Why we need Ontology Archiving

by Denis Streitmatter, InfAI
Problems of accessing Ontologies

Access

- incorrect RDF deployment (A)
- link rot / unavailable ontologies (FA)
- no/unclear versioning (FR)
- no stable citation for dependency (IR)
Archivo Monitor

- Archivo - automatic, augmented
- Ontology Archive
- Crawls 3 times a day
- Data from April 2021 to November 2021
- Covers ~ 1440 ontologies
Ontology Availability in Archivo
## Ontology Availability in Archivo

<table>
<thead>
<tr>
<th></th>
<th>all onts</th>
<th>all failing</th>
<th>temp. failing</th>
<th>[0.01:5)%</th>
<th>[5:25)%</th>
<th>[25:75)%</th>
<th>[75:100)%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.00%</td>
<td>0.50%</td>
<td>0.50%</td>
<td>0.50%</td>
<td>5.15%</td>
<td>26.87%</td>
<td>75.12%</td>
</tr>
<tr>
<td>Q1</td>
<td>0.00%</td>
<td>1.00%</td>
<td>1.00%</td>
<td>0.50%</td>
<td>6.47%</td>
<td>32.84%</td>
<td>88.56%</td>
</tr>
<tr>
<td>Med</td>
<td>0.50%</td>
<td>4.98%</td>
<td>3.72%</td>
<td>1.00%</td>
<td>7.46%</td>
<td>36.32%</td>
<td>88.56%</td>
</tr>
<tr>
<td>Q3</td>
<td>5.97%</td>
<td>12.19%</td>
<td>7.96%</td>
<td>1.99%</td>
<td>10.45%</td>
<td>69.40%</td>
<td>89.90%</td>
</tr>
<tr>
<td>Max</td>
<td>100.00%</td>
<td>100.00%</td>
<td>99.00%</td>
<td>4.98%</td>
<td>24.88%</td>
<td>74.62%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Avg</td>
<td>10.64%</td>
<td>19.67%</td>
<td>12.20%</td>
<td>1.59%</td>
<td>9.17%</td>
<td>47.27%</td>
<td>88.90%</td>
</tr>
<tr>
<td>#</td>
<td>1439</td>
<td><strong>775</strong></td>
<td>709</td>
<td>394</td>
<td>224</td>
<td>51</td>
<td><strong>40</strong></td>
</tr>
</tbody>
</table>
Studying effects on linked open data

- How huge is this impact on the LOD Cloud?
- Using LOD-a-lot HDT dump (2017)
- 28.36 billion triples
- 524 GB of compressed HDT data
Ontology Reusability in Archivo

- access even already unavailable ontologies
- cite a certain version of an ontology (identified with a timestamp)
- persistent snapshots of any ontology version

One REST request:

http://archivo.dbpedia.org/download? o={ontology-URI} v={version} f={format}

Conclusion and Countermeasures

Conclusion

- ~50% of ontologies measured had no problems
- ~7% of ontologies are at least often failing
- ~32% of all triples in LOD-a-lot use terms which are at least often unavailable

Countermeasures

- Follow the best practises for publishing RDF vocabs
- Submit your ontology to an archiving service, e.g. Archivo or LOV