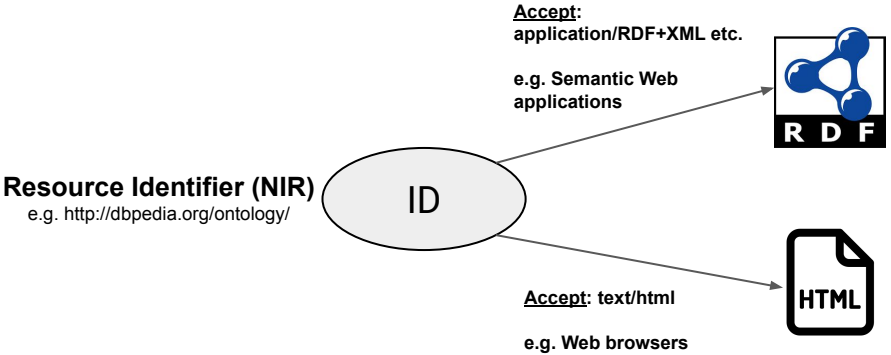
Why we need Ontology Archiving

by Denis Streitmatter, InfAI

Problems of accessing Ontologies



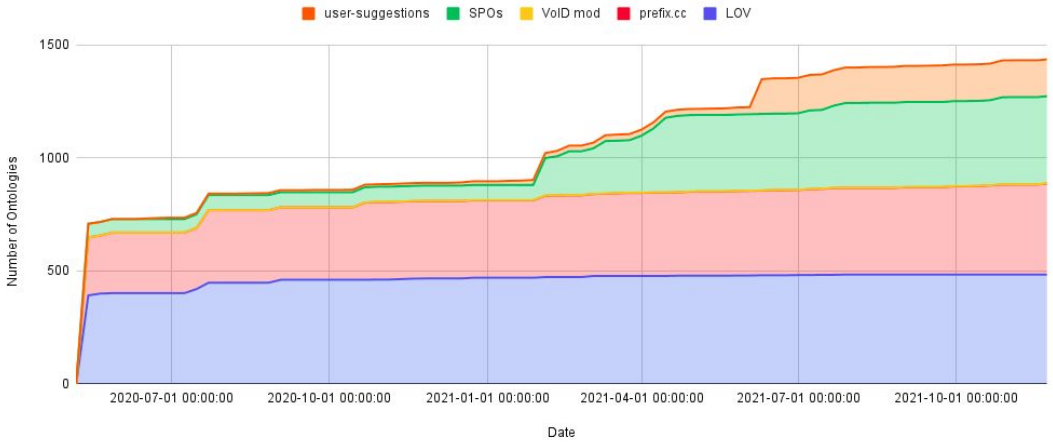
Access

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

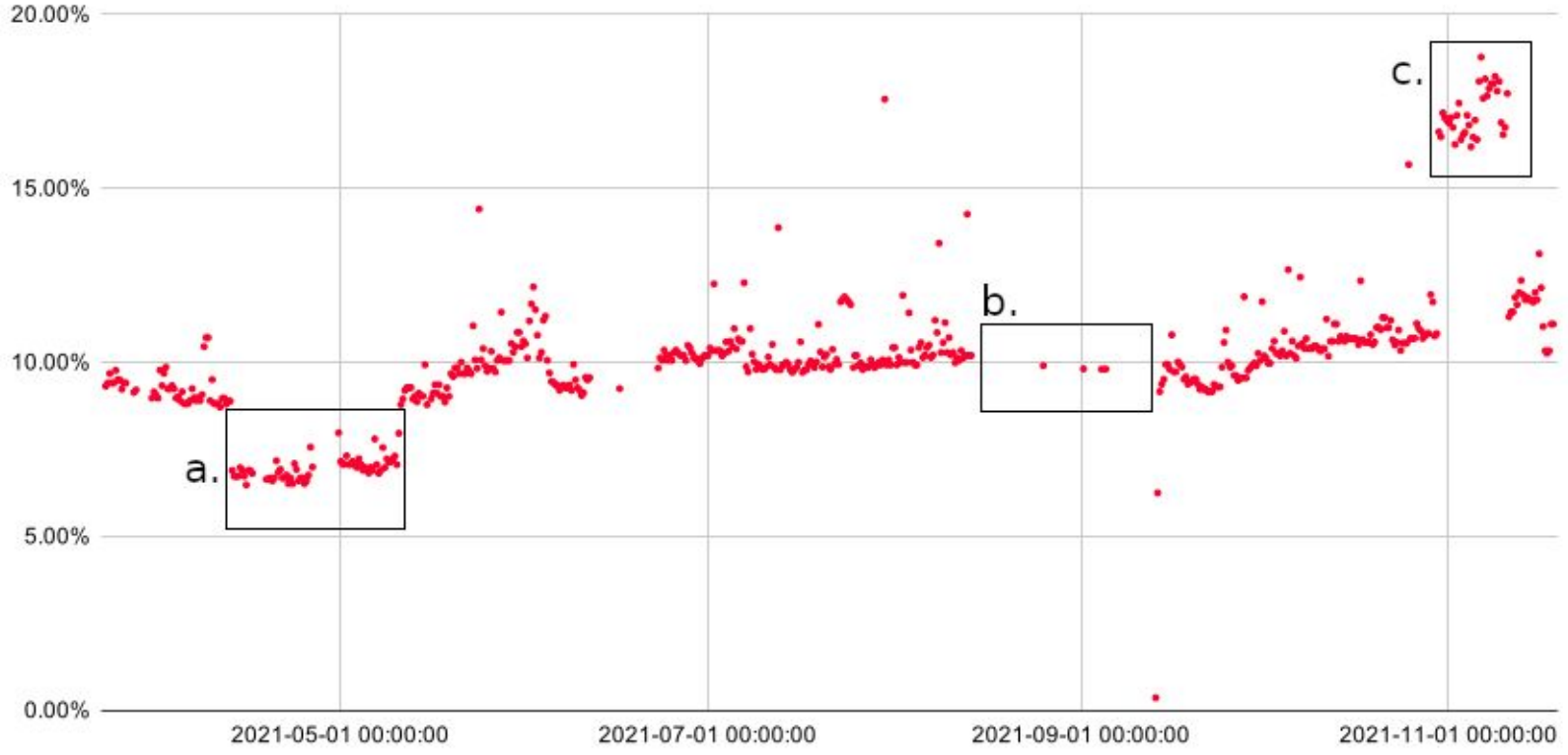
Findable **A**ccessible **I**nteroperable **R**eusable

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



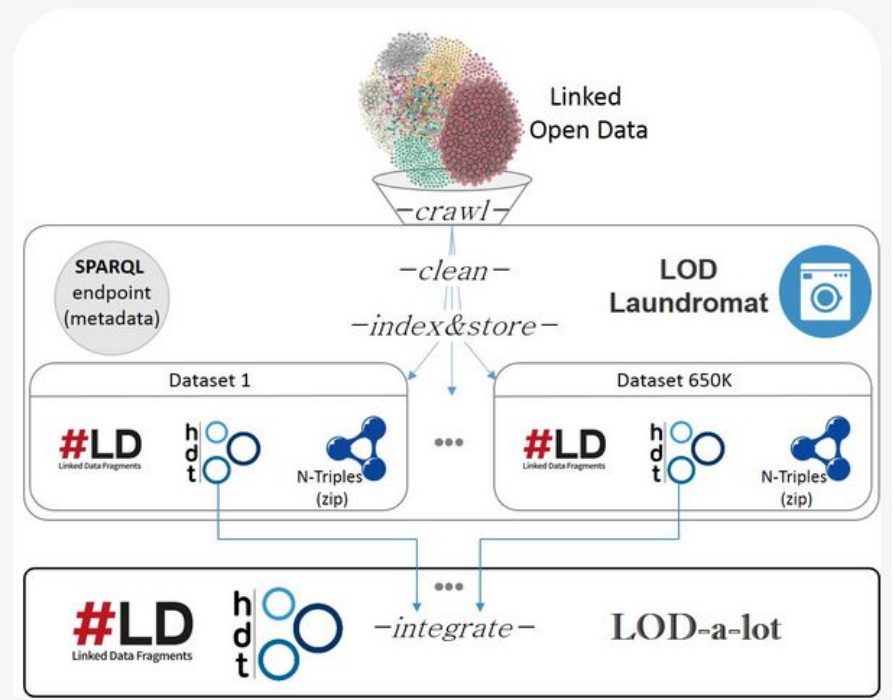
Ontology Availability in Archivo

DBpedia

	all onts	all failing	temp. failing	[0.01:5)%	[5:25)%	[25:75)%	[75:100)%
Min	0.00%	0.50%	0.50%	0.50%	5.15%	26.87%	75.12%
Q1	0.00%	1.00%	1.00%	0.50%	6.47%	32.84%	88.56%
Med	0.50%	4.98%	3.72%	1.00%	7.46%	36.32%	88.56%
Q3	5.97%	12.19%	7.96%	1.99%	10.45%	69.40%	89.90%
Max	100.00%	100.00%	99.00%	4.98%	24.88%	74.62%	99.00%
Avg	10.64%	19.67%	12.20%	1.59%	9.17%	47.27%	88.90%
#	1439	775	709	394	224	51	40

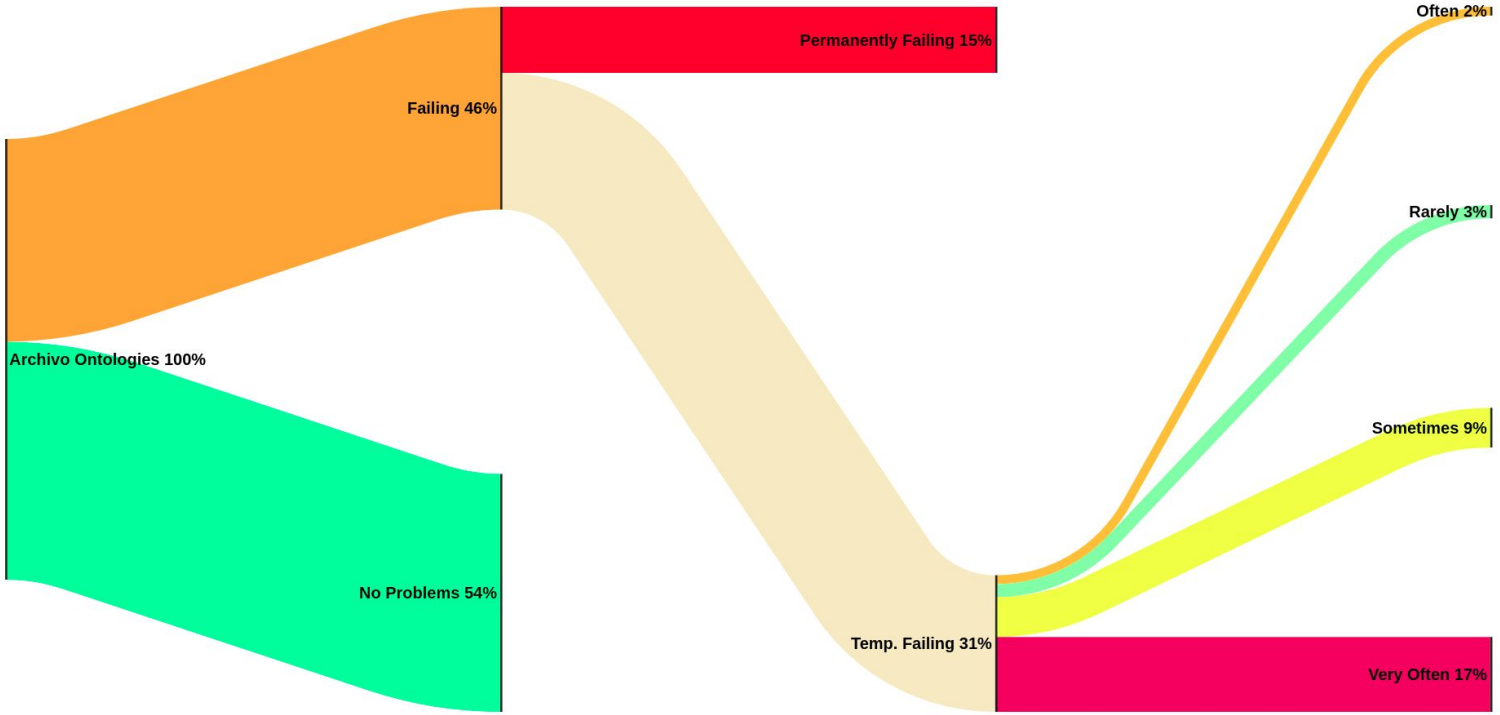
Studying effects on linked open data

- 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



LOD-a-lot overview and data flow

LOD-a-LOT Availability



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}

e.g. https://archivo.dbpedia.org/download?o=http%3A//www.georss.org/georss/&v=2020.08.10-110000&f=ttl

Conclusion and Countermeasures

DBpedia

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