex : un ensemble d'état d'anxiété extrême vécus par le patient durant une semaine). Ces ensembles correspondent à des *Collections*, telles que nous les avons définies en Section 4.1²⁹. Les *Qualities* en jeu sont des *Cardinals* et le *Space* leur correspondant est une échelle ordinale. Les échelons de cette échelle (*Qualia*) sont dès lors des localisations de cardinalité (ex : A cardinality of 3). Selon notre conceptualisation des *Values*, ceci revient à associer (*rValue*) à ces positions de cardinalité des *Numbers* comme *Values*.

Au-delà du rôle de communication d'un résultat de mesure, la reconnaissance de l'entité *Value* est un moyen de rendre compte du fait que deux *valeurs* reposant sur des unités de mesure différentes peuvent être considérées, en un certain sens, comme identiques, par exemple que $324\text{ mt} = 1063.044\text{ ft}$ ou que $5\text{ m per s} = 18\text{ km per h}$. Une telle identité, que nous conceptualisons au moyen de la relation *eqT*, s'explique par le fait qu'il s'agit de *Values* associées à une même *Region*. Une même équivalence peut tenir entre des *Qualitative values* et des *Quantitative values* comme l'illustre en Figure 4 l'exemple de l'échelle de mesure proposée par Morris (1983) pour évaluer la sévérité d'un état de démence : l'échelon de l'échelle *Mild severity* est codé à la fois par la *Qualitative value* : *Mild* et le *Number* : 1. On notera que de telles correspondances peuvent être établies entre des *Regions* structurant un même type de *Quality*. Supposons qu'on utilise une échelle de mesure à 5 points pour juger de la hauteur des bâtiments de la ville de Paris : <Very low, Low, Medium, High, Very high>. Dans ce cas, des correspondances peuvent être établies entre des échelons de cette échelle et des *Regions* d'un autre *Space* structuré selon une unité de mesure, le mètre ou toute autre unité. Les *Regions* concernées n'ayant pas forcément de *Value* associée, il apparaît que ces correspondances entre *Regions* apportent des

²⁹ On notera, en considérant les exemples venant d'être énoncés, l'importance de considérer des *Collections* générales dont les membres peuvent être indifféremment des *Objects* ou des *Events* et dont la cardinalité peut être nulle. Cette généralité est à contraster avec la proposition de *collection* faite par Bottazzi *et al.* (2006).

informations complémentaires par rapport aux correspondances entre *Values*. Par contre, si les correspondances ne sont dues qu'à l'usage d'unités de mesure différentes et ne dépendent pas d'un type de *Quality*, il paraît préférable de les gérer au niveau des *Values*.

En introduisant dans DataTop la notion de *valeur*, au sens du SI, nous ne faisons que suivre l'option prise par les ontologies courantes de quantités et d'unités de mesure définies depuis EngMath (Gruber and Olsen, 1994)³⁰ (ex : OM (Rijgersberg *et al.*, 2011), QTD (Hodgson and Keller, 2011)). Ces ontologies considèrent la catégorie *Scalar value* en lui associant une magnitude (un nombre) et une unité de mesure. Dans DataTop, ces associations sont prises en charge (cf. Figure 4) par les relations *magnitude* et *unitOfMeasure*. Pour la conceptualisation des nombres, nous proposons de réutiliser une classification consensuelle présente dans nombre d'ontologies³¹. Pour la conceptualisation des unités de mesure, nous choisissons de doter DataTop d'une ontologie minimale, que nous définissons ci-dessous.

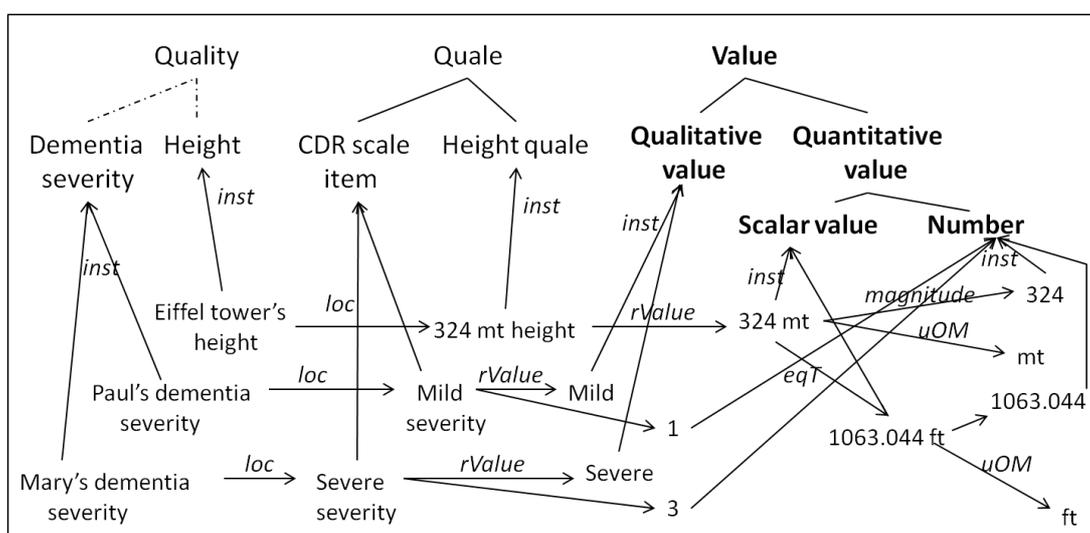


Figure 4. Modélisation des *Values*. La relation *rValue(x,y)* tient pour « *x* (une *Region*) a pour valeur *y* (une *Value*) » ; la relation *eqT(x,y)* tient pour « *x* (une *Value*) est équivalente à *y* (une autre *Value*) » ; la relation *magnitude(x,y)* tient pour « *x* (une *Scalar value*) a pour magnitude *y* (un *Number*) » ; la relation *uOM(x,y)* tient pour « *x* (une *Scalar value*) a pour unité de mesure – *unitOfMeasure* – *y* (une *Unit of measure*) ».

Pour définir cette ontologie minimale, il s'agit tout d'abord de préciser la nature d'une *unité de mesure*. Les définitions données dans la littérature sont largement consensuelles. Nous citerons, comme exemple, la définition du SI assimilant une unité de mesure à (ICGM, 2008, p. 7) : « une grandeur scalaire réelle, définie et adoptée par convention, à laquelle on peut comparer toute autre grandeur de même nature pour exprimer le rapport de deux grandeurs sous la forme d'un nombre ». Une unité de mesure est ainsi une grandeur, autrement dit (selon DOLCE-CORE) une *Quality*. Suivant cette définition, nous modélisons l'unité de mesure (*Unit of measure*) comme une *Quality* choisie conventionnellement et servant de référence pour exprimer la valeur de toute *Quality* de même nature (cf. Figure 5). Des *Units of measure* comme le mètre (*Meter*) ou le kilogramme (*Kilogram*) sont des qualités individuelles inhérentes à des entités (*Objects* et *Events*) de

³⁰ Rappelons toutefois que EngMath assimile *qualités individuelles* et *valeurs*.

³¹ Par exemple, dans l'ontologie SUMO (Suggested Upper Merged Ontology) : <http://www.ontologyportal.org/>.

référence³². Servant à comparer les magnitudes d'autres qualités individuelles du même type, les Units of Measure sont dans le même temps des qualités de type Length, Mass, etc., ce qui conduit à définir les catégories Unit of length, Unit of mass, etc., (cf. Figure 5). Le choix de ces catégories dépend bien sûr de la classification retenue pour les Qualities. Par ailleurs, une Unit of Measure est une grandeur scalaire, comme nous l'illustrons en Figure 5 avec le Gram. Ainsi, la Quality: Gram a pour valeur (qValue) la Scalar Value: 1 Gr ayant pour magnitude (magnitude) le Number: 1 et pour unit de mesure (uOM) elle-même.

Cette ontologie minimale peut être complétée par des relations entre Units of measure ou entre Qualities et Units of measure (telles qu'on peut les trouver dans des ontologies comme OM (Rijgersberg *et al.*, 2011) ou UO (Gkoutos *et al.*, 2012)) permettant d'exprimer que le Meter est l'unité de base dans le SI pour mesurer des Lengths ou bien que le Kilogram est une unité multiple du Gram.

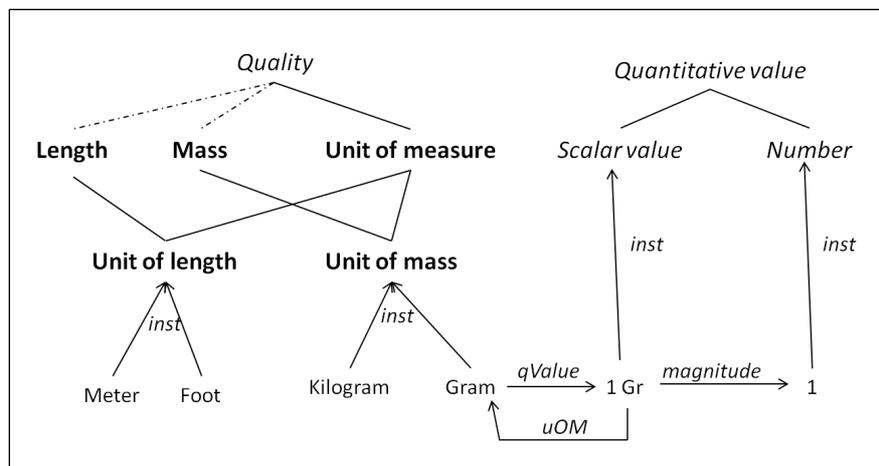


Figure 5. Modélisation des Units of Measure.

4.3 Conceptualisation de l'action d'observation

Comme nous l'avons vu en Section 2, les actions qualifiées dans un sens large d'« observation » sont diverses. Certaines consistent en une perception directe de l'entité observée, d'autres comportent l'utilisation d'instruments de mesure. Par ailleurs, certaines observations donnent lieu à l'interprétation de données, à l'agrégation de plusieurs données, et peuvent comporter des observations intermédiaires d'autres entités ou des observations de la même entité observée selon plusieurs dimensions. Quelle que soit la forme prise par ces actions, celles-ci créent des situations faisant intervenir diverses entités (observateur, donnée, instrument de mesure, etc.). Pour conceptualiser ces situations d'actions et mettre en relation les entités mises en jeu, nous dotons DataTop d'une ontologie minimale d'actions.

La catégorie centrale est ainsi celle d'Action. Conformément aux conceptions courantes en philosophie de l'action (ex: (Searle, 1983)), nous définissons une Action comme un processus intentionnel contrôlé par une entité capable d'intentions (il s'agit là de processus visant le monde dans lequel se réalise l'action). Ces Actions (voir Figure 6) s'opposent aux Happenings n'ayant pas de cause

³² Le kilogramme, par exemple, correspond à la masse d'un prototype constitué d'un alliage de 90% de platine et de 10% d'iridium, conservé au Bureau international des poids et mesures à Sèvres (France). Pratiquement, la conceptualisation de ces entités de référence n'est pas nécessaire, d'autant que d'autres modes de définition, ne reposant pas sur des artefacts, sont proposés pour la plupart des unités de mesure.

intentionnelle³³. En philosophie de l'action, plusieurs catégories d'intentions sont habituellement distinguées, conduisant à distinguer plusieurs catégories d'actions en fonction de la nature du contrôle exercé (Pacherie, 2008). C'est ainsi que sont distinguées les actions délibérées (initiées et contrôlées par une *intention préalable*³⁴) des automatismes, ou routines (contrôlées par une *intention en action*). Pour des raisons de place, dans ce document, nous ne détaillons pas la définition de ces différentes catégories d'actions. En Figure 6, nous nous contentons de faire apparaître l'Observation, comme spécialisation du concept Action et l'action Measurement comme une Observation spécifique. Selon les applications, d'autres Actions sont à définir.

Des entités participent temporairement aux Actions. Pour rendre compte de la manière dont ces entités participent aux Actions (ex : une entité joue le rôle d'agent, une autre le rôle d'instrument) nous introduisons autant de relations (appelées dans la littérature « relations casuelles » ou « rôles thématiques ») spécialisant la relation de participation de DOLCE (entre Objects et Events). Ces sous-relations (ex : hasForAgent, hasForInstrument, hasForResult) peuvent ensuite être utilisées dans une conceptualisation pour introduire des concepts dynamiques – Agent, Instrument, Result – et des spécialisations de ces concepts. Dans OntoNeuroLOG, nous avons ainsi modélisé le concept Score comme spécialisant le concept Result (Batrancourt *et al.*, 2010). À noter, en Figure 6, à propos de l'utilisation de la relation hasForTemporalLocation associant à l'action sa localisation temporelle, que cette dernière ne correspond pas forcément aux temps auxquels des positions de qualités sont associées à l'entité observée : un délai d'acquisition d'une mesure par un instrument peut conduire à un décalage entre l'instant d'une position d'une qualité et le moment de l'observation.

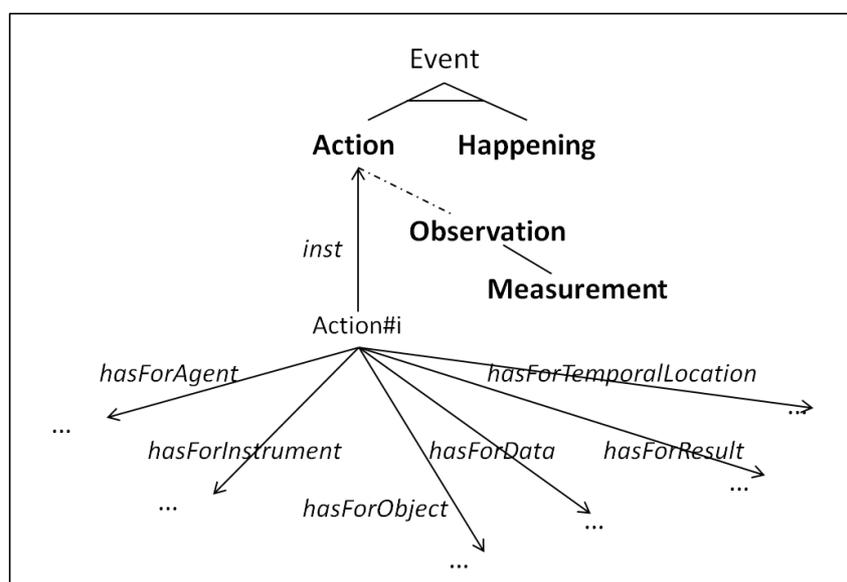


Figure 6. Modélisation d'un schéma d'Action : les « ... » tiennent pour des instances correspondant aux entités participant à l'action Action#i.

Nous venons de voir comment associer des entités mobilisées dans le cadre d'une Action à cette dernière en tenant compte de façons dont ces entités participent à l'Action. Ces relations conceptuelles ne disent toutefois pas grand-chose de

³³ Une extension de DOLCE analogue à celle que nous présentons a également été définie par (Trypuz, 2008).

³⁴ Nous reprenons ici la terminologie de (Searle, 1983).

l'essence des entités en question. Jouer le rôle d'Agent, par exemple, suppose de posséder une disposition (ou capacité) à contrôler intentionnellement des événements, cependant une telle disposition peut être possédée par des entités très différentes : un être humain, un animal, une organisation, un robot, un capteur sophistiqué³⁵. Dans le cas du rôle d'Instrument, les entités jouant ce rôle sont généralement des artefacts techniques ayant été intentionnellement produits pour remplir ce rôle, mais il peut s'agir aussi bien de capteurs physiques que de procédures de mesure conceptuelles. Par ailleurs, dans le cas d'une observation incluant la transmission langagière du résultat, l'entité jouant le rôle Result est une expression langagière, mais, comme il a été rappelé à plusieurs reprises, un contenu conceptuel (une Value) peut également jouer le rôle de Result. Ces remarques mettent en lumière la nécessité de disposer d'un cadre ontologique plus large permettant de rendre compte, notamment, de la nature des artefacts et d'entités « porteuses » d'informations. La conceptualisation de ces classes d'entités fait l'objet des sections suivantes.

4.4 Conceptualisation des entités jouant le rôle d'Instrument

Des entités peuvent occasionnellement être utilisées pour mesurer une qualité d'une autre entité et jouer ainsi le rôle d'Instrument, sans pour autant avoir été conçues à cet effet. Ainsi, au jeu de pétanque, une brindille d'herbe peut être utilisée (faute de mieux) pour estimer la distance de boules au but afin d'identifier la boule la plus proche du but. De même, un objet dont on connaît déjà la masse peut nous servir à estimer celle d'un autre objet. Toutefois, comme l'atteste le sens courant du terme « instrument de mesure », des entités existent qui sont intentionnellement créées pour jouer ce rôle d'aide à la mesure.

Pour rendre compte de l'intentionnalité de la production de ces entités et de la fonction qui leur est attribuée, nous avons étendu DOLCE au moyen d'une sous-ontologie formelle d'*artefacts* (Kassel, 2010). Cette ontologie introduit un ensemble minimal de concepts, dont les concepts *Artefact*, *Functional entity* et *Technical artefact*. Le concept *Artefact* (voir Figure 7) modélise la notion philosophique courante d'une « entité intentionnellement produite pour une certaine raison ». (Hilpinen, 2011). Le concept *Functional entity* désigne pour sa part toute entité à laquelle un agent attribue une *fonction*. Finalement, les *Technical artefacts* sont des entités combinant les deux propriétés, d'être un *Artefact* et une *Functional entity*³⁶. L'apport de ce module ontologique est d'offrir un ensemble de primitives conceptuelles pour rendre compte de l'origine de l'existence d'un *Technical artefact* et de sa fonction.

Comme le montre la Figure 7, les *Technical artefacts* peuvent être des *Physical objects*, comme dans le cas des capteurs (*Sensor*), ou des *Non-physical objects*, comme pour les *tests* et *questionnaires* utilisés en psychométrie et qui peuvent être assimilés à des contenus de documents (Batrancourt *et al.*, 2010). Dans tous les cas, ces *Artefacts* possèdent, comme

³⁵ Rappelons que nous considérons différents niveaux de contrôle.

³⁶ Cette conceptualisation suppose qu'il existe des *Artefacts* qui ne sont pas des *Technical Artefacts*. Nous classons dans cette catégorie, par exemple, des œuvres d'art intentionnellement produites pour transmettre une émotion mais non pour permettre à un agent de réaliser des actions, ce qui correspond à la notion de *fonction* que nous retenons (Kassel, 2010). De façon duale, nous considérons qu'il existe des *Functional entities* qui ne sont pas des *Artefacts*. Par exemple, un galet ramassé sur une plage peut faire office de presse-papier. À noter également qu'un *Artefact* peut se voir attribuer une fonction non conventionnelle, comme lorsqu'une agrafeuse est utilisée en tant que presse-papier.

tout objet, des qualités qui leur sont propres. Nous représentons en Figure 7 quelques qualités de capteurs (*Sensor quality*). La description des *Instruments* (plus exactement, des entités jouant le rôle d'*Instrument*) est un composant important de la sémantique des données. Connaître, par exemple, la résolution ou la plage de mesures d'un capteur ou bien le paramétrage d'un outil d'acquisition d'images, sont des informations précieuses pour l'interprétation des données.

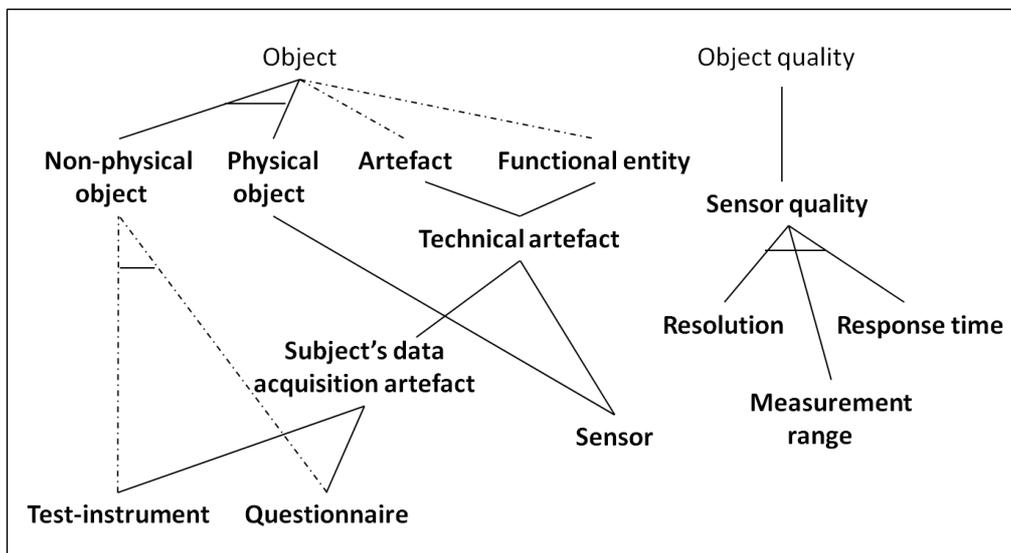


Figure 7. Modélisation d'artefacts physiques (ex : des capteurs) et non-physiques (ex : des tests et questionnaires utilisés pour acquérir des données d'un sujet).

4.5 Conceptualisation des données et d'entités « porteuses » d'informations

Dans cette section, nous complétons finalement notre cadre ontologique pour intégrer les données et rendre compte de leur contenu informationnel. En Section 2.4, nous avons identifié les *données* à des *enregistrements physiques de formes possédant un contenu informationnel*. La conceptualisation de tels contenus, de leur codification au moyen de langages en des expressions textuelles ou graphiques et de la matérialisation physique de ces expressions, correspond à la finalité du module ontologique I&DA (Fortier and Kassel, 2004). Ce module étend DOLCE en introduisant un ensemble de concepts relevant du domaine de la sémiotique, dont nous présentons informellement ici les principaux (cf. Figure 8) :

- Une *Conceptualization* est un contenu conceptuel. I&DA distingue deux catégories de *Conceptualizations*, à savoir les *Concepts* et les *Descriptions*. Ce dernier concept est à prendre dans le sens large d'un contenu utilisant des *Concepts* pour se référer à des objets d'un monde. Parmi les *Descriptions*, sont distinguées les *Propositions* – correspondant à une description factuelle à laquelle une valeur de vérité peut être affectée, et les *Instructional descriptions* – correspondant à des descriptions impératives (« faire ceci, plus faire cela ») ou conditionnelles (« si telle condition est satisfaite, faire ceci, sinon faire cela »).
- Une *Expression* est une expression ou codage d'une *Conceptualization* dans un langage de communication, comme pour l'expression linguistique « la température de la pièce est de 20°C » exprimant une *Proposition*. Le langage peut être textuel, graphique, gestuel. Il peut

être un langage de « bas niveau » non compréhensible (ou difficilement compréhensible) par un être humain, mais interprétable par une machine.

- Une `Inscription` est la matérialisation physique, au moyen de matière, de la forme d'une entité. La création d'une `Inscription` peut se faire directement à partir de l'entité, comme pour une trace de pas ou une empreinte digitale, ou bien indirectement en matérialisant une `Expression` se référant à une entité, comme pour une impression encrée d'une image ou d'un texte, ou comme une cellule de picots en écriture braille. Nos données scientifiques correspondent à cette dernière sous-classe d'`Inscriptions` que sont les `Inscriptions of expression`.
- Un `Physical support` est un objet physique servant de support à des `Inscriptions` qui y adhèrent, leur conférant ainsi une stabilité (même courte) dans le temps. Certains `Physical supports` jouent incidemment ce rôle (ex : une table sur laquelle est inscrite un graffiti, une pellicule de buée/poussière dans laquelle est un mot dessiné). D'autres, au contraire, les `Information medium`, sont intentionnellement produits pour réaliser cette fonction de support et sont donc d'artefacts techniques (ex : une feuille de papier, un tableau, un espace en mémoire interne d'un ordinateur).

Comme nous venons de le voir, les données scientifiques correspondent à des `Inscriptions of expression`. Elles participent du monde physique public en matérialisant des `Expressions` qui, par voie de conséquence, deviennent publiques et « porteuses d'informations » pour des agents. L'expression « porteur d'information » que nous venons d'utiliser repose sur une certaine conception de l'information qui mérite d'être précisée.

Le terme « porteur », tout d'abord, est entendu au sens où une donnée/expression est *source* d'informations pour des agents plutôt qu'elle ne *transporte* une information. En d'autres termes, il est considéré que l'information n'est pas contenue dans la donnée, qu'elle n'est pas un objet véhiculé par une donnée (comme le laisse entendre l'expression « flux d'information »), mais qu'elle résulte d'interprétations par des agents. Ces interprétations sont contextuelles et dépendent des intérêts, compétences et états de connaissances d'agents interpréteurs³⁷. Pour traduire le fait que des `Expressions` soient sources d'informations pour des agents et qu'elles informent de *quelque chose*, I&DA fait usage de la relation sémiotique générale `refersTo` (communément nommée « `isAbout` ») : une `Expression` se réfère à une entité (un `Particular` en général), pour un agent.

Qu'en est-il plus précisément de la notion d'*information* ? Le point de vue de la philosophie de l'information est que ce terme, largement polysémique (comme le terme « donnée ») et utilisé dans de nombreuses disciplines, désigne de fait plusieurs notions (Floridi, 2011a)(Krebs, 2011). Compte tenu de notre propos dans ce rapport, nous nous contentons d'évoquer une catégorie particulière d'information, à savoir l'information *sémantique* conceptuelle, telle que conceptualisée par (Floridi,

³⁷ Cette conception relationnelle plutôt qu'objective de l'information constitue aujourd'hui un pilier de la philosophie de l'information (Krebs, 2011, p ; 239) : « ... *In taking seriously the fact that the same data can become informative in different ways, depending on the interests and competences of a given interpreter, one can see how uninformative the metaphor of transportation is in the end.* »

2011b). Selon cette conception, les *Descriptions* de I&DA, qu'elles soient factuelles ou prescriptives, relèvent d'informations : elles résultent d'interprétations de données par des agents et possèdent pour ces agents une dimension aléthique dépendant de leur état de connaissances³⁸. Dans le module I&DA, deux relations portent sur les *Descriptions* : lors d'actes de discours, les *Descriptions* sont codées dans un langage en des *Expressions* (*isExpressedBy*) ; à l'instar des *Expressions*, les *Descriptions* se réfèrent à des entités (*refersTo*), ce qui revient à considérer qu'elles sont elles-mêmes porteuses d'informations.

En conclusion de cette section, résumons l'apport de DataTop pour rendre compte de la sémantique des données. D'une part, un lien (de la relation *isExpressedBy*) peut être établi entre une donnée – plus exactement l'*Expression* matérialisée par la donnée – et une *Description* pour rendre compte du contenu sémantique de la donnée. D'autre part, des liens (de la relation *refersTo*) peuvent être établis avec des entités auxquelles la donnée se réfère. Ces possibilités permettent de rendre compte du contenu informationnel de la donnée. Par ailleurs, des propriétés peuvent caractériser la donnée elle-même en tant qu'enregistrement physique, pour renseigner sur son contexte de production, sa localisation, son support, etc. Comme nous l'avons indiqué en Section 2.3, ces métadonnées viennent compléter la sémantique des données.

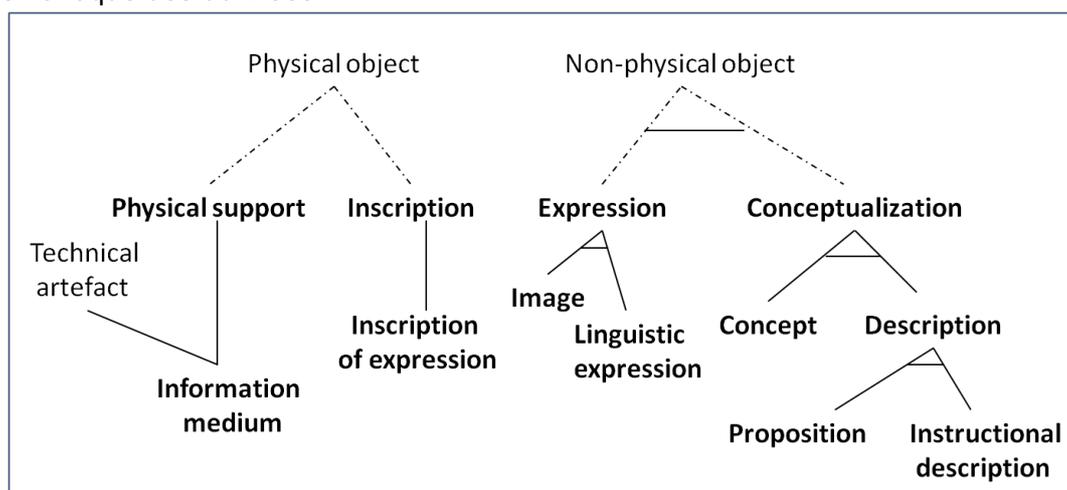


Figure 8. Principaux concepts de I&DA (Fortier and Kassel, 2004).

5 Brève discussion

Dans cette section, nous souhaitons discuter brièvement de l'état d'avancement de DataTop, selon deux dimensions : son contenu et son niveau de spécification.

Le domaine de DataTop a été défini de façon à couvrir l'ensemble des catégories utiles à la conceptualisation des données scientifiques, leur contenu et leur contexte de production. Parmi ces catégories, certaines nécessitent encore d'être

³⁸ La dimension aléthique est une dimension de classement des informations. Dans le cas d'une *Description* factuelle, il s'agit de la véridicité de la *Proposition*. Dans le cas d'une *Description* prescriptive, si celle-ci décrit par exemple un procédé pour réaliser une action, un agent va estimer si ce procédé lui paraît pertinent. En lien avec cette modalité, la question de savoir si une information factuelle possède nécessairement comme valeur de vérité la valeur *Vrai* pour un agent, continue de faire l'objet de débats (Scarantino and Piccinini, 2010).

approfondies, y compris au plus haut niveau (DOLCE-CORE) comme le souligne un article récent du LOA (Guarino, 2013). Par ailleurs, des extensions de DataTop s'avèrent nécessaires, tout en conservant son caractère générique, indépendant de tout domaine d'application.

Concernant le modèle de base de DOLCE-CORE, à savoir – des entités (*Objects* et *Events*) occupent temporairement des positions (*Regions*) selon des dimensions (*Qualities*) – des questions restent en suspend. Notamment, une classe importante d'entités mesurées sont les dispositions ou capacités humaines à réaliser des actions, comme la *capacité à mémoriser des mots* ou la *capacité à marcher sur une certaine distance sans assistance* (Bétrancourt *et al.*, 2010), et la question du classement de ces entités se pose : sont-elles des *Events* ou bien des *Qualities* ? Dans la littérature, des réponses divergentes ont été apportées (cf. (Kassel, 2010)(Borgo *et al.*, 2011)) et ces divergences s'expliquent tout autant par un manque de précision et un manque de consensus dans la définition des catégories *Object* et *Event*. Notons à ce propos que l'ontologie BFO (Smith, 2012) considère que les qualités peuvent être acquises et perdues, ouvrant la possibilité de classer les dispositions comme des qualités. À titre d'autre exemple de question, le mode d'attribution des *Qualities* à ses hôtes est-il unique, comme le laisse penser la relation d'*inhérence* de DOLCE-CORE ? Suivant Guarino (2013), des distinctions sont à considérer. Ainsi, par exemple, si on considère un vase conique, cela a un sens de lui attribuer *globalement* une hauteur tandis que la largeur varie suivant la partie du vase considéré. Guarino (2013) s'appuie sur cette différence d'attribution (globale vs locale) pour prôner l'introduction d'une nouvelle catégorie ontologique, le *champ de qualité*, en complément des *Qualities*.

Concernant les extensions apportées à DOLCE-CORE et présentées en Section 4, là également, des questions restent en suspend. À propos de la notion de *valeur* (*Value*), notamment de valeur qualitative, plusieurs questions se posent : quelles sont les valeurs à prendre en considération, par exemple « moyennement » ou « moyennement profond » ? Par ailleurs, ces valeurs sont-elles indépendantes des *Qualities* mesurées (ex : une armoire ou un cours d'eau) ? Enfin, quelles sont les *Regions* auxquelles des valeurs sont associées et, notamment, toutes les *Qualia* ont-elles une valeur ? Côté actions, nous disposons d'un schéma général de modélisation comportant des relations thématiques rendant compte de rôles joués temporairement par des entités vis-à-vis de l'action et nous nous sommes contentés de positionner les actions d'observation et de mesure. Nous estimons toutefois qu'une analyse plus fine des actions en jeu, notamment pour distinguer l'*observation* de l'*interprétation* et introduire l'*acquisition de donnée*, reste à conduire. Enfin, côté données de l'observation, le module I&DA présenté en section 4.5 positionne ces entités dans un système sémiotique permettant de rendre compte de leur format et de contenus qui leur sont attribués par des agents. Par contre, nous ne disposons pas de moyens conceptuels pour décrire leurs conditions d'usage (confidentialité, accessibilité) et leur cycle de vie (ex : le fait que des données soient extraites ou dérivées d'autres données). Par ailleurs, toujours concernant les données, le cadre ontologique doit être étendu pour permettre de prendre en compte des collections spatiales et temporelles de données.

Venons-en à évoquer l'état de formalisation de DataTop. L'ontologie OntoNeuroLOG, déjà citée, exploite une version partielle de DataTop basée sur DOLCE, spécifiée en OntoSpec et codée dans un langage d'ontologie pour le Web (un dialecte de OWL). Le cadre ontologique présenté dans ce document, par contre, est fondé sur un noyau révisé de DOLCE, DOLCE-CORE, pour lequel nous ne disposons pas d'axiomatique complète (une partie seulement de cette axiomatique a été publiée dans (Borgo and Masolo, 2009)). Nous sommes de fait en attente de

disposer de cette axiomatique complète pour formaliser à nouveau l'ensemble de DataTop présenté de façon discursive dans ce document.

6 Conclusions et perspectives

Dans ce document, nous avons proposé un cadre ontologique générique (indépendant de tout domaine scientifique) permettant de rendre compte au « niveau connaissance » à la fois de données de l'observation générées et exploitées dans un contexte d'études scientifiques et de leur sémantique. Quels que soient le mode d'enregistrement des données, leur format ou leur structuration, ce cadre fournit un ensemble d'entités conceptuelles de base permettant de représenter des connaissances portant sur ces données, leur contenu informationnel, leur contexte de production et les conditions de leur utilisation.

DataTop se prête ainsi à associer des métadonnées sémantiques à des données. Dans le contexte des Bases de Données une telle association est motivée par le fait de plonger le modèle d'un domaine, que constituent le schéma et les enregistrements d'une Base de Données, dans un modèle plus riche sémantiquement, prenant en compte davantage de connaissances du domaine concerné (Borgida & Mylopoulos, 2005). Pratiquement, une telle association suppose de peupler l'ontologie avec des instances. Mais, avant cela, l'application portant sur un domaine scientifique particulier, l'ontologie nécessite d'être complétée pour intégrer les concepts spécifiques de ce domaine.

Pour en revenir à DataTop lui-même, le travail entrepris dans ce document est loin d'être finalisé. Comme nous l'avons vu dans la section précédente, ce cadre ontologique nécessite d'être approfondi et complété. À cette fin, trois perspectives de travaux sont dégagées. L'intention des auteurs du présent document est de le diffuser largement auprès de chercheurs travaillant dans le domaine de l'ontologie formelle (si possible, l'ontologie formelle de l'observation) et auprès de scientifiques concernés par la sémantique des données, de sorte à évaluer notre cadre ontologique. Sa mise en œuvre est également prévue dans différents contextes applicatifs de sorte également à l'évaluer et à le consolider. Par ailleurs, le cadre DataTop sera aligné avec d'autres ontologies de référence, de haut niveau ou dans des domaines scientifiques particuliers. Compte tenu du domaine visé par le projet CrEDIBLE, il est prévu à court terme un alignement avec l'ontologie fondatrice BFO utilisée dans le domaine biomédical, dans sa nouvelle version 2.0 (Smith, 2012).

7 Références

- B. Batrancourt, M. Dojat, B. Gibaud, and G. Kassel. A core ontology of instruments used for neurological, behavioral and cognitive assessments. In A. Galton and R. Mizoguchi (eds.), *Proc. of the Sixth International Conference on Formal Ontology and Information Systems (FOIS 2010)*, pages 185-198, IOS Press, 2010.
- J. Bogen. Theory and Observation in Science. In E.N. Zalta (ed.), the *Stanford Encyclopedia of Philosophy (Spring Edition)*, 2010. URL = <http://plato.stanford.edu/archives/spr2010/entries/science-theory-observation/>
- J. Bogen, and J. Woodward. Saving the phenomena. *The Philosophical Review*, 97(3):303-352, 1988.
- A. Borgida, and J. Mylopoulos. Data Semantics Revisited. In *Proc. of the Second International Conference on Semantic Web and Databases*, LNCS 3372, pages 9-26, Springer-Verlag Berlin, Heidelberg, 2005.
- S. Borgo, and C. Masolo. Foundational choices in DOLCE. In S. Staab, and R. Studer (eds.), *Handbook on Ontologies (Second Edition)*, pages 361-382, Springer Verlag, 2009.

- S. Borgo, R. Mizoguchi, and B. Smith. On the ontology of functions. *Applied Ontology*, 6(2):99-104, 2011.
- E. Bottazzi, C. Catenacci, A. Gangemi, and J. Lehman. From collective intentionality to intentional collectives: An ontological perspective. *Cognitive Systems Research*, 7(2-3):192-208, 2006.
- E. Bottazzi, R. Ferrario, and C. Masolo. The mysterious appearance of objects. In M. Donnelly, and G. Guizzardi (eds.), Proc. of the *Seventh International Conference on Formal Ontology in Information Systems (FOIS 2012)*, pages 59-72, IOS Press, 2012.
- R. Carnap, and Y. Bar-Hillel. *An Outline of a Theory of Semantic Information*. MIT (Research Laboratory of Electronics) Technical report n°247, October 1952. Accessible à : <http://dspace.mit.edu/bitstream/handle/1721.1/4821/RLE-TR-247-03150899.pdf>.
- L. Floridi. *The Philosophy of Information*. Oxford, Oxford University Press, 2011(a).
- L. Floridi. Semantic Conception of Information. In E.N. Zalta (ed.), the *Stanford Encyclopedia of Philosophy (Spring Edition)*, 2011(b). URL = <http://plato.stanford.edu/archives/spr2011/entries/information-semantic/>.
- J.Y. Fortier, and G. Kassel. Managing knowledge at the information level: an ontological approach, In Proc. of the *ECAI'2004 workshop on knowledge management and organizational memories*, pages 39-45, Valencia (Spain), 2004.
- A. Galton, and R. Mizoguchi. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Applied Ontology*, 4(2):71-107, 2009.
- P. Gärdenfors. *Conceptual Spaces: the Geometry of Thought*. MIT Press, Cambridge, Massachusetts, 2000.
- P. Gärdenfors. Conceptual Spaces as a Framework for Knowledge Representation. *Mind and Matter*, 2(2):9-27, 2004.
- C. Germain-Renaud, F. Fürst, T. Jacob, M. Jouvin, G. Kassel, J. Nauroy, and G. Philippon. The Green Computing Observatory: a data curation approach for green IT. In Proc. of the *EGI Community Forum 2012 / EMI Second Technical Conference*, Munich (Germany), 2012.
- J.J. Gibson. *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1979.
- G.V. Gkoutos, E.C.J. Green, A.-M. Mallon, J.M. Hancock, and D. Davidson. Using ontologies to describe mouse phenotypes. *Genome Biology*, 6:R8, 2004.
- G.V. Gkoutos, P.N. Schofield, and Robert Hoehndorf. The Units Ontology: a tool for integrating units of measurement in science. *Database*, 2012.
- P. Grenon, and B. Smith. SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition and Computation*, 4(1):69-104, 2004.
- T.R. Gruber, and G.R. Olsen. An Ontology for Engineering Mathematics. In J. Doyle, P. Torasso, and E. Sandewall (eds.), proc. of the *Fourth International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, 1994.
- N. Guarino. Local Qualities, Quality Fields, and Quality Patterns: A Preliminary Investigation. In O. Kutz, M. Bhatt, S. Borgo, and P. Santos (eds.), Proc. of the *Second Interdisciplinary Workshop on SHAPES*, Rio de Janeiro (Brazil), March 3-4, 2013. Available online at ceur-ws.org.
- P.M.S. Hacker. Events and objects in space and time. *Mind*, 91:1-19, 1982.

- R. Hilpinen. Artifact. In E.N. Zalta (ed.), the *Stanford Encyclopedia of Philosophy (Winter Edition)*, 2011. URL = <http://plato.stanford.edu/archives/win2011/entries/artifact/>.
- R. Hodgson, and P.J. Keller. QUDT – Quantities, Units, Dimensions and Data Types in OWL and XML. QUTD v.1.1, URL = <http://www.qutd.org/>.
- JCGM (Joint Committee for Guides in Metrology). International vocabulary in metrology – Basic and general concepts and associated terms (VIM). 3rd edition, 2008.
- G. Kassel. A formal ontology of artefacts. *Applied Ontology*, 5(3-4):223-246, 2010.
- G. Kassel, M. Turki, I. Saad, and F. Gargouri. From collective actions to actions of organizations: an ontological analysis. In Proc. of the *Symposium Understanding and Modelling Collective Phenomena (UMoCoP 2012)*, Birmingham (England), July 2012.
- J. Krebs. Philosophy of Information and Pragmatic Understanding of Information. *Ethics & Politics*, 13(2):235-245, 2011.
- W. Kuhn. A Functional Ontology of Observation and Measurement. In K. Janowicz, M. Raubal, and S. Levashkin (eds.), Proc. of the *Third International Conference on GeoSpatial Semantics (GeoS 2009)*, pages 26-43, Springer Verlag, 2009.
- J. Kurtzke. Rating neurological impairment in multiple sclerosis: an expanded disability scale. *Neurology*, 33:1444-52, 1983.
- C. Masolo. Founding properties on measurement. In A. Galton and R. Mizoguchi (eds.), Proc. of the *Sixth International Conference on Formal Ontology and Information Systems (FOIS 2010)*, pages 89-102, IOS Press, 2010(a).
- C. Masolo. Objects, Events, Qualities: An introduction to formal ontological distinctions in DOLCE. Advanced course at ESSLI, Copenhagen, 2010(b). (Slides available at: <http://www.loa.istc.cnr.it/Teaching.html>).
- C. Masolo, and S. Borgo. Qualities in Formal Ontology. In P. Hitzler, C. Lutz, and G. Stumme (eds.), Proc. of the *Workshop on Foundational Aspects of Ontologies (FOnt 2005)*, Koblenz (Germany), pages 2-16, 2005.
- C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology. WonderWeb Deliverable D18, Final Report, vr. 1.0, 2003.
- C. Masolo, L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi, and N. Guarino. Social roles and their descriptions. In D. Dubois, C. Welty and M.-A. Williams (eds.), Proc. of the *Ninth International Conference on the Principles of Knowledge Representation and Reasoning*, pages 267-277, 2004.
- J. Morris. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43:2412-2414, 1983.
- NIST (National Institute of Standards and Technology). The International System of Units (SI). B.N. Taylor and A. Thomson (eds.), NIST Special Publication 330, 2008.
- E. Pacherie. The phenomenology of action: A conceptual framework. *Cognition*, 107:179-217, 2008.
- F. Probst. Observations, measurements and semantic reference spaces. *Applied Ontology*, 3(1-2):63-89, 2008.
- Z.W. Pylyshyn. *Things and Places. How the Mind Connects with the World*. MIT Press, 2007.
- H. Rijgersberg, M. van Assem, and J. Top. Ontology of Units of Measure and Related Concepts. *Semantic Web*, 4(1):3-13, 2011.
- A. Scarantino, and G. Piccinini. Information without truth. *Metaphilosophy*, 41(3):313-330, 2010.

- H.R. Schiffman. *Sensation and Perception*. Wiley, New York, 1982.
- J.R. Searle. *Intentionality*. Cambridge: Cambridge University Press, 1983.
- A. Sheth. Data Semantics: *what, where and how?* In R. Meersman and L. Mark (eds.), *Data Semantics*, IFIP Transactions, pages 601-610, Chapman and Hall, London, 1996.
- B. Smith. Fiat Objects. *Topoi*, 20(2):131-148, 2001.
- B. Smith. On Classifying Material Entities in Basic Formal Ontology. In Proc. of the *Third Interdisciplinary Ontology Meeting*, Tokyo: Keio University Press, pages 1-13, 2012.
- S.S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677-680, 1946.
- L. Temal, M. Dojat, G. Kassel, and B. Gibaud. Towards an ontology for sharing medical images and regions of interest in neuroimaging. *Journal of Biomedical Informatics*, 41:766-78, 2008.
- R. Trypuz. *Formal Ontology of Action: A Unifying Approach*. Wydawnictwo Kul, 2008.
- B.C. van Fraassen. *Scientific Representation: Paradoxes of Perspective*. Clarendon Press, 2008.
- Z. Wood, and A. Galton. A taxonomy of collective phenomena, *Applied Ontology*, (3-4):267-292, 2009.
- J.F. Woodward. Data and phenomena: a restatement and defense. *Synthese*, 182:165-179, 2011.