



**HAL**  
open science

## Construction collective de standards de données en écologie : standards de données en écologie

Marie-Angélique Laporte, Eric Garnier, Isabelle Mougenot

### ► To cite this version:

Marie-Angélique Laporte, Eric Garnier, Isabelle Mougenot. Construction collective de standards de données en écologie : standards de données en écologie. JFO: Journée Francophone sur les Ontologies, Jun 2011, Montréal Canada. hal-01832497

**HAL Id: hal-01832497**

**<https://hal.umontpellier.fr/hal-01832497>**

Submitted on 7 Jul 2018

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

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

---

# Construction collective de standards de données en écologie

## Standards de données en écologie

Marie-Angélique Laporte<sup>\*,\*\*</sup> — Eric Garnier<sup>\*\*</sup> — Isabelle Mougenot<sup>\*</sup>

\* LIRMM (Université Montpellier 2 - CNRS), 161, rue Ada 34095 Montpellier, France  
mougenot@lirmm.fr

\*\* CNRS, Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, 1919, Route de Mende, FR-34293 Montpellier Cedex 5, France  
marie-angelique.laporte@cefe.cnrs.fr, eric.garnier@cefe.cnrs.fr

---

*RÉSUMÉ. L'écologie fonctionnelle manque de standards de données, pour pouvoir pleinement faire usage des données, issues des observations menées sur le long terme. Ces standards ne seront acceptés par l'ensemble de la communauté, que si ils reflètent la perception du plus grand nombre. Nous proposons une approche centrée sur les modèles, qui exploite les ressources généralistes du web de données afin de guider une communauté d'experts dans la construction collective d'un thésaurus. La démarche est mise en pratique au travers de la construction du thésaurus dédié aux traits fonctionnels chez les plantes. Le thésaurus, une fois stabilisé, sera à son tour mobilisable sur le web et viendra amorcer la construction d'une ontologie de domaine en écologie fonctionnelle.*

*ABSTRACT. Lack of data standards in functional ecology inhibits the most efficient use of long-term observational data. Such data standards however, would be accepted by the whole community if they reflect the opinion and experience of the majority. Here we propose an approach to this problem that is based on models, which integrate web of data resources to guide a community of experts in the collaborative and joint construction of a generally accepted thesaurus. Our practical approach is based on the design of a vocabulary that is dedicated to functional traits in plants. Once the thesaurus has been stabilized, it will be mobilized on the web of data, with the expectation that this approach will lead to a domain ontology in functional ecology.*

*MOTS-CLÉS : écologie fonctionnelle, web de données, thésaurus, modèles de données*

*KEYWORDS: functional ecology, web of data, thesaurus, data models*

---

## 1. Introduction

En écologie, les données et les expertises se sont accumulées au cours du temps, à partir d'études menées souvent de façon indépendante. L'un des enjeux de la discipline est maintenant d'aborder une nouvelle phase, qui consiste à pouvoir combiner et exploiter des jeux de données acquis et traités jusqu'alors de manière autonome (Reichman *et al.*, 2011). L'écologie fonctionnelle manque de standards de données pour pouvoir véritablement partager les données du domaine. Ces standards ne seront acceptés par l'ensemble de la communauté, que si ils reflètent la perception du plus grand nombre. Dans ce travail, nous assimilons la notion de thésaurus à la notion de standard puisque un thésaurus portant sur un domaine particulier, reflète en quelque sorte un consensus de la communauté sur les termes consacrés de ce domaine et sur l'organisation de ces termes. Dans ce sens, le thésaurus est ici perçu, à la fois, comme un processus d'organisation, conduit par une communauté, autour de la structuration des termes d'intérêt du domaine et comme le résultat de ce processus d'organisation. Nous en explorons les aspects organisationnels, et présentons notre approche centrée sur les modèles, qui va guider une communauté d'experts, dans la construction collective d'un thésaurus.

L'article est construit de la manière suivante. La section 2 dresse un panorama des référentiels de savoir sur le web, et souligne les manques avérés de l'existant en matière de standards de données en écologie. La section 3 est consacrée aux activités de conceptualisation au travers de la notation UML. Les modèles généralistes du web de données, habituellement définis pour l'annotation sémantique, sont envisagés, dans notre travail, comme des modèles de construction de consensus. La section 4 est consacrée à la présentation des vocabulaires qui sont une dérivation logique des modèles conceptuels présentés. La section 5 résume les résultats obtenus et en donne quelques perspectives.

## 2. Etat de l'art

Les écologues ont accumulé, sur le long terme, des volumes importants de données au cours de multiples observations et expérimentations. Ces données sont parfois acquises pour répondre à des objectifs particuliers propres à des études spécifiques. Leur stockage et leur représentation ne présentent alors pas un enjeu majeur. Ces données peuvent également s'inscrire dans des contextes d'étude pouvant être très variés, et sont alors décrites en des termes parfois très spécifiques. En conséquence, les données produites en écologie fonctionnelle ne sont que très partiellement réutilisées ou partagées au sein de la communauté. Face à ces limites, des référentiels de contenus reposant sur les standards du web de données, peuvent apporter des solutions, avec l'objectif premier de faire émerger de nouvelles connaissances.

### 2.1. *référentiels de savoir et web de données*

Notre préoccupation première, qui porte sur l'accès, le partage et la diffusion de l'information au sein d'une communauté d'experts, nous oriente tout naturellement vers le web de données ou encore le web sémantique (Shadbolt *et al.*, 2006). Le web de données fournit les concepts, les méthodes et les outils, qui permettent de glisser progressivement d'un web qui favorise majoritairement le partage de documents, à un web qui privilégie le partage des données et leur exploitation concertée par des agents logiciels adaptés. Le web de données s'appuie en priorité sur les principes clés des métadonnées et des vocabulaires contrôlés voire des ontologies (Gruber, 1993), qui sont mis en complémentarité. Les vocabulaires contrôlés comme les ontologies seront généralisés, dans la suite de l'article, par la notion de schémas de concepts. Les lots de métadonnées et les schémas de concepts voient leurs descriptions unifiées au travers de langages de représentation standards à l'image de RDF (Manola *et al.*, 2004), RDFS (RDF Schema) (Brickley *et al.*, 2004) et OWL (Grau *et al.*, 2008). L'idée phare est, de considérer chaque source de données comme une collection de grains de connaissance. Chaque grain de connaissance est un triplet {sujet, prédicat, objet} et est exprimé au travers du langage de description de ressources RDF. Tout triplet consiste en une ressource qui en est le *sujet*, et qui est décrit par une valeur, en l'occurrence l'*objet*, au travers d'une propriété ou *prédicat*. La notion de complémentarité entre les métadonnées et les schémas de concepts, prend tout son sens au niveau du couple (*prédicat, objet*) qui vient enrichir la ressource. Une métadonnée (Neuroth *et al.*, 2003) est envisagée, ici, au travers de ses usages, à savoir comme une information structurée, qui en s'appliquant à une ressource du web, va non seulement l'enrichir d'informations additionnelles, mais aussi en simplifier l'exploitation par des agents logiciels. Concrètement, une métadonnée s'apparente à un couple (*propriété, valeur*) ; la *propriété* rend compte de la structure de la métadonnée et la *valeur* révèle son contenu informationnel. Ce contenu informationnel peut, en outre, provenir d'un schéma de concepts et correspondre typiquement à un élément décrit dans ce schéma. Cette double structuration des ressources, à la fois par la structure de la métadonnée et par la structure de la valeur de la métadonnée, offre de nombreuses perspectives en terme de partage et d'intégration de sources de données.

### 2.2. *Normaliser les descripteurs désignant les concepts du domaine*

Les thésaurus font partie des schémas de concepts et viennent lever les ambiguïtés des langues naturelles pour contrôler et clarifier la diffusion et les échanges d'information et ainsi faciliter la communication. Dans ce sens, les thésaurus reflètent les choix délibérés des communautés, pour la désignation des termes clés, dans leur domaine d'expertise. Il existe deux grandes catégories de termes : les descripteurs qui désignent les concepts de manière préférentielle, et les non-descripteurs qui sont les synonymes ou quasi-synonymes des descripteurs. En raison de leur utilité dans de nombreux domaines, notamment dans les systèmes documentaires, les thésaurus ont fait l'objet de travaux de normalisation (ANSI, 2005) et sont élaborés au travers de

principes directeurs. Ainsi les descripteurs et non descripteurs vont être reliés les uns aux autres au travers de relations sémantiques (hiérarchiques, associatives et d'équivalence). Les normes ne font aucune préconisation en ce qui concerne l'implantation des thésaurus. Au sein du web de données, le standard SKOS (Simple Knowledge Organisation System) (Isaac *et al.*, 2008) a été conçu pour compléter les normes en fournissant un format commun pour l'exploitation, non seulement des thésaurus mais de tout schéma de concept. SKOS permet tout à la fois de faciliter la construction proprement dite de thésaurus mais aussi d'anticiper la mise en place de passerelles entre thésaurus. Ainsi, tout thésaurus peut s'envisager comme une ressource disponible sur le web, et est potentiellement exploitable de concert avec d'autres thésaurus. A cet effet, SKOS vient, au même titre que d'autres ressources généralistes du web de données à l'image du standard de méta-données Dublin Core (The DCMI Usage Board, 2003) ou du format des réseaux sociaux FOAF (Dumbill, 2002), jouer le rôle d'un référentiel interopérable qui va faciliter la mise à disposition de ressources autonomes, mais cependant, normalisées et complémentaires.

### **2.3. Initiatives communautaires dans les sciences du vivant autour des schémas de concepts**

De nombreuses communautés, majoritairement en biologie moléculaire, se sont dotées de schémas de concepts de domaine, sur ces dernières années, facilitant ainsi la diffusion et les échanges d'information intra ou inter-communautaires. Le vocabulaire contrôlé Gene Ontology (GO) (Gene Ontology Consortium, 2010) est l'exemple même d'une des toutes premières initiatives. A l'origine, la motivation était de normaliser les descripteurs associés aux gènes et à leurs produits d'expression, afin de privilégier la comparaison des génomes, et d'en dégager les mécanismes généraux régulant l'expression des gènes. De nombreux vocabulaires ont ensuite été définis en génomique et en protéomique sur les mêmes principes que GO. Les communautés, à l'origine de l'ensemble de ces vocabulaires, ont notamment fédéré leurs efforts autour du format partagé OBO (Open Biomedical Ontologies) (Smith *et al.*, 2007) qui donne un cadre normatif aux descripteurs clés de la biologie et à leurs relations spécifiques. Afin de contribuer à rendre ces vocabulaires totalement interopérables, des travaux récents portent sur la conversion de vocabulaires OBO en vocabulaires SKOS (Jupp *et al.*, 2007). Différents vocabulaires contrôlés, définis en OBO, présentent de l'intérêt dans le domaine de l'écologie. Nous pouvons citer Plant Ontology (PO), Trait Ontology (TO) ou Crop ontology (CO) (Shrestha *et al.*, 2010). Néanmoins, PO, TO, et CO sont plutôt centrés sur les aspects moléculaires des plantes et, dans ce sens, ne sont pas directement exploitables en écologie fonctionnelle. Hors du contexte d'OBO, d'autres initiatives, à l'exemple d'AGROVOC - le thésaurus agricole multilingue de la FAO (Food and Agriculture Organisation) (Soergel *et al.*, 2004) - commencent également à faire leur apparition et exploitent le formalisme SKOS.

#### 2.4. Manques avérés de standards de données en écologie fonctionnelle

L'écologie est une discipline ancienne, qui couvre un large panel de thématiques scientifiques et qui, de ce fait, a fait l'objet d'acquisitions de nombreuses données provenant d'observations et d'expérimentations (e.g. campagnes de relevés floristiques, mesures de caractéristiques des espèces, de processus écosystémiques...). La grande majorité des jeux de données collectés demeure toutefois inaccessible en raison, par exemple, des modes de gestion de ces données, de leur nature hautement diverse ou encore de la propriété intellectuelle assujettissant ces données. De manière additionnelle, des données secondaires, pouvant porter, par exemple, sur le climat ou la nature des sols, sont parfois nécessaires pour révéler pleinement le potentiel des données en écologie.

L'écologie fonctionnelle s'intéresse aux mécanismes de contrôle s'exerçant sur les fonctions des systèmes écologiques, des organismes au globe terrestre, en passant par les écosystèmes et les paysages. Ces fonctions se manifestent par des flux entrants ou sortants de ces systèmes, mais aussi par des flux internes d'allocations des ressources e.g. entre organismes ou, au sein des organismes, entre des processus tels que la croissance ou la reproduction. Dans ce contexte, la notion de trait, qui porte sur les entités biologiques des systèmes écologiques, va offrir un des moyens de typer fonctionnellement les individus. Dans ce sens, un trait fonctionnel est une *caractéristique mesurable* à l'échelle de l'individu (Violle *et al.*, 2007), clairement lié à une fonction biologique, et va permettre d'appréhender directement ou indirectement la valeur adaptative de cet individu. En guise d'exemple, les feuilles d'une plante peuvent être caractérisées par leur durée de vie *Leaf Life-span*, ou encore leur surface spécifique *Specific Leaf Area*, alors que les graines peuvent être décrites au travers de leur masse *Seed Mass* ou de leur taille *Seed Size*. Une description détaillée et des méthodes de mesure pour 28 traits fonctionnels des plantes, considérés comme fondamentaux pour leur fonctionnement, sont présentées par (Cornelissen *et al.*, 2003). Cette publication fait référence dans la communauté d'experts et y fait même office de mémento, démontrant de l'importance de se doter de standards.

Différentes bases de données organisent l'information sur les traits fonctionnels, à l'image de Leda Traitbase, PlanTraits, SID (Seed Information Database), BioPop ou Try database (Kattge *et al.*, 2010). Ces différentes bases ont été mises en place pour répondre au volume sans cesse croissant de données collectées, et à la volonté de partager l'ensemble de ces données pour répondre à des questions portant par exemple sur la distribution des espèces en relation avec les conditions du milieu (climat, usage des terres, etc) ou sur les relations entre diversité des plantes et fonctionnement des écosystèmes. Ces bases de données sont toutefois contraintes à la fois par des accès limités et par une couverture partielle de l'information présentant un intérêt.

En parallèle, des standards de métadonnées viennent renforcer l'effort de structuration déjà conduit par les bases de données. Ecological Metadata Language (EML) (Michener *et al.*, 1997) fournit un modèle d'organisation des données qui vient compléter un jeu de données en écologie et qui va permettre d'apporter des réponses aux questions qui auront une formulation de type : *Où, Quand, Qui, Quoi, Comment*. Le propos est donc non seulement de décrire les données elles même mais aussi les jeux

de données qui reflètent une unité de collecte, sous la responsabilité d'un scientifique. Un des objectifs actuels de l'écologie fonctionnelle vise non seulement l'étude de la variabilité d'un trait fonctionnel entre individus d'une même espèce, ou entre espèces, mais aussi l'étude concertée de la variabilité de plusieurs traits fonctionnels clés permettant de définir certaines stratégies adaptatives correspondant à la façon dont les plantes gèrent leur adéquation avec le milieu. A cette fin, il devient indispensable de disposer d'un standard de données dédié aux traits, qui vienne en normaliser les termes descriptifs de la manière la plus exhaustive possible. Ce standard doit recevoir l'agrément de l'ensemble de la communauté. Il ne s'agit donc pas seulement de construire le thésaurus *trait* sur le socle actuel des standards du web de données mais également de prendre en charge les perceptions de chaque expert de la communauté et de mutualiser ces perceptions au sein du thésaurus.

Nous avons pris le parti d'articuler notre démarche autour d'une activité de modélisation conceptuelle qui nous apparaît comme prépondérante et qui confère le niveau d'abstraction nécessaire à l'exploitation concertée de plusieurs modèles de données. Il s'agit notamment de proposer une représentation adaptée des éléments qui viennent structurer la construction collective d'un thésaurus dédié. Dans cette optique, la perception de chaque expert est retranscrite sous la forme de métadonnées qui, bien que subjectives, seront à même de faire évoluer progressivement le thésaurus.

### **3. Approche conceptuelle**

#### **3.1. *Préalable à la conception***

La modélisation va revêtir un rôle clé puisqu'elle va apporter les artéfacts nécessaires à l'intégration des standards que nous réutilisons à notre profit. De fait, nous ne construisons que les formats de données qui font défaut. Les activités de conceptualisation vont donc nous permettre de guider l'exploitation concertée de l'ensemble de ces formats. Ainsi le thésaurus co-construit par les experts du domaine, va être structuré au travers du standard SKOS pour ce qui relève de la structure du thésaurus, proprement dite, du standard FOAF pour ce qui relève de l'expression de l'information relative aux acteurs faisant vivre le système, et du standard de méta-données Dublin Core pour ce qui relève de la prise en charge des contributions des experts, vues comme autant d'annotations, venant enrichir le thésaurus. Les avantages d'une telle approche, sont à notre sens, nombreux. Ainsi, nous pouvons partager des standards de données avec l'ensemble des communautés qui appuient leurs travaux sur les technologies du web de données, et également bénéficier des évolutions régulières de ces standards ainsi que de leurs outils associés. De cette manière, nous pouvons envisager toute sorte d'extension possible, comme par exemple d'établir des relations entre les descripteurs du thésaurus de traits fonctionnels, et des descripteurs provenant de vocabulaires contrôlés définis pour les sciences du vivant, à l'image des vocabulaires de Plant Ontology (PO) (The Plant Ontology Consortium, 2002) ou du vocabulaire PATO (Phenotype And Trait Ontology).

Nous avons fait le choix du langage de modélisation objet UML (Unified Modeling

Language) (OMG, 2007), pour mener à bien nos activités de modélisation conceptuelle. UML va nous permettre d’analyser la problématique associée au domaine d’intérêt dans son intégralité, et d’en simplifier la réalité, tout en faisant émerger les aspects clés. Il s’agit de construire les modèles structurels, à même de proposer des solutions, en terme de système à mettre en œuvre. UML est également, doté d’un méta-modèle, ce qui en fait un langage de modélisation particulièrement adapté à la transformation de modèles. Cette dernière qualité nous importe tout particulièrement puisqu’il s’agit ensuite pour nous, de plonger nos modèles dans les formalismes associés au web de données.

### 3.2. Efforts de modélisation

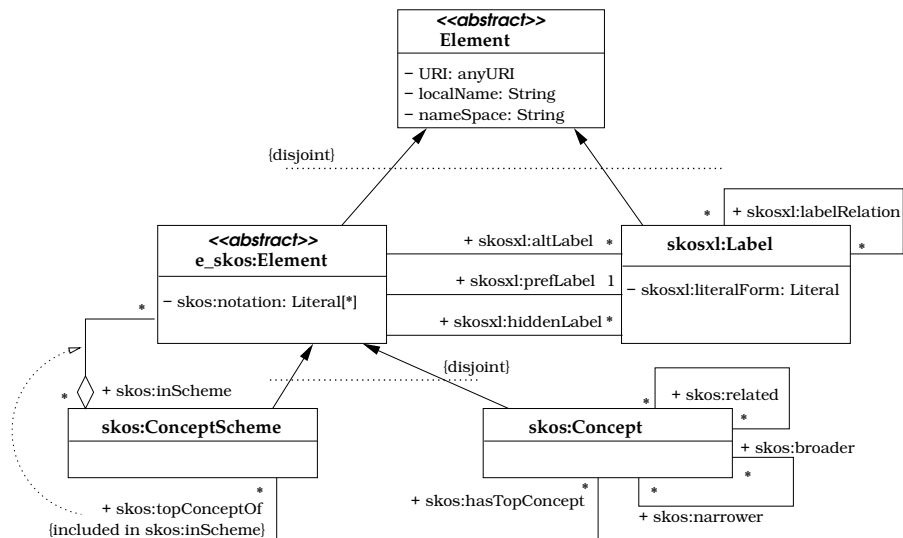
Nous nous limitons à la description des éléments de modélisation qui fournissent le socle de notre approche qui est orientée vers de la réutilisation de modèles d’annotation de ressources, au bénéfice de modèles d’adaptation mutuelle de contenus.

#### 3.2.1. Réflexion autour de SKOS et de ses extensions

SKOS (Isaac *et al.*, 2008) est un modèle, qui peut structurer tout schéma d’organisation de savoir. Nous réalisons un travail partiel de rétro-conception mené autour de SKOS, à partir du schéma SKOS défini en OWL (W3C Working Group, 2009). Le diagramme de classes UML présenté dans la figure 1, exploite conjointement les modèles SKOS et SKOS XL. Une classe abstraite `e_skos:Element` est définie pour les besoins de la modélisation, et généralise les classes `skos:Concept` et `skos:ConceptScheme`. SKOS XL (Recommendation Edition, 2009) est une extension de SKOS, et offre une représentation adaptée à la complexité lexicale des termes qui désignent les éléments SKOS. Ainsi, SKOS XL permet de gérer de manière autonome, les descripteurs comme les non descripteurs de tout élément SKOS, au travers de la classe `skosxl:Label` et d’associer ensuite, ces descripteurs et non descripteurs vus, comme autant de ressources indépendantes. Ce point est essentiel ; de nombreux désaccords entre les experts du domaine portent, en effet, sur la désignation des termes consacrés à la description des traits fonctionnels. La modélisation de ces termes en tant qu’instances de la classe `skosxl:Label` va nous permettre de les faire évoluer indépendamment des concepts désignés et de gérer les choix communautaires faits par les experts portant sur ces termes. La figure 1 donne, également une description des classes `Concept` et `ConceptScheme` et de leurs associations. Ainsi, la classe `ConceptScheme` permet de modéliser de concert plusieurs schémas d’organisation de concepts et donc d’aller vers le partage et l’intégration de ces schémas. Cette qualité va nous permettre d’anticiper les passerelles possibles du thesaurus `trait` vers d’autres standards de données. Il est, également, à noter les associations réflexives de `skos:Concept` vers `skos:Concept` qui vont permettre d’organiser les concepts SKOS au travers de relations hiérarchique et de voisinage.

En matière de documentation, le modèle SKOS offre la possibilité d’enrichir les instances de ses classes au travers de différentes propriétés souvent descriptives



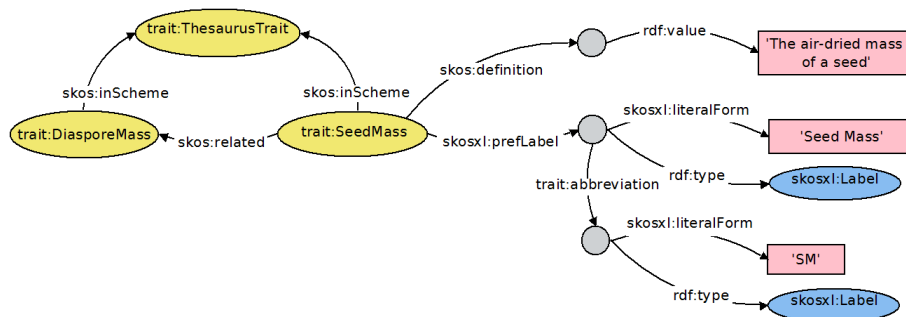


**Figure 1.** Diagramme de classes : articulation de SKOS et de son extension SKOS XL

à l'exemple de `skos:definition` ou encore de `skos:changeNote` pour ce qui concerne le vocabulaire SKOS, ou de `dct:relation` pour ce qui concerne le vocabulaire Dublin Core. Cette notion d'annotation va nous permettre de nous rapprocher des travaux menés autour du standard Dublin Core et autour des métadonnées de manière générale, et va également œuvrer en direction du partage des concepts.

### 3.2.2. Réflexion autour du standard de métadonnées Dublin Core

La notion d'annotation est généralisée par celle de métadonnées. Le standard Dublin Core est un ensemble de schémas de métadonnées qui vont à la fois décrire et associer tout type de ressource. Dublin Core est activement porté par une communauté qui a construit un large ensemble de fonctionnalités facilitant son utilisation à large échelle. Dans cette optique, un métamodèle, nommé modèle abstrait du Dublin Core (DCAM) (Baker, 2005; Nilsson *et al.*, 2006) propose une abstraction de la notion de *ressource décrite*, sous la forme d'une collection de métadonnées. Chaque métadonnée est un couple (propriété, valeur). La propriété est issue d'un des vocabulaires standards appartenant au web de données. La valeur peut être à son tour envisagée comme une ressource et décrite à son tour. Ce métamodèle nous convient tout particulièrement puisqu'il va autoriser l'attachement de métadonnées à toute sorte d'objet du thesaurus : descripteur, non descripteur, concept, schéma de concepts et même annotation. De sorte, nous pouvons envisager un modèle prenant en charge la contribution de chaque expert sous la forme de lots de métadonnées venant enrichir les éléments du thesaurus. Nous nous inspirons donc très volontairement du métamodèle du Dublin Core, nous généralisons tout élément SKOS et SKOS XL en *ressource décrite*



**Figure 2.** Portion RDF du thésaurus trait décrivant le concept Seed Mass

du Dublin Core, de manière à les enrichir au travers d'agrégats de métadonnées. La valeur d'une métadonnée est aussi une *ressource* et va pouvoir être décrite, à son tour, par un agrégat de métadonnées. Le canevas de représentation d'adaptation mutuelle de contenu, ainsi défini, est exploité au sein de la figure 3.

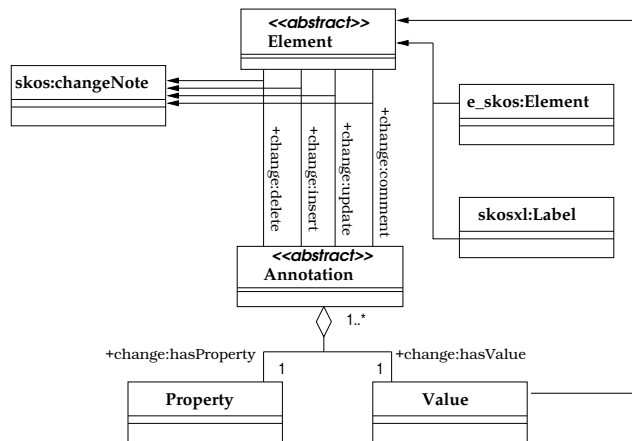
#### 4. Mise en œuvre

Un premier regard porte, sur la construction du thésaurus proprement dit à partir des formats SKOS et SKOS XL. Nous abordons, dans un second temps, les aspects relatifs à l'adaptation mutuelle du thésaurus.

##### 4.1. Thésaurus de traits fonctionnels chez les plantes

Les modèles SKOS et SKOS XL sont envisagés à la manière de modèles descripteurs du modèle *trait*. Il est ainsi possible au sein du modèle *trait*, de proposer des éléments de description additionnels, qui vont rendre compte de la spécificité du domaine des traits fonctionnels. Une portion de graphe RDF illustre le vocabulaire *trait* à la figure 2. Le schéma de concepts nommé `trait:ThesaurusTrait`, et les concepts voisins de traits fonctionnels `trait:DiasporeMass`<sup>1</sup> et `trait:SeedMass`<sup>2</sup> viennent étendre les classes du modèle SKOS. Le concept `trait:SeedMass` est ensuite enrichi de descripteurs qui étendent la classe `skosxl:Label` du modèle SKOS XL. Les labels vont pouvoir ainsi être enrichis, à leur tour. Dans notre exemple, nous illustrons cette faculté par l'exploitation d'une relation `trait:abbreviation` qui vient raffiner l'association `skosxl:labelRelation` de SKOS XL. Les écologues ont l'habitude de désigner les traits par des abréviations et la propriété `trait:abbreviation` va, de fait, associer une abréviation au label préféré d'un trait (SM pour le trait Seed Mass).

1. Masse de la diaspore  
2. Masse de la graine

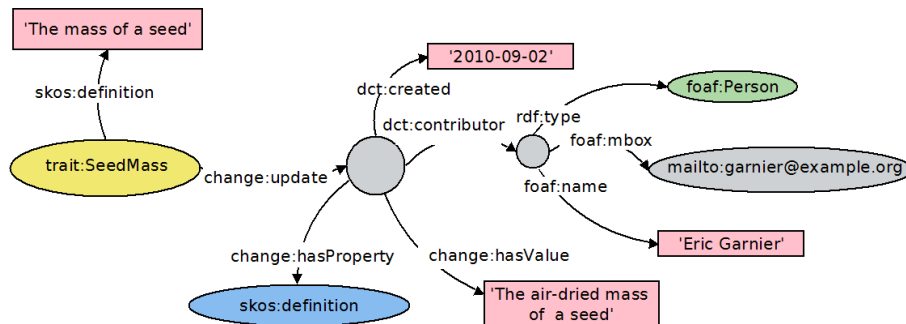


**Figure 3.** Diagramme de classes du vocabulaire change

#### 4.2. Annotation du thésaurus

Il s'agit maintenant de prendre en charge les adaptations au vocabulaire pouvant être apportées par chacun des experts. A cet effet, un modèle présenté à la figure 3, conforme au méta-modèle du Dublin Core, est construit. Ce modèle va guider la construction du vocabulaire d'annotation prenant en charge les aspects collaboratifs. Parmi les différentes finalités des métadonnées, nous avons retenu leur fonction originelle dans la gestion des données, et plus précisément dans la maintenance, et la mise à jour des données. Les métadonnées vont dans ce sens, informer des derniers changements appliqués à un jeu de données. Notre approche tire pleinement parti de cette dernière qualité, les modifications apportées aux éléments du thésaurus par les experts du domaine, sont assimilées à des métadonnées susceptibles de venir mettre à jour le jeu de données initial. Toutes les mises à jour voulues par les experts ne vont pas être directement appliquées sur le jeu de données mais conservées et structurées comme autant de métadonnées. Nous faisons une analogie avec les fichiers de journalisation dans le monde des bases de données, chaque instruction s'exécutant au sein d'une transaction va y être tracée, afin de pouvoir rejouer ou bien déjouer la transaction en cas de panne. La différence porte sur la non mise en concurrence des activités englobées au sein des transactions, puisque aucune modification n'est véritablement appliquée sur les données, du moins dans un premier temps.

Dans ce sens, et pour permettre une construction collaborative du thésaurus, nous avons développé le vocabulaire nommé *change*. Ce vocabulaire définit les éléments nécessaires aux mises à jour du thésaurus et plus largement à tout ce qui va relever des aspects d'adaptation mutuelle entre les experts du domaine. Ce vocabulaire se compose uniquement de propriétés. Trois de ces propriétés `change:insert`, `change:update`, `change:delete` spécialisent la propriété du vocabulaire SKOS `skos:changeNote` et correspondent à la typologie habituelle des mises à jour pou-



**Figure 4.** Exemple de contribution d'expert portant sur `trait:SeedMass`

vant être opérées sur des données. La propriété `change:comment` étend également `skos:changeNote` afin de prendre en charge tout commentaire sous forme de texte libre pouvant être ajouté par un expert. Le vocabulaire *change* se compose également de deux propriétés supplémentaires, `change:hasValue` et de `change:hasProperty` qui vont permettre de valuer les modifications effectuées au niveau d'une ressource par un couple de propriété-valeur. La figure 4 offre un exemple d'exploitation du vocabulaire *change*, le trait *Seed Mass* est décrit par sa définition *the mass of a seed*. Un expert du domaine, trouvant cette définition trop imprécise, propose de la modifier par la valeur *the air-dried mass of a seed*. L'identité de l'expert est conservée et typée et sa contribution est enregistrée. Toutes ces informations se rapportent à une annotation qui est associée par une propriété `change:update` au concept `trait:SeedMass`.

## 5. Discussion : conclusion et perspectives

Nous répondons aux besoins exprimés par une communauté d'experts en écologie par des activités de modélisation qui ont permis de dégager une double structuration portant sur l'organisation de thésaurus et des aspects de construction collective sous-jacents. Cette double organisation exploite la complémentarité, en terme de finalité, des schémas de concepts et des métadonnées et met à profit les ressources généralistes du web de données SKOS et Dublin Core, qui normalisent respectivement les schémas de concepts et les métadonnées. Une première perspective est à court terme, et porte sur la construction en cours, au travers d'un environnement informatique dédié, du thésaurus portant sur les traits fonctionnels. Le souhait est de collecter et de concentrer les perceptions des experts au sein du thésaurus et de mettre en place un système participatif dans lequel la valeur ajoutée provient de l'agrégation des contributions des différents experts. Un travail précurseur en terme d'ingénierie de connaissances participative est présentée dans (Giboin *et al.*, 2008). La première version du thésaurus en cours d'acquisition, une fois stabilisé, sera rendue disponible par les écologues à l'ensemble de la communauté, sous forme d'une ressource mobilisable sur le web. Une seconde perspective est à plus long terme, et porte sur l'exploitation du thésaurus

construit, comme préalable à la construction d'une ontologie portant sur les traits fonctionnels. L'objectif reste, pour les écologues, lié à la capacité de disposer de modèles de connaissances intégrateurs qui vont permettre de trouver des réponses pertinentes à des questions posées, nécessitant par exemple, d'agréger des données associées à différents traits fonctionnels.

## 6. Bibliographie

- ANSI, « Guidelines for the Construction, Format and Management of Monolingual Thesauri {(ANSI/NISO Z39.19-2005)} », , American National Standard, 2005.
- Baker T., « A Common Grammar for Diverse Vocabularies : The Abstract Model for Dublin Core », in E. Fox, E. Neuhold, P. Premssmit, V. Wuwongse (eds), *Digital Libraries : Implementing Strategies and Sharing Experiences*, vol. 3815 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, p. 495, 2005.
- Brickley D., Guha R. V., RDF Vocabulary Description Language 1.0 : RDF Schema, W3c recommendation, W3C, 2004.
- Cornelissen J. H. C., Lavorel S., Garnier E., Díaz S., Buchmann N., Gurvich D. E., Reich P. B., Steege H. T., Morgan H. D., Van Der Heijden M. G. A., Pausas J. G., Poorter H., « A handbook of protocols for standardised and easy measurement of plant functional traits worldwide », *Australian Journal of Botany*, vol. 51, n° 4, p. 335, 2003.
- Dumbill E., « Finding friends with XML and RDF : The Friend-of-a-Friend vocabulary can make it easier to manage online communities », *IBM's XML Watch*, 2002.
- Gene Ontology Consortium, « The Gene Ontology in 2010 : extensions and refinements. », *Nucleic acids research*, vol. 38, n° Database issue, p. D331-335, 2010.
- Giboin A., Durville P., Gandon F., « Ingénierie ontologique participative : essai de mise en oeuvre avec l'éditeur collaboratif ECCO », *Atelier IC 2.0, joint aux IC2008*, 2008.
- Grau B. C., Horrocks I., Motik B., Parsia B., Patelschneider P., Sattler U., « OWL 2 : The next step for OWL », *Web Semantics : Science, Services and Agents on the World Wide Web*, vol. 6, n° 4, p. 309-322, 2008.
- Gruber T., « A Translation Approach to Portable Ontology Specification », *Knowledge Acquisition*, vol. 5, n° 2, p. 199-220, 1993.
- Isaac A., Summers E., « SKOS Simple Knowledge Organization System Primer », , W3C Technical Report, 2008.
- Jupp S., Bechhofer S., Kostkova P., Stevens R., Yesilada Y., « Document Navigation : Ontologies or Knowledge Organisation Systems ? Sealife Use Case », *NETTAB'2007*, 2007.
- Kattge J., Ogle K., Bonisch G., Diaz S., Lavorel S., Madin J., Nadrowski K., Nollert S., Sartor K., Wirth C., « A generic structure for plant trait databases », *Methods in Ecology & Evolution*, 2010.
- Manola F., Miller E., RDF Primer, {W3C} recommendation, World Wide Web Consortium, 2004.
- Michener W. K., Brunt J. W., Helly J. J., Kirchner T. B., Stafford S. G., « Nongeospatial metadata for the ecological sciences », *Ecological Applications*, vol. 7, n° 1, p. 342, 1997.

- Neuroth H., Bargheer M., « Metadata Models, International Developments and Implementation », in F. Bai, B. Wegner (eds), *Electronic Information and Communication in Mathematics*, vol. 2730 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, p. 112-121, 2003.
- Nilsson M., Johnston P., Ambjörn N., Powell A., « Towards an interoperability framework for metadata standards », *Proceedings of the 2006 international conference on Dublin Core and Metadata Applications : metadata for knowledge and learning*, Dublin Core Metadata Initiative, p. 13-23, 2006.
- OMG, « Unified Modeling Language (OMG UML), Infrastructure, V2.1.2 », 2007.
- Recommendation Edition, « SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document », 2009.
- Reichman O. J., Jones M. B., Schildhauer M. P., « Challenges and Opportunities of Open Data in Ecology », *Science*, vol. 331, n° 6018, p. 703-705, February, 2011.
- Shadbolt N., Berners-Lee T., Hall W., « The Semantic Web Revisited », *IEEE Intelligent Systems*, vol. 21, n° 3, p. 96-101, 2006.
- Shrestha R., Arnaud E., Mauleon R., Senger M., Davenport G. F., Hancock D., Morrison N., Bruskiwich R., McLaren G., « Multifunctional crop trait ontology for breeders' data : field book, annotation, data discovery and semantic enrichment of the literature », *AoB Plants*, 2010.
- Smith B., Ashburner M., Rosse C., Bard J., Bug W., Ceusters W., Goldberg L. J., Eilbeck K., Ireland A., Mungall C. J., Leontis N., Rocca-Serra P., Ruttenberg A., Sansone S.-A., Scheuermann R. H., Shah N., Whetzel P. L., Lewis S., « The OBO Foundry : coordinated evolution of ontologies to support biomedical data integration. », *Nature biotechnology*, vol. 25, n° 11, p. 1251-5, November, 2007.
- Soergel D., Lauser B., Liang A., Fisseha F., Keizer J., Katz S., « Reengineering thesauri for new applications : the AGROVOC example. », *Journal of Digital Information*, 2004.
- The DCMI Usage Board, {DCMI} Type Vocabulary, {DCMI} recommendation, Dublin Core Metadata Initiative, 2003.
- The Plant Ontology Consortium, « The plant ontology consortium and plant ontologies. », *Comparative and Functional Genomics*, vol. 3, n° 2, p. 137-142, 2002.
- Violle C., Navas M.-L., Vile D., Kazakou E., Fortunel C., Hummel I., Garnier E., « Let the concept of trait be functional ! », *Oikos*, n° 116, p. 882-892, 2007.
- W3C Working Group, « SKOS Simple Knowledge Organization System Primer », 2009.