Gulliver's Travels and the Future of Repositories

David Corbly

May 2018

OU UNIVERSITY LIBRARIES
The University of Oklahoma

We See With Great Exactness….

"…nature has adapted the eyes of the Lilliputians to all objects proper for their view: they see with great exactness, but at no great distance." - Jonathan Swift, *Gulliver's Travels* (1726-1727)

Like the Lilliputians and Blefuscudians, we're so busy arguing (but politely!) over how to crack an egg—this repository platform, that metadata standard, this interoperability—that we've forgotten about the slightly bemused giant in our midst: Google.

Great Minds, Different Approaches

Rick Luce, Herbert Van de Sompel and Paul Ginsparg helped kick-start open access repositories in 1999 by collaborating on the Open Archives Initiative for sharing scholarly content via pre-print archives.

Herbert now advocates for decentralized repositories – "Researcher Pods"—not too unlike Stevan Harnad's 1994 "subversive proposal" that launched the Open Access movement—or the "Domain of One's Own" initiative that took shape in 2012.

Clifford Lynch coined the term "institutional repositories" in 2003 and advocated for local control. By 2016, he agreed with Eric Van de Velde and others that perhaps that didn't work out so well.
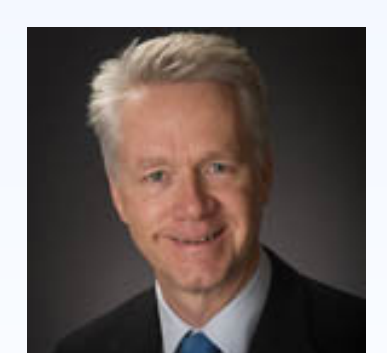
In 2017-2018, Kenning Arlitsch and Carl Grant began suggesting fewer separate repositories, perhaps even establishment of a unified national repository.
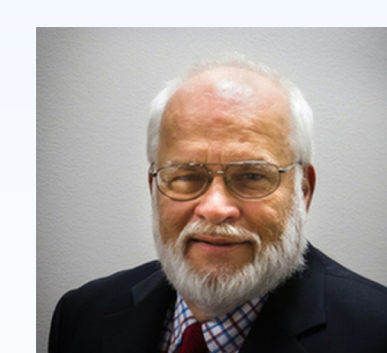
Sarven Capadisli (csarven.ca) got a shout-out from Herbert in 2017 for his work on a decentralized article publishing platform that demonstrates the viability of linked "Researcher Pod" repositories.
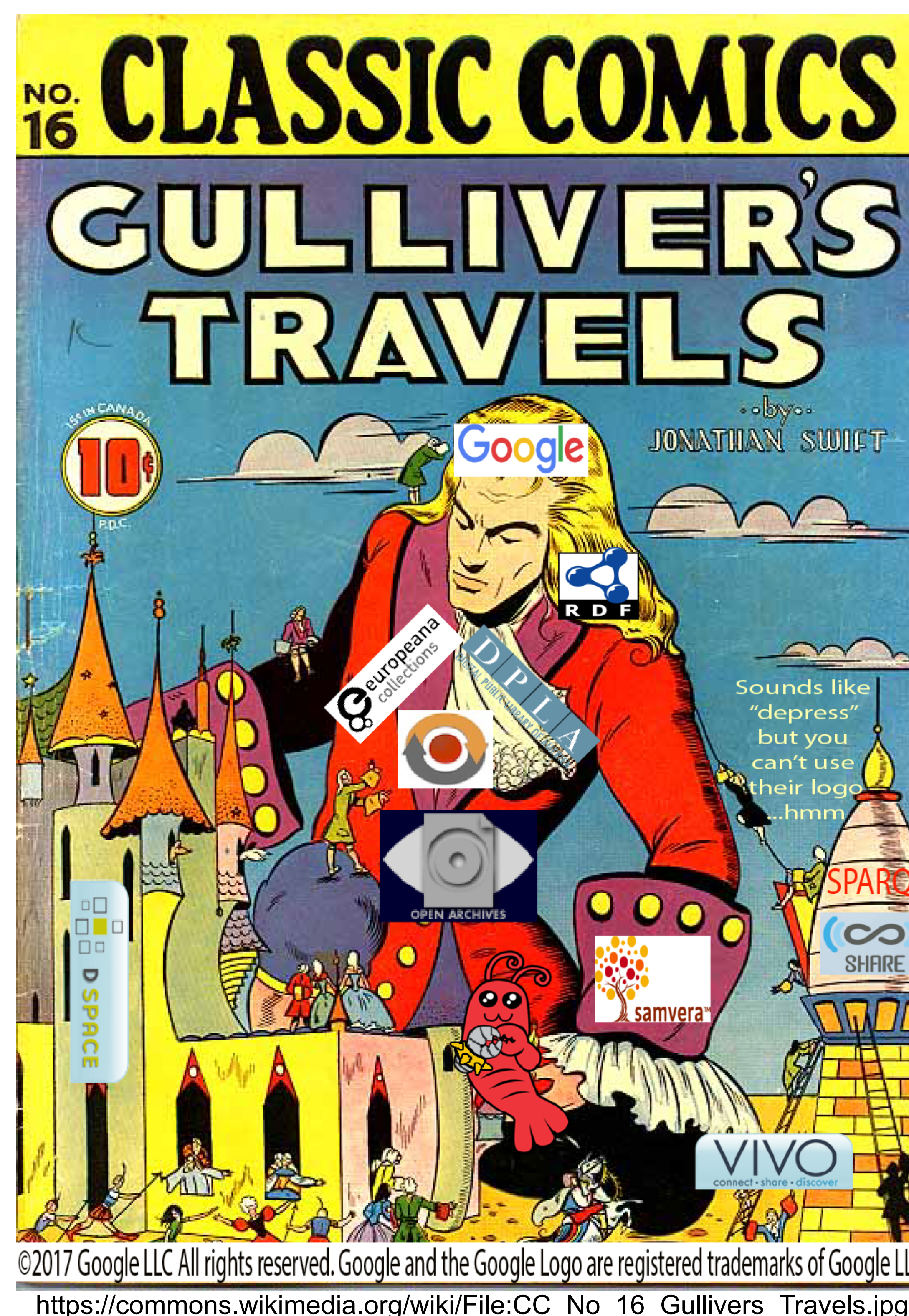
What We Give Google

- Library-centric standards and metadata it doesn't understand: MARC, OAI-PMH, Dublin Core
- Unstructured data with poor context
- Mobile-unfriendly content like big PDFs
- Wasted effort on custom search interfaces

What Google Wants

- Revenue (by making customers happy)
- Trusted, structured, backlinked and updated content and metadata
- Mobility and localization
- Fast and reliable serving of content
- Words and context that its RankBrain artificial intelligence can understand

Google is Our Customer

Ask someone searching Google if they prefer their results as:

- ❑ RDFa on Fedora via the Valkyrie gem of Samvera?
- ❑ Dublin Core via OAI-PMH on DSpace XMLUI?
- ❑ Just straight up microdata or JSON-LD?

You know the look you're going to get! Sorry, most people—and Google—don't care about the things repository managers and librarians do.

Great Minds + Google = Success

SEO — First, make Google happy with search engine optimization wizardry on our repositories. Patrick O'Brien and Kenning Arlitsch have been telling us how to do this for half a decade now.

Cooperate — Leverage cooperatives and aggregators like Duraspace, SHARE, DPLA and Europeana to work with Elsevier, Clarivate, Digital Science and CRIS systems to drive traffic to repositories—the commercial interests aren't going away anytime soon.

Unite — Kenning and Carl and COAR are all on the right track—we need fewer separate repositories AND more linked ones. The current model is too expensive and fragmented. Budget the 2.5% commitment to shared initiatives advocated by David Lewis.

Liberate — Herbert and Sarven are also on the right track. Encourage researchers to run their own portable repositories—perhaps using Tim Berners-Lee's Solid ("social linked data") initiative at MIT—but also give them institutional infrastructure and interoperability support.