

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 4.4.1: Utilisation de la sémantique publiée pour les données agronomiques



Photo par [Michael \(Mikey\) Cantor](#) Sous licence CC BY SA 2.0

Objectifs et résultats d'apprentissage

Cette leçon a pour objectif de fournir des conseils et des exemples sur la façon de codifier les (méta)données en utilisant la sémantique publiée, avec des exemples spécifiques pour les données agronomiques.

A la fin de cette leçon, vous devrez en être en mesure ;

- De comprendre comment identifier les normes de données appropriées pour partager les données agronomiques
- De comprendre comment la sémantique est intégrée dans les (méta)données
- De (guider les développeurs vers ou choisir des outils qui) adoptent les vocabulaires les plus appropriés pour les données d'un type spécifique



Sommaire

Module 4 : Partage des données ouvertes.....	2
Leçon 4.4.1: Utilisation de la sémantique publiée pour les	
Données agronomiques.....	2
Objectifs et résultats d'apprentissage.....	2
Liste des illustrations.....	4
1. Introduction	5
2. Identification des vocabulaires les plus appropriés.....	5
3. Utilisation des vocabulaires existants pour vos données agronomiques.....	7
3.1. Vocabulaires pour la structure des (méta)données.....	8
3.1.1. Sémantique pour les métadonnées des ensembles de données.....	9
3.1.2. La sémantique des données.....	10
3.2. Vocabulaires pour les valeurs de (méta)données.....	14
Lecture complémentaire.....	17

Illustrations

- Illustration 1 Exemple de codage RDF de métadonnées d'ensemble de données et de structure de données à l'aide de Data Cube9
- Illustration 2 Exemple d'observation au format XML en utilisant le schéma O&M..... 10
- Illustration 3 Exemple d'observation au format JSON en utilisant le modèle O&M..... 11
- Illustration 4 Exemple d'une expérience décrite au format JSON en utilisant la norme ICASA. 12
- Illustration 5 Exemple d'observation dans RDF Turtle à l'aide de l'ontologie SSN du W3C..... 12
- Illustration 6 Exemple d'une observation codée en RDF (Turtle) avec l'ontologie OM Lit..... 13
- Illustration 7 Observation codée en RDF (Turtle) à l'aide du Data Cube..... 13
- Illustration 8 Utilisation de vocabulaires pour les valeurs dans l'exemple du Data Cube.....15
- Illustration 9 A partir de l'exemple de la illustration 6..... 16

1. Introduction

Dans la leçon 4.4, nous avons donné un aperçu de l'interopérabilité sémantique. Dans cette leçon-ci, nous donnerons des exemples plus précis de sa mise en œuvre, en particulier l'identification et la réutilisation de la sémantique publiée pour les données agronomiques.

Les exemples pratiques que nous utiliserons tout au long de cette leçon sont liés aux observations et aux expériences agronomiques. Nous fournirons également des exemples reliés à l'ajout de sémantique aux métadonnées.

2. Identification des vocabulaires les plus appropriés

Nous avons déjà lu dans la leçon 4.4 quelques conseils utiles pour choisir les vocabulaires les plus appropriés. Voyons maintenant quelques exemples pratiques utilisant des ensembles de données d'observations/expériences agronomiques (vous pouvez consulter tous les vocabulaires mentionnés dans cette leçon sur la carte GODAN déjà mentionnée - <http://vest.agrisemantics.org>).

- En général, vous aurez besoin d'un vocabulaire qui vous aidera à décrire l'ensemble de données qui contient vos données, pour faciliter la découverte de l'ensemble de données. Dans ce cas, vous devriez chercher des vocabulaires qui décrivent aussi précisément que possible un ensemble de données ou votre type d'ensemble de données spécifique. Par exemple, dans l'Agrisemantics Map of standards, vous pouvez rechercher des vocabulaires qui décrivent des métadonnées de ressources d'information (vocabulaires génériques, pour tout type de ressource) et vous trouverez déjà des vocabulaires comme Dublin Core ou schema.org qui sont compris par de nombreux outils et peuvent coder des métadonnées comme titre, auteur, sujet, format, espace et temps, métadonnées etc.

Vous pouvez ensuite restreindre votre recherche aux normes de données qui décrivent un " ensemble de données ". Vous y trouverez des vocabulaires comme le catalogue de données du W3C et le Data Cube du W3C, ainsi que des formats de données comme NetCDF et HDF5. Le choix entre ces normes de données dépend de la façon dont vous voulez exposer vos données: en lisant les descriptions, vous découvrirez que NetCDF et HDF5 sont des structures de données sans sémantique et non extensibles avec des schémas, donc si vous les utilisez, vous ne pouvez utiliser aucune autre sémantique publiée pour enrichir les données. D'un autre côté, vous pouvez aussi lire qu'ils sont particulièrement adaptés à de grandes quantités d'observations, donc dans le cas d'observations agronomiques, vous voudrez peut-être les considérer.

Vous pouvez même décider d'exposer vos données en utilisant plus d'une norme: voir section 3.

- Les vocabulaires ci-dessus fournissent les métadonnées pour décrire l'ensemble de données, mais vous voudrez peut-être aussi identifier des vocabulaires de valeur à utiliser comme vocabulaires contrôlés pour certaines de vos métadonnées d'ensemble de données. Par exemple, pour l'élément 'Dublin Core dc:subject' ou l'élément DCAT 'dcat:theme', vous pouvez indiquer le sujet en utilisant un terme contrôlé à partir d'un thésaurus ou d'une classification. Vous pouvez ensuite rechercher des vocabulaires de type thésaurus, classification ou titre de rubrique, et peut-être limiter la recherche à un domaine ou un type de données spécifique. Si vous souhaitez utiliser un thésaurus général et bien connu pour l'agriculture, vous pouvez choisir d'utiliser la terminologie d'AGROVOC pour identifier votre sujet, ou si vous partagez des données agronomiques dans une communauté spécialisée, vous pouvez limiter la recherche aux ontologies des plantes ou aux observations ou expériences agronomiques et vous trouverez peut-être que l'Ontologie des expériences agricoles INRA a les termes dont vous avez besoin. Vous pouvez voir comment utiliser ces valeurs dans vos métadonnées dans la section 3.
- Que vous partagiez vos métadonnées d'ensemble de données et vos données dans le même fichier (par exemple, un fichier XML ou RDF) ou séparément (par exemple, un fichier XML pour les métadonnées et un CSV ou NetCDF pour les données), vous pouvez rechercher des sémantiques supplémentaires pour vos données. Par exemple, si votre ensemble est un ensemble de données d'observations agronomiques, vous pourriez vouloir
- Utiliser une norme largement adoptée pour le partage des observations (indépendamment du type d'observation) : vous pouvez rechercher des schémas ou des ontologies décrivant des observations et vous trouverez la norme ISO/OGC Observations & Mesures ('O&M'), formalisée comme schéma XML et schéma JSON.
- Utiliser des noms de variables convenus pour chaque phénomène ou propriété observée : vous pouvez rechercher des vocabulaires de type dictionnaire de données décrivant les observations sur le terrain et vous trouverez par exemple la liste des variables principales de l'AgMIP ICASA.
- Utiliser des valeurs standardisées pour les valeurs de certaines propriétés (comme la forme ou d'autres propriétés phénotypiques d'une plante) : vous pouvez rechercher des listes de codes ou des ontologies de propriétés phénotypiques et vous trouverez l'ontologie de la qualité phénotypique OBO pour les propriétés phénotypiques générales (comme les formes) et plusieurs ontologies spécifiques de culture.

Au-delà du choix d'un vocabulaire adapté aux données que vous gérez, d'autres critères sont utiles pour choisir le vocabulaire le plus approprié:

- La mise en œuvre idéale du " Web sémantique " passe par les espaces de nommage, les URI et l'approche des données liées (voir leçon 4.3) : pour pouvoir exploiter ces technologies, il est préférable de choisir des vocabulaires qui ont été publiés en RDF (XML peut aussi être une bonne option) et de suivre l'approche des données liées (utiliser les URI, lien vers les autres vocabulaires).
- Il est préférable de choisir des vocabulaires qui sont largement utilisés, afin que vos données soient interopérables avec le plus d'ensembles de données et de systèmes possibles.
- Si les vocabulaires existants ne répondent pas à vos besoins, vous pouvez décider de créer votre propre vocabulaire et de le publier (idéalement, en collaboration avec d'autres partenaires dans votre communauté afin d'assurer une large adoption).

Une fois que vous avez identifié les vocabulaires pertinents à votre type de données, voyons comment les utiliser.

3. Utilisation des vocabulaires existants pour vos données agronomiques

Dans la leçon 4.4, nous avons expliqué comment codifier les données en utilisant un schéma XML ou un vocabulaire RDF, et nous avons vu quelques exemples. Dans cette leçon, nous donnerons d'autres exemples potentiellement pertinents dans le domaine de l'agronomie.

Mise en garde importante : nous donnons les exemples ci-dessous en supposant que le gestionnaire de données a la liberté de choisir le format de données de son choix et les vocabulaires de son choix. En général, comme l'implémentation des exemples ci-dessous ou d'exemples similaires n'est pas quelque chose que vous feriez normalement manuellement, il y a deux façons pour un gestionnaire de données d'implémenter la sémantique:

- À l'aide de générateurs ou de convertisseurs d'ensembles de données ou de métadonnées ad hoc, écrits par leurs développeurs et configurables à leur gré;
- l'utilisation d'un outil de gestion d'ensemble de données qui supporte la sémantique: il n'est pas facile de trouver un outil qui vous permette d'exposer des (méta)données en utilisant n'importe quel vocabulaire de votre choix; en particulier pour le modèle de métadonnées, les outils viennent normalement avec leur propre modèle, qui peut (devrait) correspondre à un vocabulaire largement utilisé publié, de sorte que le gestionnaire des données ne peut choisir parmi différents outils que celui qui possède le meilleur modèle de métadonnée; certains outils peuvent autoriser le gestionnaire à utiliser des valeurs de valeurs contrôlées issues de certaines métadonnées spécifiques.

Par conséquent, certains cas, tel qu'illustré ci-dessous (et des exemples similaires) ne peuvent pas être mis en œuvre librement, car il peut y avoir des contraintes techniques dues aux outils utilisés. Cependant, voir comment la sémantique pourrait être idéalement intégrée dans les (méta)données peut aussi aider à choisir les meilleurs outils et augmenter les attentes des gestionnaires de données, les rendant ainsi plus exigeants en termes d'outils et conduisant peut-être à l'amélioration des outils existants.

Dans les deux sections suivantes, nous appliquerons la distinction, comme mentionné dans la leçon 4.4, entre la sémantique pour la structure des données (vocabulaires de description, schémas) et la sémantique pour les valeurs (vocabulaires de valeurs).

3.1. Vocabulaires pour la structure des (méta)données

Comme vocabulaires pour la structure des (méta)données, vous utiliseriez normalement un schéma ou une ontologie. Les schémas et ontologies sont l'encodage d'un modèle de (méta)données, alors que le type d'encodage (schéma XML, schéma RDF, ontologie OWL) rend le vocabulaire particulièrement adapté à un format de (méta)données spécifique (généralement XML ou RDF); les modèles de métadonnées peuvent également être adoptés pour des (méta)données dans différents formats, pour autant que ces formats supportent des schémas ou noms personnalisés. Ainsi, les schémas et ontologies peuvent être utilisés d'une manière ou d'une autre dans JSON et CSV (voir la section sur les formats de données dans la leçon 2.3), comme structure et étiquettes JSON ou comme noms de colonnes CSV, comme nous l'avons vu dans la leçon 4.4.

D'autre part, certaines normes de données sont conçues comme des modèles agnostiques de format et sont ensuite formalisées sous forme de schémas de différents types (schéma XML, ontologie) et des spécifications sont fournies pour les utiliser dans d'autres formats (CSV, JSON).

3.1.1. Sémantique pour les métadonnées des ensembles de données

Comme indiqué dans la section 2, lors de la recherche de vocabulaires pour les métadonnées d'ensembles de données, vous trouverez les vocabulaires DCAT et Data Cube du W3C RDF ainsi que le schéma XML ISO 19115 (Information géographique - Métadonnées¹) Schéma XML (ISO/TS 19139²). En lisant plus loin et en examinant les vocabulaires, vous remarquerez:

¹ <https://www.iso.org/standard/53798.html>

² <http://www.isotc211.org/schemas/2005/>

- DCAT possède des propriétés de base pour décrire l'ensemble de données " ressource " et ses distributions, mais aucune pour décrire la structure des données;
- Data Cube est spécialement conçu pour les ensembles de données statistiques, mais possède des classes et des propriétés qui peuvent décrire la structure des données de tout ensemble de données d'observation (DataStructureDefinition, dimension, mesure) et coder les observations elles-mêmes;
- La norme ISO 19115 contient de nombreuses métadonnées pour décrire l'ensemble de données en tant que ressource et pour décrire la structure des données.

Disons que vous voulez sérialiser vos métadonnées d'ensemble de données en RDF pour une meilleure interopérabilité et que vous voulez exposer toutes les métadonnées sur la structure des données afin que les outils logiciels puissent analyser automatiquement les données. Dans un tel cas, le vocabulaire du Data Cube peut être le bon choix.

Comme nous l'avons brièvement mentionné dans la section précédente, vous pouvez utiliser plus d'un vocabulaire dans le même ensemble de données. Il suffit de déclarer les espaces de noms de tous les vocabulaires utilisés dans l'en-tête et d'utiliser ensuite leurs classes et propriétés dans le fichier. Dans l'exemple ci-dessous, nous sautons l'en-tête avec tous les préfixes et les espaces de noms: il suffit de dire que 'eg:' est juste un préfixe pour l'espace de noms local, 'qb:' est le préfixe pour l'espace de noms du vocabulaire W3C Data Cube, et 'dct:' est un préfixe pour le vocabulaire Dublin Core.

L'exemple ci-dessous décrit une structure de données conçue pour contenir une mesure ("température minimale quotidienne de l'air, moyenne", indiquée par une URI conventionnelle de la liste des variables maîtresses de l'ICASA, non encore publiée) et trois dimensions : zone, période et identifiant du champ où la mesure est effectuée. Cela signifie qu'il s'agira des données de cet ensemble de données et que les noms et attributs utilisés dans les données seront ceux indiqués ici. Voir la sous-section suivante pour un exemple des données de cet ensemble de données sérialisées avec Data Cube.

```

eg:dataset-le3 a qb:DataSet;
  dct:title           "Minimum daily air temperature 2004"@en;
  rdfs:label          "Minimum daily air temperature 2004"@en;
  dct:description     "Minimum daily air temperature 2004 - From fields..."@en;
  dct:publisher       eg:organization ;
  dct:issued          "2010-08-11"^^xsd:date;
  dct:subject         <http://aims.fao.org/aos/agrovoc/c_230> ;
  qb:structure        eg:dsd-le3 ;

eg:organization a foaf:Organization;
  foaf:name           "Test organization";

eg:dsd-le3 a qb:DataStructureDefinition;
  # The dimensions
  qb:component        [ qb:dimension  eg:refArea;          qb:order 1 ];
  qb:component        [ qb:dimension  eg:refPeriod;        qb:order 2 ];
  qb:component        [ qb:dimension  eg:field_id;         qb:order 3 ];
  # The measure(s): "Minimum daily air temperature, average" from ICASA variables
  qb:component        [ qb:measure    <http://purl.org/icasa/variables#tmina>];
  # The attributes
  qb:component        [ qb:attribute  sdmx-attribute:unitMeasure;
                        qb:componentRequired "true"^^xsd:boolean; ].

```

Illustration 1 Exemple de codage RDF de métadonnées d'ensemble de données et de structure de données à l'aide de Data Cube

ISO 19115 est également un vocabulaire approprié couvrant des dizaines d'éléments de métadonnées pour un ensemble de données. Alors que l'ISO/TS 19139 fournit un schéma XML pour l'ISO 19115, une représentation RDF (OWL) de l'ISO 19115 a été développée³ par le CSIRO (Australie).

Vous pouvez codifier les métadonnées d'un ensemble de données dans un fichier distinct de l'ensemble de données réel ou dans le même fichier que les données. Cela dépend de plusieurs facteurs: votre modèle de (méta)données et le vocabulaire que vous choisissez respectivement pour les métadonnées de l'ensemble de données et les données comme tel (certains vocabulaires couvrent les deux choses dans la même structure de données, d'autres non), votre architecture de référentiel/catalogue, la taille des données, le format de vos données (si le format des données ne permet pas les métadonnées d'ensemble, ou ne vous permet pas de créer un fichier séparé pour les métadonnées, vous devrez peut-être en utiliser le vocabulaire de votre choix). Dans la section suivante, nous verrons comment utiliser la sémantique dans la section des données (que ce soit dans le même fichier ou dans un fichier séparé, selon les conditions ci-dessus).

3.1.2. Sémantique pour les données

Commençons par l'exemple d'observations issues d'un ensemble de données résultant d'expériences agricoles. Lors de votre recherche dans la section 2, vous avez trouvé qu'il existe une norme générale de données ISO pour les observations et mesures⁴ (O&M), qui, à partir d'un modèle UML, fournit à la fois un schéma XML et un schéma JSON du modèle.

³ <http://def.seegrid.csiro.au/static/isotc211/iso19115/2003/metadata.html>

⁴ <http://www.opengeospatial.org/standards/om>

L'exemple ci-dessous montre comment utiliser les éléments du schéma XML O&M pour décrire une observation simple : la mesure de la masse du fruit à la température de 22,3 °C.

```
<om:OM_Observation <!-- namespaces hidden -->
<gml:description>Observation test instance: fruit mass</gml:description>
<gml:name>Observation test 1</gml:name>
<om:type xlink:href="http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_Measurement"/>
<om:phenomenonTime> <!-- hidden --> </om:phenomenonTime>
<om:resultTime xlink:href="#ot1t"/>
<om:procedure xlink:href="http://www.example.org/register/process/scales34.xml"/>
<om:parameter>
  <om:NamedValue>
    <om:name xlink:href="http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature"/>
    <om:value xsi:type="gml:MeasureType" uom="Cel">22.3</om:value>
  </om:NamedValue>
</om:parameter>
<om:observedProperty xlink:href="http://sweet.jpl.nasa.gov/2.0/phys.owl#Mass"/>
<om:featureOfInterest xlink:href="http://wfs.example.org?request=getFeature&featureid=fruit37f "/>
<om:result xsi:type="gml:MeasureType" uom="kg">0.28</om:result>
</om:OM_Observation>
```

Illustration 2 Exemple ⁵ d'observation au format XML en utilisant le schéma O&M

L'exemple suivant montre comment utiliser le schéma O&M JSON pour décrire une observation simple : la mesure de la température de l'air à un point spécifique. Les étiquettes utilisées ('observedProperty', 'featureOfInterest'...) et la structure d'imbrication ('uom' sous 'result') montrent clairement que même dans un format différent, le schéma est le même. Puisque dans le fichier XML le préfixe 'om :' serait mappé à l'espace de noms du schéma XML, et dans le fichier JSON le @contexte serait l'URL du schéma JSON, tout logiciel analysant des ensembles de données dans les deux formats interprétera les éléments/labels de la même manière.

```
{
  "id": "measure-instance-test",
  "type": "Measurement",
  "phenomenonTime": { "instant": "2011-05-11T00:00:00+10:00" },
  "observedProperty": { "href": "http://environment.data.gov.au/def/property/air_temperature" },
  "procedure": { "href": "http://www.opengis.net/def/waterml/2.0/processType/Sensor" },
  "featureOfInterest": { "href": "http://waterml2.csiro.au/rgs-api/v1/monitoring-point/419009/" },
  "resultTime": "2011-05-12T09:00:00+10:00",
  "result": {
    "value": 3.2,
    "uom": "http://qudt.org/vocab/unit#DegreeCelsius" }
}
```

Illustration 3 Exemple d'observation au format JSON en utilisant le modèle O&M

Si vous faites davantage de recherches, vous constaterez qu'il existe également une norme de données spécifique aux expériences agricoles: la norme de données ICASA⁶. Un schéma XML pour cette norme est encore en cours de développement, mais la norme a été sérialisée en JSON.

⁵ http://portal.opengeospatial.org/files/?artifact_id=41510

⁶ http://dssat.net/wp-content/uploads/2014/02/White2013ICASA_V2_standards.pdf

L'exemple ci-dessous illustre une expérience codée en JSON à l'aide du modèle de données et des variables ICASA. Vous pouvez voir qu'au lieu d'URI, ICASA utilise des noms de variables codés courts: tous les codes sont dans la liste des variables principales de l'ICASA⁷, qui définit la signification de toutes les variables et constitue la ressource sémantique ICASA, où vous trouverez que 'fielelele' signifie 'altitude du champ' et 'icpccr' signifie 'résidu, crop code for previous crop'.

Puisque les variables sont identifiées par des codes et non par des URI, et que les codes ne sont pas associés à des définitions dans un fichier lisible par machine, les outils logiciels ne peuvent pas rechercher la signification et ne peuvent pas déduire la sémantique de référence derrière le code. Par conséquent, même si les variables ICASA sont probablement la liste la plus complète pour les expériences agricoles et sont utilisées dans d'autres systèmes, leur utilisation actuelle ne garantit pas une interopérabilité sémantique totale. Il y a du travail à effectuer pour exprimer les variables ICASA dans une ontologie⁸.

```
'exname': "UFGA8201MZ 1",
'local_name': "NIT X IRR, GAINESVILLE 2N*3I",
'people': "BENNET,J.M. ZUR,B. HAMMOND,L.C. JONES,J.W.",
'institution': "UNIVERSITY OF FLORIDA, GAINESVILLE, FL, USA",
'site': "IRR.PARK,UF.CAMPUS 29.63;-82.37;40.;FLA",
'tr_name': "RAINFED LOW NITROGEN",
'id_field': "UFGA0002",
'wst_id': "UFGA", /* weather station id */
'soil_id': "IBM2910014", /* soil id */
'fl_lat': "29.63", /* latitude */
'fl_long': "-82.37", /* longitude */
'fiele': "40", /* field elevation */

'initial_condition': [
  "icpccr": "MAZ", /* previous crop was maize */
  "icdat": "19820225", /* date for initial conditions */
  "icrt": "100", /* initial root residue kg/ha */
  "icrag": "1000", /* initial surface residue kg/ha */
  "icrn": ".8", /* initial residue N % */
  "soilLayer": [ /* init soil layer data: */
    /* depth (cm), moisture (frac), NH4 (ppm), NO3 (ppm) */
    {"icb1": "5", "ich2e": ".086", "icnh4": ".5", "icno3": ".1"},
    {"icb1": "15", "ich2e": ".086", "icnh4": ".5", "icno3": ".1"},
    {"icb1": "180", "ich2e": ".258", "icnh4": ".5", "icno3": ".1"}
  ]
]

'management':{
  "events":[
    {"event": "fertilizer", /* Fertilizer application */
      "date": "19820225", /* Feb 25, 1982 */
      "fecd": "FE001", /* Ammonium nitrate */
      "feacd": "AP001", /* Broadcast, incorporated */
      "fedep": "10", /* 10 cm deep */
      "feamn": "27"}, /* 27 kg[N]/ha */
  ]
}
```

Illustration 4 Exemple⁹ d'une expérience décrite au format JSON en utilisant la norme ICASA

Si vous recherchez un vocabulaire RDF pour une interopérabilité complète, alors vous pouvez considérer le 'W3C'Semantic Sensor Network Ontology' (SSN), une ontologie construite sur la base des normes OGC SensorML et O&M. Les classes

⁷ <http://research.agmip.org/display/dev/ICASA+Master+Variable+List>

⁸ <https://github.com/craig-willis/icasa/blob/master/docs/design.md>

⁹ Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards.

http://dssat.net/wpcontent/uploads/2014/02/White2013ICASA_V2_standards.pdf

et propriétés les plus pertinentes pour les observations se trouvent sous l'espace de noms <http://www.w3.org/ns/sosa/> (Sensor, Observation, Sample, and Actuator – SOSA). Vous remarquerez que le vocabulaire suit le modèle OGC Observation- FeatureOfInterest-ObservedProperty. Notez que cette description utilise également d'autres vocabulaires, en particulier @prefix qudt-1-1 : <http://qudt.org/1.1/schema/qudt#>, un vocabulaire pour les unités de mesure.

```
<observation/1087> rdf:type sosa:Observation ;
rdfs:label "observation #1087"@en ;
sosa:hasFeatureOfInterest <tree/124> ;
sosa:observedProperty <tree/124/height> ;
sosa:hasResult [
  qudt-1-1:unit qudt-unit-1-1:Meter ;
  qudt-1-1:numericalValue "15.3"^^xsd:double ] .

<tree/124> rdf:type sosa:FeatureOfInterest ;
rdfs:label "tree #124"@en .

<tree/124#height> rdf:type sosa:ObservableProperty , ssn:Property ;
rdfs:label "the height of tree #124"@en .
```

Illustration 5 Exemple¹⁰ d'observation en RDF Turtle en utilisant l'ontologie SSN du W3C

Faisant toujours partie de la famille des normes de données pour les observations issues de l'O&M, le CSIRO a mis au point une version OWL de l'O&M ("OWL for Observations"¹¹). Cette ontologie est également alignée avec l'ontologie SSN mentionnée ci-dessus.

```
my:obsTest1 a oml:Measurement ;
rdfs:comment "Observation test instance: fruit mass"^^xsd:string ;
rdfs:label "Observation test 1"^^xsd:string ;
oml:featureOfInterest <http://wfs.example.com?request=getFeature&featureid=fruit37f> ;
oml:observedProperty <http://sweet.jpl.nasa.gov/2.0/phys.owl#Mass> ;
oml:phenomenonTime [ a w3time:Instant ;
  w3time:inXSDDateTime "2005-01-11T16:22:25.00"^^xsd:dateTime ] ;
oml:procedure my:Scales1 ;
oml:result [ a oml:Measure ;
  rdf:value "0.28"^^oml:Number ;
  oml:uom <http://www.opengis.net/def/uom/UCUM/0/kg> ] ;
oml:resultTime "2005-01-11T16:22:25.00"^^xsd:dateTime
```

Illustration 6 Exemple¹⁴ d'une observation codée en RDF (Turtle) avec l'ontologie OM Lite

Ou encore, en continuant avec l'exemple de la section précédente, en utilisant le vocabulaire du Data Cube pour les métadonnées de l'ensemble de données, les observations peuvent également être codées à l'aide du Data

¹⁰ Semantic Sensor Network Ontology. <https://www.w3.org/TR/vocab-ssn/#Observationoverview>

¹¹ <http://www.essepuntato.it/lode/owlapi/http://def.seegrid.csiro.au/ontology/om/om-lite>

¹⁴ Cox, Simon J D. Basic Observations and Sampling Feature Ontology. <http://www.semanticweb-journal.net/system/files/swj890.pdf>

Cube, bien que d'autres vocabulaires soient nécessaires dans l'entité Observation.

L'exemple ci-dessous code la mesure de la température minimale journalière (suivant la structure de données définie dans l'exemple d'en-tête du jeu de données Data Cube dans la section précédente). En plus des espaces de noms utilisés dans l'exemple précédent, nous avons ici deux espaces de noms supplémentaires pour les éléments de métadonnées, un pour les attributs statistiques (SDMX) et un pour les variables (les variables ICASA):

```
@prefix sdmx-attribute : <http://purl.org/linked-data/sdmx/2009/attribute#>
```

```
@prefix icasa-var : <http://purl.org/icasa/variables#>
```

```
eg:o1 a qb:Observation;  
  qb:dataSet eg:dataset-le1 ;  
  eg:refArea      ex-geo:newport_00pr ;  
  eg:refPeriod    <http://reference.data.gov.uk/id/gregorian-interval/2004-01-01T00:00:00/P3Y> ;  
  eg:field_id     1234 ;  
  sdmx-attribute:unitMeasure qudt-unit:DegreeCelsius ;  
  # tmina: "Minimum daily air temperature, average" from ICASA variables  
  icasa-var:tmina 1.7 ;
```

Illustration 7 Observation codée en RDF (Turtle) à l'aide du Data Cube

Comme nous l'avons dit dans la leçon 4.4, le niveau d'interopérabilité plus élevé 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 faits: (a) les éléments de métadonnées (classes ou propriétés en RDF) sont identifiés par les URI et ces URI sont déréférencés vers des pages Web qui contiennent des informations lisibles par machine sur la classe/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 il peut continuer à les suivre et obtenir plus et mieux; et (b) comme nous l'avons dit dans la leçon 4.3, un fichier RDF a moins d'ambiguïtés potentielles que des fichiers XML (non-RDF) et est interprété plus sûrement.

L'approche du Web sémantique a été conçue en partant de l'hypothèse que les technologies de données liées sont utilisées, de sorte que l'approche idéale de l'interopérabilité sémantique est l'utilisation de RDF et des caractéristiques essentielles des données liées (URI et liens vers d'autres URI).

3.2. Vocabulaires pour les valeurs de (méta)données

Un cas différent se présente lorsque vous voulez utiliser les valeurs d'un vocabulaire existant comme valeurs de certaines métadonnées, par exemple si vous voulez utiliser un terme sans ambiguïté pour "température de l'air" ou pour identifier une espèce végétale. Dans ce cas, vous ne cherchez pas un vocabulaire (un ou plusieurs) qui possède les classes et propriétés dont vous avez besoin pour décrire/structurer/modéliser vos données; vous cherchez des vocabulaires qui contiennent les termes, les concepts que vous voulez utiliser

pour des valeurs spécifiques, parce que vous voulez les identifier sans équivoque (aucune chaîne arbitraire), avec des termes/codes faisant autorité, indépendants des langues et pratiques locales. Par exemple, vous pouvez utiliser le terme AGROVOC pour " température de l'air " ou l'identificateur d'ontologie végétale pour une espèce végétale.

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 pas 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 connaissent le vocabulaire et peuvent faire correspondre le littéral avec l'URI peuvent déjà en tirer des avantages.

Idéalement, vous devriez utiliser l'URI du terme auquel vous voulez vous référer. Voyons comment les URI des ressources sémantiques publiées sont utilisées comme valeurs dans les exemples fournis dans la section précédente.

Dans l'exemple de métadonnées de l'ensemble de données de Data Cube de l'illustration 4.4.1.1., l'objet de l'ensemble de données est indiqué par l'URI d'un concept dans AGROVOC: un outil lié et sensible aux données consulte l'URI http://aims.fao.org/aos/agrovoc/c_230 et trouverait une description RDF du concept indiquant que l'étiquette en anglais est " air temperature ", qu'il s'agit d'un concept plus étroit de " température " et qu'il existe des concepts identiques dans d'autres vocabulaires, dont le thésaurus USDA NAL et celui du GEMET. L'outil pourrait exploiter cette information pour combiner cet ensemble de données avec d'autres ensembles de données similaires.

De plus, la propriété observée / mesurée est définie par un URI : dans ce cas, l'URI est conventionnel, car il provient de la version expérimentale RDF ICASA Variables Master List, non publiée (l'URI est donc encore non résoluble). Lorsque la liste maîtresse de l'ICASA sera publiée dans RDF, un outil logiciel pourra voir que <http://purl.org/icasa/variables#tming> signifie "température minimale quotidienne de l'air, moyenne, semis à récolter".

L'attribut obligatoire de la mesure, qui est l'unité de mesure, est également indiqué par un URI : le préfixe sdmx-attribute est associé à l'URI <http://purl.org/linked-data/sdmx/2009/attribute#>, donc sdmxattribute:unitMeasure est l'abréviation de <http://purl.org/linked-data/sdmx/2009/attribute#unitMeasure>: en cherchant cette URI, un outil trouverait que le nom du concept en anglais est 'Unit of Measure', qu'il signifie 'The unit in which the data values are measured' et a une définition complète dans un certain fichier PDF.

```

eg:dataset-le3 a qb:DataSet;
dct:title "Minimum daily air temperature 2004"@en;
dct:subject <http://aims.fao.org/aos/agrovoc/c_230> ;
qb:structure eg:dsd-le3 ;

eg:dsd-le3 a qb:DataStructureDefinition;
[...]
qb:component [ qb:measure <http://purl.org/icasa/variables#tmina>;
qb:component [ qb:attribute sdmx-attribute:unitMeasure; ] .

```

Illustration 8 Utilisation de vocabulaires de valeurs de l'exemple du Data Cube de l'illustration 1

Dans l'exemple OM Lite de la figure 6, la caractéristique d'intérêt est indiquée par l'URL d'une API qui devrait renvoyer le RDF d'une " caractéristique " à partir d'un registre de caractéristiques (différentes communautés de recherche créent leurs propres registres de caractéristiques pertinentes ; l'URL utilisée ici est seulement un exemple).

La propriété observée (la masse) est indiquée par une URI du (...) <https://sweet.jpl.nasa.gov/2.0/phys.owl#Mass>: en cherchant cet URI, une machine trouverait que pour cette propriété, l'unité de mesure par défaut est le kilogramme (indiqué à son tour par un un autre URI). L'unité de mesure est à nouveau spécifiée dans la propriété 'uom', cette fois en utilisant une URI de l'Open Geospatial Consortium (qui n'est pas entièrement résoluble) ; l'URI du service japonais de la NASA utilisé ci-dessus aurait également bien fonctionné (<https://sweet.jpl.nasa.gov/2.0/sciUnits.owl#kilogram>) ; encore mieux l'URI du kilogramme de l'ontologie QUDT (...) : <http://qudt.org/vocab/unit#Kilogram>.

La propriété observée 'masse de fruits' aurait également pu être indiquée par l'URI pour 'Poids frais de fruits' à partir des variables ICASA (qui, espérons-le, sera bientôt publiée): <http://purl.org/icasa/variables#ffad>.

Pour une interopérabilité accrue, tous ces URI auraient pu être utilisés. Idéalement, tous les différents URI dans différentes ressources sémantiques se référant au même concept seraient liés les uns aux autres par une propriété owl:sameAs, de sorte que dans vos données vous n'avez besoin que d'un seul URI et les outils sémantiques pourraient rechercher tous les liés. Cependant, il n'existe pas encore de bons mappages fiables, donc plus vous pouvez établir des liens avec des URI externes, mieux c'est.

```

my:obsTest1 a oml:Measurement ;
rdfs:comment "Observation test instance: fruit mass"^^xsd:string ;
rdfs:label "Observation test 1"^^xsd:string ;
oml:featureOfInterest <http://wfs.example.org?request=getFeature&featureid=fruit37f> ;
oml:observedProperty <http://sweet.jpl.nasa.gov/2.0/phys.owl#Mass> ;
oml:procedure my:Scales1 ;
oml:result [ a oml:Measure ;
rdf:value "0.28"^^oml:Number ;
oml:uom <http://www.opengis.net/def/uom/UCUM/0/kg> ] ;

```

Illustration 9 A partir de l'exemple de l'illustration 6

Les exemples de cette leçon, aussi spécifiques soient-ils aux observations sur le terrain, peuvent être facilement reproduits pour d'autres types de données, en utilisant des vocabulaires qui suivent un modèle de données adapté à ce type de données. Les exemples ont pour but de vous aider à comprendre les mécanismes d'ajout à vos données, d'une sémantique partagée ou convenue.

Lecture complémentaire

- Antoniou, G., and van Harmelen, F. (2003). *A Semantic Web Primer*. MIT Press, Cambridge, MA, USA. <http://www.csd.uoc.gr/~hy566/SWbook.pdf>
- White, J. W. *et al.* (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and Electronics in Agriculture* **96**, 1–12.
<https://doi.org/10.1016/j.compag.2013.04.003>