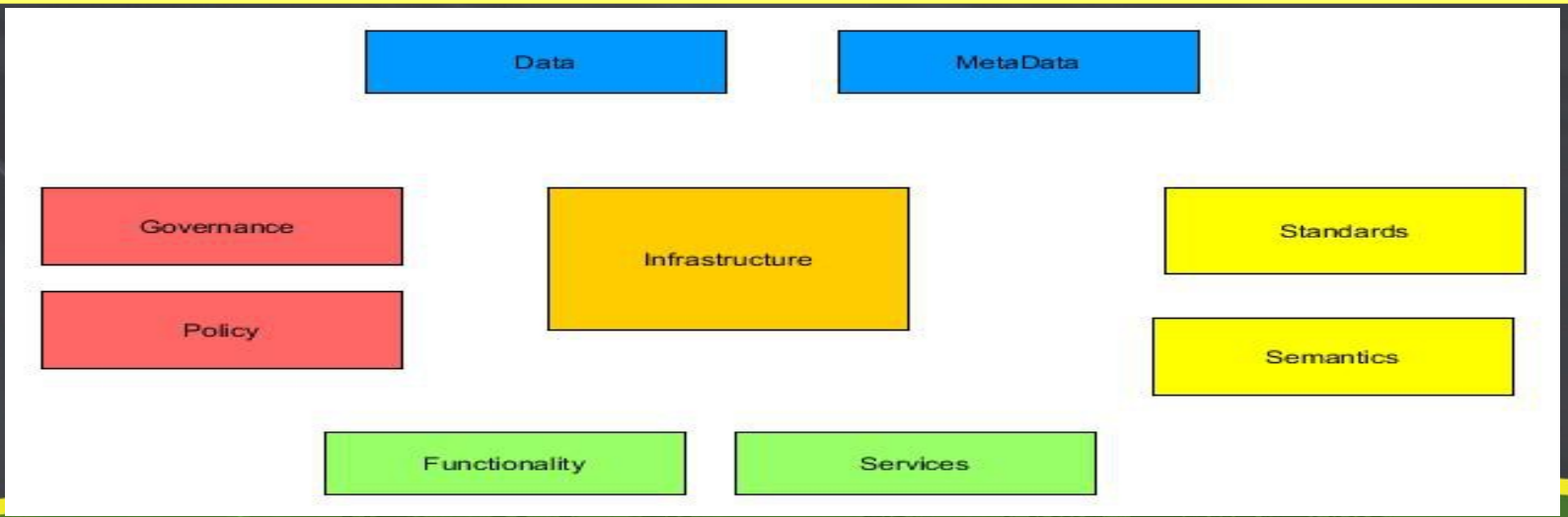
RESEARCH DATA ALLIANCE

Data Foundations And Terminology (DFT) IG



research data sharing without barriers

rd-alliance.org

DFT IG Breakout Session P 12 Breakout - Tuesday, 6 Nov., 11:30 – 1:00

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

Overview of Objectives for P-12

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

- Who is completing work and has vocabularies?
- How do they relate to each other?
- How do we organize our defined, termed concepts?
- 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. Organization of Termed Concepts

4. "Continued" 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.

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. Examples of RDA work

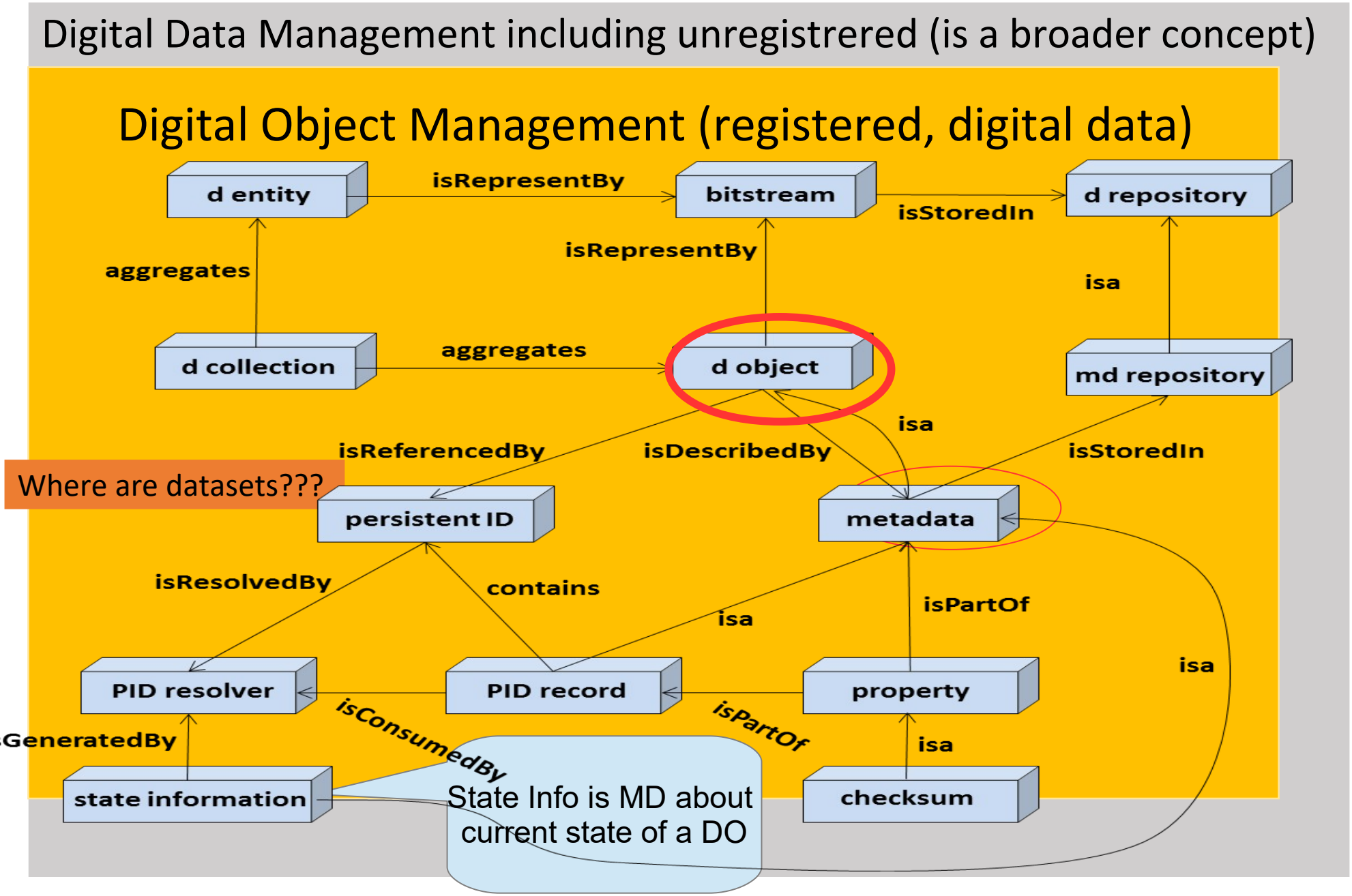
4. Issues and Interested Parties Discussion

    What is the best way to organize the RDA data vocabulary?

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

6. Interest in broader vocabulary effort and Next Steps?

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



Some Vocabulary around P8

- *Collections* and various types of metadata were actively discussed
  - *Descriptive metadata* is a type of metadata that describes a resource for purposes such as discovery and identification and **Contextual Metadata**
  - *Curation* metadata describe who supports a curated resource and its availability.
  - *Data Citation Metadata* is a type of metadata/Administrative metadata that plays the role of citing a dataset in an analogous way that books or journals or a computer, access date, version number, and a persistent identifier or locator. Metadata that maps to DataCite schema or Dublin Core Terms etc.
  - *Domain Metadata* or domain-specific metadatais non-general metadata used to capture domain information as reflected in domain vocabularies and when possible domain-specific metadata should maps to metadata standards used within a scientific domain.
  - *Rich Metadata* describes data with enough accurate and relevant attributes to make it easily findable.
  - *Key Metadata* is information associated with a digital object (or entity) that are required for discovery.
  - *Metadata Catalogue* A type of data catalog (catalogue) used to access informatation about data
- Data Transparency; Fair Use; Patent; Copyright Infringement; Rights Statement; etc.

Metadata types: Some Preliminary DFT non-mutually exclusive definitions

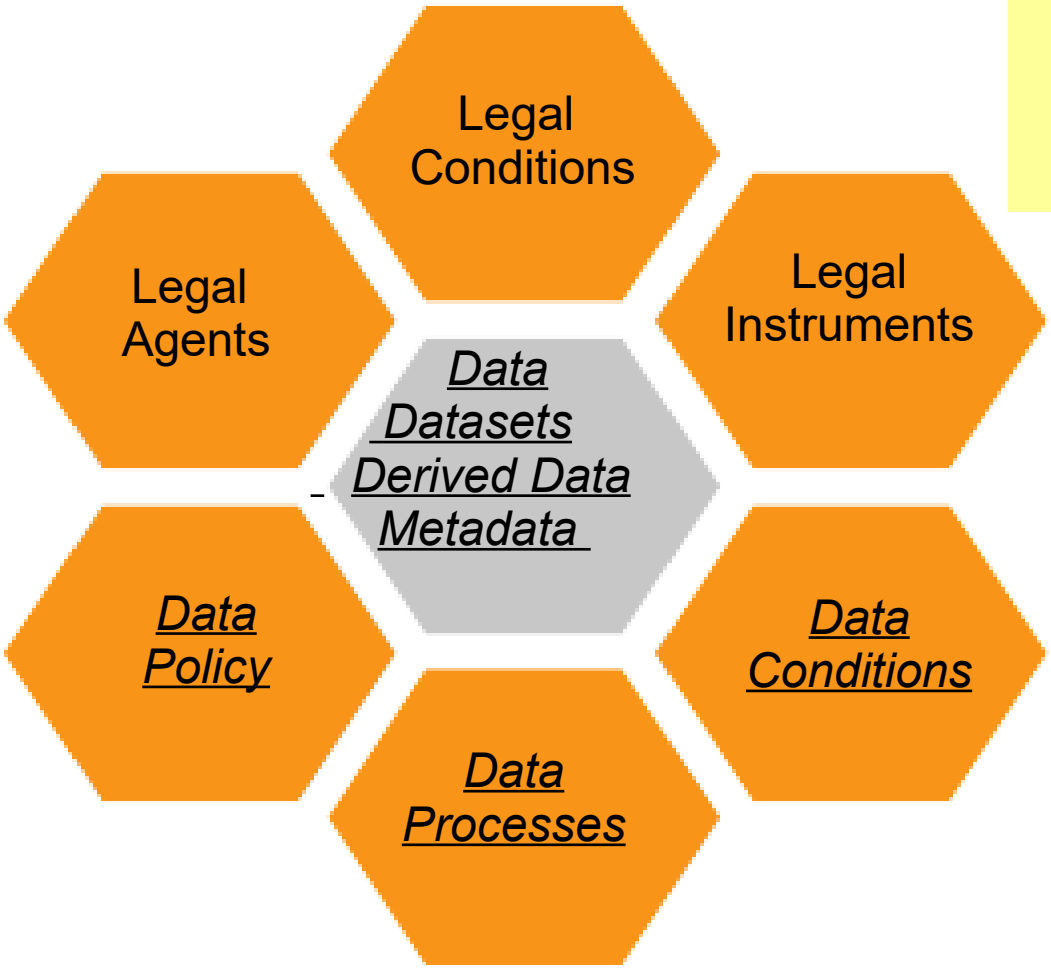
1. Administrative Metadata is a type of Metadata that provides information to help manage a resource, such as when and how data was created, a file type and other technical information, and who can access it.
2. Authenticity Metadata is a type of metadata that conveys information needed to link a data 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 …..

Datacentric Uptake Example: “Legal Interoperability” Conceptual Areas

Rights
Individual Contracts
Use Agreements

Executive Agreement
Policy Regulation
Legislation
Rights Statement

Data Producers
Data Users
Government
Agency
Research
Organization
Rights Holder



Research Policy
Inter-Governmental Policy
Non-Governmental Policy

Access
Open Data
Harmonization
Dissemination
Reuse
Fair Use
Restriction

Recent Additions/Edits Concepts that RDA is "talking about"

September

Primary data; Value lists; Controlled Vocabulary; Data end users;

Reference linking; Metadata schema; XSD

July and August

RDF Dataset Profile; Dataset Profile Feature; Publisher-facing policy, Packaging;

FAIR-compliant; Repository certification

June

Data Sharing (added principles); Metadata schema; Discoverability;

Data standards; Preservation; Data Curation

May

Data Publication; (from the NLM data thesaurus.)

Linked Data; Data Catalog;

C2CAMP – had many new terms.....

Science Ecosystem, digital object model, tightly associated metadata, Object typing, Mapping services, Global Digital Object Cloud, virtually aggregated digital objects, Solution Space ?????

RESEARCH DATA ALLIANCE

Main page
Add Term
Add Category
Stable Releases
Download in various formats

Browse Term Collection
All Terms - Hierarchical
All Terms - List
List by scope
Recent populated terms
Ted-T Graph

Help

Tools

Category:RDA Term Collection

This category uses the form **RDA**.

Subcategories

This category has the following 4 subcategories, out of 4 total.

I

▶ Infratructure (22 P)

O

▶ Organizational (1 C, 142 P)

P

▶ Procedural (1 C, 62 P)

U

▶ Uncategorized (95 P)

Pages in category "RDA Term Collection"

The following 17 pages are in this category, out of 17 total.

A

- Administrative metadata

B

D cont.

- Data Librarian
- Data policy
- Data Professional

R

- Raw Data
- Representation object
- Research Stakeholder

Current Organization of Terms

Uncategorized

Mashup

Procedural

Access Control

Infrastructure

Authentication
System

Aggregations

Active Collection

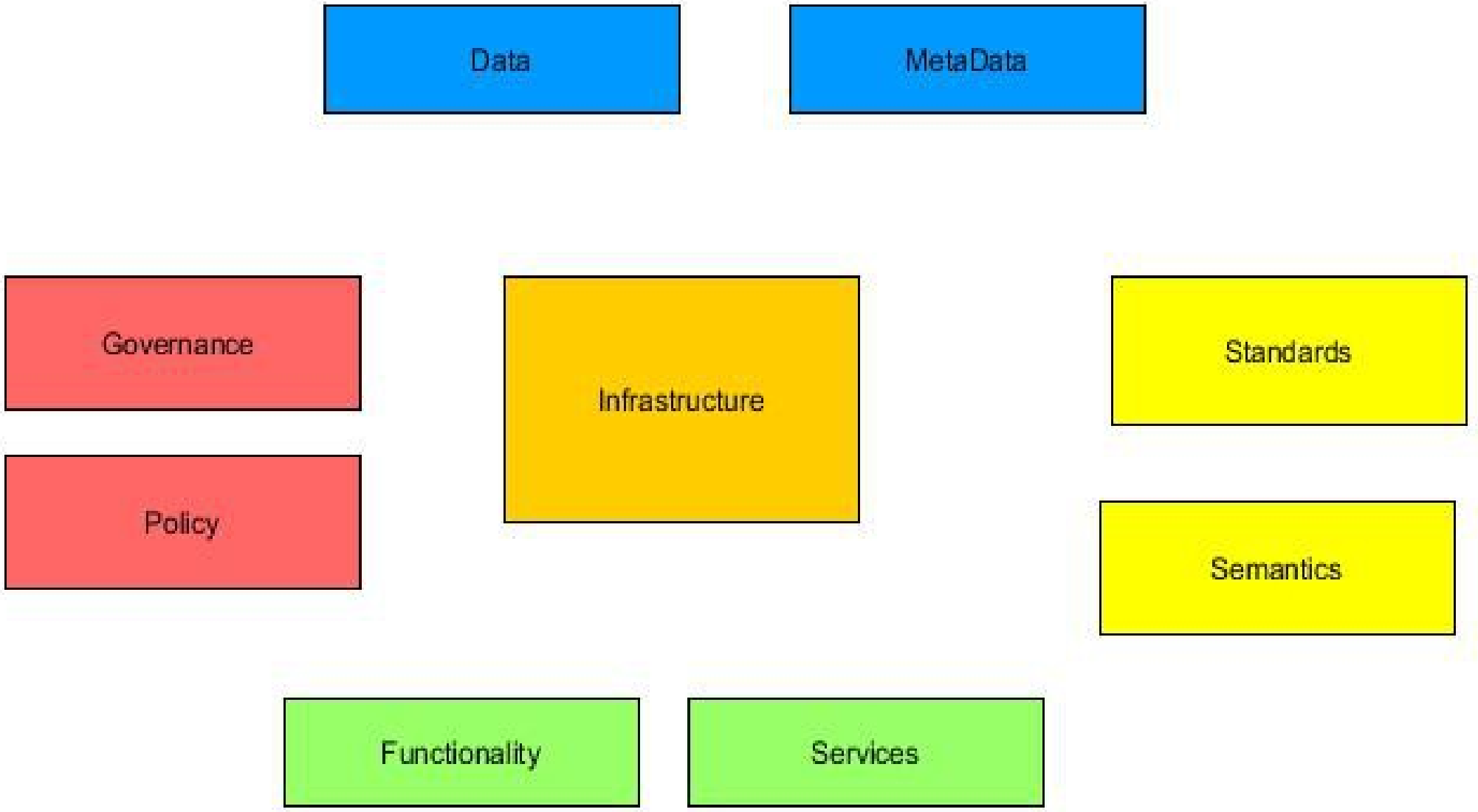
Operations

Operation

Organizational

Digital Data Object

Alternative Organization Of Termed Concepts



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.

Spawn a Wider Interest Group on Vocabularies?

Raised in BoF on best practices for use of vocabularies.

DFT IG seen as one place to discuss creation of a new, wider group.

Thoughts??

Backup Slides

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...

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