



**WG/IG Collaboration Meeting
January 3, 2018 NIST
Text Analysis Overview
Gary Berg-Cross**

Data Type
Registries WG

registries

data-types

metadata

cross-domain

interoperability

reusability

research data sharing without barriers
rd-alliance.org

Background from Göteborg

“The start-up phase of RDA is over..”

“RDA progress and future direction should be made more explicit through an annual Road-mapping exercise, developed in coordination between TAB and the Chairs, with feedback solicited from OAB and Council. “

Some of our Meeting Qs were :

- ◆ Where have we gotten in 4 years and what is missing?
- ◆ What is actionable/adoptable, how do they fit together?
 - ◆ What is missing?
 - ◆ Does RDA have a core? Does it have many cores?

Context: we have some data about RDA (WGs, IGs, members, organisation, etc.).....

- ◆ *But can we answer these Qs as a volunteer effort?*

Ad Hoc Analysis Group Leveraged “Available” Data

Some Topics for Analysis to Cover

3

- ★ Overview and history of RDA groups,
- ★ Interactions between groups (current, past, potential),
- ★ Some idea of trends in the above, such as who is a lineage parent of a WG?

- ★ Supported by:
 - ★ A high-level topical view of what RDA is about. (word topics)
 - ★ A medium level view that shows the terms defining the topics. (DFT)
 - ★ A high-level group view of RDA. (Group history graph)
 - ★ A high-level view that summarizes what topics RDA groups are pursuing. (word topics)
 - ★ View that shows shared topical relations between RDA groups.
 - ★ Medium level view that shows joint meetings among RDA groups.
 - ★ Medium level views that show influence of prior groups on later .
 - ★ Trending analysis for important initiatives such as FAIR.....

- **Earlier Background - RDA Vision Statements**
What are we trying to achieve?
- **Earlier Analysis & the 6 words effort - what we say and show**
- **Diagrams**
 - Words and making sense of what we understand
- **Just a Start some Deeper Analysis?**

What is the RDA Vision? Improving what we say ⁵

The very general RDA vision (for improved branding)

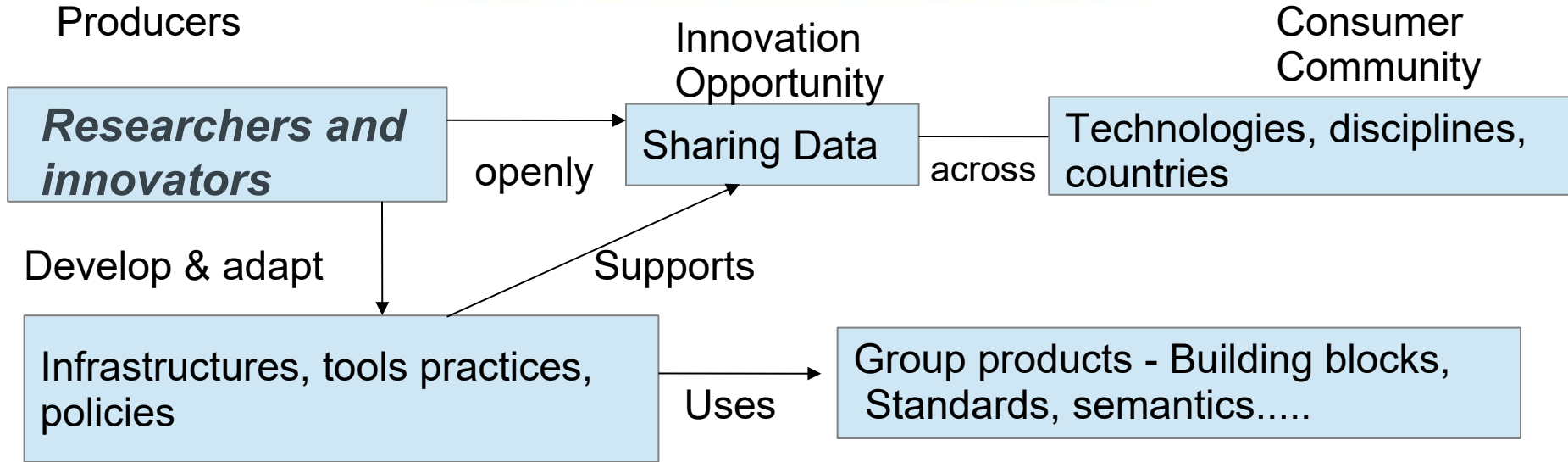
We say we are about

- 1. researchers and innovators openly **sharing data** across technologies, disciplines, and countries to address the grand challenges of society*
- 2. develop and adopt **infrastructure** that promotes data-sharing and data-driven research, and accelerate the growth of a cohesive data community that integrates contributors*
- 3. "Enabling infrastructure provides the building blocks for data-driven innovation,..." The tools, policy, and practice that make data usable and **interoperable** are critical"*

From Rda in a_nutshell_january_2017 etc.....

What we say and show...

Addressing grand challenges of society



Predication??

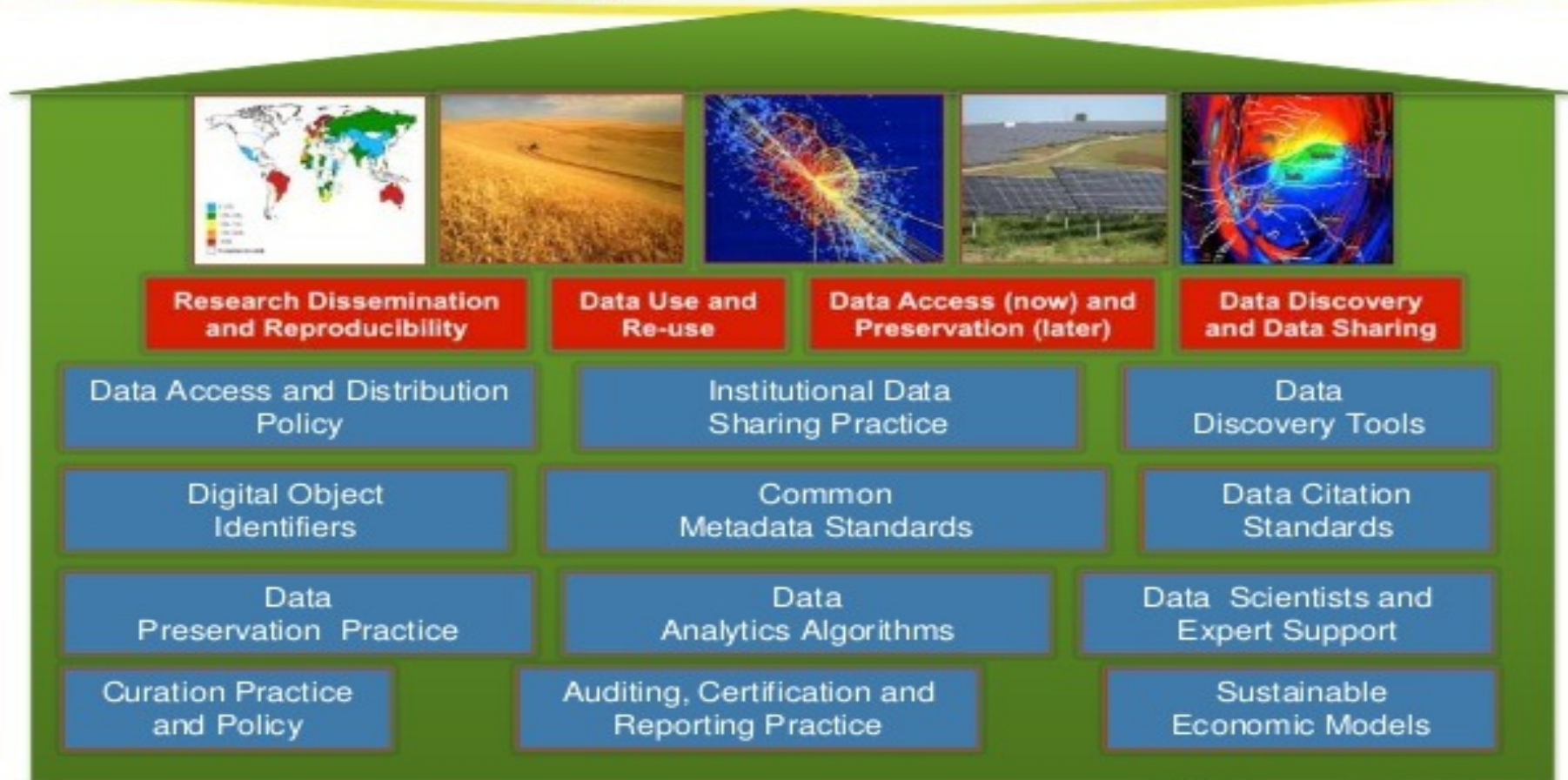
Tagging???

Summarization?

(Argument Network)

More Detail? We Show Loose Infrastructure Area Populated by Building Blocks

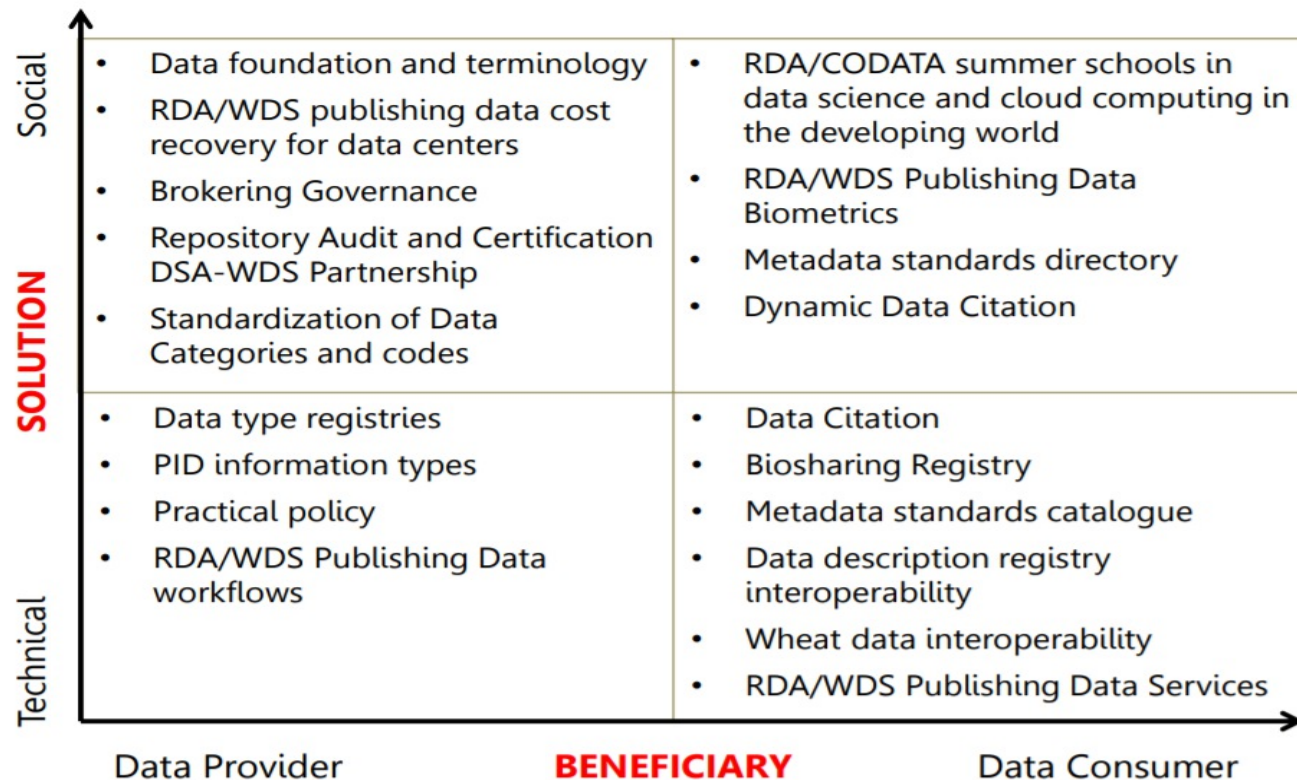
Many Infrastructure Building Blocks Needed to Accelerate Progress



Attempts to Categorize Groups

- Socio-Tech & Provider-Consumers*

Focus of RDA Working Groups: Build infrastructure and support its adoption by communities of use



* Also use Domain & Community or Reference vs Partnership Groups

Early Use of Key Words for Groups

WG IG Name	Keywords
Community Capability Model IG	; Data Management, Data Literacy, Education
Data for Development	; Discovery, Knowledge Organization / Representation, Education, Data Literacy
Development of cloud computing capacity and education in devel	Education, Research Practices
Long tail of research data IG	; Interoperability, Data Fabric, Knowledge Organization / Representation
Quality of Urban Life Interest Group) Governance, Education, Values / Ethics
RDA/CODATA Summer Schools in Data Science and Cloud Com	Education, Data Literacy, Research Practices
Agriculture Data Interest Group (IGAD)) Interoperability, Data Fabric, Knowledge Organization / Representation
Big Data IG	; Big Data, Interoperability, Discovery, Values / Ethics
Biodiversity Data Integration IG	; Interoperability, Data Brokering, Knowledge Organization / Representation
Brokering IG	; Interoperability, Data Brokering, Knowledge Organization / Representation
Data Description Registry Interoperability (DDRI)	; Interoperability, Brokering, Registry
ELIXIR Bridging Force IG	Interoperability, Data Brokering, Infrastructure
Marine Data Harmonization IG) Interoperability, Data Brokering, Infrastructure
Metadata IG) Knowledge Organization / Representation, Discovery, Metadata
Structural Biology IG	; Data Brokering, Knowledge Organization / Representation, Metadata
Toxicogenomics Interoperability IG	; Interoperability, Data Brokering, Data Management, Metadata
Wheat Data Interoperability WG) Interoperability, Data Brokering, Data Management, Metadata

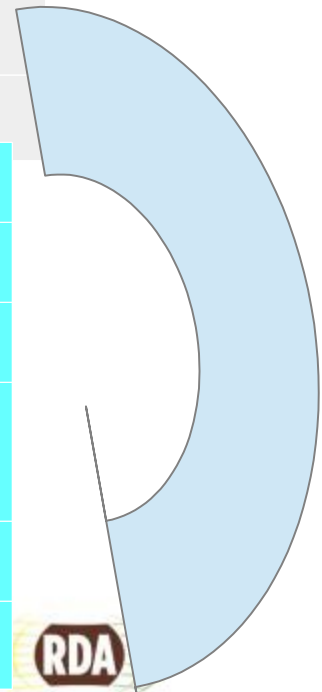
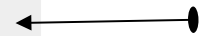
Term Assignment. Orange: social/consumer; Blue: technical/consumer. Terms chosen by group to describe activity more precisely than name alone.

The 6 Word Exercise -41 Groups

Research Data Repository Interoperability, BioSharing Registry, Data Discovery Paradigms, Linguistic Data, AgSemantics, Agricultural Data, Data Type Registries, Research Data Collections, International Indigenous Data Sovereignty, Global Water Information, Vocabulary Services, Libraries for Research Data, Metadata IG, Data in Context, Metadata Standards Catalog, Data for Development, TDWG Metadata Standards for attribution of physical and digital collections stewardship, ELIXIR Bridging Force, Small Unmanned Aircraft Systems, Chemistry Research Data, CODATA Materials Data, Infrastructure & Interoperability, CODATA Materials Data, Infrastructure & Interoperability, Early Career and Engagement, Repository Platforms for Research Data, Marine Data Harmonization, Data Citation, Rice Data Interoperability, WDS Assessment of Data Fitness for Use, Education and Training on handling of research data, Federated Identity Management, PID Kernel Information, PID, CODATA Legal Interoperability, WDS Scholarly Link Exchange, NDS, Active Data Management Plans, Archives and Records Professionals for Research Data, Ethics and Social Aspects of Data

Example - 6 Word Results

<u>RDA Group</u>	<u>List of Words</u>
Research Data Repository Interoperability WG	repository-platform
Research Data Repository Interoperability WG	research-data
Research Data Repository Interoperability WG	<i>interoperability</i>
Research Data Repository Interoperability WG	data-packages
Research Data Repository Interoperability WG	replication
Research Data Repository Interoperability WG	migration
Data Citation WG	data-citation
Data Citation WG	dynamic-data
Data Citation WG	data-versioning
Data Citation WG	persistent-identifiers
Data Citation WG	timestamps
Data Citation WG	reproducibility



Some Interpretation of 6 Word Results

<u>RDA RDR Interop Group</u>	<u>List of Words</u>
Research Data Repository Interoperability WG	repository-platform
Research Data Repository Interoperability WG	research-data
Research Data Repository Interoperability WG	<i>interoperability</i>
Research Data Repository Interoperability WG	data-packages
Research Data Repository Interoperability WG	replication
Research Data Repository Interoperability WG	migration

RDA Interoperability is about research data (maybe in a domain) and its packages. It involves research infrastructure/platforms and supports the functions or services of replication and migration (or reproducibility)

General Idea: enumerate the forms of data – packages (sets, collections..)
types of infrastructure
types of services

Agrisemantics WG:

is oriented towards Interoperability using better semantics for its Agriculture metadata vocabularies.

implied is the functions or services ,
for example replication and
migration



semantics

agriculture

interoperability

vocabularies

ontologies

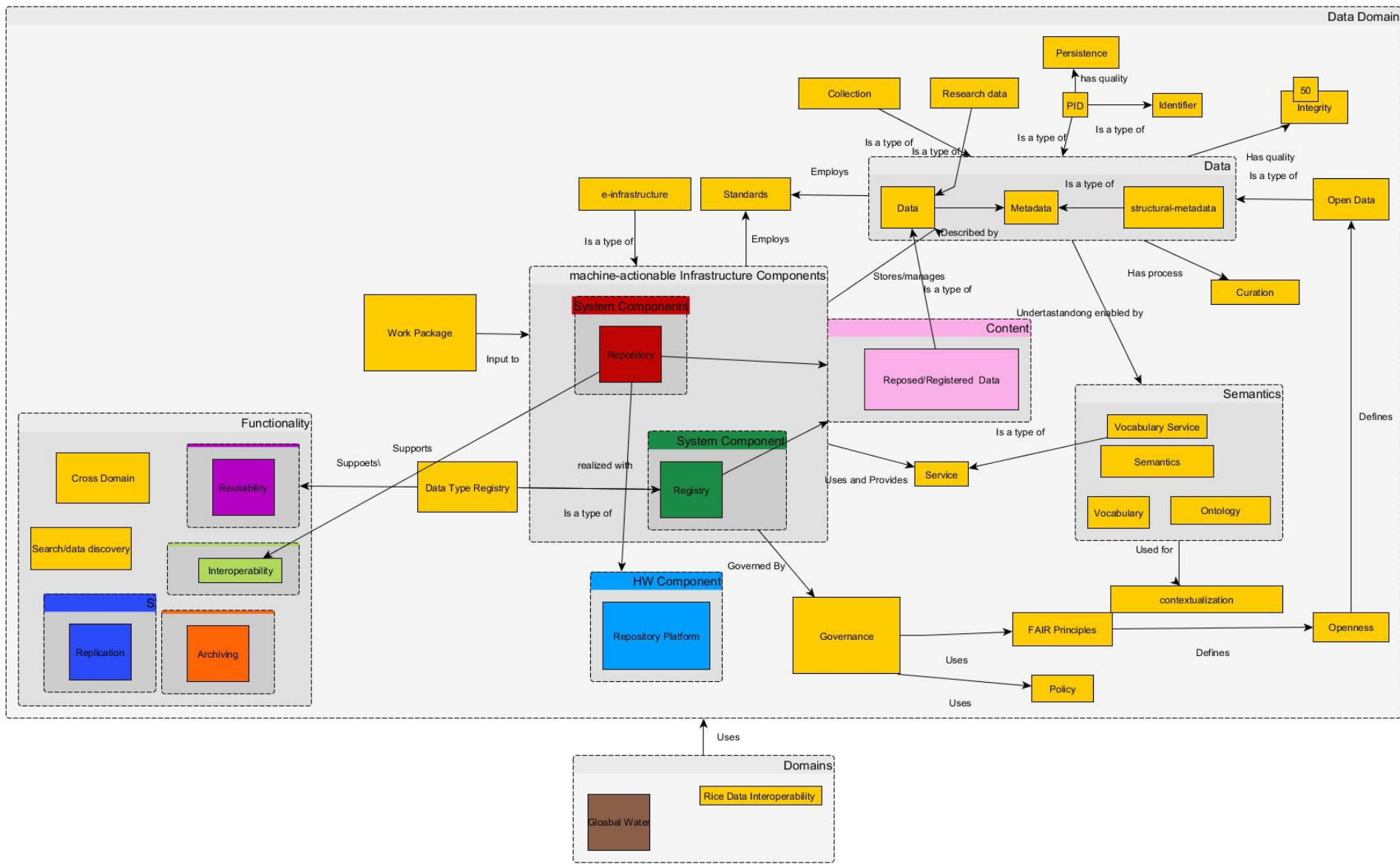
metadata

Adding Definitions - BioSharing Registry: connecting data policies, standards & databases in life sciences

Made sense to look at how many of the 6 words were defined as part of DFT work

interoperability	Interoperability describes the extent to which systems and devices can work together.....
standards	???? Too General
databases	A collection of inter-related data often with controlled redundancy, organized according to a scheme to serve one or more applications; the data are stored so that they can be used by several programs without concern for data structures or organization. (ANSI X3-172/ISO 11179-1)
repository	Repository (aka Data Repository or Digital Data Repository) is a searchable and queryable interfacing entity that is able to store...
policies	Specification of how a property of a digital entity will be controlled. (Also Very general)
fair	FAIR Data Principles are a set of guiding principles to make data Findable.....

Diagrams, Words and making sense of what we understand



Notional Organization of the Words into Categorical topics

16

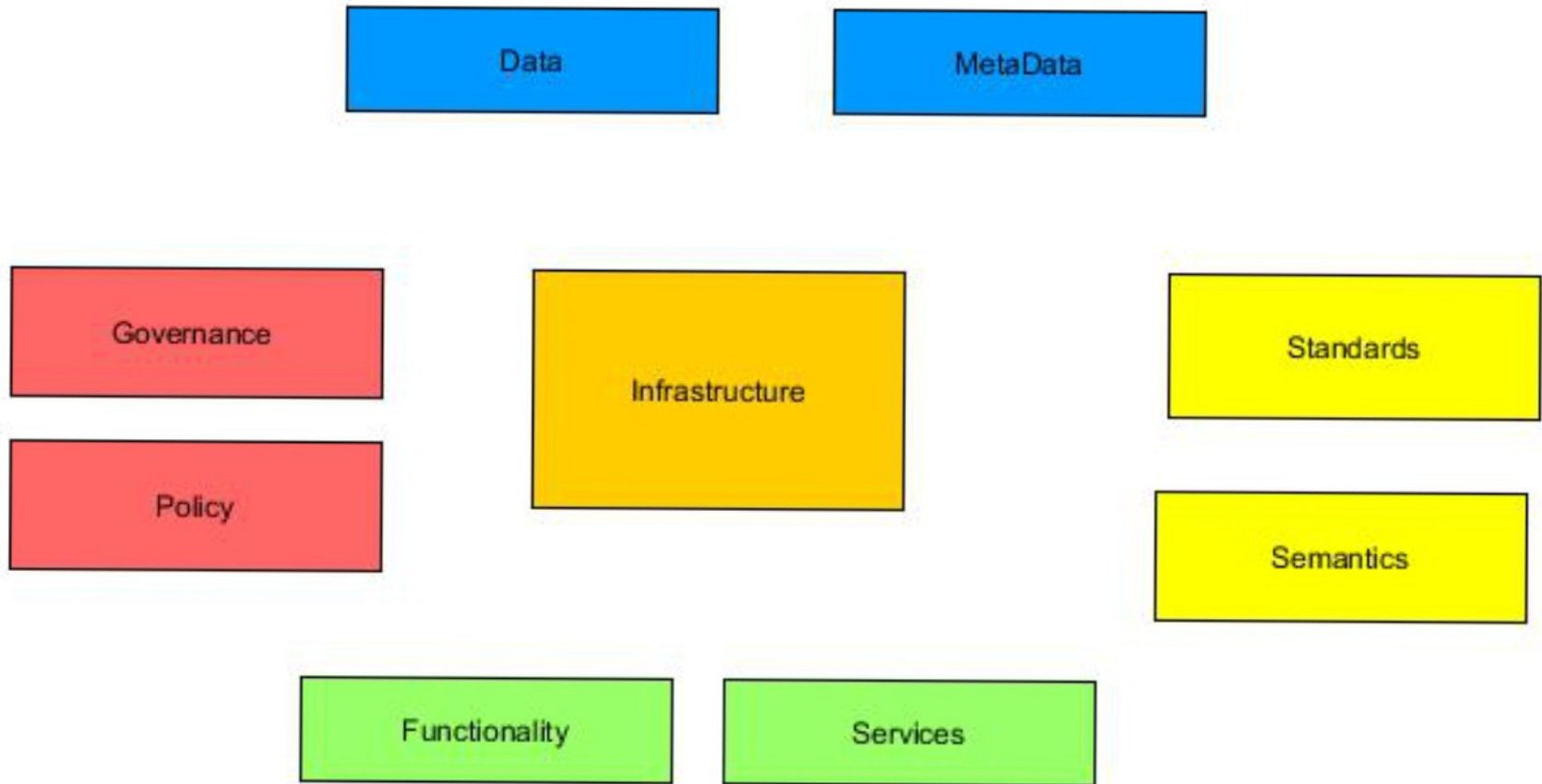
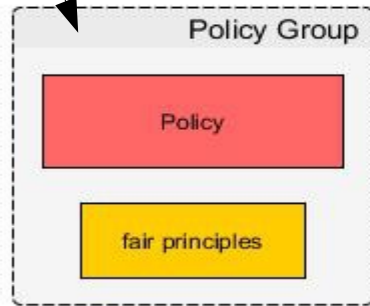
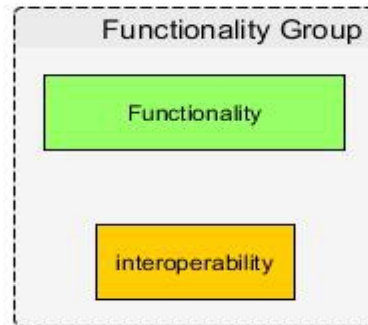
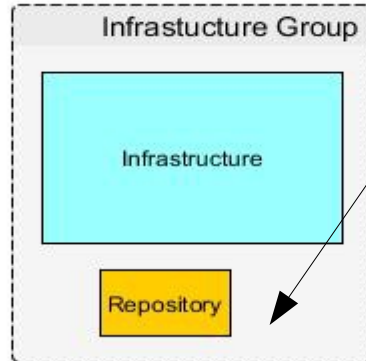


Diagram Illustrating relations with Group & Categorical Topics 17

Policy is a general concepts. In orange is FAIR principles group term which may belongs to Policy



Policies



BioSharing Registry Words: connecting data policies, standards & databases in life sciences

Databases relates to Data and Repository is in Infrastructure.

Standards

Standards

We'd like deeper analyses to show some of these things & help address what people say they need:

History of RDA groups & Gp Interactions (current, past, potential):
Form and leverage partnerships – *use joint meeting as evidence*

Develop mechanisms that encourage and reward good practice:
*WG Data Description Registry Interoperability, IG PID, WG RDA/WDS
Publishing Data Services*
*experience of 3 RDA Gps to tease out specific issues related to identifiers
and interoperability*

Build common infrastructure and incentive structures:
*RDA/CODATA Materials Data, Infrastructure & Interoperability, WG Data
Type Registries, WG PID Information Types, IG Data Fabric & IG Domain
Repositories -> Data Type Registries (DTR), & DEMO PROJECT*