Wikidata and EaaSI

Katherine Thornton

6 November, 2018

Overview

- Wikidata and EaaSI
- Wikidata
- WikiDP
- Resources

Wikidata and EaaSI

- Collaborative modeling of the domain of computing
- EaaSI metadata will be interoperable with data models in Wikidata
- We are contributing to open knowledge

Happy Birthday, Wikidata!

Wikidata went live on 29 October 2012.

Wikidata Birthday Celebrations

Figure 1: Map of 6th birthday celebrations

Wikidata: A Knowledge Base of Structured Data

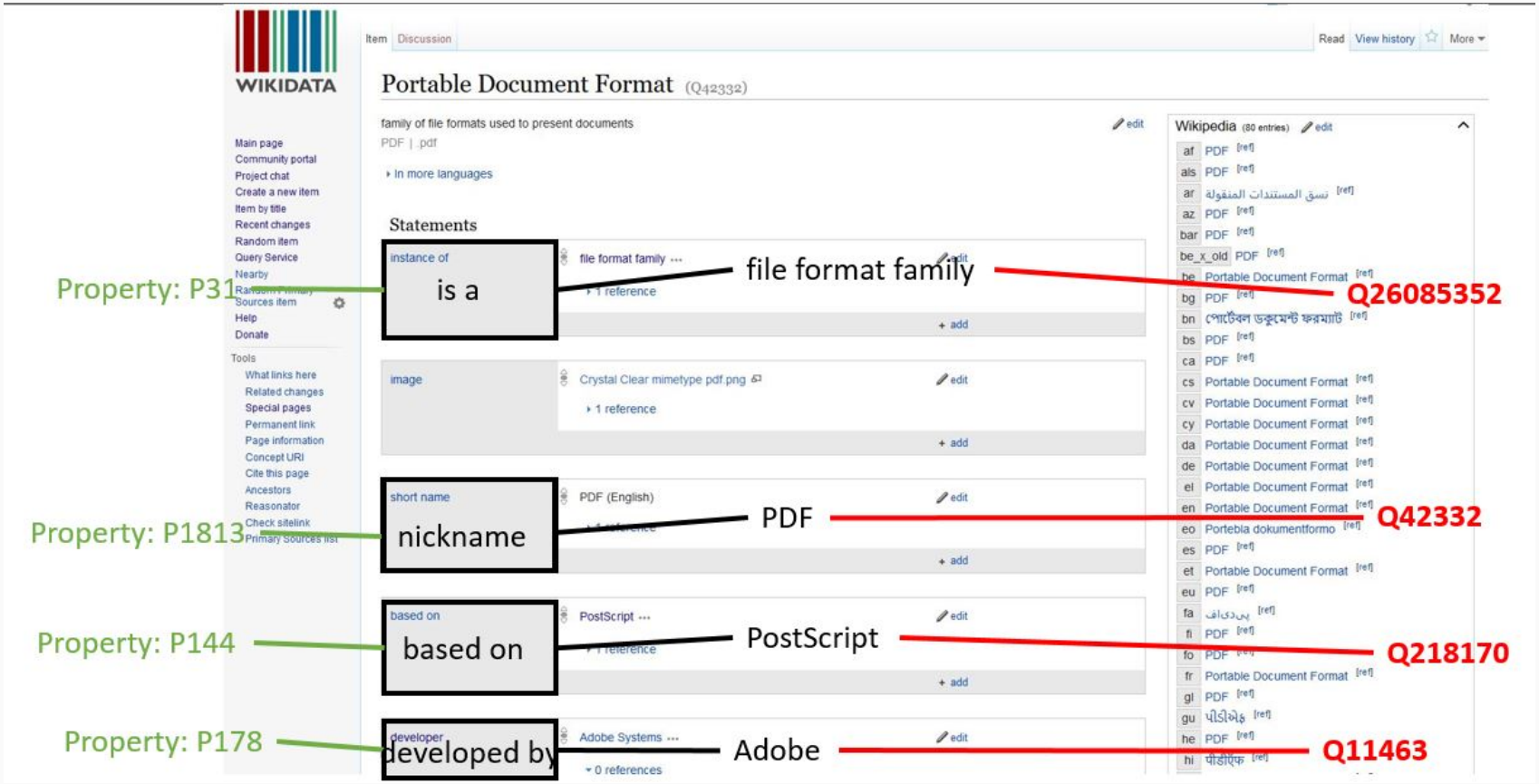
This knowledge base of structured data is:

- Machine-readable linked open data
- Editable by anyone with Internet access
- Designed to support both human and algorithmic curation
- Fully-versioned wiki
- around 300 human languages

Wikidata Stats

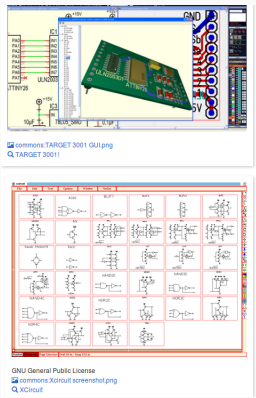
- over **19,000** active editors per month
- over **51 million** items have been created
- more than **260** bots

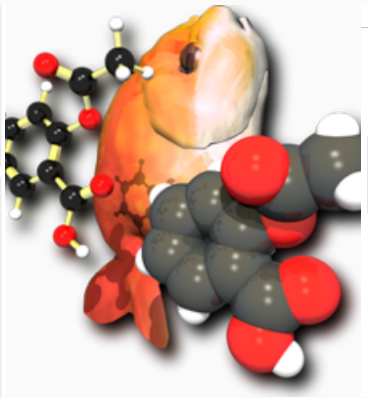
Items, properties, unique identifiers

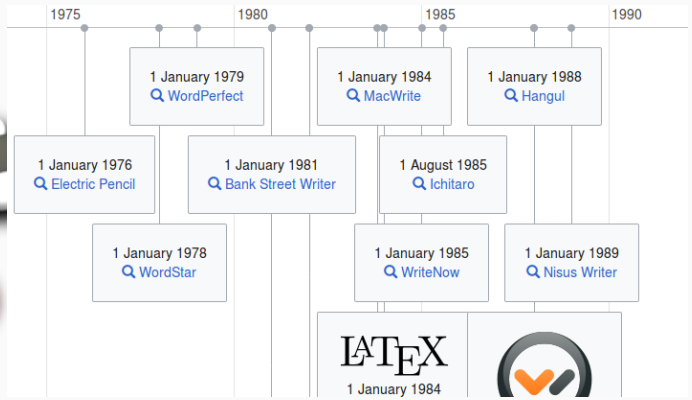


Status of software data in Wikidata

- **85,000** instances of software in Wikidata today
- commercial software
- research software
- free and open source software (FLOSS)







Timeline of discontinued Google products

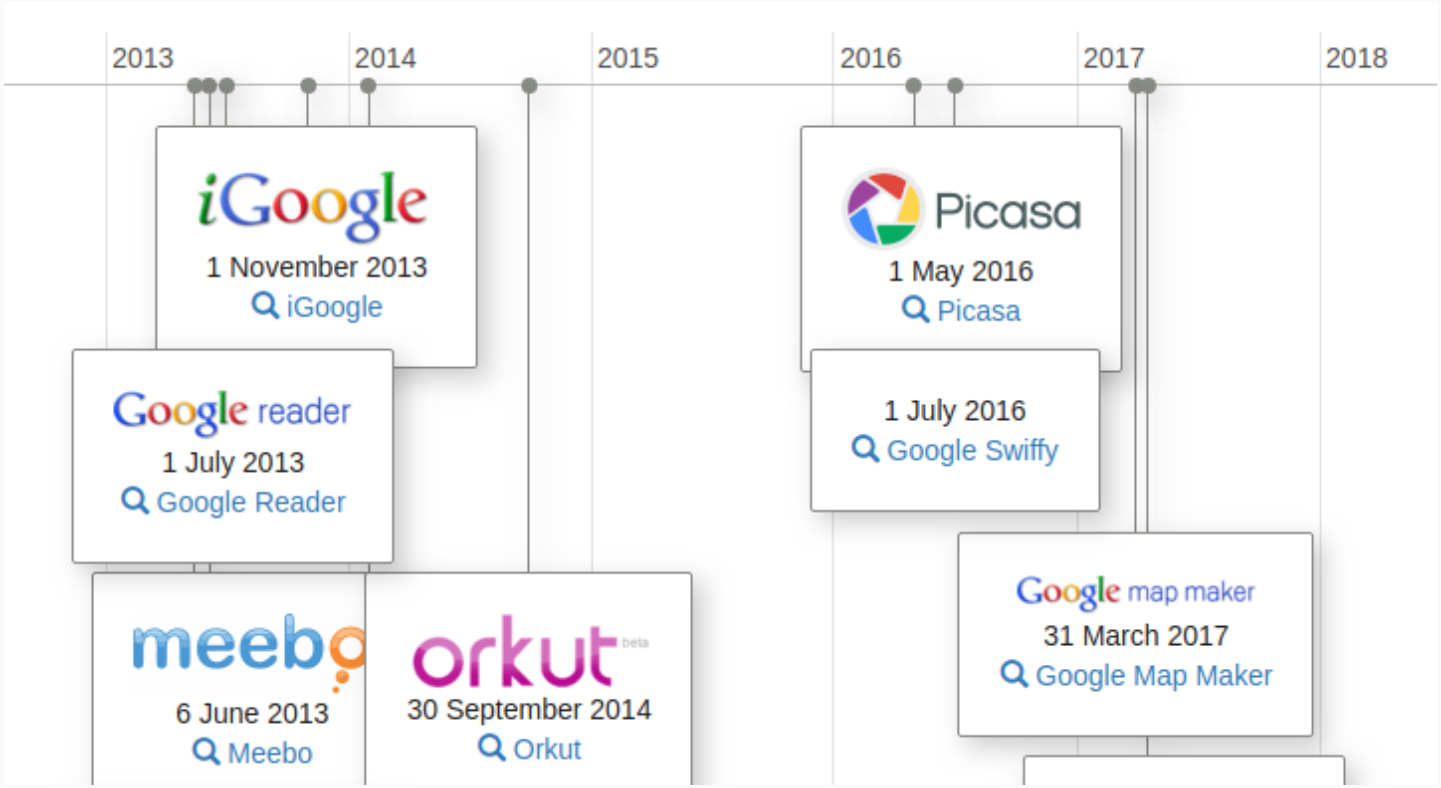


Figure 2: Try this query!

What software available under a free software license can I use to open .obj files?

```
SELECT DISTINCT ?app ?appLabel ?logo WHERE {
  ?app (wdt:P31/wdt:P279*) wd:Q7397.
  ?app wdt:P1072 wd:Q2119595.
  ?app wdt:P275 ?lic.
  ?lic (wdt:P31/wdt:P279*) wd:Q3943414.
  OPTIONAL {?app wdt:P154 ?logo.}
  SERVICE wikibase:label { bd:serviceParam wikibase:language
}
```

Figure 3: Try this query!

Wikidata is a linking hub for external IDs

- External IDs have their own data type
- 58 percent of WD properties are external ids 2570/4439
- Joachim Neubert's paper on this topic

Figure 4: All external ids for NumPy

Status of file format data in Wikidata

- **3,389** instances of file format in Wikidata today
- PRONOM has 1,553 entries: of these we have **1,208** file formats with PUID external ids
- **2,629** items connected to Just Solve the File Format Problem ids

Links between descriptive and technical metadata

Bubble chart of software titles by number of readable file formats

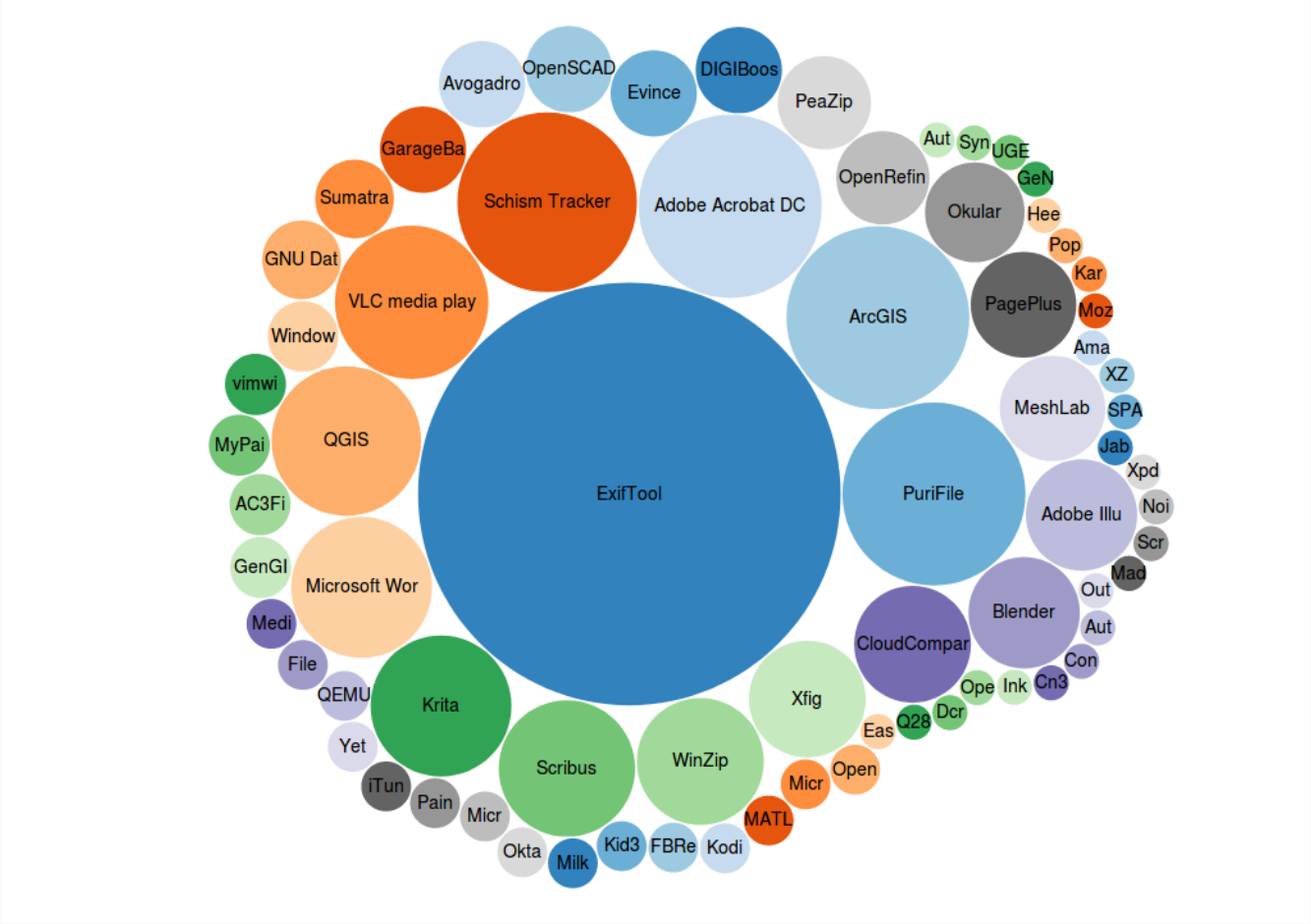


Figure 5: Try this query!

What does LoC get from the creation of P2366 LoCFDD external id?

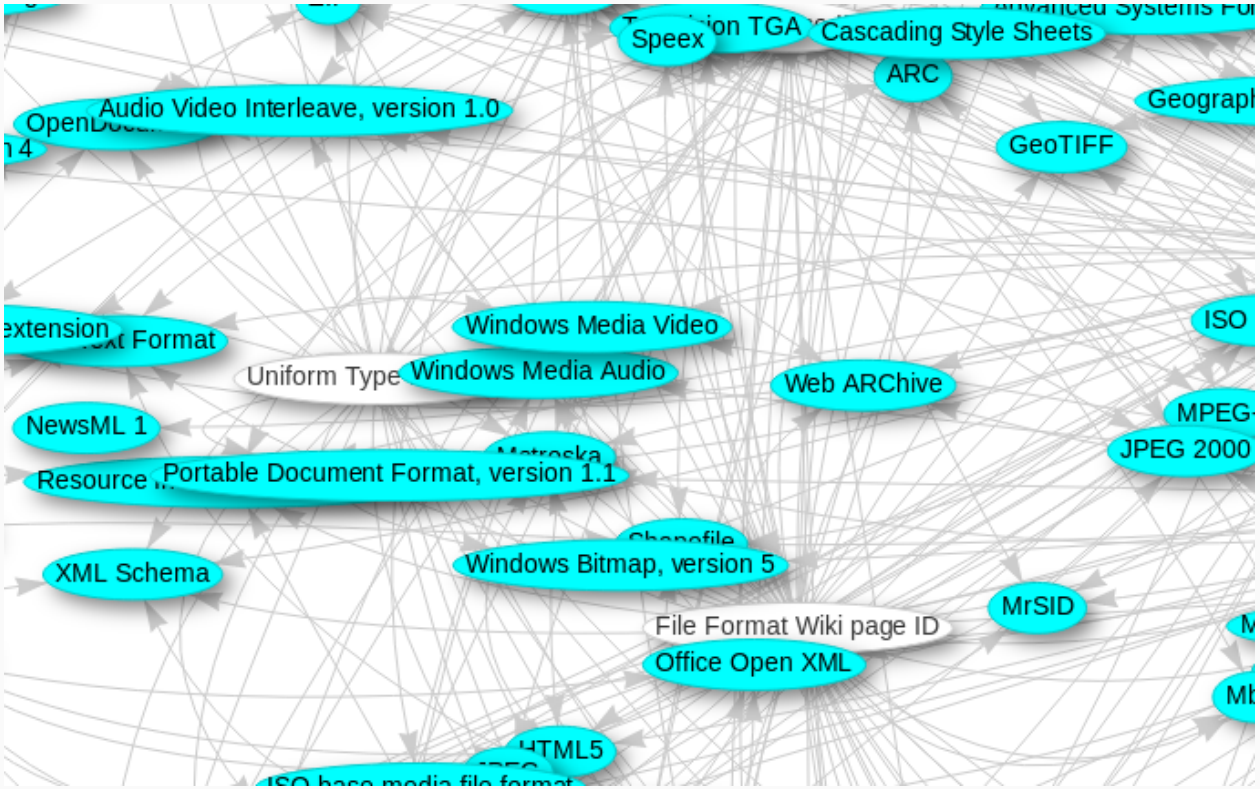


Figure 6: Try this query!

Cross-domain knowledge base

- developers, organizations, events, places

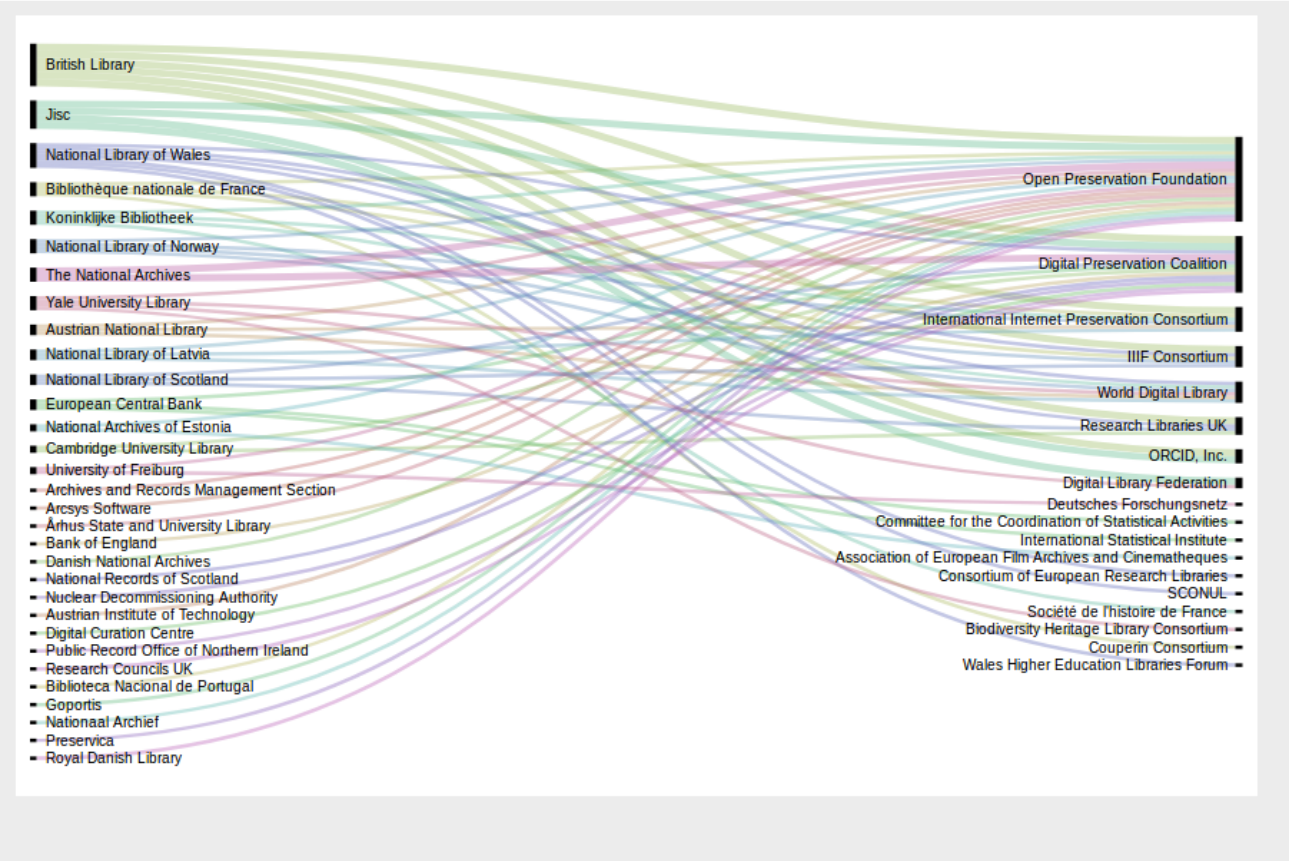


Figure 7: Orgs that are members of OPF and DPC plus other memberships

What will we do with URIs for resources in the domain of computing?

- leverage structured data in digital preservation system
- software citation
- unambiguous identifiers for all parts of a pre-configured emulated computing environment

Software titles for 3D graphics and file formats

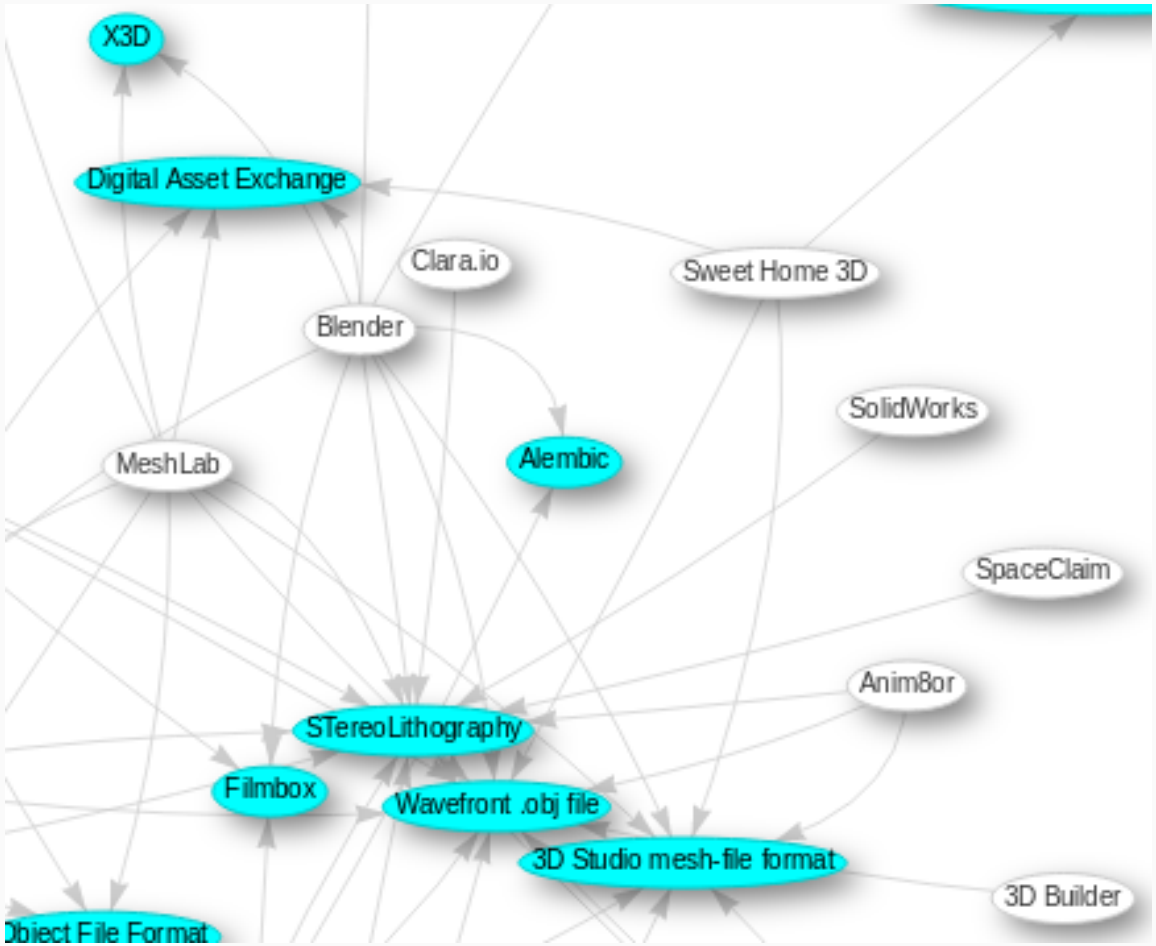


Figure 8: Try this query!

Machine-Readable “alternative to” website powered by Wikidata

- multiple serialization formats (.ttl, .rdf, .json)
- reverse look-up tool for file formats to software
- unique Wikidata URIs so we can discuss software without confusion

Data Curation Stats as of Oct 1, 2018

Items	Jan 18	Oct 18	net change
Software Items	64,925	85,548	20,623
File Format Items	2,834	3,405	571
File Format Items with PUIDs	777	1194	417
File Format Signatures	167	183	16
Emulators	106	115	9
File Systems	146	153	7
Device Drivers	17	27	10
Plugins	155	176	21

Wikidata for Digital Preservation

Figure 9: wikidp.org

Development Team

- **Kenneth Seals-Nutt**: software engineer
- **Katherine Thornton**: data curation, data models, SPARQL queries
- **Carl Wilson**: technical mentor
- **Euan Cochrane**: digital preservation program of work

WikiGenomes

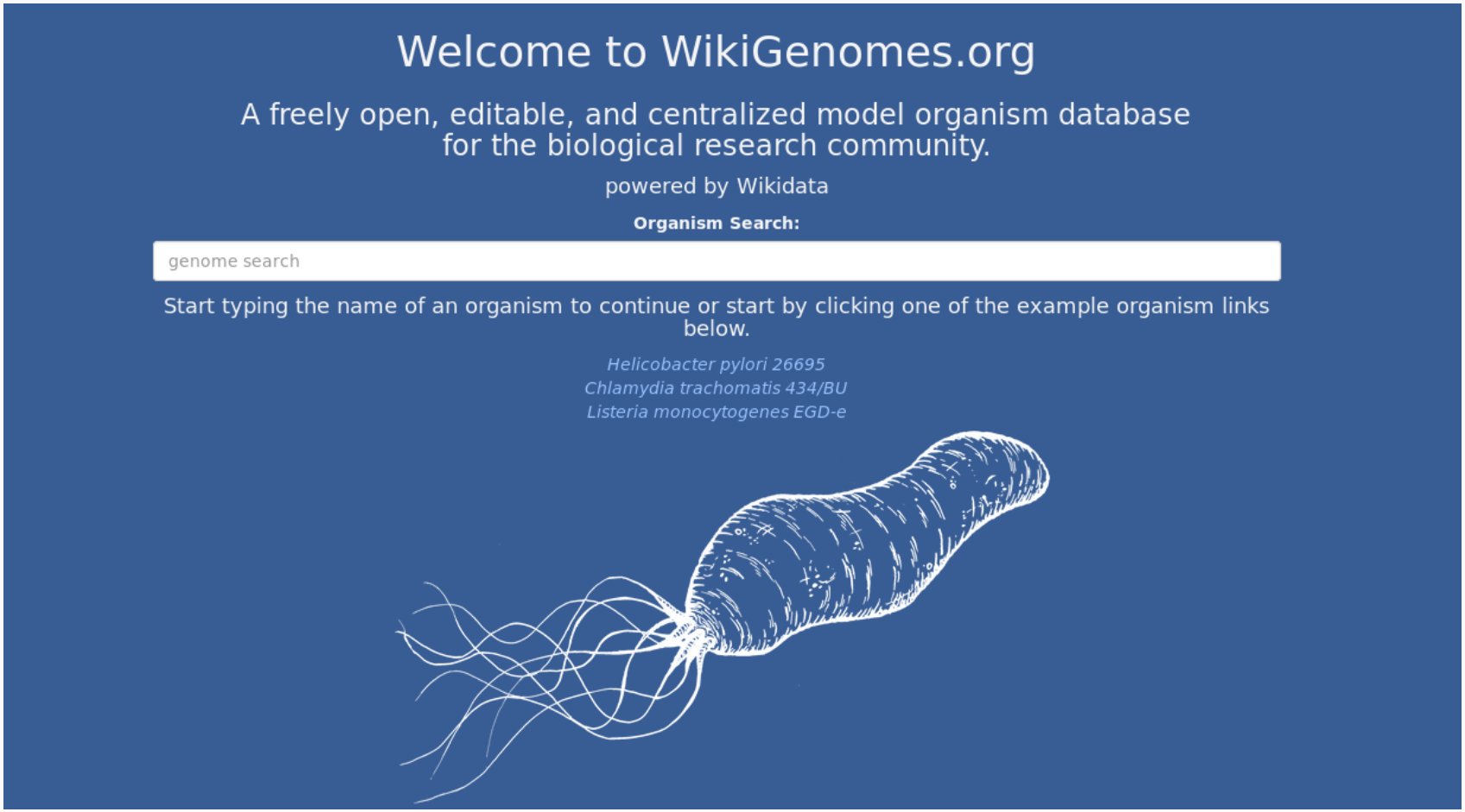


Figure 10: wikigenomes.org

Role of Portals

- About **5,000** properties in Wikidata
- Data models are not pre-defined
- Portal has a domain-specific property checklist

Technologies

- Python
- Flask
- SPARQL
- Wikidata Integrator
- Wikimedia API

wikidp.org

Figure 11: Screenshot of search results in the WikiDP portal

Web App Interactions

- Custom search
- View data from Wikidata
- Contribute data from Wikidata
- Property checklists

Node Host Communities

- Wikidata
- WikiDP
- reuse metadata from within EaaSI

Wikidata Resources

- Getting started with Wikidata video
- Wikidata Editing Cheatsheet
- Tours of the Wikidata site
- WikiProject Informatics

Wikidata-Related blog posts related to authorities

- OCLC Hanging Together
- ZBW
- ODNB

Connect with this work

- Github repository for source code
- SPARQL queries to reuse Wikidata data
- Kat's tweets full of SPARQL-y data from Wikidata @wikidigi

Thank you!

katherine.thornton@yale.edu

Acknowledgements

- Andrew W. Mellon Foundation
- Alfred P. Sloan Foundation
- Software Preservation Network
- Open Preservation Foundation
- Wikidata community