How can DCAT be used to address the needs of databases and complex visualisation tools?

Beata Lisowska (@lisssek)
Development Initiatives (devinit.org)
- Online resource for the discovery of financial and resource flow data.
- The Data Hub would benefit from a additional **machine-readable layer of context/metadata**.
- Development Initiatives (DI) is **committed to improving the interoperability** of all development-related and humanitarian data.
- **Can DI adhere to the W3C principles (Data on the Web Best Practices) of publishing metadata in a human- and machine-readable format?**
Development Data Hub & Data Warehouse
This growing complex of interconnected data serving a range of digital products poses three problems:

1. How does DI, and how do Data Hub users, keep track of what data is available and whether it is up to date?

1. The intellectual credibility of DI’s work depends on metadata that explains the provenance and methodology of its analysts’ calculations.

1. The joined-up ‘raw’ databases in the warehouse will, in future, become a public good with an open API. How will third-party developers wanting to make use of this repository access the metadata they will need to accompany the data they extract?

Can DCAT be the answer to these problems?
Step 1

Treat front-end (Data Hub) as a data catalog!

Pros: each dataset can be described by dcat:Dataset in a conventional use.

Problem: how to select only the datasets that were used to build a visualisation?

Solution: Describe all datasets but pull out the relevant ones using dct:Identifier

This solution does not describe the Data Warehouse that is the heart of the tool!
How to capture the complex journey of a data through a data warehouse?

**Solution:** Treat each data series as dct:Dataset and each data source as a data catalog.

**Pros:** Using dct:Source each original data source could be mapped to its correct origin such as World Bank WDI or OECD DAC data.
The bigger challenge

How to merge step 1 and 2?

Could/should the Development Data Hub (left) be linked using **dct:Source** to the Data Warehouse (right)?
Conclusions

Data on its own, without contextual information or links to other similar sources, often proves difficult to analyse or interpret.

Application of metadata standards, such as DCAT, can provide a challenge if applied to complex systems.

Metadata should provide a machine-readable map to make information on a journey of a single data point from its origin to its final destination available and traceable across platforms and data producers.

A data warehouse drives the data published on the web and as such should also be comprehensively described by a metadata standard.
Thank you!

Contact me:
Email: beata.lisowska@devinit.org or Twitter: @lisssek

Our project:
Email: info@joinedupdata.org or Twitter: @Joinedup_Data

We are currently seeking comments on our consultation paper:
http://juds.joinedupdata.org/consultation-paper/