Certifying the interoperability of RDF database systems

Karima Rafes, BorderCloud/Inria-Saclay
Julien Nauroy, Inria-Saclay
Cécile Germain, University Paris Sud and CNRS
Plateforme CDS de Paris-Saclay
Plateforme "Data as a Service"
du Center for Data Science
Participants

Roughly 200 researchers in 32 laboratories

CDS is a transversal interdisciplinary laboratory of the Paris Saclay University. Part of the strategic plans for scientific computational environments currently developed by the French agency for scientific research (CNRS) and Ministry of Higher Education and Research.
Convert ‘raw’ to Linked Data in the laboratories
Building in parallel a Linked Data platform for hosting their data

Publish neuroscience
Why Linked Data? Because in theory, any service or tools of scientists can reuse the data.
Demo, how discover the datasets of scientists?


**Herschel Space Observatory** (Q209630)

List of datasets in relation with this article

- **HESIOD**: The Herschel IdOc Database is delivering photometric maps and spectral cubes from the PACS and SPIRE instruments (IR domain), reprocessed at IAS with the latest ESA pipelines and with high level customized pipelines. Virtual Observatory compatible. *(source)*

**Herschel Space Observatory** was a space observatory built and run by the European Space Agency (ESA). It was active from 2009 to 2013. It was the largest infrared telescope ever launched, carrying a single-mirror telescope and instruments sensitive to the far infrared wavebands (55–672 μm). **Herschel** was the fourth mission in the ESA science programme, along with **Rosetta**, **Planck** and **Giotto**. NASA is a partner in the **Herschel** mission, with US participation to the mission: providing mission-enabling instrument...
First result:
Front and back office to publish their datasets

Search an Open Dataset at Paris-Saclay
Linked Data, in practice
Interoperability is not optional
Lack of interoperability causes two complications

• Migration between databases and their updates
  • The last version is always the better in the science I want… inference, velocity, etc.
    ➔ To accelerate the science
  • Needs in the new platform IaaS and PaaS
    To linking data to results of science
    ➔ Reproducibility as the ultimate goal

• Development…
  The same code doesn’t work with another endpoint SPARQL!
  ➔ But, SPARQL is also a protocol? No?
A first solution:

We check the database systems.

We give our results to scientists or developers before to deploy a new RDF database in the cloud of the university.
A better solution: Certifying the interoperability?

A user can add tests, i.e., his query in production.
Test framework named TFT (Tests for Triple stores)
https://github.com/BorderCloud/TFT

We push our results on the web site : Sparqlscore.com
Everybody can reproduce the same tests with TFT.
**THIS SOFTWARE SCORES 380 OUT OF 459 POINTS**

Triplestore tested: Fuseki v1.1.1-SNAPSHOT  
Testing software used: TFT v0.2  
[Back to the list of triplestores](#)

<table>
<thead>
<tr>
<th>grid observatory tests</th>
<th>sparql 1.1 tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grid Observatory 3.0</strong></td>
<td><strong>SPARQL 1.1 Entailment Regimes</strong></td>
</tr>
<tr>
<td>10 Subtract date</td>
<td>Literal with language tag test</td>
</tr>
<tr>
<td>10 Subtract date : Test the protocol.</td>
<td>Pass</td>
</tr>
<tr>
<td>10 Subtract date : Test the response.</td>
<td>Pass</td>
</tr>
<tr>
<td>View the original test suite</td>
<td>bind01 - BIND fixed data for OWL DL</td>
</tr>
<tr>
<td>Results:</td>
<td>Pass</td>
</tr>
<tr>
<td>View the errors</td>
<td>bind02 - BIND fixed data for OWL DL</td>
</tr>
<tr>
<td>50 Filter date &amp; order by date</td>
<td>bind03 - BIND fixed data for OWL DL</td>
</tr>
<tr>
<td>60 Query to calculate ERT-ART</td>
<td>bind04 - BIND fixed data for OWL DL</td>
</tr>
<tr>
<td>Partial</td>
<td>Pass</td>
</tr>
<tr>
<td>Fail</td>
<td>Pass</td>
</tr>
<tr>
<td>Pass</td>
<td>Pass</td>
</tr>
</tbody>
</table>
Conclusion:
Who will check in the Semantic Web the Linked Data Quality about the protocol?

• Benchmarking the velocity but without the protocol is insufficient

• An open benchmark is possible and can help to converge

• The SPARQL 1.1 is a recommendation but not the tests.
  • There are again works… How to pay that?

• How help to create a really interoperable ecosystem?
  • Like HTML5 with "Test the Web Forward" or?
Thanks

Questions ?

Karima.rafes@gmail.com