



RESEARCH DATA ALLIANCE

Data Foundations and Terminology Interest Group

**Breakout Session at 8th RDA Plenary
16th September 2016, 15.30-17.00**

Data Foundations and Terminology Interest Group Breakout Session 8th RDA Plenary

Friday 16th September 2016, 15.30-17.00 p.m.

Gary Berg-Cross

- Gary reports about the former of the working group. After the WG, there was still interest in the DFT work and the DTF Interest Group got founded
- The IG DFT works together with other RDA Working / Interest groups in joint meetings
- Domain specific groups are also interested in sharing their vocabulary: there is a close collaboration between the DFT IG and the VSIG and BoF Domain Vocabularies
- Since the last plenary, there were updates in the vocabulary, especially legal Interoperability issues were addressed

Raphael Ritz

- Reports on the last changes in the TeD-T tool, especially various download formats (RDF, PDF, HTML, Plain Text, TreeTagger annotations)
- PID assignment is the next function to be implemented in TeD-T: Still open questions are how to handle granularity and versioning.
- Further open issues: connections to other projects in RDA (DTR, PIT, Collections WG, Dynamic Data ...)

Ahmed Eleish

- Project: Data driven vocabulary development, using native language processing tools (NLP) to create domain specific vocabularies. Eventually interesting terms can be identified automatically from RDA or other domain specific documents
- Discussion:
 - Definitions of the chosen terms still has to be done manually, maybe some semantic approaches can be interesting
 - Maybe possible to map identified terms against existing definitions
 - Existing tool suites and models can be used to do the task
 - Terms in the middle of the freq. distribution are of interest
- ISO activities: official ISO vocabulary will be published as SKOSS. Relations to RDA: terms were adopted from DFT vocabulary

Misc

- It should be possible to define relations between terms
- PIDs should be assigned to every individual definition so that linking between different vocabularies is possible
- Mutability: things are changing over time, it should be possible to mark definitions or terms as deprecated
- Interoperability between terms: sometimes definitions are very close to each other but still not the same: how to make links between definitions?
- Translating terms and their definitions to other languages than English
- Versioning of stable snapshots of the DFT vocabulary is necessary because term definitions can change over time

Tom Zastrow, thomas.zastrow@mpcdf.mpg.de