
Pour une gestion des données de recherche en environnement : l'ontologie méta-observatoire

**Marie-Laure BETBEDER¹, Sylvie DAMY², Bénédicte
HERRMANN¹**

{marie-laure.betbeder, sylvie.damy, benedicte.herrmann}@univ-fcomte.fr

1 Institut FEMTO-ST/DISC UMR 6174 Université Bourgogne Franche-Comté

2 Laboratoire Chrono-Environnement UMR 6249 Université Bourgogne Franche-Comté

RESUME. L'objectif global du projet décrit dans cet article, est de concevoir un environnement et des outils pour aider les chercheurs ou acteurs du domaine environnemental à construire des systèmes d'information adaptés et interopérables afin de gérer, partager, publier et valoriser leurs données. Nous présentons ici la première phase de ce projet à savoir l'étude des propositions existantes et l'ontologie proposée qui permet de décrire les concepts de base de ces systèmes d'information.

ABSTRACT. The main goal of the project described in this paper is to design an environment and tools to help researchers or environmental actors to build adapted and interoperable information systems. These information systems must be adapted and interoperable in order to manage, share, publish and value their managed data. We present here the first step of this project, the study of the existing proposals and the proposed ontology which allows to describe the basic concepts of these information systems.

MOTS-CLES : Méta-modèle, Gestion et valorisation des données de recherche, Ontologie, Observation, Observatoire, Environnement.

KEYWORDS: Meta-model, Management and valorisation of research data, Ontology, Observation, Observatory, Environment

1. Introduction

L'ouverture des données de la recherche scientifique est devenue un enjeu majeur au fil de ces dernières années avec en particulier les directives européennes, les contraintes du programme européen H2020 (H2020, 2017a; 2017b; 2016) et la récente loi pour une république numérique. Il est donc primordial de développer des systèmes d'information pour permettre un accès et une valorisation des données de recherche. De plus, pour que ces données soient pérennes, réutilisables et partageables, il est nécessaire que ces données soient compréhensibles, interprétables et interopérables (King, 2007) (Reffay *et al.*, 2012).

Nous présentons dans cet article l'ontologie "*méta-observatoire*", résultat de la première phase de notre projet qui a pour objectif d'aider les chercheurs du laboratoire Chrono-environnement de l'université de Bourgogne Franche-Comté à gérer leurs données de recherche. Ce travail repose sur une expérience d'aide à la conception de modèles de données issues de leurs projets de recherche.

Dans la deuxième partie de l'article nous présentons l'expérience de création de bases de données au laboratoire puis, le constat que nous avons pu en faire et notre proposition de mise en œuvre d'un environnement d'aide à la conception de bases de données dans le domaine environnemental (partie 3). Après une présentation des ontologies existantes du domaine (partie 4), nous détaillons l'ontologie « méta-observatoire » (partie 5).

2. Contexte : Gestion des données de recherche dans un laboratoire

Certaines disciplines scientifiques telle que l'astronomie ont déjà une pratique de l'ouverture des données avec l'observatoire virtuel (Observatoire virtuel, 2017) qui propose un ensemble de protocoles et d'outils permettant de partager des données complexes et de les valoriser à travers des portails reconnus par la communauté. Dans le domaine environnemental le projet DataOne (DataOne, 2017), (Tenopir C., 2015) propose l'accès à des données via des entrepôts de données généralistes ou spécialisés et surtout un ensemble de guides de bonnes pratiques pour la gestion des données de recherche tout au long de leur cycle de vie.

La mise en œuvre d'une gestion des données de la recherche au niveau d'un laboratoire de recherche, qui ne dispose pas forcément des personnes ressources adéquates, s'avère difficile. Au laboratoire Chrono-environnement de l'Université de Bourgogne Franche-Comté, l'axe transversal Bases de Données, créé en 2012, a pour objectif d'aider les chercheurs à gérer et valoriser les données issues de leurs travaux de recherche. Ce laboratoire pluridisciplinaire regroupe des disciplines très variées allant de la géologie à l'archéologie en passant entre autres par la médecine et l'écologie. Ces données de recherche peuvent être très différentes. Certaines données proviennent de capteurs, d'autres d'observations sur le terrain, ou encore de traitements plus ou moins complexes. Les formats des fichiers servant à stocker les données sont hétérogènes (feuille de calcul, fichier pdf, image, ...). Les données

peuvent être accessibles soit par un groupe restreint de chercheurs (pouvant être de disciplines différentes) soit par une communauté scientifique au niveau national ou international.

L'axe Bases de Données soutient les chercheurs dans la gestion de leurs données à travers la création de base de données, la sauvegarde et le partage des données.

Les bases de données sont un outil essentiel pour structurer et organiser les données. Leur création passe par trois étapes essentielles : la modélisation des données, leur intégration dans la base de données, et le développement des interfaces de manipulation. Le travail réalisé par l'axe transversal s'intéresse principalement à la modélisation des bases de données (Charbonnier *et al.*, 2014). Actuellement plus d'une dizaine de bases de données ont été créées. Citons par exemple les bases de données Paleofire (PaleoFire, 2016) et EWET (EWET, 2017).

3. Problématique et Proposition

La conception des bases de données au laboratoire Chrono-environnement a fait ressortir a priori une très grande hétérogénéité des données que ce soit au niveau de leur format, de leur origine ou de leur nature. Cependant nous avons mis en évidence des concepts communs apparaissant dans la majorité de ces bases de données et surtout le concept d'observation. Dans les bases de données créées on repère notamment des éléments communs qui décrivent le contexte des observations en répondant aux questions : Qui a fait l'observation ? Quelle observation ? Quand ? Où ? Comment ? Pourquoi ? Dans le cadre de quel projet ? ...

Actuellement, la conception de chaque base de données suppose la redéfinition de ces concepts.

L'objectif de notre projet est de proposer un environnement d'aide à la conception de bases de données décrivant des observatoires environnementaux. Nous définissons un observatoire environnemental comme un système d'information créé dans le cadre d'un projet de recherche, pour regrouper, stocker et gérer l'ensemble des observations permettant d'étudier un phénomène environnemental. Notre environnement d'aide doit rationaliser et faciliter la conception des bases de données mais aussi permettre d'assurer la qualité, le partage et la réutilisation des données. Pour cela nous nous proposons d'intervenir dès la conception des modèles de données grâce à l'utilisation d'un méta-modèle. Lors de la création d'un observatoire, ce méta-modèle sera spécialisé, adapté pour définir un modèle adéquat pour le projet de recherche et enfin transformé en un modèle relationnel. Ainsi les bases de données développées indépendamment les unes des autres mais issues du même modèle reposeront sur les mêmes concepts et pourront ainsi être interoperables. Les données pourront être exportées, interprétées, interrogées et échangées dans et entre les différentes bases de données.

La première étape de ce projet, présentée dans cet article, consiste en la définition d'un méta-modèle générique, appelé "méta-observatoire" qui définit les concepts de base d'un observatoire environnemental.

Pour la formalisation du méta-observatoire nous avons choisi l'approche des ontologies. Une ontologie, « est une spécification explicite d'une conceptualisation » (Gruber, 1993) qui permet de représenter un ensemble d'objets (dans notre cas les concepts de base de l'observatoire) en y ajoutant la dimension sémantique. Le choix de ce formalisme permet de réutiliser des ontologies existantes, concept inhérent aux ontologies et de bénéficier de l'interopérabilité apportée par les ontologies. Le rôle clé de notre ontologie est de proposer un niveau d'abstraction commun au-dessus des conceptions (logiques ou physiques) spécifiques des bases de données et systèmes informatiques dans le domaine environnemental. L'idée est de caractériser chaque « observatoire » par une ontologie spécialisée puis de la transformer en un modèle relationnel.

Ce méta-modèle doit être suffisamment générique pour prendre en compte, par exemple tous les différents types d'observations. Lors de la modélisation d'un observatoire, les chercheurs seront accompagnés dans la démarche de description de leurs méthodes, outils, observations dans le but de sélectionner et/ou spécialiser les éléments du méta-modèle cohérents avec leurs données et contextes. Le méta-modèle est ainsi considéré comme un noyau qui peut être restreint, spécialisé ou étendu (si un élément n'a pas été prévu dans le méta-modèle).

4. Modèles d'observation environnementale

Notre première étape est l'étude des ontologies existantes dans le même domaine. Nous présentons ci-après trois modèles décrivant l'observation dans le domaine environnemental.

4.1 SERONTO

SERONTO (Socio-Ecological Research and Observation oNTology) (Werf *et al.*, 2008) est une ontologie créée en 2009 par le réseau européen ALTER-Net. Le but de cette ontologie est de gérer et d'intégrer les données socio-écologiques de recherche liées à la biodiversité à long terme. SERONTO est utilisée dans divers domaines de l'écologie et des sciences sociales pour décrire des données de surveillance et des données expérimentales. Cette ontologie d'observation générique propose les concepts de base pour représenter des observations, des mesures et des stratégies d'échantillonnage des données.

Elle peut être étendue pour créer des ontologies de domaine par extension des concepts de base pour les adapter à des besoins spécifiques. SERONTO peut être utilisée avec la plupart des standards de données écologiques.

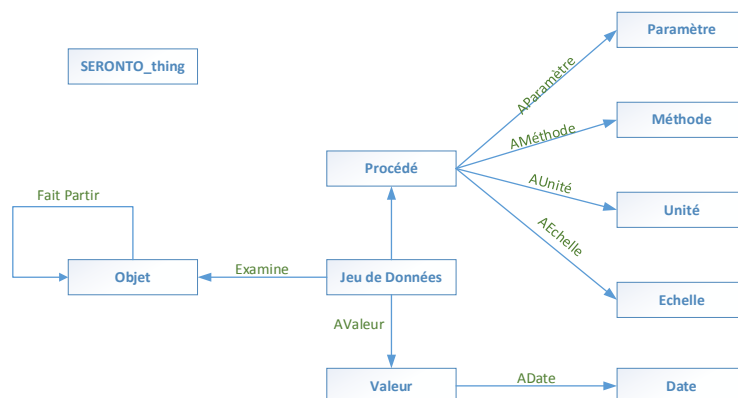


Figure 1 : Concepts de base de SERONTO

Dans SERONTO (cf. figure 1), une observation est réalisée sur un *objet*, elle est caractérisée par une *Valeur* collectée à une *date* donnée suivant un *Procédé*. Ce *Procédé* est décrit au moyen de *Paramètre*, d'*Unité*, d'*Echelle* et de *Méthode*. Les informations temporelles sont traitées de façon explicite, ainsi une *date* est associée à chaque observation. Le concept de base de SERONTO est le *Jeu de Données* qui regroupe plusieurs observations : différentes observations peuvent être réalisées sur un même *Objet*, ou un même *Procédé* d'observation peut être utilisé sur différents *Objets*.

La classe de base *SERONTO_thing* permet de spécifier des propriétés administratives, comme par exemple *CrééPar*, *EntréPar*, *AUnProjet*. Ces propriétés sont héritées par toutes les classes de l'ontologie. SERONTO est développée avec le langage OWL-DL.

Dans SERONTO, certains éléments de contexte de l'observation, telle que la date, sont introduits. L'ontologie propose la possibilité de regrouper des observations. Par contre, la notion d'observation n'est pas vraiment explicite. A notre sens, elle est un peu différente de celle de jeu de données.

4.2 ODM

ODM (Observations Data Model) a été développé en 2008 par le groupe CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science), consortium de laboratoires de recherche (Horsburgh *et al.*, 2008). ODM est conçu pour représenter les données des observations, systèmes d'observation et leurs métadonnées. Les domaines d'application visés sont des systèmes d'observations hydrauliques, mais ODM peut être utilisé pour d'autres domaines. Les concepts et leurs relations sont présentés dans la figure 2.

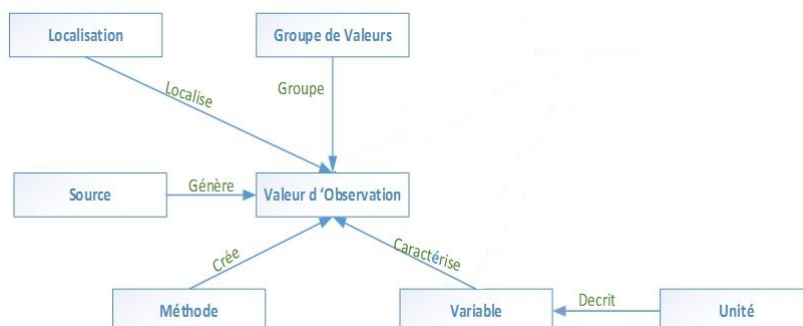


Figure 2 : Concepts de base d'ODM

Dans ODM, l'élément central est la *valeur de l'observation* qui est relié aux informations considérées comme des métadonnées. La valeur de l'observation est ainsi reliée aux caractéristiques fondamentales telles que le lieu où les observations ont été faites (*localisation*), la date et l'heure à laquelle les observations ont été faites et le type de variable qui a été observée (par exemple le débit - *variable*). D'autres données nécessaires à l'interprétation sont reliées à l'élément central comme les méthodes utilisées pour faire les observations (*méthode*), des commentaires qualificatifs sur l'observation (*données qualifiées*) et les informations sur l'organisation qui a fait l'observation (*source*). Il est possible de regrouper des *valeurs d'observation* (*groupe de valeurs*).

Les concepts utilisés dans ODM sont peu extensibles et se limitent à des listes de termes que l'on peut ajouter aux différentes classes. ODM ne gèrent pas très bien les observations complexes. Cependant, ODM propose un élément de description du contexte de l'observation en introduisant la notion de localisation.

4.3 OBOE

OBOE (Extensible Observation Ontology) (Madina *et al.*, 2007) est développée et maintenue dans le cadre du projet SEEK (Science Environment for Ecological Knowledge). Cette ontologie sert de base pour décrire et représenter des données d'observation dans les domaines écologiques et environnementaux. Elle propose pour cela un noyau développé en OWL avec des classes de base qui peuvent être étendues pour définir une ontologie spécifique à un projet d'observation.

4.3.1 Présentation

Pour OBOE, une observation (cf. Figure 3) exprime des mesures de caractéristiques sur une entité. Une entité représente les objets observables de façon

générique. Ces objets peuvent ainsi être spécialisés pour un projet d'observation spécifique.

La partie la plus développée de cette ontologie est la mesure qui décrit le résultat de l'observation : la valeur, mais aussi les éléments qui précisent comment cette valeur a été obtenue au moyen du protocole et le standard de mesure qui définit l'unité de mesure ou la classification utilisée (classification taxonomique par exemple). De plus, une précision peut être associée à la valeur obtenue.

Cette ontologie très conceptuelle et bien documentée offre un modèle extensible. La majorité de ses classes peuvent être étendues. Mais surtout, OBOE propose une très bonne description de l'observation.

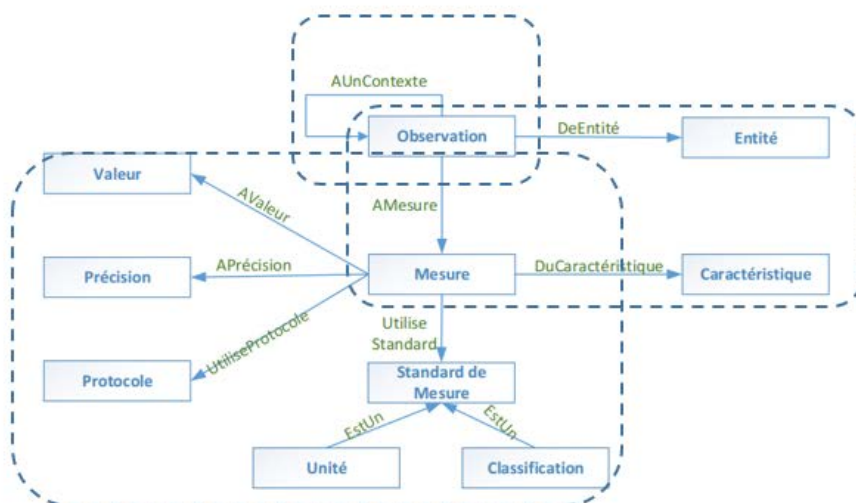


Figure 3 : Concepts d'observation de OBOE

4.3.2 Instanciation d'un jeu de données dans OBOE

La figure 4 présente l'instanciation dans OBOE d'un ensemble concret d'observations, issu d'un projet de recherche, qui a été simplifié. Elle décrit une entité de type micromammifère qui a comme taxon "arvicola" et qui a été capturée à la date du 10 mars 2013 sur la parcelle 3 par M. Dupont dans le cadre du projet SCAPE. On mesure son poids et on effectue un prélèvement sanguin pour réaliser des analyses permettant de doser différentes molécules (epoxiconazole et fenpropidine).

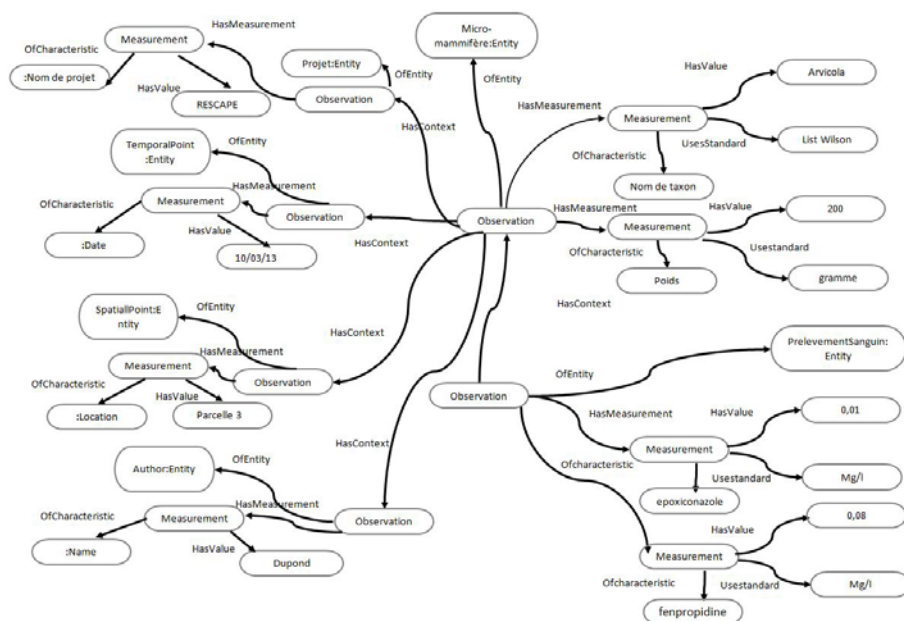


Figure 4 : Exemple d'instanciation dans OBOE

Dans la figure 4 les informations temporelles et spatiales liées à l'entité observée (date et lieu) sont au même niveau que les informations concernant le prélèvement sanguin et les dosages de molécules qui ici représentent le cœur de l'information. Toutes ces différentes informations sont au même niveau.

Dans OBOE « tout est observation ». Cela permet de tout décrire mais la représentation est peu lisible et ne permet pas de hiérarchiser l'information. La généralité d'OBOE est à la fois une force, tout peut être décrit, et une faiblesse : (1) Il est assez difficile de représenter ses données, (2) le fait de tout représenter en observation implique une lourdeur dans la tâche (nous n'avons représenté qu'un exemple simplifié!) (3) tous les éléments de contexte d'une observation sont présentés comme une nouvelle observation au moyen de la relation *HasContext*.

5. L'ontologie « méta-observatoire »

L'expérience de création de bases de données, l'étude des différentes ontologies existantes et l'exercice d'instanciation d'un jeu de données dans OBOE nous ont permis de définir des éléments importants pour la conceptualisation de notre ontologie :

Il faut que le contexte d'une observation soit défini clairement et toujours de manière identique afin de pouvoir exprimer pour une observation un Qui (acteur),

Quand (Date), Où (Lieu). Le contexte d'une observation est un élément indépendant de l'observation en elle-même que l'on observe le taxon d'un micromammifère capturé sur une parcelle, le nombre d'occurrence d'un taxon palynologique dans un prélèvement de carottage ou la taille d'un quercus. Pour pouvoir pérenniser ces données, mais surtout s'assurer de leur qualité, il est important de garder une certaine traçabilité de l'observatoire et donc de savoir : qui la faite, quand, où et comment.

La notion d'observatoire doit être introduite en définissant le cadre général pour permettre un meilleur référencement et une meilleure visibilité. On peut ainsi préciser que l'observatoire regroupe un ensemble de données portant sur l'étude d'un site archéologique, dirigé par Mr Dupont et que cet observatoire a été créé en 2016.

Il faut définir un ensemble de concepts obligatoires pour permettre de disposer d'informations fiables et pérennes et réutilisables. Un jeu de données de type observatoire ne peut pas être considéré comme satisfaisant si certaines informations ne sont pas présentes. Ainsi une observation dont on ne connaît pas les conditions de création ne pourra pas être reproduite. Nous proposons donc de définir dans notre ontologie un ensemble de concepts obligatoires. Tout observatoire construit à partir de « méta-Observatoire » devra obligatoirement contenir une représentation de ces concepts.

Il est nécessaire de spécifier deux niveaux d'observation, l'unité d'étude et le prélèvement. Un prélèvement provient d'une étude d'étude. Notre expérience dans la conception des bases de données au laboratoire nous a amené à proposer une distinction dans les entités observées. Nous définissons 2 niveaux (1) l'unité d'étude qui correspond à une entité observée ou relevée sur le terrain. Il s'agit par exemple d'un campagnol piégé sur une parcelle, (2) le prélèvement ; cette entité est issue d'une unité d'étude. Il s'agit par exemple d'un prélèvement sanguin sur le campagnol, de poils récupérés sur le campagnol.

Nous présentons dans cette section une première version du modèle conceptuel de l'ontologie « méta-observatoire » basé sur le noyau de l'ontologie OBOE et étendu par des concepts issus des constats explicités précédemment. Ce modèle conceptuel (cf. figure 5) représente les classes de base de l'ontologie et leurs relations.

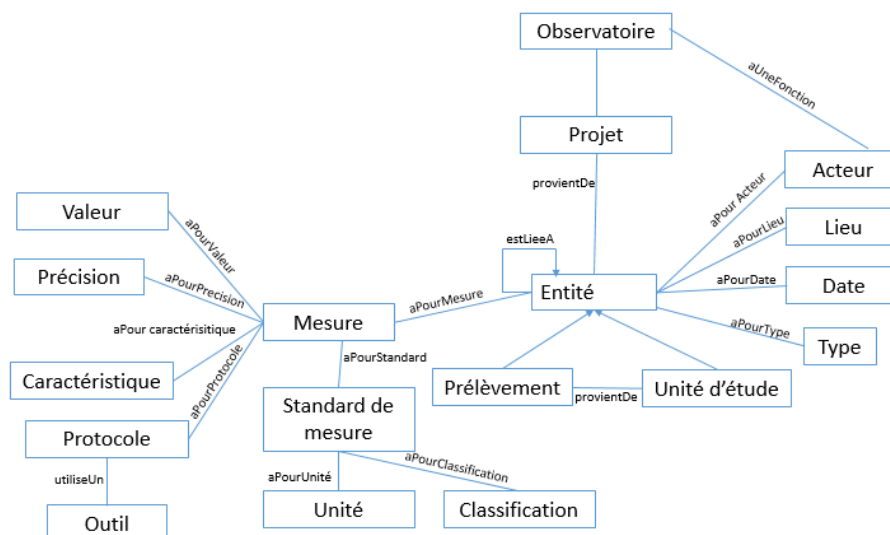


Figure 5 : l'ontologie Méta-Observatoire

5.1 Classes issues d'OBOE

Les classes de l'ontologie méta-observatoire issues d'OBOE constituent la partie de l'ontologie qui modélise les mesures. Les observations sont les données de base d'un observatoire. Une observation est la mesure d'une caractéristique sur une entité cible de l'observation. Les classes de base d'OBOE réutilisées dans l'ontologie sont *Mesure*, *Caractéristique*, *Standard de Mesure*, *Unité*, *Classification* (qui permettra d'intégrer les référentiels), *Précision*, *Valeur* et *Protocole*. Leurs propriétés et leurs interrelations sont maintenues. Seule la classe *Protocole* a été adaptée pour prendre en compte les nouvelles classes de l'ontologie Meta-Observatoire. La relation «*aUnContexte*» est supprimée car l'objectif de l'ontologie est d'explicitier les éléments de contexte comme la date, l'acteur, le lieu, etc., qui sont exprimés dans des classes spécifiques. Ces classes sont extensibles avec des sous classes et représentent, comme c'est le cas dans OBOE, des points d'extension de l'ontologie méta-observatoire.

Lors de la création d'un modèle d'observatoire, les chercheurs pourront ajouter des classes et sous-classes correspondant aux spécificités de leurs données d'observations et à leurs besoins de représentation.

5.2 Classe *Observatoire*

La classe *Observatoire* permet de décrire l'ensemble des informations contenues dans la base de données et les métadonnées qui décrivent l'observatoire (en vue du référencement des observatoires dans les portails d'accès aux données). Ces

informations ont un rôle descriptif et d'affichage pour la valorisation de l'observatoire. Cette classe contient des informations d'ordre administratif comme le nom de l'observatoire, ses dates clés, ses origines et des propriétés d'ordre technique comme sa plateforme et son état d'hébergement. Les contributeurs de l'observatoire comme les financeurs, les administratifs sont représentés avec les relations à la classe *Acteur*.

5.3 Classe *Projet*

La classe *Projet* permet de relier l'observatoire à des groupes de travail et des moyens de financement. Un observatoire est en général financé par un ou plusieurs projets de recherche. Un projet peut être national, régional, de recherche ou autre. Il peut avoir un financeur et plusieurs structures ou laboratoires partenaires. Cette classe permet de faire des regroupements d'observations. Ce concept de regroupement est présent dans ODM. Cette classe est aussi un point d'extension possible. Les observations peuvent faire partie d'un projet ou directement d'un observatoire. L'utilisation de cette classe n'est pas obligatoire dans l'ontologie.

5.4 Classe *Acteur*

La classe *Acteur* représente les différents intervenants dans l'observatoire. Ces intervenants peuvent être de différentes natures et genres. Ceci fait de la classe *Acteur* un point d'extension de l'ontologie. Cette classe est reliée aux classes *Unité d'Etude* et *Prélèvement* afin de représenter l'information de qui a fait l'observation, collecté l'unité d'étude ou a fait le prélèvement. Les liens entre les acteurs et les projets sont modélisés par la relation « *participeA* » ceci permet d'avoir des informations sur les membres d'équipes et projets. Les autres intervenants comme les financeurs, administrateurs et autres sont aussi représentés par la classe *Acteur*, la relation « *aUneFonction* » précise le type de lien entre ces acteurs et l'observatoire. Dans OBOE, l'information sur qui a fait l'observation est considérée comme une observation.

5.5 Classes *Entité – Unité d'Etude – Prélèvement – Type*

La classe *Entité* représente les entités cibles de l'observation. Dans l'ontologie méta-observatoire, nous distinguons deux types d'entités observées et proposons donc deux classes : *Unité d'étude* et *Prélèvement* qui héritent de la classe *Entité*. Les *Unités d'étude* sont les éléments collectés sur le terrain qui feront l'objet d'observations, cela peut être par exemple un carottage ou un animal. Les *Prélèvements* sont réalisés à partir d'une unité d'étude comme le prélèvement sanguin réalisé dans l'exemple sur l'arvicola. Ces deux classes sont des points d'extensions de l'ontologie. Une relation « *ProvientDe* » lie les classes *Unité d'étude* et *Prélèvement*. La classe *Type* permet de spécifier le type de l'entité, par exemple prélèvement sanguin, ou micromammifère. Elle sera particulièrement utile

dans le cas d'observatoires composés d'entités de nature différente. Une entité peut être liée à une autre entité.

5.6 Classes *Lieu* – *Date*

Les informations spatio-temporelles dans l'ontologie méta-observatoire sont modélisées par les deux classes *Lieu* et *Date*, comme dans ODM et SERONTO et contrairement à OBOE qui considère ces informations comme des observations.

La classe *Date* est utilisée pour spécifier la date d'observation, de collecte d'une unité d'étude ou d'un prélèvement.

La classe *Lieu* contient les informations spatiales sur les unités d'études dans l'observatoire comme par exemple des coordonnées GPS, des nomenclatures administratives, une surface, d'autres types de coordonnées.

Ces deux classes constituent aussi des points d'extensions. Les extensions de la classe *Date* peuvent être des informations comme Date de Début ou Date de Fin. La classe *Lieu* peut être étendue avec des informations géographiques prédéfinies ou des types de données géographiques.

5.7 Classes *Protocole* – *Outil*

Un *Protocole* explique les étapes ou méthodes de travail, les analyses faites sur l'entité ou les actions de préparation pour réaliser la mesure. Ces informations permettent de décrire comment l'observation a été réalisée de manière à permettre sa validation, à apprécier sa qualité et à la rendre reproductible dans la mesure du possible. La classe *Protocole* est une classe d'OBOE, elle constitue un point d'extension de l'ontologie méta-observatoire comme c'est le cas dans OBOE. La classe *Outil* modélise les outils utilisés par chaque protocole, comme les capteurs, instruments, etc. Un outil est relié à un protocole avec la relation « *utiliseUn* ».

5.8 Instanciation d'un jeu de données dans Méta-Observatoire

La figure 6 reprend l'exemple précédent d'instanciation d'un jeu de données (cf.4.3.2) avec notre ontologie méta-observatoire. Nous avons effectué cette instanciation pour pouvoir d'une part comparer l'instanciation d'un même jeu de données avec OBOE et Méta-Observatoire et d'autre part éprouver notre modèle. Notons que dans la pratique, le Méta-Observatoire ne sera pas instancié (les données seront intégrées dans la base de données générée à partir de l'ontologie Méta-Observatoire).

Nous avons une instance d'*Unité d'étude* et une instance de *Prélèvement* sur lesquelles des mesures sont effectuées. L'unité d'étude correspond au micromammifère capturé et le prélèvement au prélèvement sanguin effectué sur celui-ci.

La représentation du jeu de données reste « fournie », mais ceci est normal car toutes les informations doivent être représentées. Le modèle obtenu est plus lisible et surtout il explicite les contextes. Il met en évidence le cœur de l'observation qui est la mesure et la replace dans son contexte d'observation.

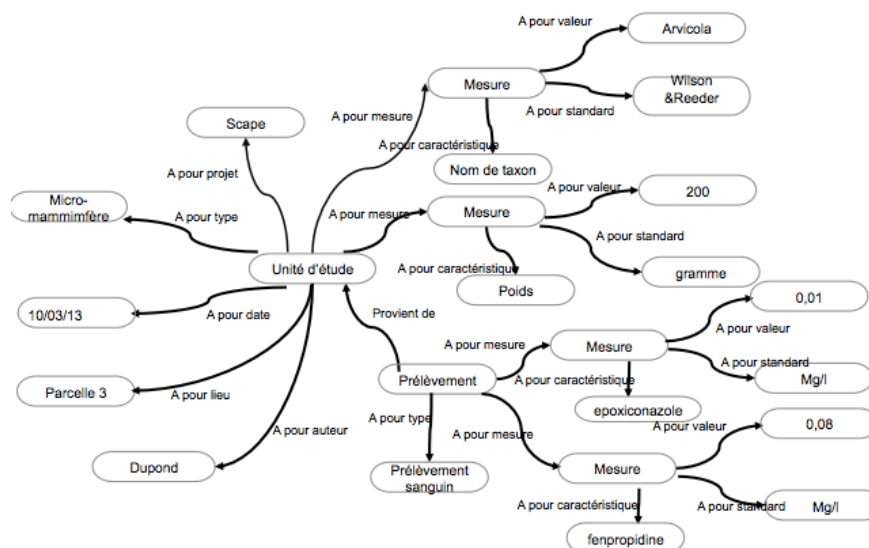


Figure 6 : Exemple d'instanciation dans Méta-Observatoire

6. Conclusion

L'ontologie méta-observatoire présentée propose la définition des concepts de base d'un observatoire en s'appuyant sur des propositions telle que OBOE. Le choix de ce formalisme réside à la fois dans la possibilité de réutiliser des ontologies existantes, concept inhérent aux ontologies et à l'interopérabilité apportée par les ontologies.

L'ontologie proposée n'est pas encore à ce stade, validée. Un travail de mise en œuvre avec des bases de données existantes permettra de conforter ou modifier l'ontologie. La classe *Protocole* proposée par OBOE devra être étendue en se basant sur des travaux actuels tels que RDA (Research Data Alliance) (RDA, 2017) et de l'IVOA (International Virtual Observatory Alliance) (IVOA, 2017).

Le développement de l'environnement d'aide à la modélisation doit être spécifié et il faut introduire des notions telles que les concepts obligatoires, les points d'extension en tenant compte des besoins de la communauté scientifique.

Notre projet à terme est de développer un outil d'aide à la modélisation de base de données. Il sera alors possible de construire des modèles spécifiques pour chaque

observatoire en “*spécialisant*” l’ontologie méta-observatoire puis en générant le modèle de base de données au format souhaité. Les différentes bases de données seront ainsi interopérables.

Notre proposition est en accord avec les objectifs H2020, elle vise une gestion de données de qualité, pérennes et interopérables. Les données seront qualifiées en utilisant les référentiels du domaine. Elles seront rendues pérennes car stockées sur des serveurs, sauvegardées (et non comme c’est encore très souvent le cas, sur le seul ordinateur du chercheur).

Bibliographie

- Charbonnier A., Knapp J., Demonmerot F., Bresson-Hadni S., Raoul F., Grenouillet F., Million L., Vuitton D.A. and Damy S. (2014), *A new data management system for the French National Registry of human alveolar echinococcosis cases*, Parasite 21 69, december 2014.
- DataOne (2017), Projet DataOne, <https://www.dataone.org/>
- EWET (2017), CHRUB, Chrono-environnement, Jenny Knapp, EWET : *Site web EmsB pour le typage d'Echinococcus*. Chrono-environnement. FR-18008901306731-2016-09-12.
- Gruber T. R. (1993). *Towards principles for the design of ontologies used for knowledge Knowledge Representation*. Kluwer Academic Publishers, Deventer, The Netherlands, 1993.
- H2020 (2017a), <http://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html>
- H2020 (2017b), *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*, 2017 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- H2020 (2016), *Guidelines on FAIR Data Management in Horizon 2020*, 2016, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Horsburgh J. S., Tarboton D. G., Maidment D. R. and Zaslavsky I. (2008), *A relational model for environmental and water resources data*, Water Resources Research, vol 44, 2008.
- IVOA (2017), *International Virtual Observatory Alliance*, <http://www.ivoa.net>
- Madina J., Bowers S., Schildhauer M., Krivovc S., Pennington D. and Villac F. (2007), *An ontology for describing and synthesizing ecological observation data*, Ecological Informatics, vol. 2, pp. 279-296, 2007.
- King G. (2007), *An Introduction to the Dataverse Network as an Infrastructure for Data Sharing*, Sociological Methods and Research, 32. pp. 173-199. [online] avail. at <http://j.mp/iHJcAa>
- Observatoire virtuel (2017), <https://www.obspm.fr/-observatoire-virtuel-.html>
- Paleofire (2016), GPWG (2016): *Global Charcoal Database - Paleofire*. Chrono-environnement. FR-18008901306731-2016-03-08.

RDA (2017), *Research Data Alliance*, <https://www.rd-alliance.org/>

Reffay, C., Dyke, G. and Betbeder, M.-L. (2012) *Data sharing in CSCR: towards in-depth long term collaboration*, in Juan, A., Daradoumis, T., Roca, M., Grasman, S., Faulin, J. (Eds.), *Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications*, IGI Global, pp. 111-134. ISBN 978-1-4666-0125-3

Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. (2015) *Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide*. PLoS ONE 10(8): e0134826. doi:10.1371/journal.pone.0134826

Werf B.v.d., Adamescu M. and Ayromlou M., et al. (2008), *SERONTO: A Socio-Ecological Research and Observation Ontology: the core ontology*, 2008. Proceedings of TDWG 2008, Fremantle, Australia, 17-25 October 2008.