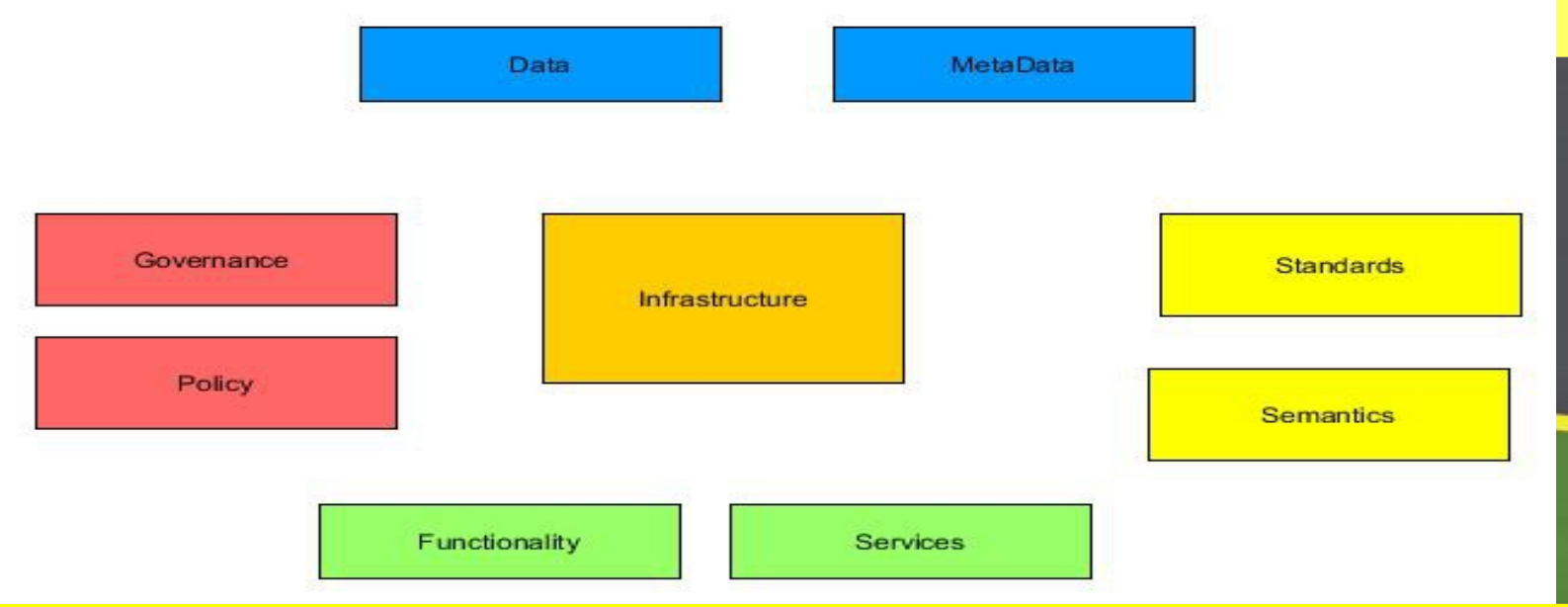
RESEARCH DATA ALLIANCE

Data Foundations And Terminology (DFT) IG



DFT IG Breakout Session P 11 Breakout - Wed, 21 March, 16:30 - 18:00

Co-Chairs DFT IG : Gary Berg-Cross & Raphael Ritz

Overview of Objectives for P-11

1. Updates and Continue IG **discussion –**

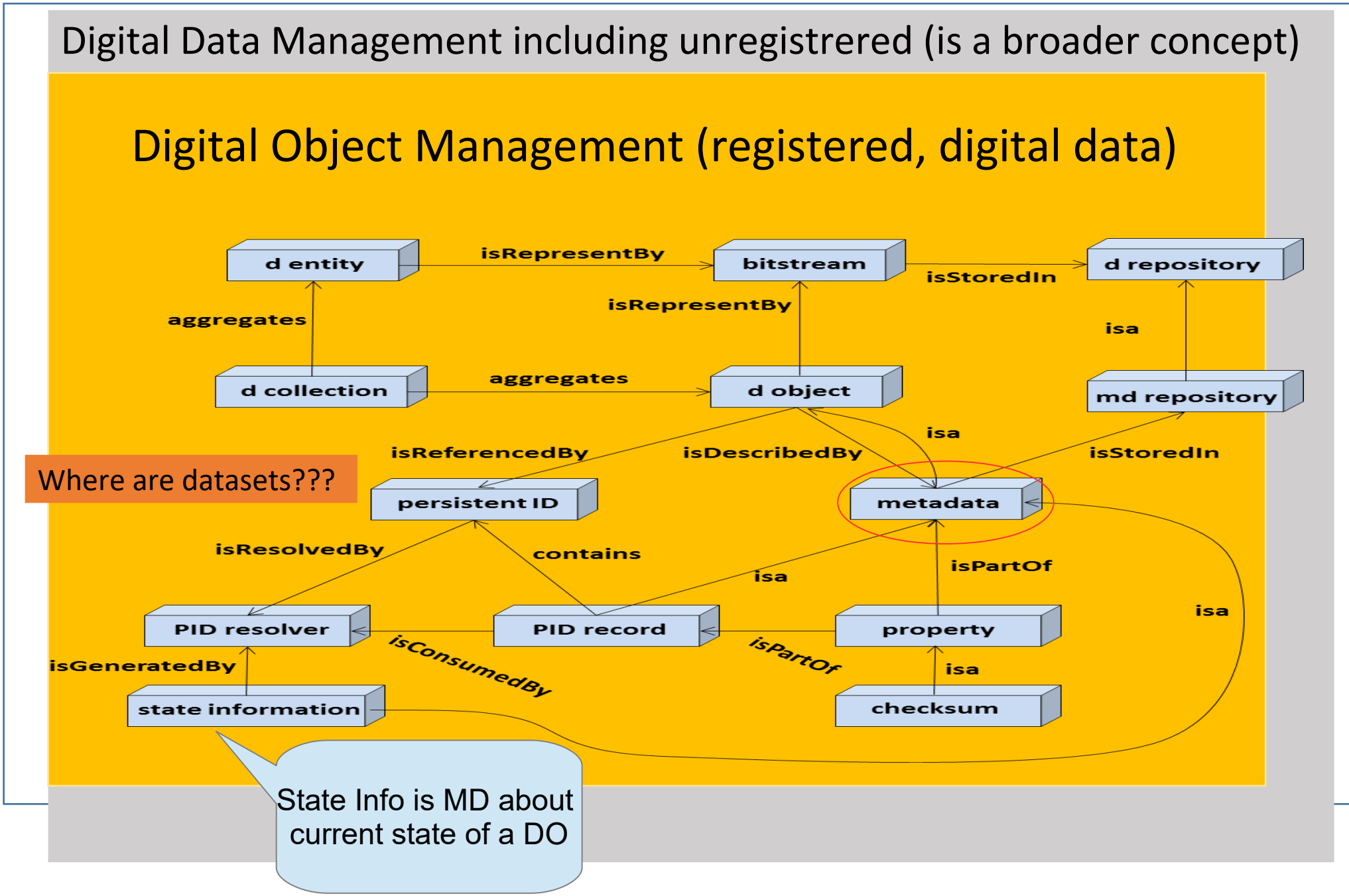
- Who is completing work and has vocabularies?
- How do they relate to each other?
- We now can handle PIDs for each term, who wants to use them?
- We have Official status under the EU public procurement legislation: "Common Technical Specification"

2. Faclitate community discussion on RDA/group core concepts
   - Building on our base help systematize the already large body of domain definition work on terms and their meaning **using a r*ationalized "consensus"*** knowledge of domain experts, especially for RDA's efforts.
   - Collaborate & coordinate with "external" vocabulary efforts – repository?

3. Exporting using our vocabulary for build informative knowledge graphs

4. New discussion of a **Vocabulary** List Registry
   - Is there interest?
   - What are the needs and what services that can be provided for a Registry?
     - Thesaurus service may help some but others need something stronger and may be able to leverage activities between different Vocab groups.

Out Base is Concept map overview of Core Terms Broadening the Discussion (Stepwise & Scope-wise)



Working Agenda

1. DFT Objectives, Overview & Update (Gary Berg-Cross - handout)

    » Vocabulary Updates & liaison relation to other RDA Groups for candidate vocabulary items.

    - We have Official status under the EU public procurement legislation: “Common Technical Specification”
    - Comply with Regulation No 1025/2012, Annex II
    - See https://datashare.mpcdf.mpg.de/s/0rq5kVmMlv0h41X for a briefing on this.

2. Tool Update (Raphael Ritz) – handling IDs for term concepts

3. RDA self analysis and working relation with MIG, DDRI, Collections.....

    - Examples of RDA analysis using definitions

4. Issues and Interested Parties Discussion

    – (DDRI) use of vocabularies and terms as part of group analysis

5.Liaison relation to other data vocabulary efforts to develop a common Registry

6. Next steps

Metadata types: Some Preliminary DFT non-mutually exclusive definitions

1. Administrative Metadata is a type of Metadata the provides information to help manage a resource, such as when and how data was created, a file type and other technical information, and who can access it.
2. Authenticity Metadata is a type of metadata that conveys information needed to link a data object to its original source with integrity. Authenticity is provided by appropriate metadata, within an archive & digital retention and preservation context. It results from verifying that a digital object & its state information has not changed.
3. Citation Metadata serves the role of identification and should provide an unambiguous identifier to the data cited, its location, and means of access.
4. Detailed Metadata is defined in distinction to simpler or light forms of metadata that provide some basic information about data, such as in Dublin Core, but which can supplement this simple information.
5. Discovery Metadata is metadata whose chief role is to discover relevant data.
6. Extract Descriptive Metadata works by using a given data type, to access a data type registry and identify a procedure that can be used to parse the data object and then apply a template to extract desired information from the contents of the data object.
7. Key Metadata is information associated with a digital object (or entity) that is required for discovery. Thus it is a part of Discovery Metadata.
8. Objective Metadata is based on assertions of fact about such things as authorship, date of creation, & version. Broadly they include attributes can be assigned by what is considered an objective and reproducible (perhaps automated) process.
9. Minimal Metadata descriptions with very little curation including DO name & PID - only marginally targeted at the role of discovery since there is much better infrastructure to accomplish this.
10. Payload , Provenance, Rich, Structural System, Topical metadata …..

Some Oct- Feb Vocabulary Updates

Adaptable information Object A type of information object which represents research data and includes documentation with contextual information that enables the reuse of the original data in contexts different from those for which the object had initially been created.

Data Understanding is that part of the data lifecycle in which data resources are identified along with data reuse and integration plan, datatification, an decision are made about tools to use

Data Lifecycle/Lifecycle is the sequence of processing that a data undergoes from its creation, documentation through its storage in a repository and eventual disposal.

Data Science is the scientific study of the generalizable extraction, organization and interpretation of information and knowledge from data. It works across all the steps of a data science lifecycle.

Working Relation with MIG, DF IG & Chairs Collaboration

- Held virtual meetings over the Winter and discussed vocabularies at Chairs Meeting

Metadata Element Set (continues to provides some input for DFT vocabulary but need improvements):

1. Unique Identifier (for later use including citation)
2. Location (URL)
3. Description
4. Keywords (terms)
5. Temporal coordinates ??
6. Spatial coordinates ??)
7. Originator (organisation(s) / person(s) -roles of agents, Orcid ID8. Project
9. Facility / equipment
10. Quality
11. Availability (license, persistence)
12. Provenance
13. Citations
14. Related publications (white or grey)
15. Related software
16. Schema
17. Medium / format

Other Data Management Vocabularies

opportunities for collaboration, coordination, and de-duplication of effort.

Despite decades of intensive work on controlled vocabularies (standardized sets of terms) problems remain with definitions that are central to RDM.

The important need for clear definitions of RDM terms is widely recognized

RDA's Data Foundations and Terminology (DFT) WG is one of the earlier initiatives.

Other important efforts include:

- Science Europe Data Glossary;
- Data Documentation Initiative (DDI); and
- Research Data Canada (RDC)/CASRAI RDM pilot
    - evolved into a new International Research Data Management glossary (IRiDiuM) supported by RDC, CASRAI, and CODATA.
- Update from Big Data at NIST and IEEE workshop...

A Data Vocabulary Registry

In light of so many different data vocabularies as mentioned there may be a role of a registry for data vocabularies. Such things exist for metadata and ontologies.

This could be used for many things from helping to find them to promoting discussion, sharing and exposing differences in scope as well as specific definitions.

As discussed in DFT virtual meetings RDA seems well situated to help or even take on this effort.

It might require being a WG and this so we are open to discussion at this Plenary.

Policy Components - Conceptual Fundamentals

Policy-based Data Management Concept Graph