

Manejo de Datos Abiertos en la Agricultura y Nutrición

Este curso de aprendizaje digital (e-learning) es el resultado de una colaboración entre socios de GODAN Action, incluyendo a **Investigaciones Ambientales Wageininen (WUR)**, **AgroKnow**, **AidData**, la **Organización de las Naciones Unidas para la Alimentación y la Agricultura** (FAO por sus siglas en Ingles), **El Foro Global sobre Investigaciones de Agricultura** (GFAR), y el **Instituto de los Estudios del Desarrollo** (IDS), **The Land Portal**, **el Instituto de Datos Abiertos** (IDI) y el **Centro Técnico de Agricultura y cooperación Rural** (CTA).

GODAN Action es un proyecto de tres años [por] el Departamento del Desarrollo Internacional del Reino Unido para capacitar a los que usan, producen, e intermediarios de datos para conectarse efectivamente con datos abiertos y maximizar la potencial por su impacto en los sectores de agricultura y nutrición. En particular, trabajamos para mejorar la capacitación, promover estándares comunes y mejores prácticas para medir el impacto. [www.godan.info]

Este trabajo está registrado con una licencia [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)



Unidad 4: Compartir datos abiertos

Lección 4.4.1 Uso de semántica publicada para datos agronomicas



Foto de Michael (Mikey) Cantor con licencia CC BY SA 2.0

Objetivos y resultados del aprendizaje

Esta lección tiene como objetivo proporcionar orientación y ejemplos sobre cómo codificar (meta) datos utilizando semántica publicada, con ejemplos específicos para datos después de estudiar esta lección, debería poder;

- Comprender cómo identificar estándares de datos adecuados para compartir datos agronómicos
- comprender cómo se integra la semántica en los (meta) datos
- (guiar a los desarrolladores o elegir herramientas que) adopten los vocabularios más adecuados para datos de un tipo específico

Contenidos

1. Introducción	5
2. Identificar los vocabularios más adecuados	5
3. Usar vocabularios existentes para los	6
3.1. Vocabularios para los (meta) datos estructura	7
3.1.1. Semántica para metadatos de conjuntos de datos	7
3.1.2. Semántica de datos	9

Lista de Figuras

Ilustración 1. Ejemplo de codificación RDF de metadatos y estructura de datos del dataset utilizando.....	8
Ilustración 2. Ejemplo de observación en formato XML utilizando el esquema O&M.....	9
Ilustración 3. Ejemplo de observación en formato JSON utilizando el modelo O&M ...	10
Ilustración 4. Ejemplo de un experimento descrito en formato JSON utilizando el estándar ICASA.....	11
Ilustración 5. Ejemplo de observación en RDF Turtle usando la ontología W3C SSN.	11
Ilustración 6. Ejemplo de una observación codificada en RDF (Turtle) con la ontología OM Lite.....	12

1. Introducción

En la lección 4.4, proporcionamos una descripción general de la interoperabilidad semántica. En esta lección, proporcionaremos ejemplos más precisos de cómo implementarlo, especialmente identificar y reutilizar la semántica publicada para datos agronómicos.

Los ejemplos prácticos que usaremos a lo largo de esta lección están relacionados con observaciones y experimentos agronómicos. También proporcionaremos ejemplos de cómo agregar semántica a los metadatos del dataset.

2. Identificar los vocabularios más adecuados

Ya vimos en la lección 4.4 algunos consejos útiles para seleccionar los vocabularios. Ahora veamos algunos ejemplos prácticos, para el caso de conjuntos de datos de observaciones / experimentos agronómicos (se pueden buscar todos los vocabularios mencionado en esta lección en el mapa GODAN ya mencionado en <http://vest.agrisemantics.org>).

- En general, se necesitará un vocabulario que ayude a describir el dataset que contiene los datos, para facilitar el descubrimiento del dataset. En este caso, debe buscar vocabularios que describan tan precisamente como sea posible un dataset o su tipo específico de dataset. Por ejemplo, en Agrisemantics Map of standards, se pueden buscar vocabularios que describen 'metadatos de recursos de información' (vocabularios genéricos para cualquier tipo de recurso) y se encontrarían vocabularios como Dublin Core o schema.org, que es entendido por muchas herramientas y puede codificar metadatos como título, autor, tema, formato, metadatos espaciales y temporales, etc.
- Luego, se puede limitar su búsqueda a estándares de datos que describan un 'Dataset'. Se encontrarán vocabularios como el W3C Data Catalog y el W3C Data Cube, así como formatos de datos como NetCDF y HDF5. La elección entre estos estándares de datos depende de cómo desee exponer los datos: leyendo las descripciones, se descubrirá que NetCDF y HDF5 son estructuras de datos sin semántica y no extensible con esquemas, por lo que si los usamos, no se puede usar ninguna semántica adicional publicada para enriquecer los datos. Por otro lado, son especialmente adecuados para grandes cantidades de observaciones, por lo que pueden considerarse en el caso de las observaciones agronómicas. Incluso puede decidirse exponer los datos utilizando más de un dato estándar: ver sección 3.
- Los vocabularios anteriores proporcionan los metadatos para describir el dataset, pero es posible que se desee identificar vocabularios de valor para usar como vocabularios controlados para algunos de los metadatos de los conjuntos de datos. Por ejemplo, para el elemento Dublin Core dc: subject o el elemento DCAT dcat: theme, es posible que se desee indicar el tema utilizando un término controlado de un diccionario de sinónimos o de una clasificación. A continuación, se pueden buscar vocabularios de tipo tesoro o clasificación o encabezamiento de materia, y quizás se reduzca la búsqueda a un dominio o tipo de datos específicos. Si se desea utilizar un diccionario de sinónimos general y conocido para la agricultura, se puede optar por utilizar términos de AGROVOC para identificar el tema, o si se comparte datos agronómicos en una comunidad especializada que puede reducir la búsqueda de ontologías para plantas o para observaciones agronómicas o experimentos, se puede encontrar que los INRA Agricultural Experiments Ontology tienen los términos que se necesitan. Puedes ver cómo usar estos valores en sus metadatos en la sección 3.

- Si se están compartiendo los metadatos de su dataset y sus datos en el mismo archivo (por ejemplo, un archivo XML o RDF) o por separado (por ejemplo, un archivo XML para los metadatos y un CSV o NetCDF para los datos), es posible que se desee buscar semántica adicional para los datos. Por ejemplo, si el dataset es un dataset de observaciones agronómicas, es posible que desee:
 - Utilizar un estándar ampliamente adoptado para compartir observaciones (independiente del tipo de observación): puede buscar esquemas u ontologías que describan observaciones y encontrar el estándar ISO / OGC Observations & Measurements (O&M), formalizado como esquema XML y esquema JSON.
 - Utilizar nombres de variables acordados para cada fenómeno o propiedad observada: se puede buscar vocabularios de tipo diccionario de datos describiendo observaciones de campo y se encontraría, por ejemplo, el AgMIP ICASA Master Variable List.
 - Utilizar valores estandarizados para los valores de algunas propiedades (como forma u otras propiedades fenotípicas de una planta): se puede buscar listas de códigos u ontologías de propiedades fenotípicas y se encontraría la OBO Phenotypic Quality Ontology para propiedades fenotípicas generales (como formas) y ontologías.

Además de elegir un vocabulario adecuado para los datos que gestiona, otros criterios útiles para seleccionar el vocabulario más apropiado son:

- La implementación ideal de la 'Web semántica'¹ es a través de namespaces, URI y el enfoque Linked Data² (consultar la lección 4.3): para poder explotar estas tecnologías, es preferible elegir vocabularios que se han publicado como RDF (XML puede ser una buena opción también) y seguir el enfoque de datos vinculados (usar URI, enlace a otros vocabularios).
- Es preferible elegir vocabularios que se utilicen ampliamente, de modo que sus datos son interoperables con más conjuntos de datos y más sistemas.
- Si los vocabularios existentes no satisfacen las necesidades, se puede decidir crear un vocabulario propio y publicarlo (idealmente, en colaboración con otros socios de la comunidad para asegurar una amplia adopción).
- Una vez identificados los vocabularios que son relevantes para el tipo de datos en cuestión, veamos cómo usarlos.

3. Usar vocabularios existentes para los datos agronómicos

En la lección 4.4 se explicaba cómo codificar datos usando un esquema XML o un Vocabulario RDF, y se vieron algunos ejemplos. En esta lección se proporciona más ejemplos potencialmente relevantes para el campo de la agronomía.

Una advertencia importante: a continuación, se proporcionan ejemplos asumiendo que el administrador de datos tiene la libertad de elegir el formato de datos de su elección y los vocabularios de su elección. En general, la implementación de los ejemplos siguientes o similares no es algo que normalmente se haría manualmente, sino que hay dos formas en las que un administrador de datos puede implementar la semántica:

- utilizando generadores o convertidores de metadatos / conjuntos de datos ad-hoc, escritos por sus desarrolladores y configurables a su antojo;

¹ W3C. Web semántica. <https://www.w3.org/standards/semanticweb/>

² W3C. Datos vinculados. <https://www.w3.org/standards/semanticweb/data>

- usar una herramienta de administración de conjuntos de datos que admita semántica: no es fácil encontrar una herramienta que permita exponer (meta) datos usando cualquier vocabulario elegido; especialmente para el modelo de metadatos, las herramientas normalmente vienen con su propio modelo, que puede (debería) corresponder a un vocabulario publicado y ampliamente utilizado, por lo que el administrador de datos solo puede elegir entre diferentes herramientas la que tenga los mejores modelos de metadatos; Algunas herramientas pueden permitir que el administrador de datos utilice valores controlados a partir de vocabularios de valor para metadatos específicos.

Por lo tanto, en algunos casos, los ejemplos siguientes (y otros similares) no se pueden implementar libremente, ya que puede haber limitaciones técnicas debido a las herramientas que se usen. Sin embargo, viendo cómo la semántica podría integrarse idealmente en los (meta) datos, también pueden ayudar a elegir las mejores herramientas y aumentarán las expectativas de los administradores de datos, lo que los hace más exigentes en términos de herramientas y quizá los conduzca a la mejora de las herramientas existentes.

En las siguientes dos secciones aplicaremos la distinción, como se menciona en la lección 4.4, entre semánticas para la estructura de datos (vocabularios de descripción, esquemas) y semántica para los valores (vocabularios de valores).

3.1. Vocabularios para los (meta) datos estructura

Como vocabularios para la estructura de (meta) datos, normalmente se usaría un esquema o una ontología. Los esquemas y ontologías son la codificación de modelos de (meta) datos, por lo que si bien el tipo de codificación (esquema XML, esquema RDF, ontología OWL) hace que el vocabulario sea especialmente adecuado para un formato de (meta) dato específico (típicamente XML o RDF), como modelos de metadatos también se pueden adoptar para (meta) datos en diferentes formatos, siempre que los formatos admitan esquemas o nombres de elementos personalizados. Entonces, los esquemas y ontologías se pueden usar de alguna manera en JSON y CSV (consultar la sección sobre formatos de datos en la lección 2.3), como estructura y etiquetas JSON o como etiquetas de columna CSV, como vimos en la lección 4.4.

Por otro lado, algunos estándares de datos están diseñados como modelos independientes del formato y luego se formalizan como esquemas de diferentes tipos (esquema XML, ontología) y se proporcionan especificaciones para utilizarlas en otros formatos (CSV, JSON).

3.1.1. Semántica para metadatos de conjuntos de datos

Como se indica en la sección 2, al buscar vocabularios para metadatos de conjuntos de datos, se encontrarían los vocabularios W3C RDF DCAT y Data Cube, así como el ISO 19115 (Información geográfica - Metadatos³) Esquema XML (ISO / TS 19139 XML Esquema⁴). Al leer y examinar los vocabularios, se notará que:

- DCAT tiene propiedades básicas para describir el 'recurso' del dataset y sus distribuciones, pero sin propiedades para describir la estructura de datos;

³ <https://www.iso.org/standard/53798.html>

⁴ <http://www.iso.org/iso/standards/2013/>

- Data Cube está especialmente concebido para conjuntos de datos estadísticos, pero tiene clases y propiedades que pueden describir la estructura de datos de cualquier dataset de observación (DataStructureDefinition, dimensión, medida) y codificar las propias observaciones;
- ISO 19115 tiene metadatos extensos, tanto para describir el dataset como un recurso y para describir la estructura de datos.

Digamos que se desea serializar los metadatos de un dataset en RDF para mejorar interoperabilidad y que se desea exponer metadatos completos sobre la estructura de los datos para que las herramientas de software también puedan analizar automáticamente los datos. En este caso, el vocabulario de Data Cube puede ser la elección adecuada.

Como mencionamos brevemente en la sección anterior, puede usar más de un vocabulario en el mismo dataset. Es suficiente declarar los espacios de nombres de todos los vocabularios usados en el encabezado y luego usar sus clases y propiedades en el archivo. En el siguiente ejemplo, omitimos el encabezado con todos los prefijos y espacios de nombres: basta con decir que 'eg:' es solo un prefijo para el espacio de nombres locales, 'qb:' es el prefijo para el espacio de nombres del vocabulario W3C Data Cube y 'dct:' es un prefijo del vocabulario Dublin Core.

El siguiente ejemplo describe una estructura de datos diseñada para contener una medición ('temperatura mínima diaria del aire, promedio', indicada con un URI convencional de la Lista maestra de variables de ICASA Master Variables List, aún no publicada) y tres dimensiones: área, período e identificador del campo donde se toma la medida. Esto significa que estos serán los datos de ese dataset y los nombres y atributos utilizados en los datos serán los indicados aquí. Consultar la siguiente subsección para ver un ejemplo de los datos en este dataset serializados con Data Cube también.

```
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; ] .
```

Ilustración 1: Ejemplo de codificación RDF de metadatos y estructura de datos del dataset utilizando

ISO 19115 también es un vocabulario adecuado que cubre docenas de elementos de metadatos para un dataset. Mientras que ISO/TS 19139 proporciona un esquema XML para ISO 19115, CSIRO Australia ha desarrollado una representación RDF (OWL) de ISO 19115⁵.

Se pueden codificar los metadatos del dataset en un archivo separado del dataset real o en el mismo archivo que los datos. Esto depende de varios factores: su (meta) modelo de datos y el vocabulario que se elija respectivamente para el metadatos del dataset y los datos (algunos vocabularios cubren ambas cosas en la misma estructura de datos, muchos otros no), la arquitectura del repositorio/catálogo, el tamaño de los datos, el formato de sus datos (si el formato de datos no permite metadatos de conjuntos de datos o no permite el uso de metadatos el vocabulario de su elección, es posible que se deba crear un archivo separado para metadatos). En la siguiente sección veremos cómo usar la semántica en los datos (ya sea en el mismo archivo o en uno separado, dependiendo de las condiciones anteriores).

3.1.2. Semántica de datos

Comencemos con el ejemplo de un dataset de observaciones de experimentos de agricultura. De su búsqueda en la sección 2, se encontró que hay una Norma de datos ISO para Observaciones y Medidas⁶ (O&M) que, a partir de un modelo UML, proporciona un esquema XML y un esquema JSON del modelo. El siguiente ejemplo muestra cómo utilizar elementos del esquema XML O&M para describir una simple observación: la medición de la masa de la fruta en la temperatura 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>
```

Ilustración 2: Ejemplo de observación en formato XML utilizando el esquema O&M7

El siguiente ejemplo muestra cómo utilizar el esquema JSON O&M para describir una simple observación: la medición de la temperatura del aire en un punto específico. Las etiquetas utilizadas ('ObservedProperty', 'featureOfInterest' ...) y la estructura de anidamiento ('uom' debajo de 'resultado') muestran claramente que incluso si está en un formato diferente, el esquema es el mismo. Dado que en el archivo XML se mapearía el prefijo 'om:' al espacio de nombres del esquema

⁵ <http://def.seoconz.com/def/iso/211/iso19115/2003/metadata.html>

⁶ <https://www.ogc.org/standards/om> <http://www.opengeospatial.org/standards/om>

⁷ http://portal.opengeospatial.org/files/?artifact_id=41510

XML, y en el archivo JSON el @context sería la URL del esquema JSON, cualquier software que analice conjuntos de datos en cualquiera de estos formatos interpretará los elementos/etiquetas de la misma manera.

```
{
  "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" }
}
```

Ilustración 3: Ejemplo de observación en formato JSON utilizando el modelo O&M

Si se investiga más, se encontrará que también hay un estándar de datos específico para experimentos agrícolas: el estándar de datos ICASA⁸. Aún está en desarrollo un esquema XML para este estándar, pero el estándar se ha serializado en JSON.

El siguiente ejemplo muestra un experimento codificado en JSON usando el modelo de datos y variables ICASA. Puede verse que, en lugar de URI, ICASA usa nombres cortos de variables codificadas: todos los códigos están en la lista de Variables Maestras ICASA⁹, que define el significado de todas las variables y constituye el recurso semántico de ICASA, donde se encontrará que 'fielele' significa 'elevación de campo' y 'icpcr' significa 'residuo, código de cultivo del cultivo anterior'.

Dado que las variables se identifican mediante códigos y no mediante URI, y que los códigos no están incluso asociados con definiciones en un archivo legible por máquina, las herramientas de software no pueden buscar el significado y no pueden inferir la semántica de referencia que hay detrás el código. Por lo tanto, incluso si las variables ICASA son probablemente la lista más completa para experimentos agrícolas y se utilizan en otros sistemas, en este momento, su uso no garantiza una interoperabilidad semántica completa. Falta por hacer el trabajo de expresar las variables ICASA⁸ en una ontología^{9,10}.

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

⁹ http://dssat.net/wp-content/uploads/2014/02/White2013ICASA_V2_standards.pdf

```

'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",
'wat_id': "UFGA", /* weather station id */
'soil_id': "IBM2910014", /* soil id */
'fl_lat': "29.63", /* latitude */
'fl_long': "-82.37", /* longitude */
'fielev': "40", /* field elevation */

'initial_condition': {
  'icpcr': "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) */
    {'icbl': "5", "ich2o": ".086", "icnh4": ".5", "icno3": ".1"},
    {'icbl': "15", "ich2o": ".086", "icnh4": ".5", "icno3": ".1"},
    {'icbl': "180", "ich2o": ".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 */
    }
  ]
}

```

Ilustración 4: Ejemplo de un experimento descrito en formato JSON utilizando el estándar ICASA¹⁰

Si se está buscando un vocabulario RDF para una interoperabilidad completa, es posible que consideremos la W3C 'Semantic Sensor Network Ontology' (SSN), del W3C, una ontología construida sobre la base de los estándares OGC SensorML y O&M. Las clases y propiedades que son más relevantes para las observaciones están bajo el espacio de nombres <http://www.w3.org/ns/sosa/> (Sensor, observación, muestra y actuador - SOSA). Nótese que el vocabulario sigue el modelo OGC Observation-FeatureOfInterest-ObservedProperty. Hay que tener en cuenta que esta descripción también usa otros vocabularios, en especial @prefix qudt-1-1: <http://qudt.org/1.1/schema/qudt#>, un vocabulario para unidades de la medida.

```

<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 .

```

Ilustración 5. Ejemplo de observación en RDF Turtle usando la ontología W3C SSN¹¹

¹⁰ <https://vest.agrisemantics.org/content/agmip-icasa-master-variable-list>
<http://research.agmip.org/display/dev/ICASA+Master+Variable+List>

¹¹ Descripción integrada de la producción y los experimentos agrícolas de campo: Los estándares de datos ICASA Versión 2.0. http://dssat.net/wp-content/uploads/2014/02/White2013ICASA_V2_standards.pdf

Siendo parte de la familia de estándares de datos para observaciones derivadas de O&M, CSIRO desarrolló una versión OWL de O&M ('OWL for Observations'¹²). Esta ontología también está alineada con la ontología SSN mencionada anteriormente.

```
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: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
```

Ilustración 6. Ejemplo¹³ de una observación codificada en RDF (Turtle) con la ontología OM Lite

O, continuando con el ejemplo de la sección anterior usando el vocabulario Data Cube para los metadatos del dataset, las observaciones también se pueden codificar usando Data Cube, aunque dentro de la entidad Observation se necesitan otros vocabularios. El siguiente ejemplo codifica la medición de la temperatura mínima diaria (siguiendo la estructura de datos definida en el dataset ejemplo de encabezado en Data Cube de la sección anterior).

Además de los espacios de nombres utilizados en el ejemplo anterior, aquí tenemos dos espacios de nombres adicionales para los elementos de metadatos, uno para los atributos estadísticos (SDMX) y otro para las variables (las variables ICASA):

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

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

¹² <http://www.essepodato.it/code/owlsp/> <http://def.essepodato.it/ontology/om/om.ttl>

¹³ Cox, Simon J D. Basic Observations and Sampling Feature Ontology. <http://www.semantic-web-journal.net/system/files/swj890.pdf>

```

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 ;

```

Figura 7: Observación codificada en RDF (Turtle) usando Data Cube

Como dijimos en la lección 4.4, el mayor nivel de interoperabilidad de usar un vocabulario RDF (y, por lo tanto, un dataset RDF), especialmente si sigue el enlace de patrón de datos (ver lección 4.2) se debe a dos hechos: (a) elementos de metadatos (clases o propiedades en RDF) se identifican mediante URI y esos URI son desreferenciados a páginas web que contienen información legible por máquina sobre la clase/propiedad, para que un programa de computadora pueda seguir el enlace y leer más datos, y si hay otros enlaces a otras entidades externas, puede continuar siguiéndolos, obteniendo cada vez más datos; y (b) como dijimos en la lección 4.3, un archivo RDF tiene menos ambigüedades potenciales que un archivo XML (no RDF) y se interpreta de forma más fiable.

El enfoque de la Web Semántica se ha diseñado partiendo del supuesto de que se utilizan tecnologías de datos enlazados, por lo que el enfoque ideal para la interoperabilidad semántica es utilizar RDF y las características principales de Linked Data (URI y enlaces a otros URI).

3.2. Vocabularios para valores de (meta) datos

Ocurre un caso diferente cuando se desea utilizar valores de un vocabulario como **valores de algunos de los metadatos**, por ejemplo, si desea utilizar un término inequívoco para "temperatura del aire" o para identificar una especie de planta. En este caso, no se está buscando un vocabulario (uno o más) que tenga la clases y propiedades que necesita para describir / estructurar / modelar sus datos; se está buscando vocabularios que contengan los términos, los conceptos que se desea utilizar para valores específicos, porque queremos **identificarlos unívocamente y sin ambigüedades** (sin cadenas arbitrarias), con términos / códigos autorizados, independiente de los idiomas y las prácticas locales. Por ejemplo, es posible que queramos utilizar el término AGROVOC para 'temperatura del aire' o el identificador de ontología vegetal para una especie vegetal.

Como mínimo, una vez identificado algún vocabulario adecuado, si el vocabulario no usa URI y/o los URI no se pueden usar en el dataset, al menos los valores literales de los términos se pueden usar en el dataset: los sistemas que conocen el vocabulario y pueden comparar el literal con el URI ya pueden hacer algo con esto.

Idealmente, se debería utilizar el URI del término al que se desea hacer referencia. Veamos como los URI de recursos semánticos publicados se utilizan como valores en los ejemplos que se proporcionaron en la sección anterior.

En el ejemplo de metadatos del dataset de Data Cube en la figura 4.4.1.1, el tema del dataset se indica utilizando el URI de un concepto en AGROVOC: una herramienta de linked data buscaría el URI http://aims.fao.org/aos/agrovoc/c_230 y encontraría una descripción RDF del concepto que dice que la etiqueta en inglés es 'temperatura del aire', que es un concepto más específico de

'temperatura' y que hay conceptos idénticos en otros vocabularios, entre los que se encuentran el tesoro USDA NAL y el tesoro GEMET. La herramienta podría aprovechar esta información para combinar este dataset con otros datasets similares.

Además, la propiedad que se está observando / midiendo se define con un URI: en este caso el URI es convencional, porque proviene de la versión experimental RDF de ICASA Variables Master List, que no ha sido publicada (por lo tanto la URI no está resuelta todavía). Cuando la lista maestra de ICASA se publique en RDF, una herramienta de software podrá ver que <http://purl.org/icasa/variables#tmina> significa "Temperatura mínima diaria del aire, promedio, desde la siembra hasta la cosecha".

El atributo obligatorio de la medida, que es la unidad de medida, también se indica con un URI: el prefijo `sdmx-attribute` está asociado al URI <http://purl.org/linked-data/sdmx/2009/attribute#>, por lo que `sdmxattribute:unitMeasure` es la abreviatura de <http://purl.org/linked-data/sdmx/2009/attribute#unitMeasure>: al buscar ese URI, una herramienta encontraría que el nombre del concepto en inglés es 'Unit of Measure', que significa 'La unidad en que se miden los valores de los datos' y que tiene una definición completa en un determinado archivo 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; ] .
```

Figura 8: Uso de vocabularios para valores en el ejemplo de Data Cube en la figura 1

En el ejemplo de OM Lite en la figura 6, la característica de interés está indicada por el URL de una API que debería devolver el RDF de una 'Característica' de un registro de características (diferentes comunidades de investigación están creando sus propios registros de características; la URL utilizada aquí es solo un ejemplo).

La propiedad observada (la masa) está indicada por un URI de (...) <https://sweet.jpl.nasa.gov/2.0/phys.owl#Mass> : buscando este URI, una máquina encontraría que para esta propiedad la unidad de medida predeterminada es kilogramo (indicado a su vez por otro URI). La unidad de medida se especifica de nuevo en la propiedad `uom`, esta vez usando un URI del Open Geospatial Consortium (que no se puede resolver por completo); el URI del servicio de la NASA/Japón utilizado anteriormente trabajaría también bien (<https://sweet.jpl.nasa.gov/2.0/sciUnits.owl#kilogram>); aún mejor el URI para kilogramo de la ontología QUDT (...): <http://qudt.org/vocab/unit#Kilogram>.

La propiedad observada 'masa de fruta' también podría haberse indicado con el URI 'Peso fresco de la fruta' de las variables de ICASA (que se espera que se publique pronto): <http://purl.org/icasa/variables#ffad>.

Para una mayor interoperabilidad, se podrían haber utilizado todos estos URI. Idealmente, todos los URI diferentes en diferentes recursos semánticos que se refieren al mismo concepto estarían vinculados entre sí a través de una propiedad owl: sameAs, de modo que nuestros datos que solo necesitamos proporcionar un URI y las herramientas semánticas pueden buscar todos los vinculados. Sin embargo, todavía no existen buenas asignaciones confiables, por lo que cuanto más URI externos se puedan vincular, mejor.

```
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> ] ;
```

Figura 9: Del ejemplo de la figura 6

Los ejemplos de esta lección, a pesar de ser específicos para las observaciones de un campo, se puede replicar fácilmente para otros tipos de datos, utilizando vocabularios que siguen un modelo de datos adecuado para ese tipo de datos. Los ejemplos están destinados a ayudar a comprender los mecanismos para agregar semántica compartida/acordada a nuestros datos.

Lecturas adicionales

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>