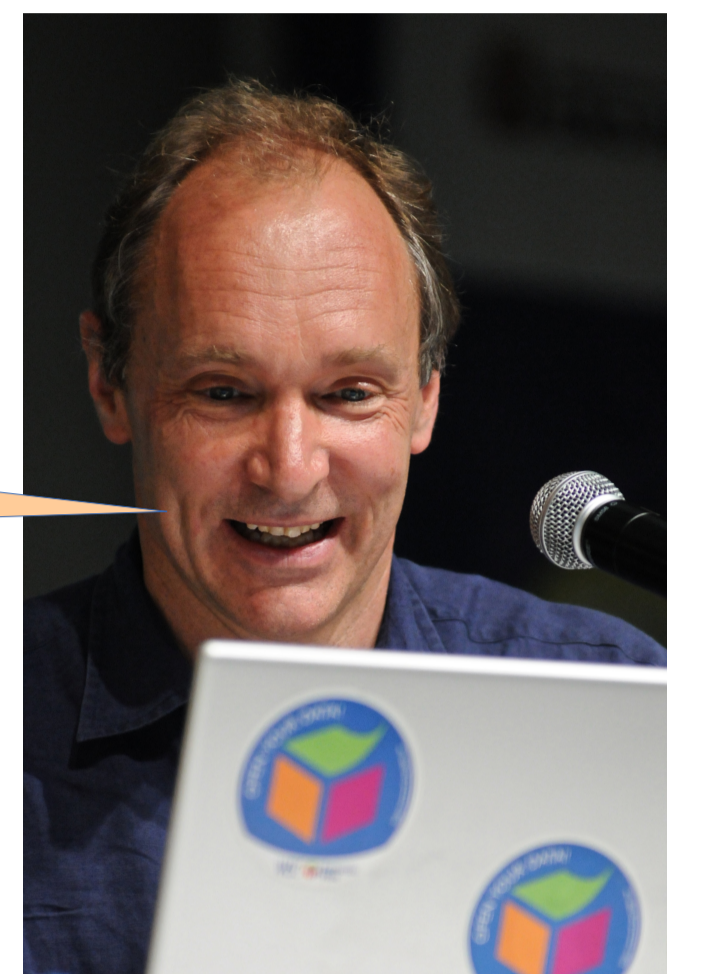


Discoverability of SPARQL Endpoints in Linked Open Data

Heiko Paulheim, University of Mannheim
Sven Hertling, University of Darmstadt

“to make the data be effectively linked, someone who only has the URI of something must be able to find their way to the SPARQL endpoint.”



Tim Berners-Lee's Linked Data Design Issues Document (2005) [1]

/.well-known/void strategy: [2]

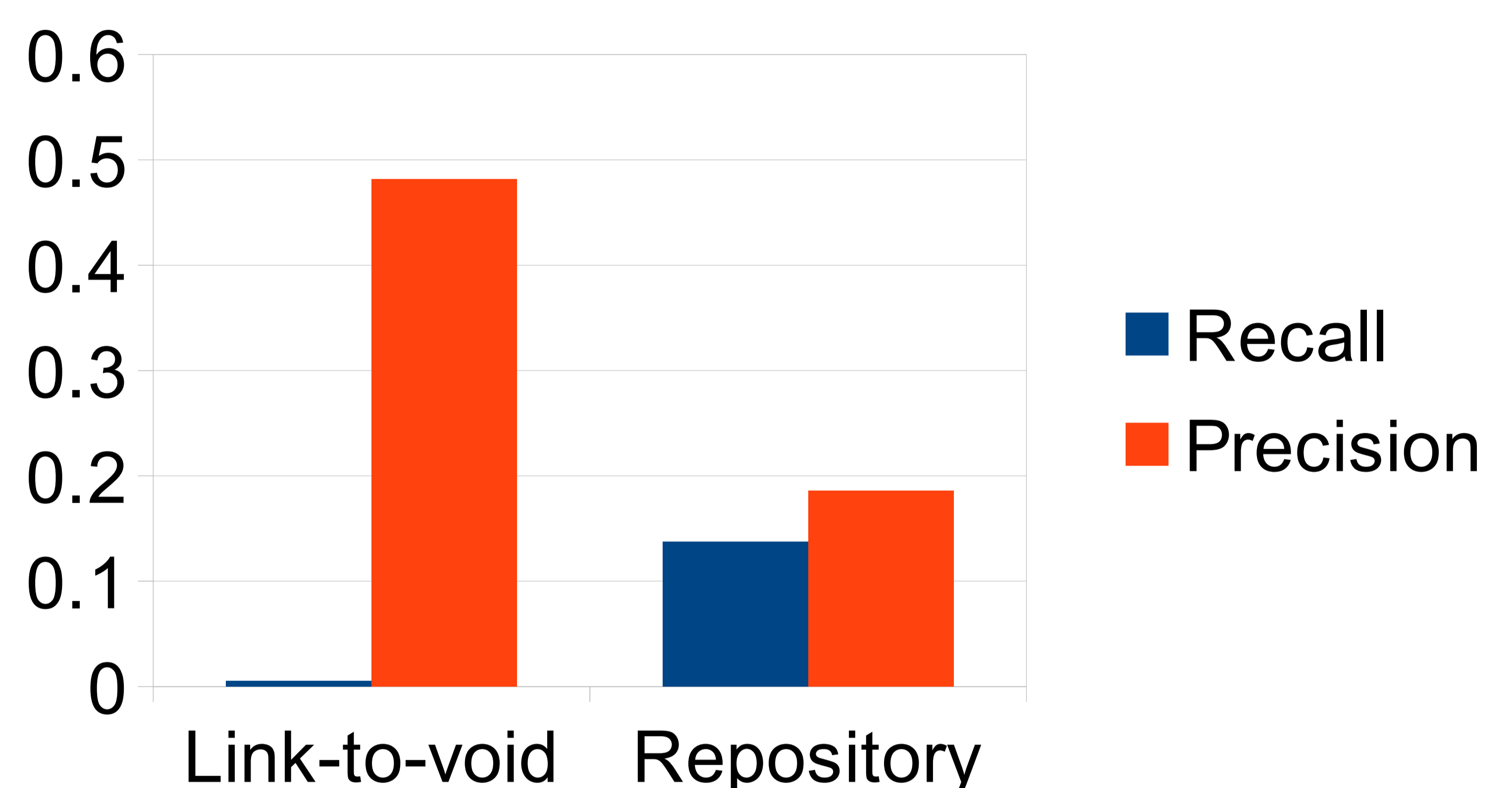
- Try to retrieve a VoID description from `http://foo.bar/.well-known/void`

.Link-to-void strategy: [2,3]

- Look for triples like
`foo:bar void:inDataset ?d .` or
`foo:bar prv:containedBy ?d .`

Repository strategy: [4]

- Look up URI host in datahub repository



Evaluation:

- performed on 10,000 random URIs from Billion Triple Challenge dataset
- two steps: retrieve endpoint and validate with `ASK` statement

Results:

- VoID and provenance are not widely adopted
- `/.well-known/void` is barely used at all
- only datahub repository provides reasonable recall

Further Findings:

- `/.well-known/void` is not used according to standards (possible reason: missing root rights)
- Moving datasets is tricky (all redirects in VoID descriptions need to be updated as well)

Prototype:

- Web service for resolving URIs to SPARQL endpoints:
`http://tinyurl.com/sparqlsenf`

[1] <http://www.w3.org/DesignIssues/LinkedData.html>.

[2] <http://www.w3.org/TR/void/>

[3] <http://trdf.sourceforge.net/provenance/ns.html>

[4] <http://datahub.io/>