

# Gestion des données ouvertes en Agriculture et Nutrition

*Ce cours en ligne est le fruit d'une collaboration entre les partenaires de GODAN Action, y compris Wageningen Environmental Research (WUR), AgroKnow, AidData, l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (FAO), le Forum Mondial sur la Recherche Agricole (GFAR), l'Institut des Etudes du Développement (IDS), le Land Portal, l'Open Data Institute (ODI) et le Centre Technique de Coopération Agricole et Rurale (CTA).*



*GODAN Action est un projet de trois ans du Département pour le Développement International du Royaume-Uni pour permettre aux utilisateurs, producteurs et intermédiaires de données de s'engager efficacement avec les données ouvertes et maximiser leur potentiel d'impact dans les secteurs de l'agriculture et de l'alimentation. Nous travaillons en particulier à renforcer les capacités, à promouvoir des normes communes et les meilleures pratiques et à améliorer la manière dont nous mesurons l'impact. [www.godan.info]*

**Ce travail est sous licence [CC BY-SA](#).**

# MODULE 4 : PARTAGE DES DONNÉES OUVERTES

## LEÇON 4.4 : Interopérabilité sémantique

---



Photo par [Kyle Spradley](#) sous licence CC BY NC 2.0

## Objectifs et résultats d'apprentissage

Cette leçon a pour objectif :

- D'expliquer les bases de l'interopérabilité sémantique
- D'introduire ce que sont les " vocabulaires ".
- De fournir des conseils sur la façon de choisir les vocabulaires les plus appropriés
- De fournir des conseils sur la façon d'utiliser les vocabulaires dans les (méta)données.

À la fin de cette leçon, vous devrez être en mesure de :

- Comprendre les bases de l'interopérabilité sémantique
- Choisir les vocabulaires qui correspondent le mieux à leurs besoins
- (Guider les développeurs à) utiliser des vocabulaires dans les (méta)données.



# Sommaire

|   |           |
|---|-----------|
| <b>Module 4 : Partage des données ouvertes.....</b>                                 | <b>2</b>  |
| <b>Leçon 4.4: Interopérabilité sémantique.....</b>                                  | <b>2</b>  |
| <b>Objectifs et résultats d'apprentissage.....</b>                                  | <b>2</b>  |
| <b>Liste des illustrations.....</b>   | <b>4</b>  |
| <b>1 Introduction .....</b>   | <b>5</b>  |
| <b>2 Structures sémantiques ou " vocabulaires ".....</b>                            | <b>6</b>  |
| <b>2.1. Types of vocabulaires .....</b>   | <b>7</b>  |
| <b>2.2 Identifier les structures sémantiques publiées les plus appropriées.....</b> | <b>10</b> |
| <b>3 Intégration de la sémantique dans les (méta)données.....</b>                   | <b>11</b> |
| <b>3.1. Utilisation d'un schéma pour vos données.....</b>                           | <b>11</b> |
| <b>3.2Vocabulaires de valeurs pour annoter/catégoriser vos données.....</b>         | <b>13</b> |

## Liste des illustrations

|   |    |
|---|----|
| Illustration 1 Exemple d'utilisation de différents types de vocabulaires pour ajouter de la sémantique aux (méta)données..... | 9  |
| Illustration 2 Exemple de XML utilisant des schémas publiés très connus.....  | 11 |
| Illustration 3 Exemple de RDF triple avec URI de termes d'autres vocabulaires.....  | 13 |
| Illustration 4 Exemple de triples RDF associant un terme local à un terme d'un vocabulaire externe.....                       | 13 |

# 1. Introduction

Dans cette leçon, nous discuterons des principes de base et des approches à utiliser pour améliorer l'interopérabilité sémantique des données.

Tous les formats de données vus dans la leçon 4.3 définissent uniquement des structures de données, comment codifier les champs/variables et les valeurs, et la seule chose qu'une machine peut faire est d'analyser la structure et extraire les variables et valeurs, sans savoir comment traiter chacune de celles-ci. Les variables et les valeurs ont une signification qui, dans de nombreux cas, ne peut être comprise que par les humains (et dans certains cas uniquement par des humains qui parlent le même langage et connaissent les conventions de la même discipline).

Les êtres humains peuvent interpréter les données à travers une sémantique lisible par l'homme qui a toujours été utilisée dans les (méta)données de différentes façons: une chaîne pour identifier le sujet d'une chose ou la couleur d'une chose (par exemple dans les descriptions phénotypiques du matériel génétique), des codes tirés d'une liste de valeurs faisant autorité (par exemple, le type de sol) ou des noms de variables classiques. Mais comme nous l'avons dit, l'interopérabilité est une question de compréhension par les logiciels informatiques: les chaînes de caractères peuvent être différentes dans chaque ensemble de données et dans différentes langues, et même les codes sans système de référence derrière eux ne signifient rien pour les ordinateurs et ne leur permettent pas de prendre des décisions sur la façon dont ils doivent traiter les valeurs.

Par contre, si les métadonnées contenaient des informations sur le système de référence (une " structure sémantique " comme un thésaurus ou une liste de codes) d'où provenaient chaque variable et chaque valeur, et si cette structure sémantique était lisible par machine et fournissait certains identifiants stables que les programmes informatiques pourraient utiliser comme valeurs stables pour concevoir leur comportement (par exemple, utiliser ces valeurs comme valeurs communes dans différents ensembles de données), nous aurions obtenu l'interopérabilité sémantique.

Ainsi, d'une part, les métadonnées doivent intégrer des informations sur les structures sémantiques de référence et indiquer les éléments exacts qu'elles utilisent à partir de cette structure ; d'autre part, ces structures sémantiques, comme les données, doivent être " sérialisées " de telle sorte que les machines puissent les lire et les traiter, et les utiliser pour interpréter ces données.

Les détails sur la façon de publier une structure sémantique, ou un "vocabulaire", dans un format lisible par machine dépassent le cadre de cette leçon. En résumé, pour les besoins de cette leçon, disons que ces vocabulaires sont publiés sous forme d'ensembles de données, avec les termes/concepts et leurs descriptions, codes et idéalement URI, dans un format lisible par machine - pour le moment supposons XML ou RDF/XML.

## 2. Structures sémantiques ou " vocabulaires "

Les vocabulaires sont des ensembles de termes convenus, éventuellement avec des relations définies entre eux. Cela inclut à la fois les termes utilisés pour les métadonnées de description, comme les noms d'éléments de métadonnées, les propriétés, les prédicats (donc les termes dans les vocabulaires de description: schémas de métadonnées, ontologies...) et les termes utilisés pour catégoriser, annoter, classer (donc les termes des vocabulaires de valeurs: thesaurus, listes de codes, classifications, listes d'autorités...).

De nos jours, suivant la voie tracée par le W3C<sup>1</sup>, le terme le plus couramment utilisé pour désigner les ressources qui définissent les éléments sémantiques est 'vocabulaires':

"Sur le Web sémantique, les vocabulaires définissent les concepts et les relations (aussi appelés " termes ") utilisés pour décrire et représenter un domaine d'intérêt. Les vocabulaires sont utilisés pour classer les termes qui peuvent être utilisés dans une application particulière, caractériser les relations possibles et définir les contraintes d'utilisation de ces termes"

Cela inclut à la fois les termes utilisés pour les éléments de description, comme les noms d'éléments de métadonnées, les propriétés, les prédicats (donc les termes dans les vocabulaires de description: schémas de métadonnées, ontologies...) et les termes utilisés pour catégoriser, annoter, classer (ainsi les termes dans les 'vocabulaires de valeurs', aussi parfois appelés 'Knowledge Organisation Systems' (KOS): thesauri, codes lists, classifications, authority list...).

D'autres termes utilisés pour définir ces ressources sont " ressources sémantiques " ou " structures sémantiques ". Les structures sémantiques sont étroitement liées aux " normes des données ", mais l'expression " normes des données " dépasse le domaine de la sémantique parce qu'il inclut également les normes syntaxiques comme les formats de données et les structures de données (voir l'interopérabilité structurelle comparée à l'interopérabilité sémantique dans la leçon 4.2).

---

<sup>1</sup> W3C. Vocabularies. <https://www.w3.org/standards/semanticweb/ontology>

Dans cette leçon, nous utiliserons ces termes de façon presque interchangeable, tendant à utiliser des " vocabulaires " ou des " structures sémantiques " lorsque nous parlons spécifiquement de sémantique et de " normes de données " lorsque nous parlons d'une combinaison de sémantique et de formats de données (certaines normes définissent les deux).

Nous éviterons également l'expression "normes sémantiques" et dirons plutôt "sémantiques publiées". Strictement parlant, une norme devrait être une spécification officiellement approuvée par différentes parties qui doivent l'utiliser pour améliorer la compatibilité d'un objet. Normalement, les normes reconnues sont créées par des organismes de normalisation. Ensuite, il y a aussi les normes de facto, des spécifications élaborées soit unilatéralement, soit avec une faible portée, qui sont largement adoptées en raison de leur popularité ou de la domination de l'industrie.

Cependant, le domaine de la sémantique, en particulier en ce qui concerne les données agroalimentaires, est encore largement expérimental et, dans de nombreux cas, les organismes de normalisation n'ont pas pu (ou ne veulent pas) se lancer dans la sémantique disciplinaire; les institutions travaillant dans des disciplines spécifiques ont donc commencé à développer leurs propres structures/vocabulaires sémantiques et très peu sont devenus des "standards". Par conséquent, ce qui est pertinent pour cette leçon, ce sont les structures sémantiques qui sont au moins: a) conçues/créées dans le but d'être largement endossées et utilisées; et b) accessibles au public, référençables ("publiés"), et éventuellement lisibles par machine.

## 2.1. Types de vocabulaires

Il n'y a pas de classification formelle des types de vocabulaires (ce qui, en soi, pourrait être un exemple utile de vocabulaire de valeurs).

L'exercice de création d'un vocabulaire de types de vocabulaire a été en partie réalisé par l'initiative Dublin Core<sup>2</sup> : leur 'Vocabulaire des types KOS' est limité à des types spécifiques du type plus général 'vocabulaire de valeur' ou KOS. Leur liste est très utile pour donner une idée de la grande variété de KOS et du mélange de caractéristiques qui sont combinées dans leur définition :

- **Système de catégorisation** : système de groupement peu structuré
- **Schéma de classification** : liste des concepts et combinaisons pré-coordonnées de concepts, classées par classification
- **Dictionnaire** : source de référence contenant des mots habituellement classés par ordre alphabétique avec des informations sur leurs formes, leurs prononciations, leurs fonctions, leurs étymologies, leurs sens et leurs utilisations syntaxiques et idiomatiques

---

<sup>2</sup> [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies#KOS\\_Types\\_Vocabulary](http://wiki.dublincore.org/index.php/NKOS_Vocabularies#KOS_Types_Vocabulary)

- **Répertoire toponymique** : dictionnaire géo-spatial des lieux nommés et dactylographiés
- **Glossaire** : collection de gloses textuels ou de termes spécialisés avec leur signification
- **Liste** : un ensemble limité de termes présentés sous la forme d'une simple liste alphabétique ou d'une autre manière évidente sur le plan logique; ne contenant aucun lien de quelque nature que ce soit
- **'Name authority list' ou fichier d'autorité**: vocabulaire contrôlé à utiliser pour nommer des entités particulières de manière cohérente.
- **Ontologie** : modèle formel qui permet de représenter les connaissances pour un domaine spécifique ; une ontologie décrit les types de choses qui existent (classes), les relations entre elles (propriétés) et les façons logiques dont ces classes et propriétés peuvent être utilisées ensemble (axiomes) [voir ci-dessous une note sur comment une ontologie peut être considérée comme un KOS mais aussi comme un vocabulaire de description, un schéma étendu].
- **Réseau sémantique** : ensemble de termes représentant des concepts, modélisés comme les nœuds d'un réseau de types de relations variables
- **Schéma des titres de sujet**: vocabulaire structuré comprenant les termes disponibles pour l'indexation des sujets, ainsi que des règles pour les combiner en chaînes de termes pré-coordonnées si nécessaire
- **Anneau synonyme** : ensemble de termes synonymes ou presque synonymes, dont chacun peut être utilisé pour désigner un concept particulier.
- **Taxonomie** : schéma de catégories et de sous-catégories qui peuvent être utilisées pour trier et organiser d'une autre manière des éléments de connaissance ou d'information.
- **Terminologie** : ensemble des désignations appartenant à une langue particulière
- **Thésaurus** : vocabulaire contrôlé et structuré dans lequel les concepts sont représentés par des termes, organisés de telle sorte que les relations entre les concepts soient explicitées et que les termes préférés soient accompagnés d'entrées d'introduction pour les synonymes ou quasi-synonymes.

Il n'existe pas de liste officielle de vocabulaires de description/modélisation, mais les plus couramment utilisés sont les suivants :

- **Schéma** (ou ensemble d'éléments de métadonnées): tout ensemble d'éléments de métadonnées, comme les schémas XML, les schémas RDF ou un ensemble de descripteurs moins formalisés.



- **Profil d'application** : un schéma constitué d'éléments de métadonnées tirés d'un ou de plusieurs espaces de nommage, combinés entre eux par des implémenteurs, et optimisés pour une application locale particulière.
- **Norme de messagerie** : normes qui décrivent comment formater syntaxiquement (et parfois sémantiquement) un message décrivant habituellement des informations relatives à un événement ou à une heure ; les messages sont déclenchés par un événement et transmis d'une manière quelconque.
- **L'ontologie**, considérée comme une forme plus puissante de schéma.

Comme on peut le voir sur les deux listes ci-dessus, les ontologies constituent un cas particulier : " En informatique et en science de l'information, une ontologie est une dénomination et une définition formelles des types, propriétés et interrelations des entités qui existent réellement ou fondamentalement pour un domaine particulier".<sup>3</sup> En tant que tel, il peut être utilisé à des fins multiples: il peut être utilisé comme un vocabulaire de description, en utilisant les relations ou même les classes définies par l'ontologie comme éléments de métadonnées/propriétés décrivant vos données (par exemple les classes "résistance aux températures extrêmes ou résistance au gel " dans l'ontologie des caractères du blé<sup>4</sup>), ou comme vocabulaire des valeurs, en utilisant des classes ou des entités comme termes pour des valeurs contrôlées (par exemple), les maladies du blé, comme *Puccinia striiformis* du Trait Ontology ou des pays du Geopolitical Ontology de la FAO<sup>5</sup>).

Parfois les frontières entre un schéma et une ontologie sont floues, mais ce qui peut être considéré comme typique d'une ontologie est peut-être la conception " fonctionnelle " plus que descriptive : les classes, les propriétés et surtout les relations sont conçues comme un modèle qui est "exploitable " et peut être utilisé pour le raisonnement. Cependant, la tendance actuelle est d'utiliser uniquement le mot "vocabulaire" et de ne pas trop s'attarder sur la définition des différents types.

La leçon 4.4.1 fournira des exemples plus spécifiques sur la façon d'identifier les vocabulaires les plus appropriés.

---

<sup>3</sup> From Wikipedia: [https://en.wikipedia.org/wiki/Ontology\\_\(information\\_science\)](https://en.wikipedia.org/wiki/Ontology_(information_science))

<sup>4</sup> <http://vest.agrisemantics.org/content/wheat-trait-ontology> <sup>5</sup>  
<http://vest.agrisemantics.org/content/geopolitical-ontology>

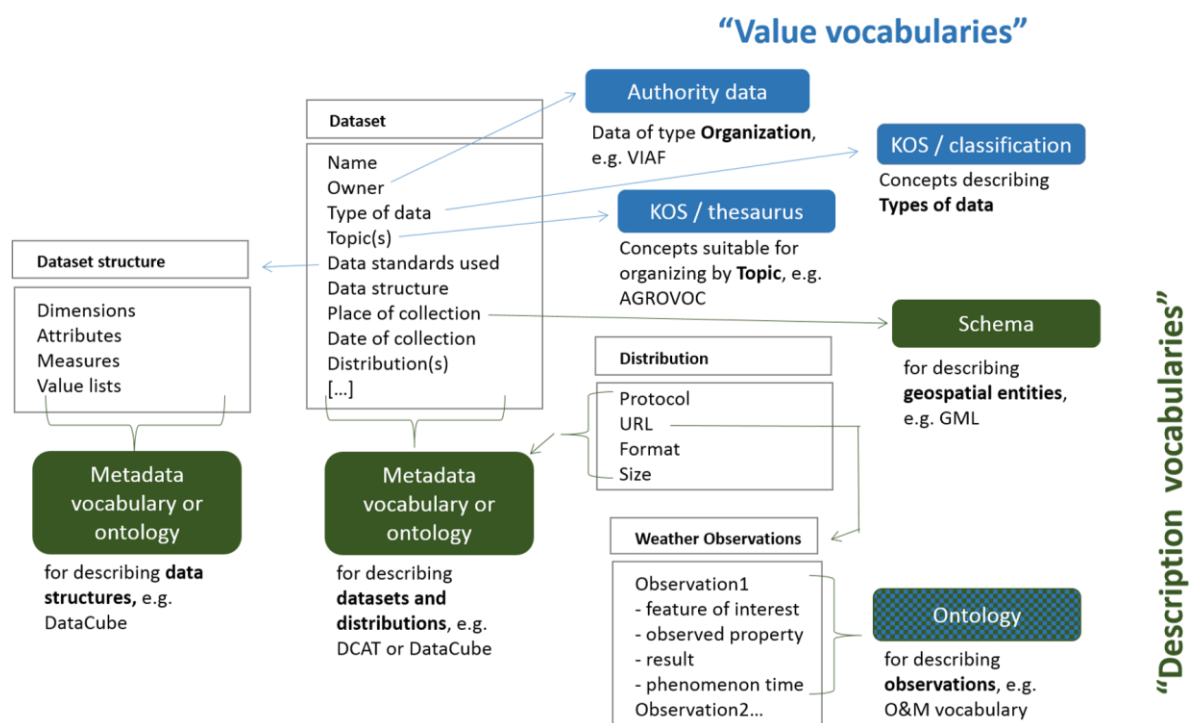


Illustration 1 Exemple d'utilisation de différents types de vocabulaires pour ajouter de la sémantique aux (méta)données

## 2.2. Identifier les structures sémantiques les plus appropriées

Les sources d'information les plus utiles sont bien sûr les catalogues dédiés au domaine agroalimentaire, mais les catalogues généraux consultables par domaine peuvent également être utiles. Une distinction importante doit être faite entre les catalogues/répertoires/registres d'une part et les dépôts d'autre part: les registres sont conçus comme des catalogues de métadonnées, fournissant des descriptions et une catégorisation des vocabulaires et des liens vers le site Web original et la sérialisation originale de la norme, tandis que les dépôts hébergent le contenu complet du vocabulaire, afin que les termes eux-mêmes puissent être consultés. Vous trouverez ci-dessous un aperçu de certains catalogues/répertoires existants de normes et de vocabulaires de données.

### Domaine agro-alimentaire

- “GODAN Action Map of data standards » - <http://vest.agrisemantics.org>  
Un catalogue de normes de données de différents types et formats pour le domaine agro-alimentaire, classées par sous-domaines, types de données, formats et autres critères.
- AgroPortal - <http://agroportal.lirmm.fr/> Un dépôt d'ontologies et de vocabulaires de valeurs, spécialisé en agronomie et en alimentation.
- Planteome - <http://browser.planteome.org/amigo>  
Un dépôt d'ontologies pour la biologie végétale.

## Général

- FAIRsharing - <https://fairsharing.org/> Tiré du répertoire Biosharing des normes pour les sciences de la vie, il s'agit maintenant d'un répertoire général des normes de données de différents types. Il dispose d'un bon système d'étiquetage, mais la couverture des normes agroalimentaires est encore faible.
- Linked Open Vocabularies (LOV) - <https://lov.okfn.org/dataset/lov> Répertoire des vocabulaires RDF (voir leçon 4.3 pour une description des RDF) couvrant toutes les disciplines. Il n'est pas organisé par domaine ou discipline et les vocabulaires ne peuvent être parcourus que par un petit nombre de balises libres.
- The Basel Register of Thesauri, Ontologies and Classifications (BARTOC) - <http://bartoc.org/> BARTOC inclut tous les types de KOS dans n'importe quel format, dans tous les domaines. La catégorisation des vocabulaires est assez générique (l'alimentation et l'agriculture relèveraient en partie de la science pure et en partie de la technologie sans autres sous-catégories).

## 3. Intégration de la sémantique dans les (méta)données

### 3.1. Utilisation d'un schéma pour vos données

Si vous identifiez un vocabulaire, un schéma ou une ontologie de métadonnées qui possède les classes et les propriétés dont vous avez besoin pour décrire vos données, vous pouvez les réutiliser pour modéliser et représenter vos données. Une chose importante à noter à propos de l'utilisation d'un schéma publié existant est qu'en faisant cela seul, vos données seront déjà plus sémantiquement interopérables, car au lieu de noms d'éléments de métadonnées arbitraires locaux qui n'ont aucun sens pour un ordinateur, vous utiliserez des noms d'éléments d'un vocabulaire publié et des outils logiciels qui connaissent ce vocabulaire pourront le mettre à profit, par exemple en faisant correspondre ces valeurs à celles de données qui sont utilisées par les mêmes schémas.

Le schéma adopté devient la 'langue' de votre structure de données. Par exemple, au lieu d'utiliser une structure XML personnalisée avec des noms d'éléments arbitraires locaux, en adoptant un schéma XML existant, vous déclarez que vous utilisez des éléments de celui-ci et pour chaque élément vous utiliserez le nom de l'élément du schéma sélectionné au lieu d'un élément local, avec un préfixe qui indique de quel schéma provient l'élément.

L'exemple ci-dessous (Illustration 2), tiré du manuel d'utilisation de la CSML<sup>5</sup>, déclare dans l'élément XML racine qu'il utilise des éléments du GML (<http://www.opengis.net/gml>) et O&M (<http://www.opengis.net/om>) du Open Geospatial Consortium (OGC) tel que prescrit dans le schéma du Climate Science Modelling Language (CSML) et attribue des préfixes à ces schémas, puis les préfixes des schémas sont utilisés dans le document pour indiquer que le nom des éléments de métadonnées utilisés provient du schéma correspondant (où sa signification est définie). Pour un ordinateur, lire <om:Phenomenon> et savoir que 'OM:' représente le schéma O&M situé sur <http://www.opengis.net/om>, au lieu de simplement lire <phenomenon>, fait une grande différence: au minimum, il sait que OM: Phenomenon dans un ensemble de données signifie la même chose et a les mêmes éléments imbriqués, mais les outils logiciels spécialisés conscients du sens du terme dans ce schéma peuvent construire des fonctionnalités avancées, visualisations, modélisation, etc.

```
<?xml version="1.0" encoding="UTF-8"?>
<Dataset xmlns="http://ndg.nerc.ac.uk/csml" xmlns:gml="http://www.opengis.net/gml"
xmlns:om="http://www.opengis.net/om" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://ndg.nerc.ac.uk/csml XMLSchemas/CSMLAppSchema.xsd" gml:id="TestDataset">
  <gml:metaDataProperty xlink:href="http://ndg.nerc.ac.uk/Metadata/TestDataset"/>
  <gml:description>Test CSML Dataset.</gml:description>
  <UnitDefinitions gml:id="UnitDefs">
    <gml:name codeSpace="http://ndg.nerc.ac.uk">TestDatasetUnitDefs</gml:name>
    <gml:definitionMember>
      <gml:UnitDefinition gml:id="psu">
        <gml:description>Conventional practical salinity units.</gml:description>
        <gml:name>practical salinity units</gml:name>
        <gml:quantityType>sea water salinity</gml:quantityType>
        <gml:catalogSymbol codeSpace="http://ndg.nerc.ac.uk/units">psu</gml:catalogSymbol>
      </gml:UnitDefinition>
    </gml:definitionMember>
  </UnitDefinitions>
  <PhenomenonDefinitions gml:id="PhenDefs">
    <gml:name>TestDatasetPhenDefs</gml:name>
    <gml:definitionMember>
      <om:Phenomenon gml:id="rainfall">
        <gml:description>Liquid precipitation measured with raingauge.</gml:description>
        <gml:name codeSpace="http://ndg.nerc.ac.uk">rainfall</gml:name>
      </om:Phenomenon>
    </gml:definitionMember>
  </PhenomenonDefinitions>
</Dataset>
```

*Illustration 2 Exemple de XML utilisant des schémas publiés très connus*

La technique de déclaration du schéma externe et d'utilisation de ses noms d'éléments avec un préfixe est la même dans le cas des documents RDF. Le niveau d'interopérabilité plus élevé résultant de l'utilisation d'un vocabulaire RDF (et donc d'un ensemble de données RDF), surtout si l'on suit le modèle Linked Data (voir leçon 4.2) est dû à deux facteurs: a) les éléments de métadonnées (classes ou propriétés dans RDF) identifiés par les URI et ces URI sont déréférencés vers des pages Web qui contiennent des informations lisibles par machine sur la classe ou la propriété, de sorte qu'un programme informatique peut suivre le lien et lire plus de données, et s'il existe d'autres liens vers des entités externes elle peut continuer à les suivre et obtenir toujours

<sup>5</sup> <http://proj.badc.rl.ac.uk/csml/browser/Documentation/trunk/CSMLUsersManual.pdf>

plus de données; b) comme indiqué précédemment, un fichier RDF a moins d'ambiguïté potentielle que le fichier XML non RDF et est interprété de façon plus fiable. Vous pouvez utiliser une ontologie de la même manière.

Une utilisation similaire des schémas existants peut être effectuée avec certaines limitations si vous utilisez le format CSV ou JSON. Un schéma plat simple peut être appliqué à un fichier CSV en utilisant les noms des éléments/propriétés comme noms de colonnes. Ceci pourrait être exploité par un logiciel qui utilise le vocabulaire et peut donc faire quelque chose de plus avec les valeurs qu'avec des chaînes de caractères personnalisées. Les schémas XML et RDF sont aussi normalement facilement représentés sous la forme d'une structure JSON, en utilisant les noms préfixés des éléments de métadonnées/propriétés comme étiquettes JSON. Un schéma XML ou RDF peut être codifié comme un schéma JSON et des fichiers JSON basés sur ce schéma peuvent être créés.

Il existe également une variante de JSON appelée JSON-LD (JSON pour Linking Data), basée sur une spécification qui fournit une méthode de codage des données liées et donc RDF utilisant JSON.

## 3.2. Utilisation de vocabulaires de valeurs pour annoter/catégoriser vos données

Un cas légèrement différent se présente lorsque vous voulez utiliser les valeurs d'un vocabulaire existant comme valeurs de certaines métadonnées, par exemple si vous voulez utiliser le terme AGROVOC pour *Oryza sativa* ou le terme identifiant un pays dans l'ontologie géopolitique de la FAO.

Au minimum, une fois qu'un vocabulaire approprié a été identifié, si le vocabulaire n'utilise pas les URI et/ou si les URI ne peuvent être utilisés dans l'ensemble de données, au moins les valeurs littérales des termes peuvent être utilisées dans l'ensemble de données. Les systèmes qui utilisent le vocabulaire et qui peuvent faire correspondre les valeurs littérales avec l'URI peuvent déjà faire quelque chose avec cela. Idéalement, vous devriez utiliser l'URI du terme auquel vous voulez vous référer.

Encore une fois, cela peut être fait de différentes façons selon le format de données que vous utilisez. En XML non-RDF, en CSV ou en JSON, vous pouvez utiliser l'URI comme valeur de l'élément / colonne / étiquette (par exemple, en XML, vous pouvez utiliser l'URI du pays d'ontologie géopolitique comme valeur de l'élément `dc:spatial`, en précisant l'attribut `schema='URI'` pour préciser qu'il s'agit d'un URI).

Cependant: (a) les analyseurs pour ces formats ne suivent généralement pas l'URI et obtiennent des données supplémentaires, comme l'étiquette du terme, de sorte qu'il peut être utile d'avoir un élément séparé pour l'étiquette du terme pour les affichages lisibles par l'homme; (b) même si l'URI contient la base terminologique URI, les analyseurs, pour ces formats, ne chercheraient normalement pas des informations supplémentaires sur ce terme, donc des métadonnées supplémentaires sur le vocabulaire dont il est extrait pourraient être utiles.

Idéalement, l'interopérabilité sémantique est entièrement réalisée en utilisant un format de sérialisation compatible RDF (XML/RDF, Turtle, N3, JSON-LD) et en utilisant l'URI du terme sélectionné comme objet d'un triple. L'avantage de l'utilisation de RDF est que les analyseurs RDF et les crawlers recherchent normalement l'étiquette à partir de l'adresse URI, vous n'aurez donc pas besoin d'ajouter un triple lorsque l'objet est la chaîne.

L'exemple ci-dessous utilise des URI de termes du thésaurus AGROVOC (pour décrire le thème de l'ensemble de données : prévisions météorologiques) et de l'ontologie géopolitique de la FAO (pour identifier le pays : Argentine)

```
<dcats:Dataset rdf:about="#item-d277fa36-beb2-48fc-afdc-e9d887d8329f">
  <dcats:title>Example dataset #1</dcats:title>
  <dcats:description>Example description</dcats:description>
  <dcats:theme rdf:resource="http://aims.fao.org/aos/agrovoc/c_8340"></dcats:theme>
  <dcats:issued>2014-01-01</dcats:issued>
  <dcats:spatial rdf:resource="http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/Argentina"></dcats:spatial>
  <dcats:temporal>2014</dcats:temporal>
  <dcats:source rdf:resource="#item-0ce19e3c-80e5-4e77-a369-ff1a93e37281"></dcats:source>
  <dcats:distribution rdf:resource="#item-19832665-75c8-4706-abcf-2f5f2009446a"></dcats:distribution>
</dcats:Dataset>
```

*Illustration 3 Exemple de RDF triple avec URI de termes provenant d'autres vocabulaires*

Même si vous ne trouvez pas le vocabulaire idéal pour répondre à vos besoins et que vous utilisez vos propres termes, vous pouvez relier votre terme local à un terme similaire ou plus large dans les vocabulaires existants. Cette technique s'appelle " cartographier " (map) les valeurs locales en les associant à des valeurs externes et, en plus d'être la pierre angulaire de l'architecture Linked Data, cette technique multiplie les possibilités de donner un sens aux données locales, car les programmes informatiques peuvent parcourir toutes les URI cartographiées et colliger davantage de métadonnées sur ce terme à partir d'autres vocabulaires.

L'exemple ci-dessous est similaire à l'exemple précédent, mais pour le thème de l'ensemble de données, un terme local plus précis est utilisé ("prévisions météorologiques"), avec une URI locale, associée à l'URI externe qui fait autorité du terme "prévisions météorologiques" dans le thésaurus AGROVOC.

```

<dcat:Dataset rdf:about="#item-d277fa36-beb2-48fc-afdc-e9d887d8329f">
  < dct:title>Example dataset #1</dct:title>
  < dct:description>Example description</dct:description>
  < dcat:theme rdf:resource="#local-6789"></dcat:theme>
  < dct:temporal>2014</dct:temporal>
</dcat:Dataset>
<skos:Concept rdf:about="#local-6789">
  <skos:preferredLabel>weather forecasts</skos:preferredLabel>
  <skos:closeMatch rdf:resource="http://aims.fao.org/aos/agrovoc/c_8340"></skos:closeMatch>
</skos:Concept>

```

*Illustration 4 Exemple de triples RDF associant un terme local à un terme d'un vocabulaire externe*

Dans la leçon 4.4.1, vous trouverez des exemples plus spécifiques sur le choix et l'utilisation des normes de données et de la sémantique pour accroître l'interopérabilité.