Linked Data for Digital Twins

Jan Top, Hannelore Heuer, December 14, 2022
Digital twin

- Digital model of something, for example a human being
- Can tell us something about this person that we do not see immediately
- Can relate to scientific knowledge
- Is used for forecasting, visualisation, diagnosis, advice, control
Data for digital twins

- For creating the twin
  - Learn from response of the real system to external stimuli
  - Learning can happen in advance or in real time
  - Example data: food intake, triglyceride and sugar levels, ...
  - Twins also use expert knowledge!

- For feeding the twin
  - Adapt twin model or parameter values in the model
  - Monitor current status of the subject and its environment
  - Example data: heart beat, blood pressure, weight, ...
  - The twin cannot (and should not try to) predict everything
Digital twin for nutritional advice

What could a digital twin of your field do for you?
Which data or knowledge would it need?
Data sharing and reuse – a complex process

- Internal and external resources
- Legacy systems
- Data quality (fit-for-use) never perfect
  - Different objectives
  - Different versions
  - ...

Manual selection and combination: laborious, costly and error prone
The ideal picture

- Direct access to primary sources
- Get it right at the source (smart devices)
- Access to what is needed for a specific task and user
- Machine readable – smart applications
How to share information?

- Know that a resource exists and where to find it
- Being able to access it and read it
- Ensure that machines can read it such that they show correct behaviour
- Ensure that it can be used in multiple contexts
The FAIR principles

- Enable sharing information on the web
- Raw data, processed data, algorithms, software, documents, ...
- FAIR is not equal to ‘open’
- FAIR can contribute to ‘fair’

But also: link to other datasets
Example: linking food product data
Linking data points as triples

- Data are measurements, observations, facts, ..., that can be structured as **triples**
- Entities and properties require **global identifiers**
- Numbers, strings remain anonymous
Play-a-LOD
Play-a-LOD
Play-a-LOD

- Take nine cards each from the stack
- Build as many triples as you can, extending the current graph
- Take new cards, until nine
Data without meaning is meaningless

How has it been created?

What is it about?

Which quantities and units?

What are limitations of its use?

What other data is needed?
Metadata

- Explanation: for proper use
  - Observed properties, quantities, units
  - Which objects, events or materials are being observed
  - External conditions, circumstances

- Provenance: for reliability, reproduction, tracing
  - ‘Library’ metadata: creator, date, institute, ...
  - Acquisition method, experiment, production method, ...
### Metadata – machine readable

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Food Nutritional Values</td>
</tr>
<tr>
<td>License</td>
<td><a href="http://purl.org/NET/rdflicense/cc-by1.0">http://purl.org/NET/rdflicense/cc-by1.0</a></td>
</tr>
<tr>
<td>Has version</td>
<td>1.2</td>
</tr>
<tr>
<td>Access URL</td>
<td><a href="http://wur.nl">http://wur.nl</a></td>
</tr>
<tr>
<td>Download URL</td>
<td></td>
</tr>
<tr>
<td>Media type</td>
<td>Excel</td>
</tr>
<tr>
<td>Part of dataset</td>
<td></td>
</tr>
</tbody>
</table>

**RDF preview**

```rdfs
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dcat: <http://purl.org/dc/terms/>.
@prefix dcat: <http://www.w3.org/ns/dcat#>.
@prefix fdp: <http://rdf.biosemantics.org/ontologies/fdp-os#>.
@prefix datacite: <http://purl.org/spar/datacite/>.

<> rdf:type dcat:Distribution;
    dcat:title "Food Nutritional Values";
    dcat:license <http://purl.org/NET/rdflicense/cc-by1.0>;
    dcat:hasVersion "1.2";
    dcat:accessURL <http://wur.nl>;
    dcat:mediaType "Excel".
```
Your metadata

How do you currently describe your data?
Metadata: use a shared vocabulary or ontology

http://www.foodvoc.org/resource/FoodTaxonomy/BlueCheese


<table>
<thead>
<tr>
<th>NDB_No</th>
<th>Shrt_Desc</th>
<th>Water (g)</th>
<th>Energ_Kcal</th>
<th>Protein (g)</th>
<th>Lipid_Tot (g)</th>
<th>Ash (g)</th>
<th>Carbohydr (g)</th>
<th>Fiber_TD (g)</th>
<th>Sugar_Tot (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01001</td>
<td>BUTTER, WITH SALT</td>
<td>15.87</td>
<td>717</td>
<td>0.85</td>
<td>81.11</td>
<td>2.11</td>
<td>0</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>01002</td>
<td>BUTTER, WHIPPED, WITH SALT</td>
<td>15.87</td>
<td>717</td>
<td>0.85</td>
<td>81.11</td>
<td>2.11</td>
<td>0</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>01003</td>
<td>BUTTER, CHEESE ANONYMOUS</td>
<td>0.24</td>
<td>876</td>
<td>0.28</td>
<td>99.48</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01004</td>
<td>CHEESE, BLUE</td>
<td>42.41</td>
<td>353</td>
<td>21.4</td>
<td>28.74</td>
<td>5.11</td>
<td>2.34</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>01005</td>
<td>CHEESE, BRICK</td>
<td>41.11</td>
<td>371</td>
<td>23.24</td>
<td>29.68</td>
<td>3.18</td>
<td>2.79</td>
<td>0</td>
<td>0.51</td>
</tr>
<tr>
<td>01006</td>
<td>CHEESE, BRIE</td>
<td>48.42</td>
<td>334</td>
<td>20.75</td>
<td>27.68</td>
<td>2.7</td>
<td>2.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>01007</td>
<td>CHEESE, CAMEMBERT</td>
<td>51.8</td>
<td>300</td>
<td>19.8</td>
<td>24.26</td>
<td>3.68</td>
<td>0.46</td>
<td>0</td>
<td>0.46</td>
</tr>
<tr>
<td>01008</td>
<td>CHEESE, CARAWAY</td>
<td>39.28</td>
<td>376</td>
<td>25.18</td>
<td>29.2</td>
<td>3.28</td>
<td>3.06</td>
<td>0</td>
<td>0.36</td>
</tr>
<tr>
<td>01009</td>
<td>CHEESE, CHEDDAR</td>
<td>37.1</td>
<td>406</td>
<td>24.04</td>
<td>33.82</td>
<td>3.71</td>
<td>1.33</td>
<td>0.28</td>
<td>0</td>
</tr>
</tbody>
</table>
Controlled vocabulary (typically for search)
Current ontologies and vocabularies

- OBO Foundry
- Ontobee
- The Ontology Lookup Service
- Bioportal
- Agroportal
Linked data versus relational data

No pre-set table structure, so easier to add just any other fact; no need to update the database design

Table headers are data as well, allowing automatic merging and reasoning without human intervention

Combining relational data with semantic wrapper: best of both worlds
Merging data sources

Shared Ontology

mapping

Local Ontology

mapping

mapping

SQL

RDF
Merging data sources
It’s happening already

Welcome to Schema.org

Schema.org is a collaborative, community activity with a mission to create, share data on the Internet, on web pages, in email messages, and beyond.

Schema.org vocabulary can be used with many different encodings, including vocabularies that cover topics, relationships between entities and actions, and documented extension model. Over 10 million sites use Schema.org to mark applications from Google, Microsoft, Pinterest, Yandex and others already using structured data.

Graph Database Market Worth $2.9 Billion by 2024 - Exclusive Report by MarketsandMarkets™

PR Newswire September 16, 2019
Digital twins need linked data and semantics