



**open (meta)data portals
from the data scientist's
point of view**

June 2017



Data in my line of work



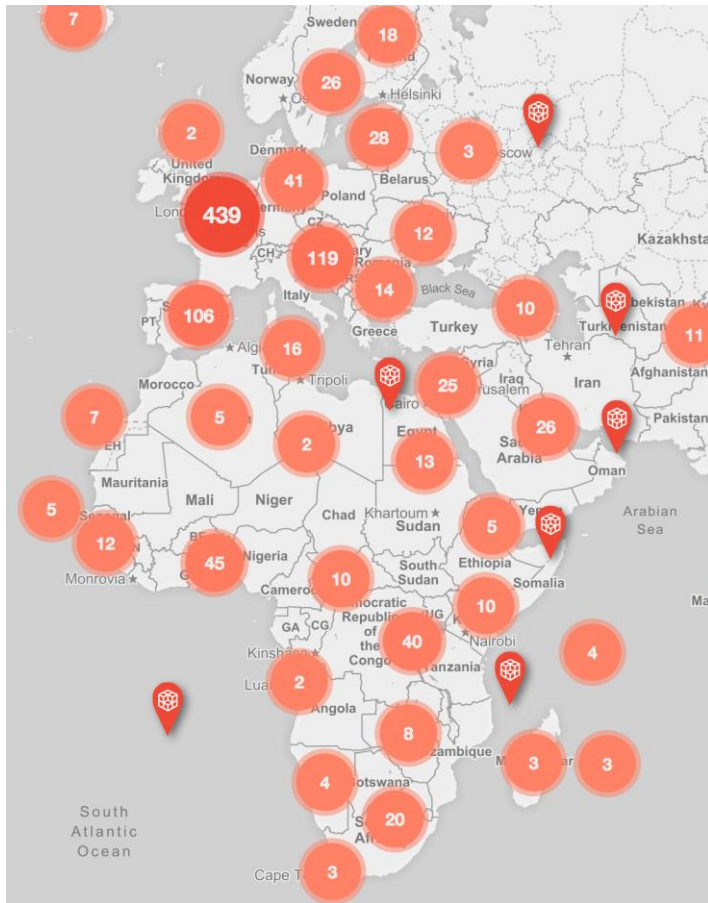
credit:John Hogg / World Bank

The screenshot shows the DATA.Ug website interface. At the top, there is a navigation bar with 'Home', 'Maps', 'Datasets', 'Request', and 'News'. Below this, the 'Datasets' section is active, displaying search results for the query 'education'. The results are categorized by Organizations, Groups, and Tags. The top result is 'Net Enrolment Rate In Secondary Education 2007' with options for WFS and WMS. Below this, there are more results for 'Net Intake Rate In Secondary Education 2007' and 'Net Enrolment Rate In Secondary Education.Csv 2004'. The bottom part of the screenshot shows the KENYA openData website, which has a similar search interface. It displays a map of Kenya and a list of results for 'education', including 'CDF Education Sector Expenditure' and 'Education Expenditures for 2002-2009'. The results are filtered by tags and source, and the order is set to Relevance.

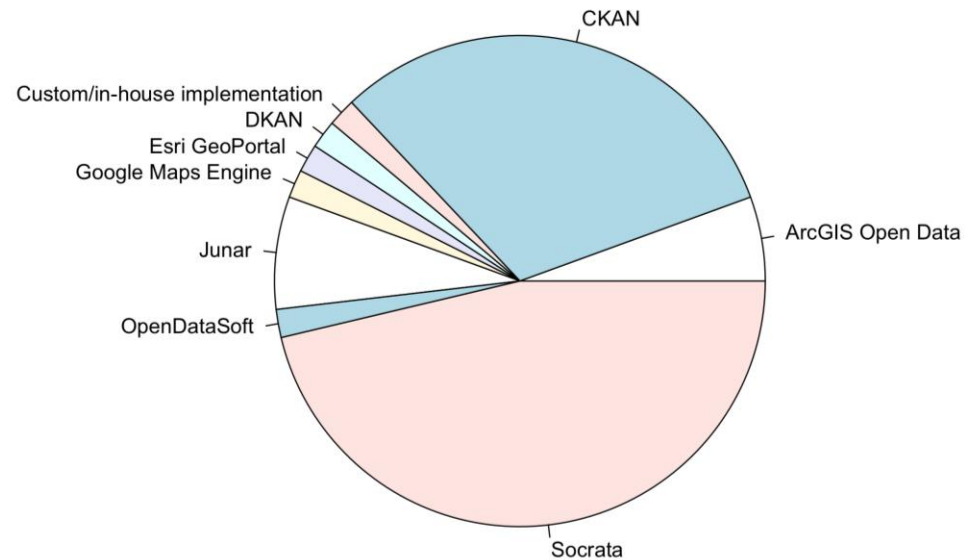
https://www.youtube.com/watch?v=wJTOo_1IFFE



Open (meta)data portals?



Source: opendatainception.io



W3C ‘Data on the Web Best Practices’ (w3.org/TR/dwbp/) encourages publishers to couple data with metadata at the time of publication. The majority of open data publishers respect this rule. However, the format in which metadata is published depends highly on the open data portal that publishes it.

How can you discover and reuse data through federated search when no common metadata schema is used across platforms?

Open data portals – what are the short-term solutions?



1. **Derivation**: when a new schema is created from an existing one, in which case the basic structure and common elements are conserved.
2. **Application profiles (AP)**: this approach ensures similar structure and common elements. APs usually emerge as a response to accommodating specific needs.
3. **Crosswalks**: entails mapping of the elements between metadata standards.
4. **Switching-across**: using switching schema as a switching mechanism among multiple schema. This is an alternative to crosswalks.
5. **Metadata framework**: a skeleton of metadata that can be used to concentrate efforts of a variety of actors towards a common standardisation in a given field.
6. **Metadata registry**: collection of metadata schemas.¹

Example: [Open Data Interoperability Platform](#) (ODIP)



Conclusions

Our July 2016 [consultation paper](http://juds.joinedupdata.org/consultation-paper/)¹ uses a simple linguistic analogy to deal with the lack of interoperability between data standards – you should either speak the same language or use a translator. The same sentiment becomes clear from the literature review of the metadata standards and the approaches to dealing with the lack of interoperability as listed above.

What is needed is a standardised way of publishing metadata by the open data portals:

- to allow users to find the context of the data in a simple and accessible way
- to allow a federated search across platforms so that the true power of metadata can be unleashed: the power of machine-readable discoverability (federated search) of data.

1. <http://juds.joinedupdata.org/consultation-paper/>

Thank You !

Beata Lisowska

Data Scientist, Joined-up Data Standards

@lissek

juds.joinedupdata.org

www.devinit.org

