WIKIDATA

i/o

Repository Fringe 2017
at the University of Edinburgh
3-4 August 2017.

Wikidata

The free and open knowledge base

Part 2: Consuming the data

Repo Fringe 2017

Ewan McAndrew - @emcandre

Navino Evans - @NavinoEvans

https://tinyurl.com/WikidataRepo2

Wikimedia's family of open knowledge projects

WIKIPEDIA
The Free Encyclopedia

WIKIBOOKS
Open books for an open world

Wikiquote

WIKIDATA

Wikimedia Tool Labs

WIKISOURCE

Wiktionary
The free dictionary

MediaWiki

WIKIVERSITY

WIKISPECIES
free species directory

WIKINEWS

WIKIMEDIA COMMONS

WIKIMEDIA INCUBATOR

WIKIMEDIA META-WIKI

wiki voyage

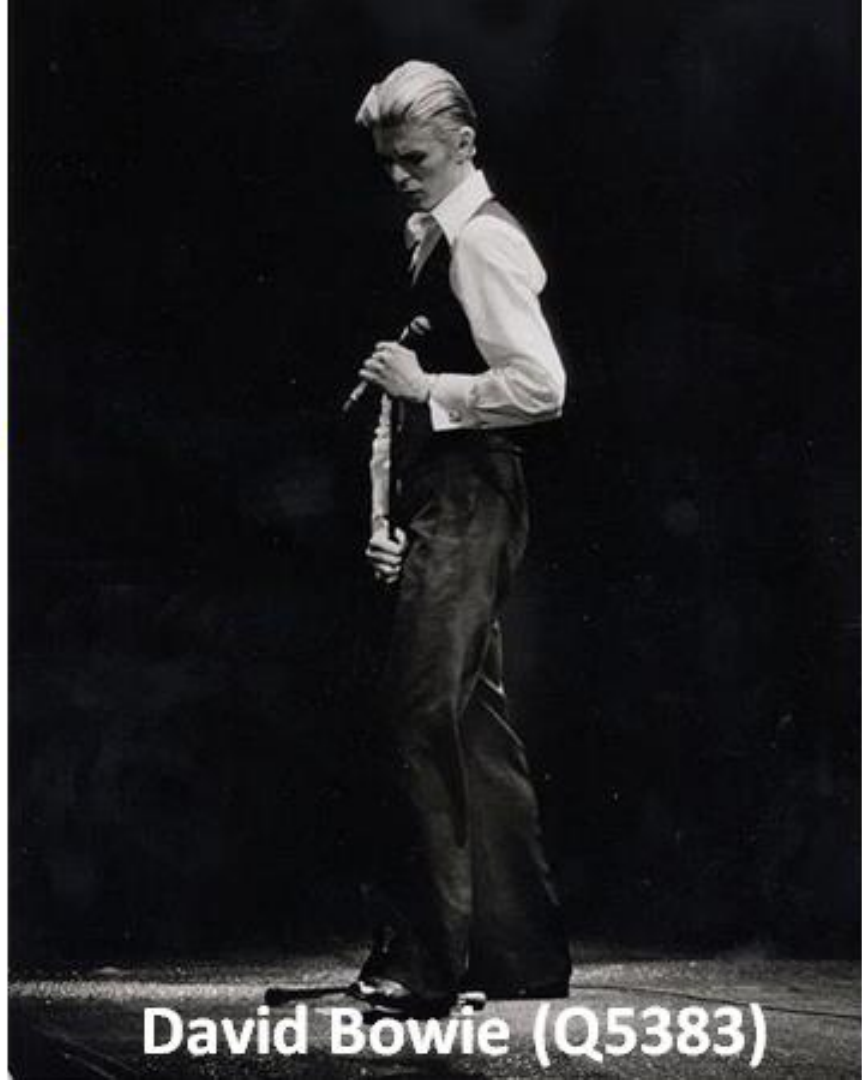https://wikimediafoundation.org/wiki/Our_projects

What is Wikidata?

Wikidata is a free linked database of secondary data that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

- Bibliographic
- Biographic
- Biomedical
- Geographic
- Taxonomic
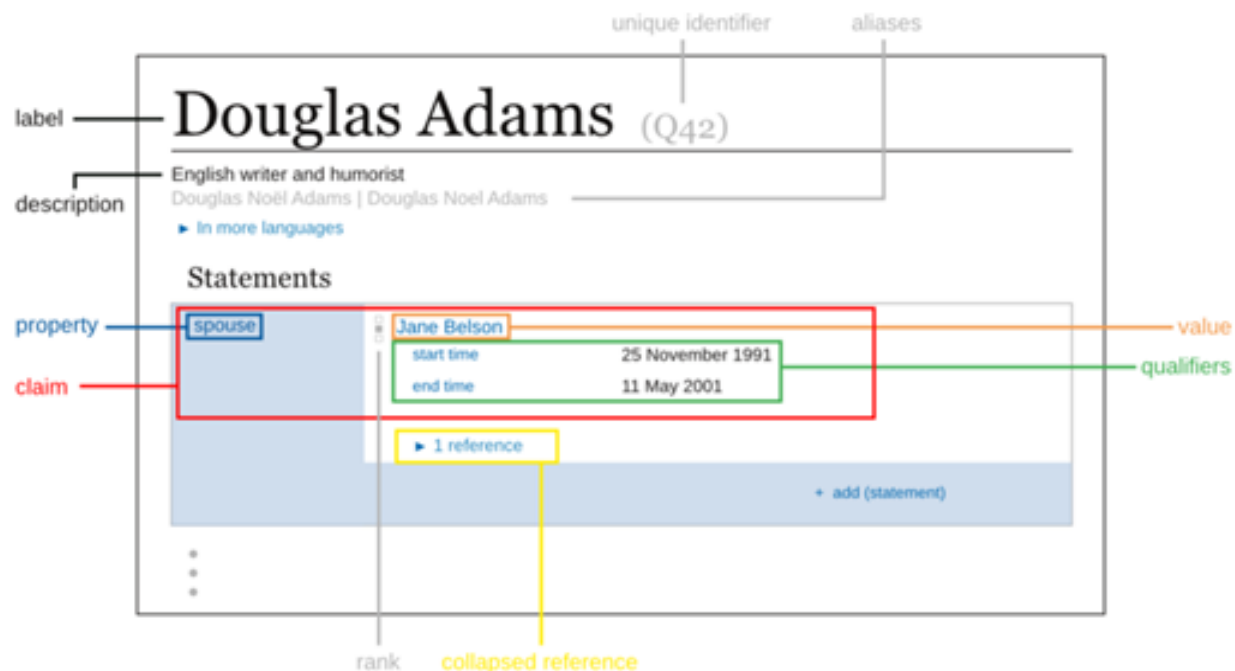- Authority file
- And more besides

date of birth (P569)

8 January 1947

David Bowie (Q5383)

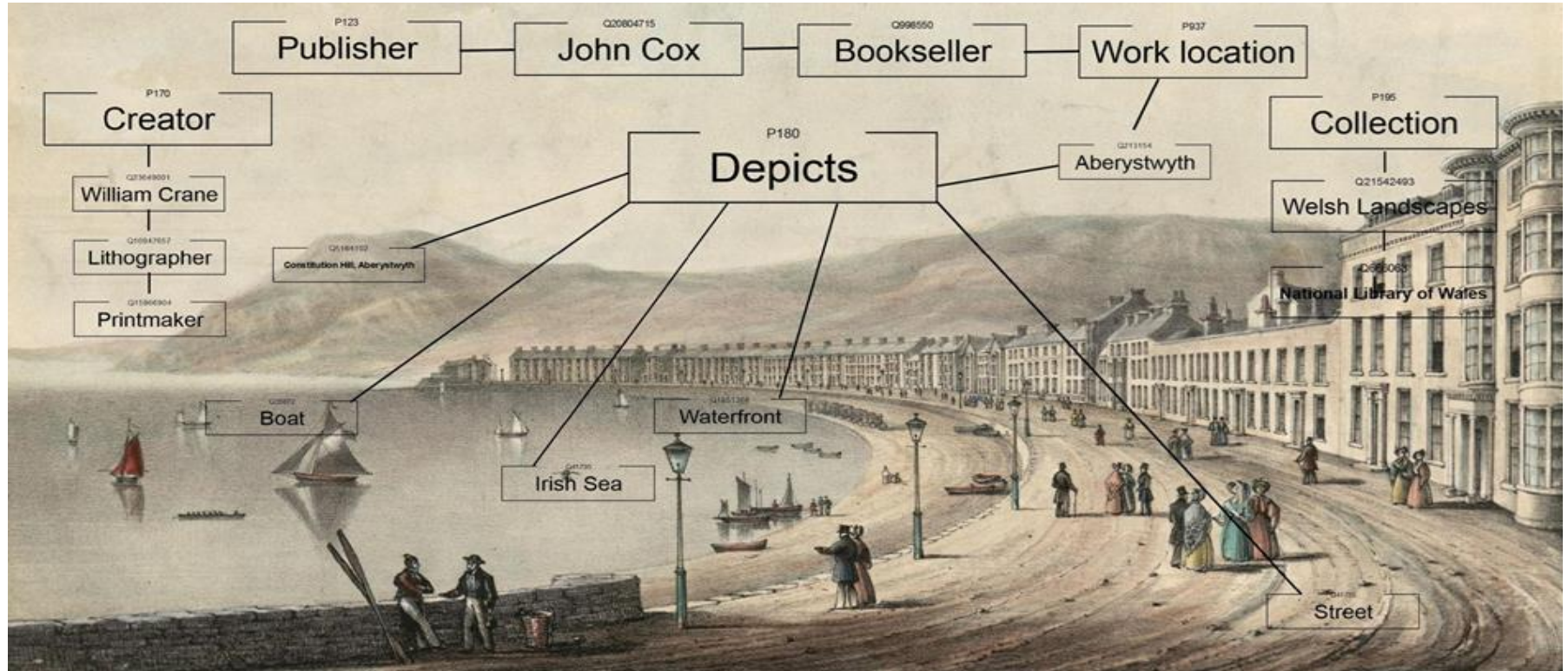# Example Wikidata item & statement

Douglas Adams is wikidata.org/wiki/Q42

Douglas Adams – in Reasonator's 'pretty' form.

SECTION 1

SHOWCASING USES OF WIKIDATA QUERIES

National Library of Wales

Sharing open knowledge about Voltaire's histories

Panama Papers

P106: occupation

P793: significant event

Q23702848: Panama Papers

https://en.wikipedia.org/wiki/User:Fniels
en/Autolists/Panama_Papers

MPs' occupations and place of education.

[Link to Wikidata query](#) - occupation.

[Link to Wikidata query](#) - education.

Image of Ken Clarke by Chris McAndrew (CC-BY)

Doctoral Thesis Metadata

Oxford Research Archive has 3237 Oxford doctoral theses on open access for anyone to download and read.
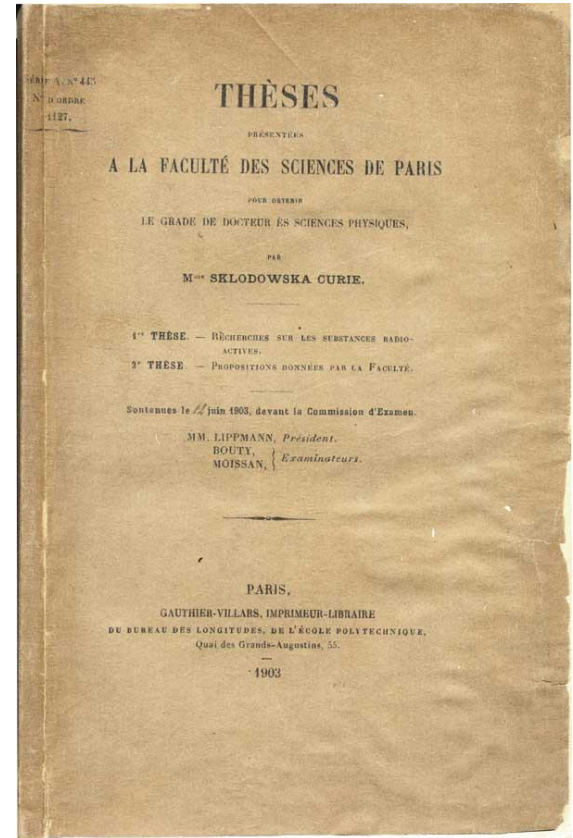
ORA are sharing their doctoral thesis metadata with Wikidata.

[Query](#) showing all doctoral theses on Wikidata.

New property: [P4101](#) - *Dissertation submitted to*

How Wikidata links the Oxford theses - [query result](#)
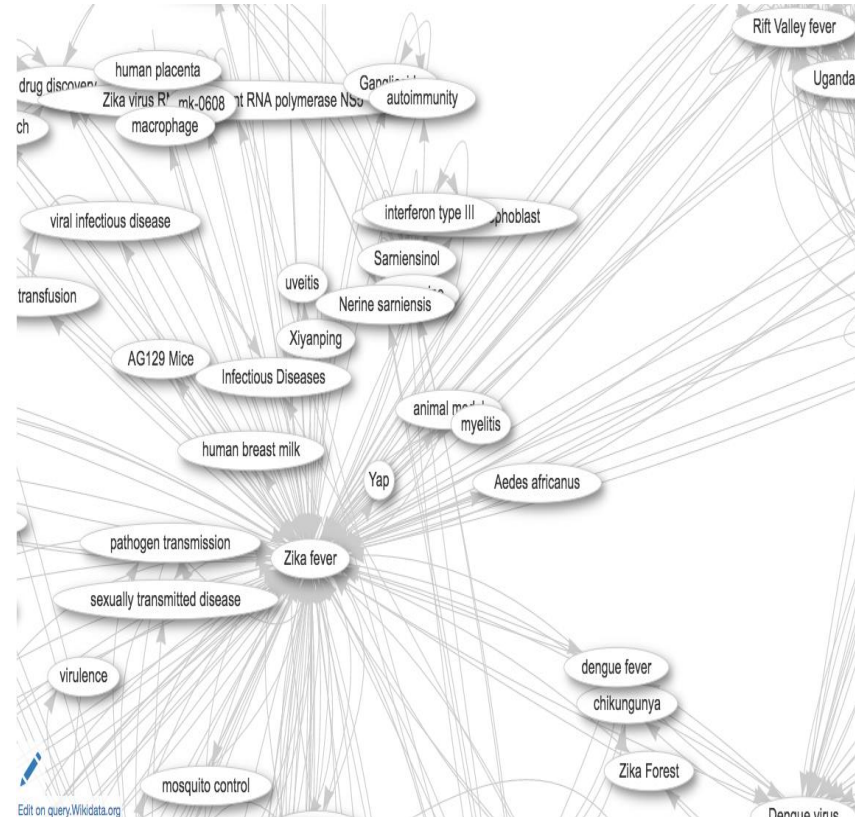
And the [query](#) itself.

THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES PHYSIQUES,

PAR

Mme SKLODOWSKA CURIE.

1re THÈSE. — RECHERCHES SUR LES SUBSTANCES RADIO-ACTIVES.
2e THÈSE. — PROPOSITIONS DONNÉES PAR LA FACULTÉ.

Soutenues le 25 juin 1903, devant la Commission d'Examen.

MM. LIPPMANN, *Président.*
BOUTY,
MOISSAN, *Examinateurs.*

PARIS,

GAUTHIER-VILLARS, IMPRIMEUR-LIBRAIRE
DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE,
Quai des Grands-Augustins, 55.

1903

[Scholia](#) - 2.3 million scientific articles

The Scholia Web service creates on-the-fly scholarly profiles for researchers, organizations, journals, publishers, individual scholarly works, and for research topics.

Among several display formats available are lists of publications for individual researchers and organizations, publications per year, employment timelines, co-author networks and citation graphs.

[Example](#), [Blog article](#) + [Video presentation](#)
[Paper on arxiv.org](#) by Finn Årup Nielsen.

Scholia

an open source, Wikidata-driven **scholarly profile application**

Uta Frith co-author graph.

Location of Turing Award recipients

The sum of all human citations

{} wikicite

SCHOLARLY DATA PROVIDERS

INTERNAL INTEGRATION

Build a structured repository of sources to serve as a foundation of free knowledge

WIKIMEDIA COMMONS

WIKIDATA

cross ref

Pub Med

iD

EXTERNAL INTEGRATION
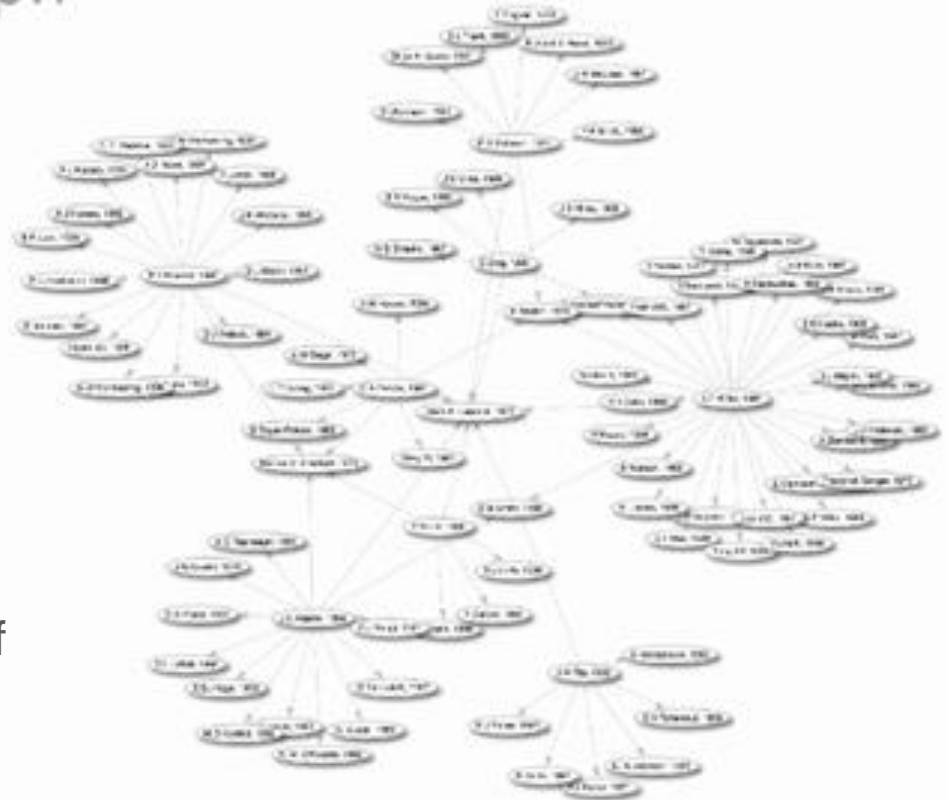
Google

ALTMETRICS SERVICES

Impactstory

PLUM ANALYTICS

DataCite

The Wikidata citation graph

3.3 million citation links
using the *cites* (P2860) property

- WikiCite project started in 2016

- Building a universal repository of sources in Wikidata.

- 500,000+ PMID references in Wikidata.

PARTIAL CITATION GRAPH FOR **ULRICH K. LAEMMLI (1970)** · http://tinyurl.com/kbzdxwh

all statements citing a *New York Times article*

most popular journals cited by statements of any item that is a subclass of *economics*

all statements citing the works of *Joseph Stiglitz*

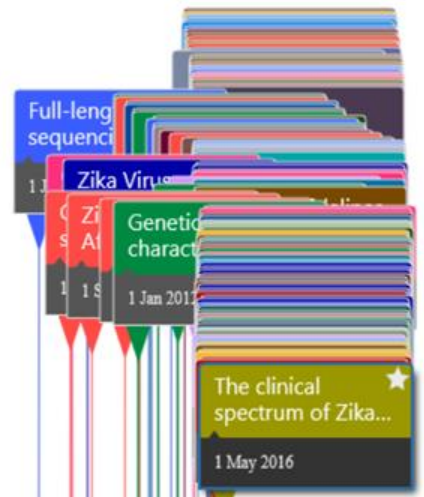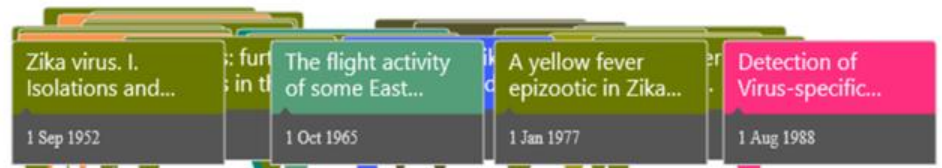all statements citing journal articles by *physicists at Oxford University in the 1970s*

all statements citing a journal article that was *retracted*

The ZikaCorpus timeline
The Zika Corpus project on Wikidata

The Zika Corpus – Histropedia timeline of scientific articles by publication date, colour-coded by publication and filtered by author and original language of work. In February 2016, the World Health Organization declared a public health emergency over the Zika virus outbreak

Other notable examples of use cases

YLE - The Finnish Broadcasting Company, Yle, has since April 1st 2016 tagged online news and feature articles with concepts from Wikidata.

Inventaire - Create an inventory of your books with Wikidata at inventaire.io

WikiGenomes - A freely open, editable, and centralized model organism database for the biological research community. Paper on WikiGenomes at Biorxiv.org

Quora - Links to Quora topics will be available through the Wikidata entities and also from Quora topic pages to Wikidata entities.

Crotos - search and display engine for visual artworks powered by **Wikidata**.

And much more besides.

SECTION 2

QUERYING WIKIDATA
PRATICAL

Federated queries

Run queries that combine data from Wikidata and other selected data sources on the web.

[List of 3rd party services supported for federated queries](#)

Simple example federated query:

[Works by Lope de Vega, retrieved from the BVMC digital library](#)

1. Lope de Vega's unique [BVMC](#) id is determined from Wikidata

2. This id is then used to retrieve works by Lope de Vega on the BVMC digital library

Getting data out of Wikidata

- **API**
  For getting data about individual Wikidata items (or groups of up to 50)

- **SPARQL Endpoint**
  Run advanced queries and get back data for up to around 200k items

- **Data Dump**
  Download all available data for large scale local processing of any size

Read more about Wikidata data access →

SPARQL TRAINING

- Go to https://query.wikidata.org/

- Try loading some example queries

Practical - Editing a query

Step 1: Load the sample query: http://tinyurl.com/ycxw4eyw

Step 2: Modify the query to find a different set of results, by:
- Changing values

- Changing properties

- Removing lines

Step 3: Share your query on Twitter and/or add to etherpad!
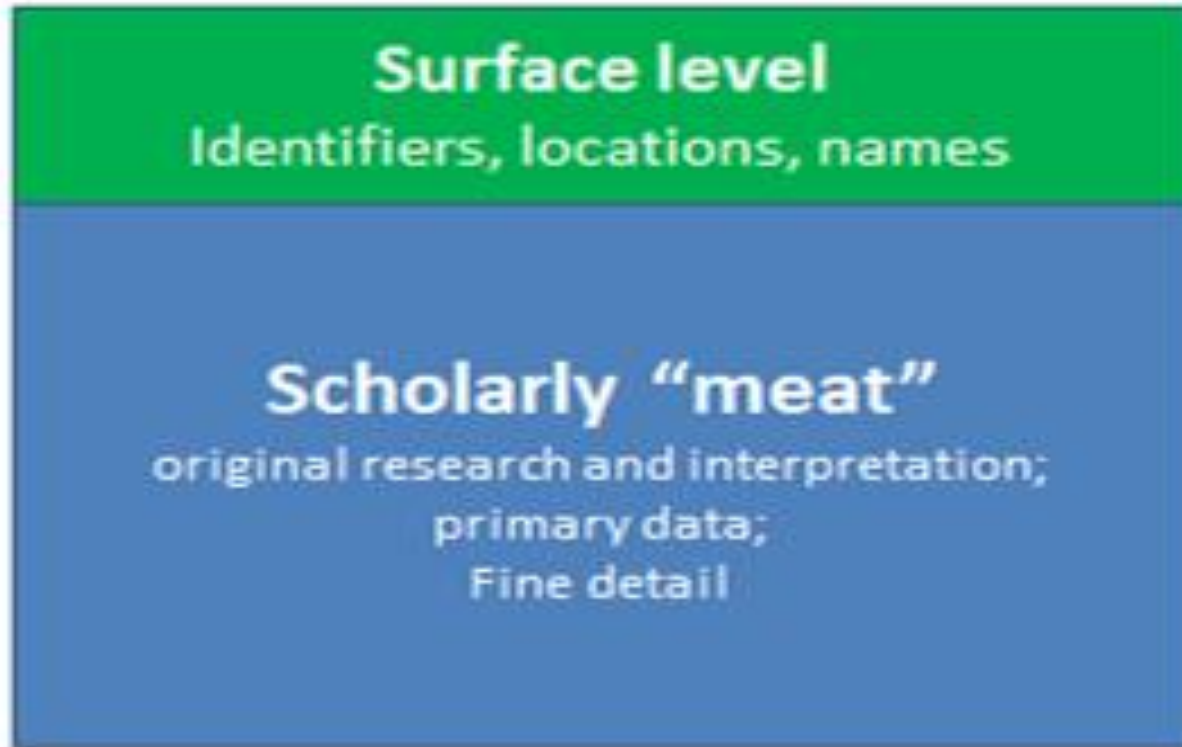
Anything is possible!

CC-0 licensed data so you can build what you want from it.
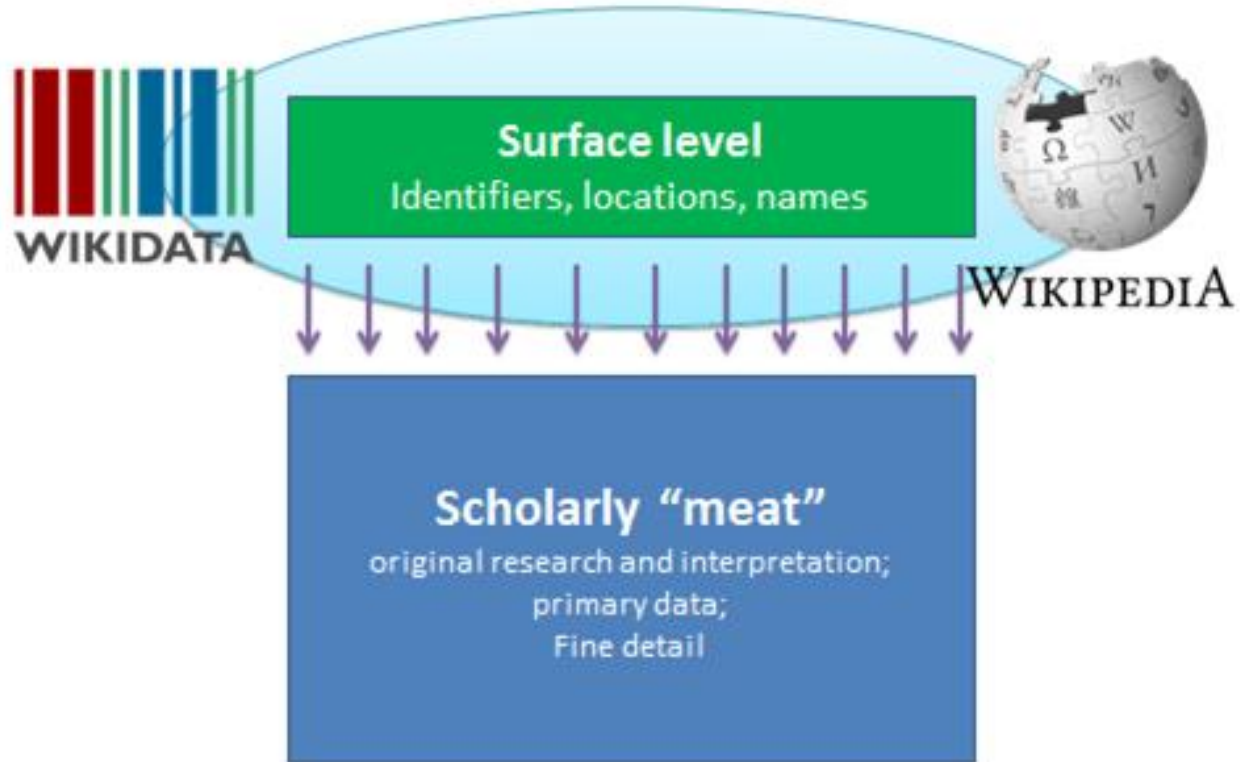
Possibilities are limitless.

Ways to take it forward.

Thanks for listening!

Why contribute to another repository?

Enrich both repositories by combining datasets.

Surface level
Identifiers, locations, names

WIKIDATA

WIKIPEDIA

Scholarly "meat"
original research and interpretation;
primary data;
Fine detail

Edinburgh - data capital of Europe

University to play pivotal role in creating Data Capital of Europe

The University of Edinburgh will be a key partner in a major initiative that seeks to make Edinburgh and its surrounding region the European leader for applying data science to products and services.

The move to transform the region into a digital powerhouse is a key strand of the recently announced Edinburgh and South-East Scotland City Deal.

Under the agreement, significant investment from the UK and Scottish Governments will be provided for major infrastructure projects in Edinburgh, the Lothians, Fife and the Borders.
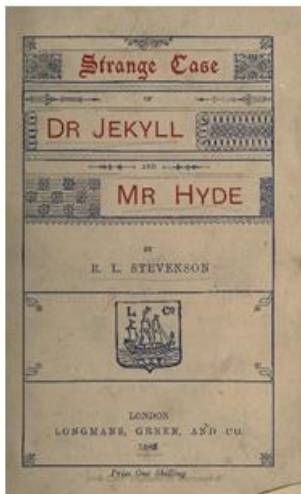
What data can we share?

flickr

Robert Louis Stevenson(Q1512)

THE UNIVERSITY of EDINBURGH

LIBRARY

Strange Case of DR JEKYLL and MR HYDE

BY R. L. STEVENSON

LONDON
LONGMANS, GREEN, AND CO.
1886

Price One Shilling

National Library of Scotland
Leabharlann Nàiseanta na h-Alba

WIKIMEDIA COMMONS

OpenStreetMap

HISTORIC ENVIRONMENT SCOTLAND

WIKISOURCE

OCLC WorldCat®

NATIONAL GALLERIES SCOTLAND

Multiple gauntlets thrown down at once!

What can you share?

- Bibliographical data
- Biographical data
- Biomedical medical
- Geographical data
- Taxonomical data
- Authority file data

The sharing of simple facts and statements costs nothing and benefits us all.

Wikidata

The free and open knowledge base

Thank you!

Repo Fringe 2017

ewan.mcandrew@ed.ac.uk

navino@histropedia.com