



Overarching RDA Activities and Landscape

Peter Wittenburg
Max Planck Data & Compute Center
former MPI for Psycholinguistics
research data sharing without barriers
rd-alliance.org

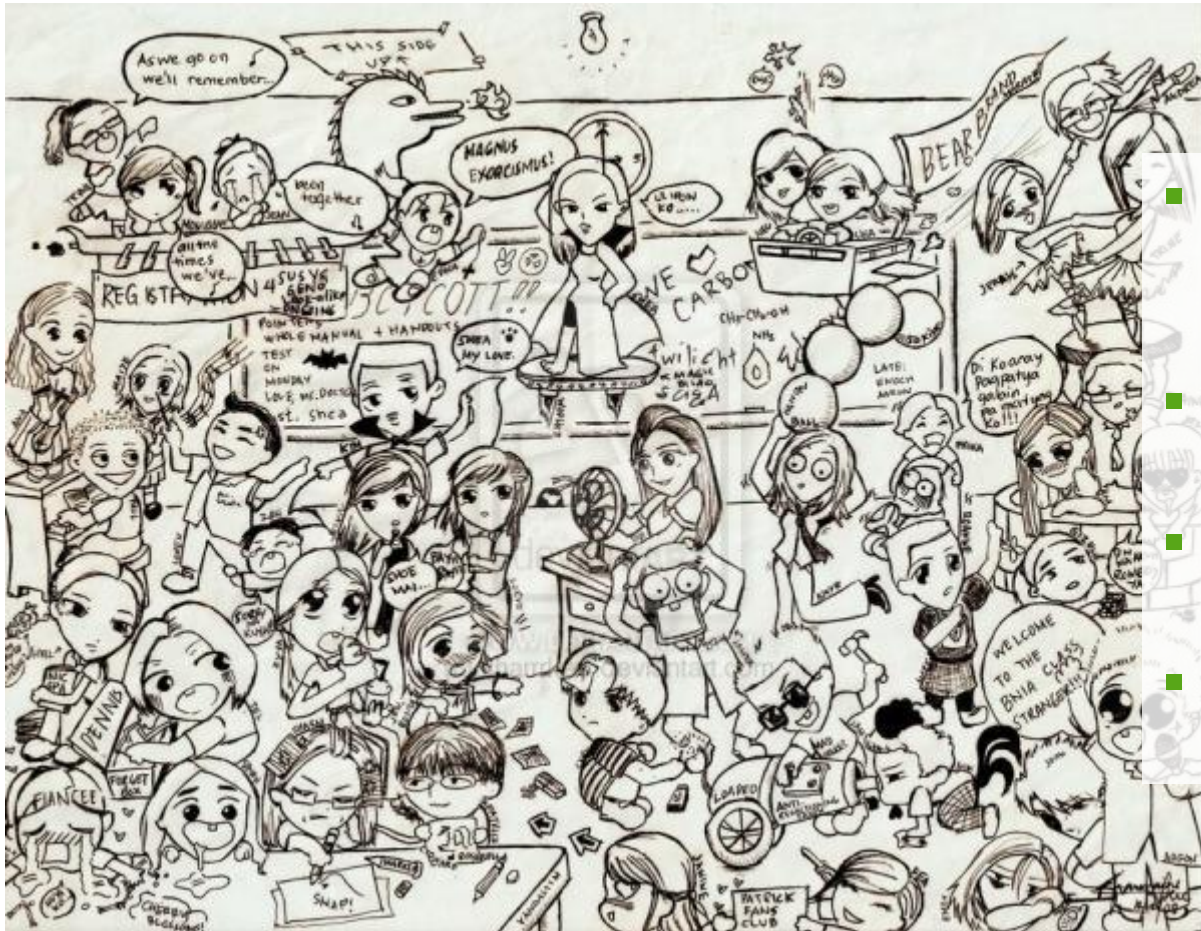
- Dynamic Development of IGs/WGs
- TAB Clustering of IGs/WGs
- Data Fabric Cluster
- Other Clusters
- Wide Agreements
- Results in Detail (-> Raphael)

Bottom-Up vs. Top-Down

- around Plenary 1 RDA started with 5 Working Groups
- now have about 39 Interest Groups, 17 Working Groups & some BoFs
- 1st Science WS demanded us to be bottom-up – guess we are

WGs and IGs – natural impression of “chaos”

- RDA as a bottom-up organization needs a “chaotic” element
 - bottom-up principles depend on active and creative minds
 - there needs to be a high degree of openness & neutrality
- now 56 groups (17 WGs and 39 IGs, some BoFs at plenaries)
 - issue 1: no one understands what is happening and how this all fits together
 - issue 2: what is the big coherent message



- creative minds start working on their babies
- people tend to work in silos
- there is overlap and contradiction
- there is confusion

Bottom-Up vs. Top-Down

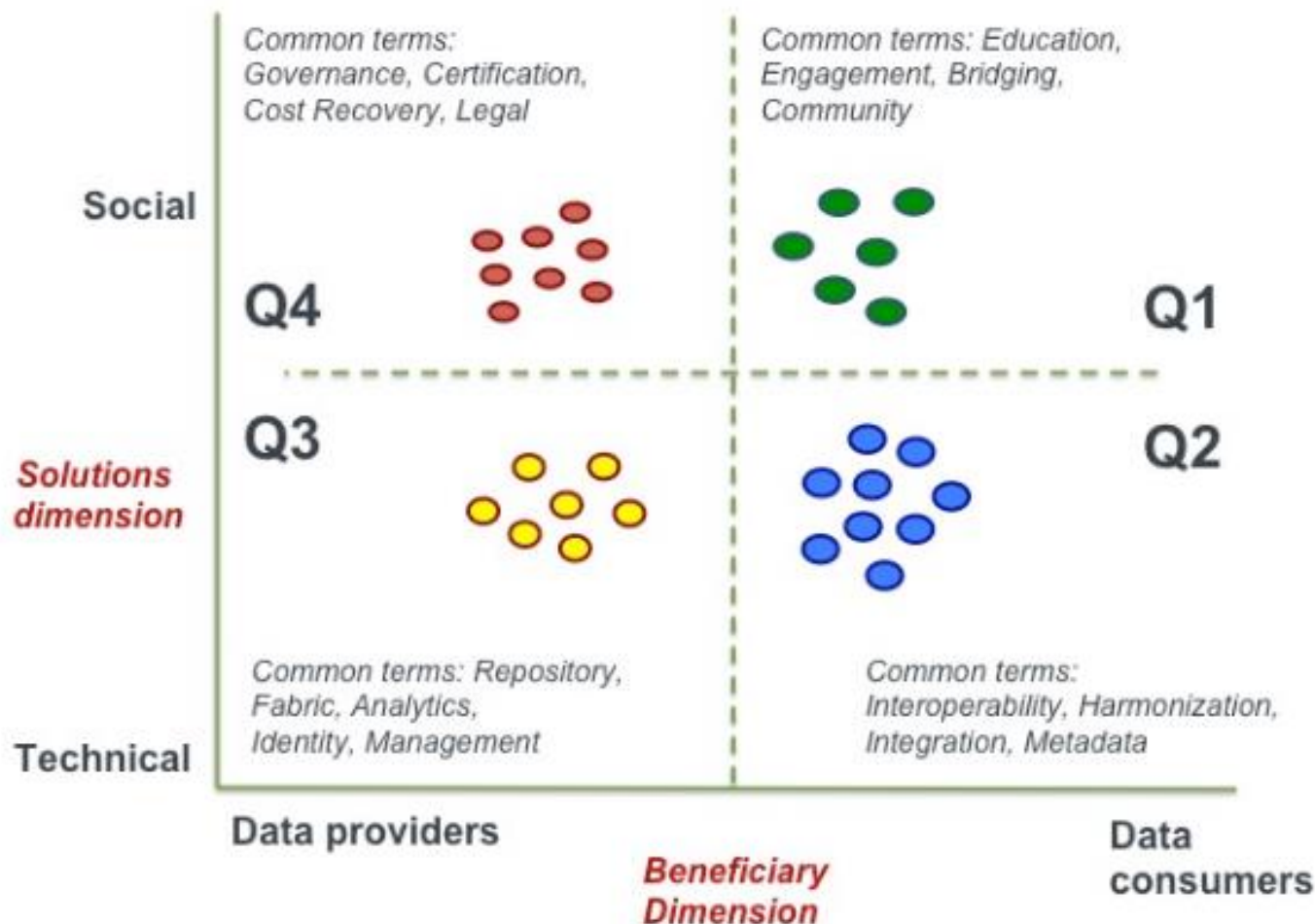
- around Plenary 1 RDA started with 5 Working Groups
- now have about 60 Interest Groups and Working Groups
- 1st Science WS demanded us to be bottom-up – guess we are
- so people started to ask: where is structure, where is coordination, etc.?
 - IETF has **area directors** to guide

- a few milestones to improve balance
 - P2: TAB started work on checking/commenting case statements
 - P3: TAB started discussion about the RDA “landscape”
 - P4: WG Chairs joined to start Data Fabric discussion (one example)
 - P4-5: TAB came up with a “Clustering” proposal

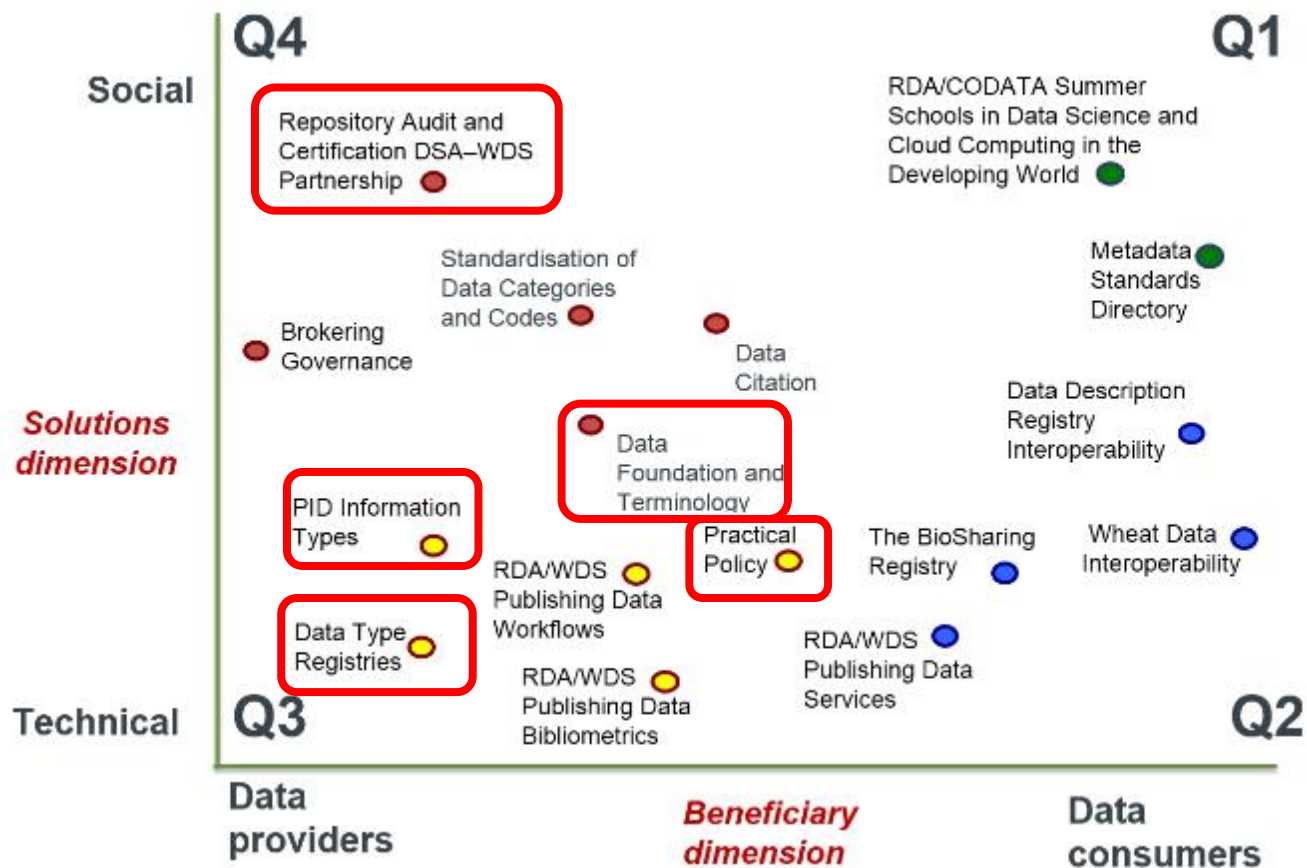
- Dynamic Development of IGs/WGs
- TAB Clustering of IGs/WGs
- Data Fabric Cluster
- Other Clusters
- Wide Agreements
- Results in Detail (-> Raphael)

- start discussing various concepts for structuring and balancing
- which purpose
 - less confusion for newcomers, how to find activities, etc.
 - organize coordination/guidance work from TAB
 - understand overlaps, gaps etc.
 - forced communication between chairs (NO would be pure overhead)
- which way
 - functional vs. life cycle vs. “meaningful dimensions”
 - no classification perfect – so give a pragmatic start
- chairs are using functional approach to come interact

TAB Landscape Proposal

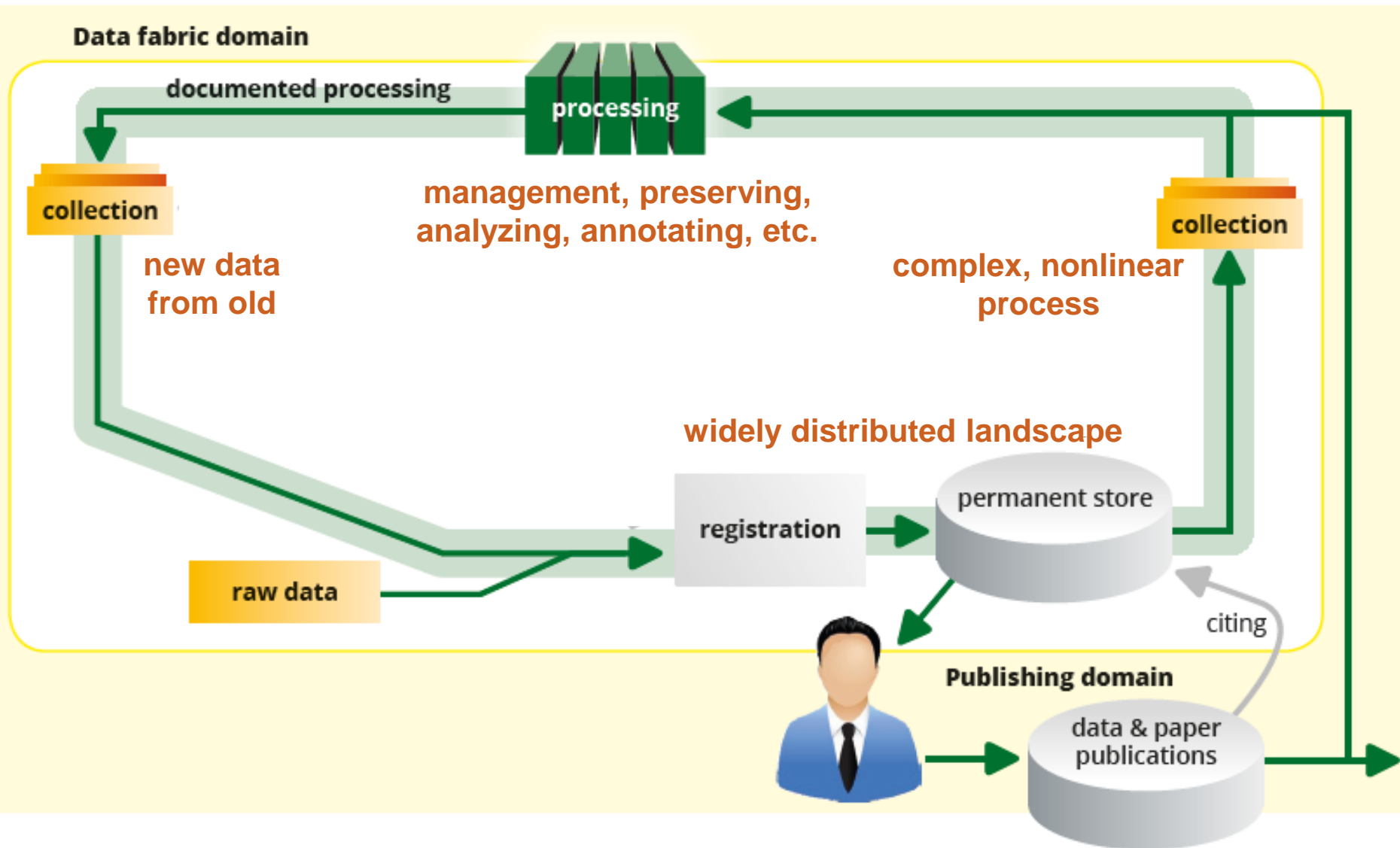


TAB Landscape Classification – intuitive?

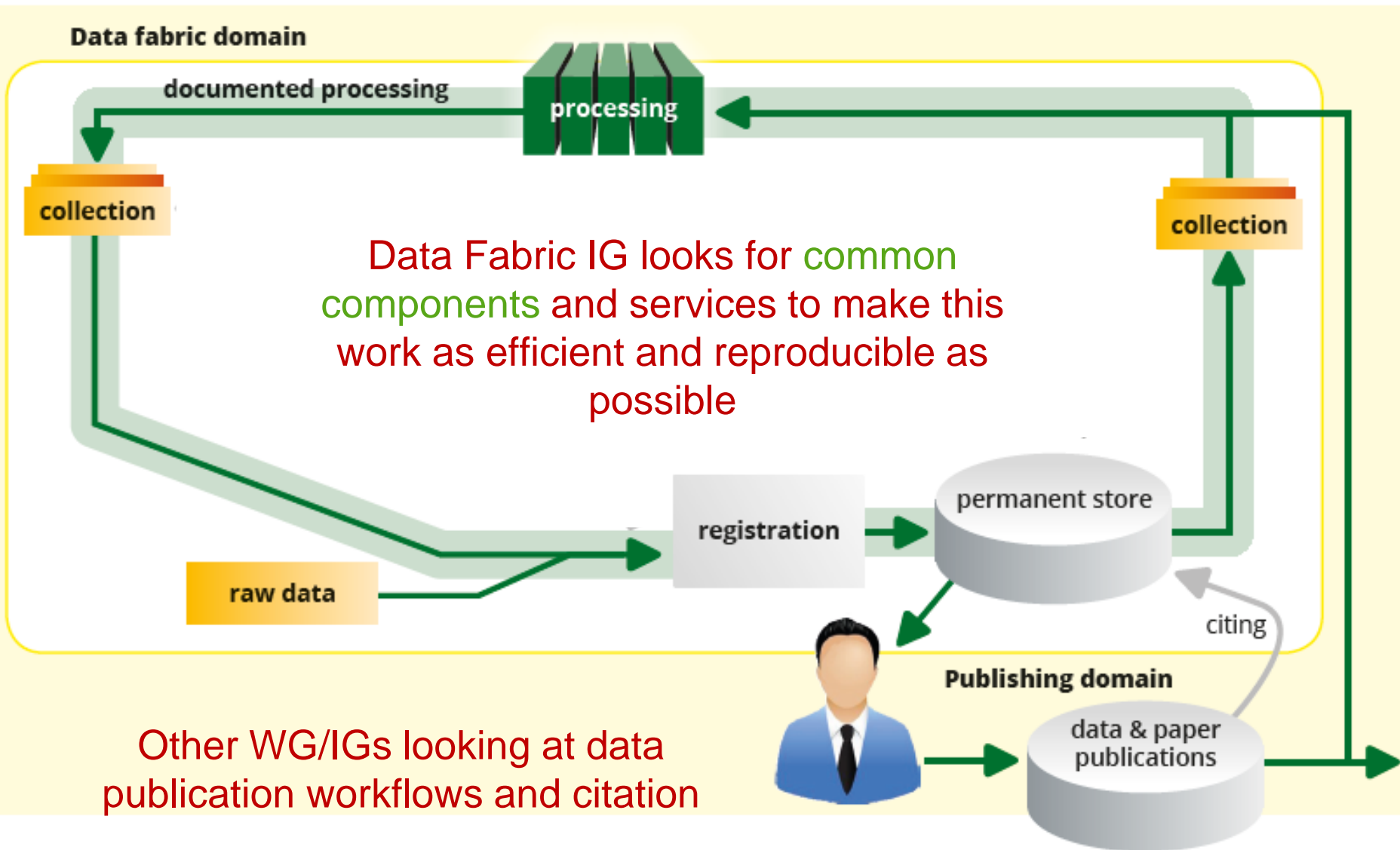


- Dynamic Development of IGs/WGs
- TAB Clustering of IGs/WGs
- **Data Fabric Cluster**
- Other Clusters
- Wide Agreements
- Results in Detail (-> Raphael)

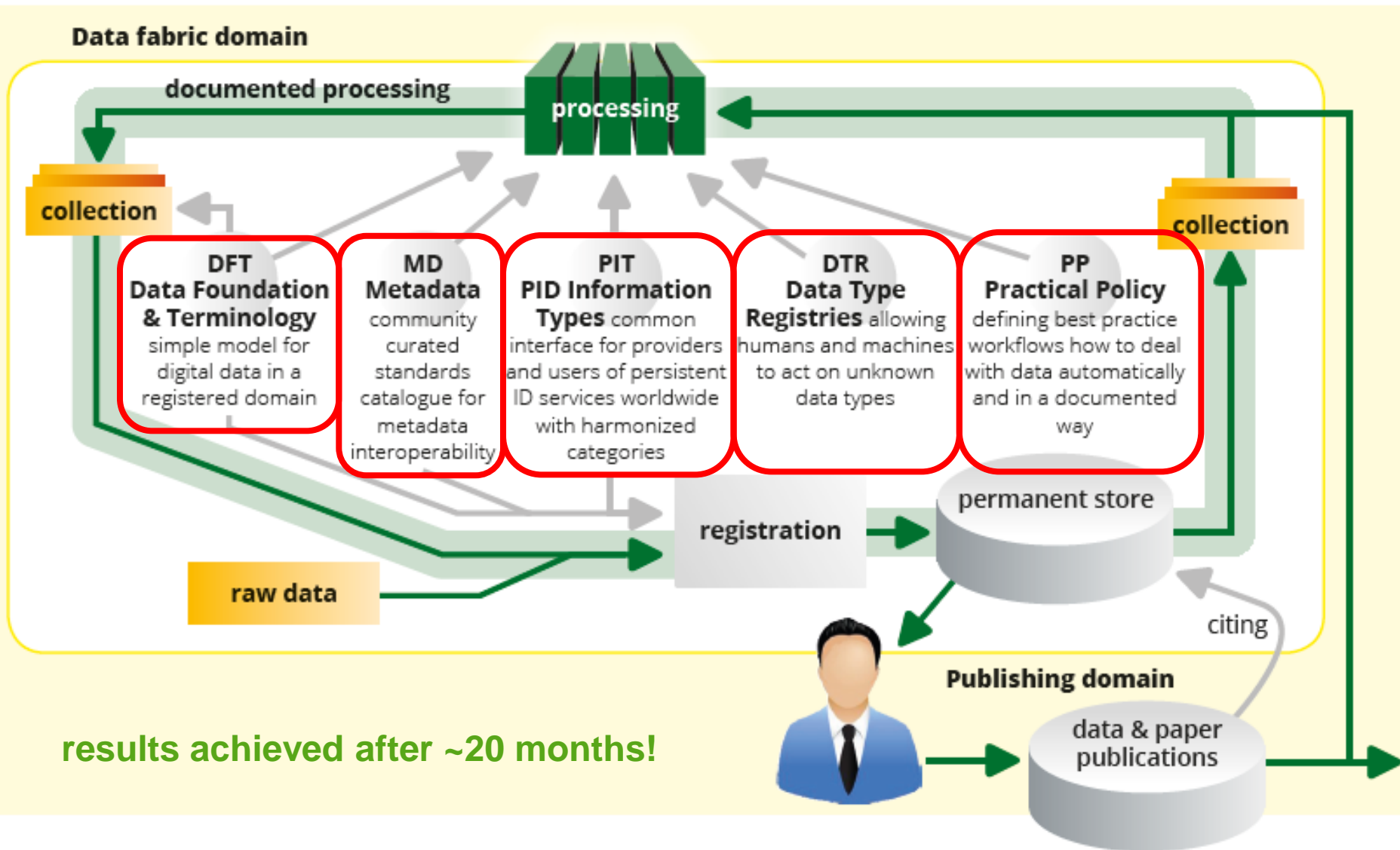
Data Creation and Re-Use Cycle in the Labs



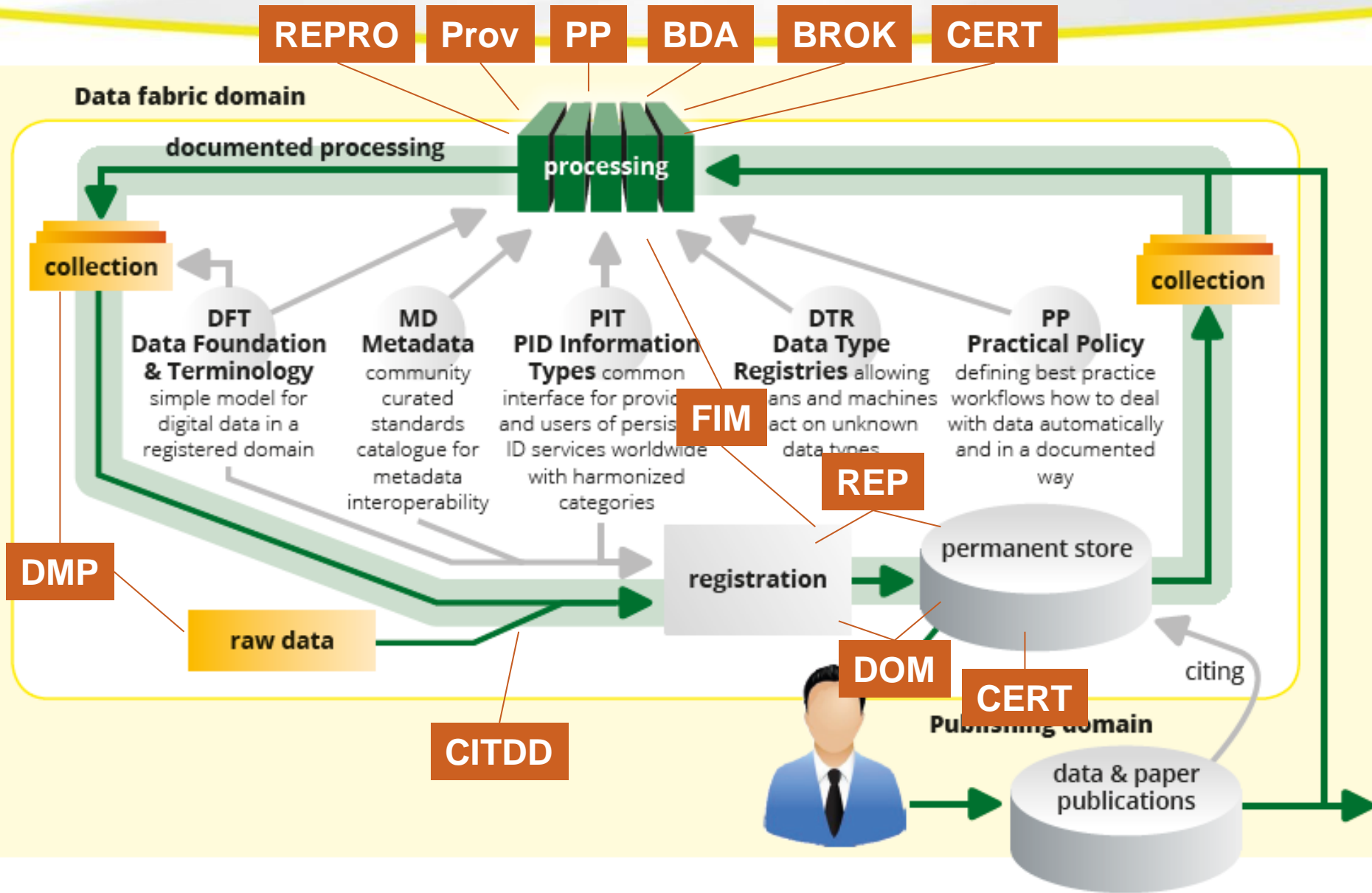
Data Fabric Interest Group



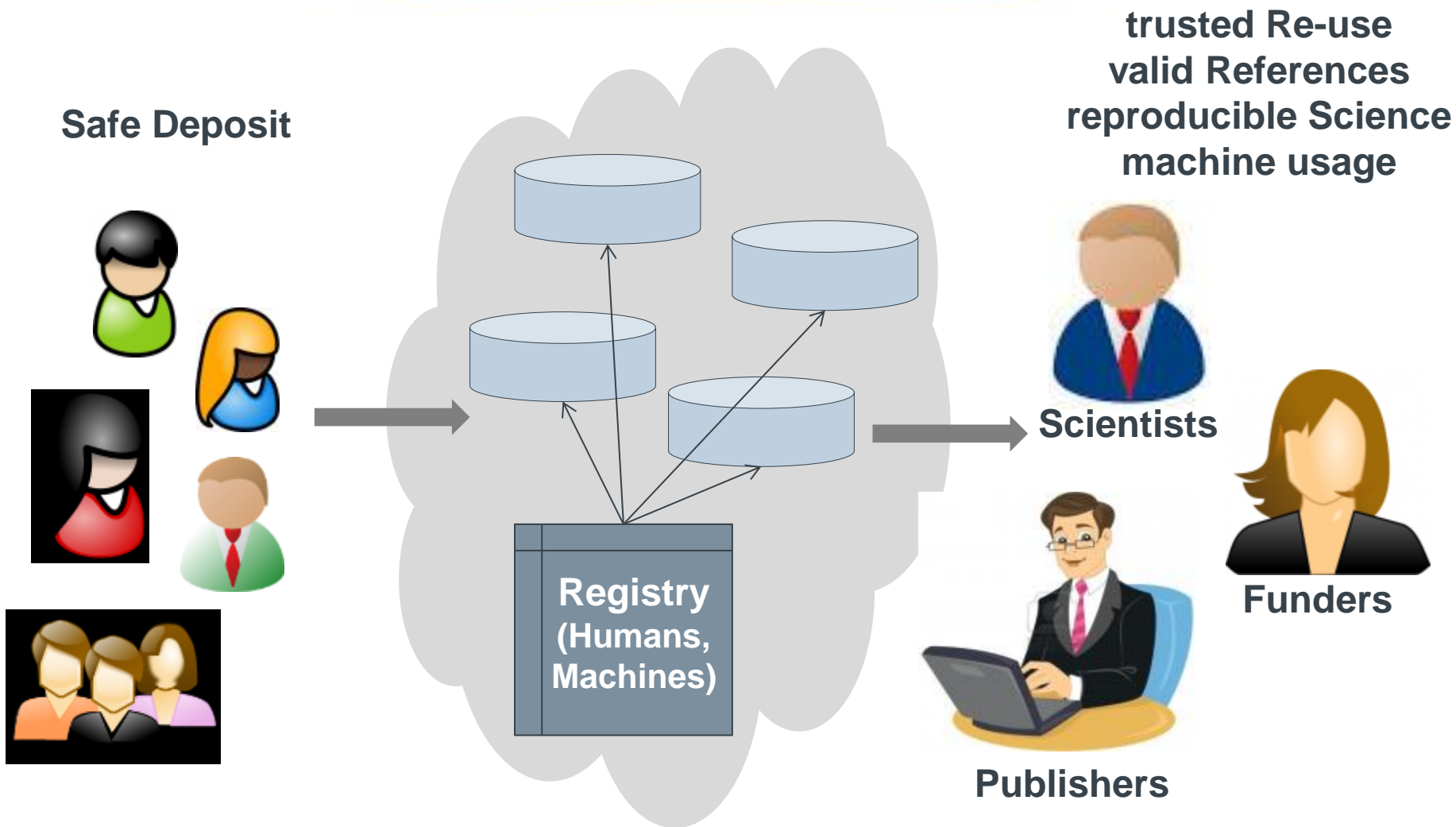
RDA – first Working Group results



DFIG – grouping of WG/IGs

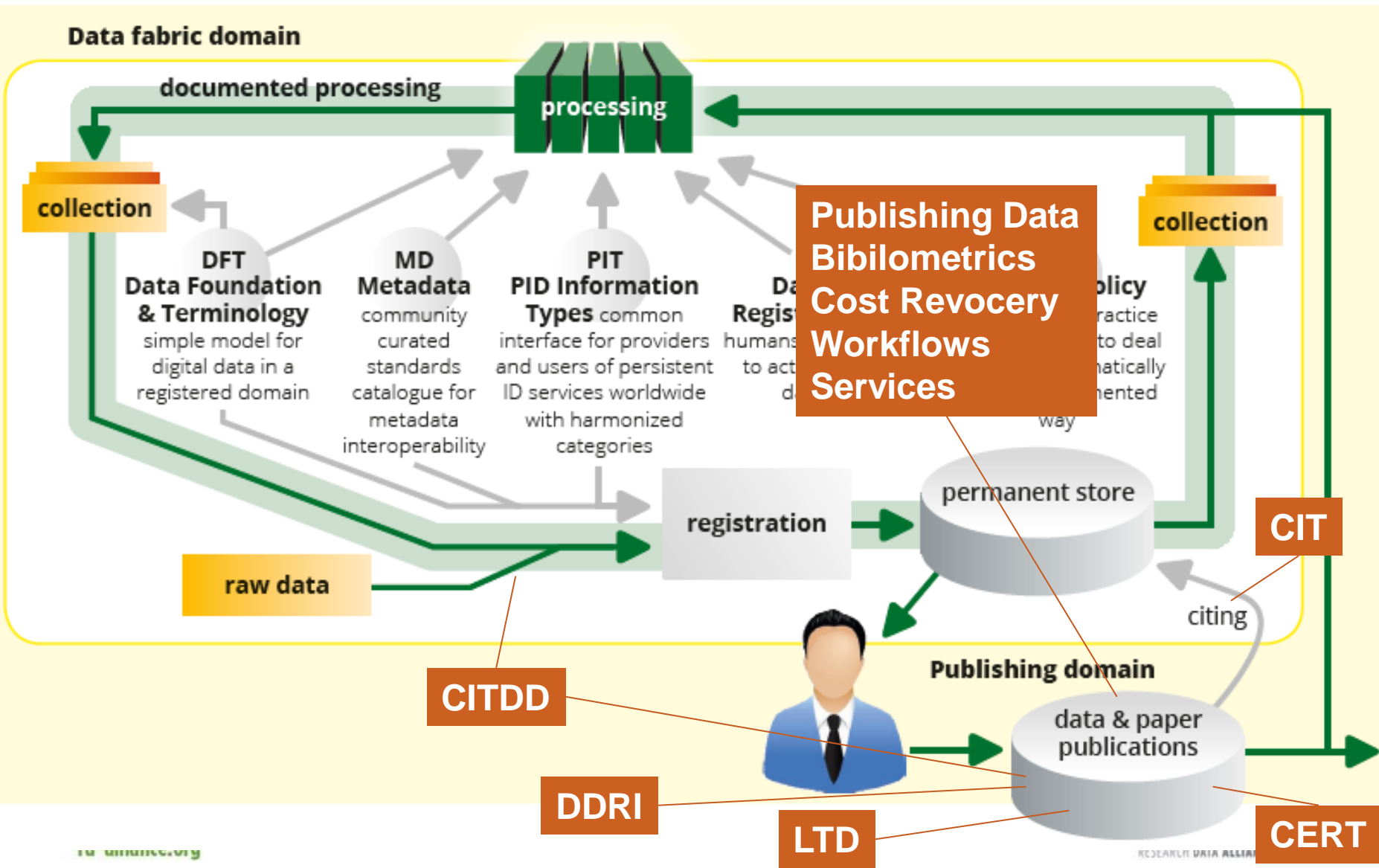


new in DFIG – Repository Registry



- Dynamic Development of IGs/WGs
- TAB Clustering of IGs/WGs
- Data Fabric Cluster
- **Other Clusters**
- Wide Agreements
- Results in Detail (-> Raphael)

Publishing Cluster



Community Groups

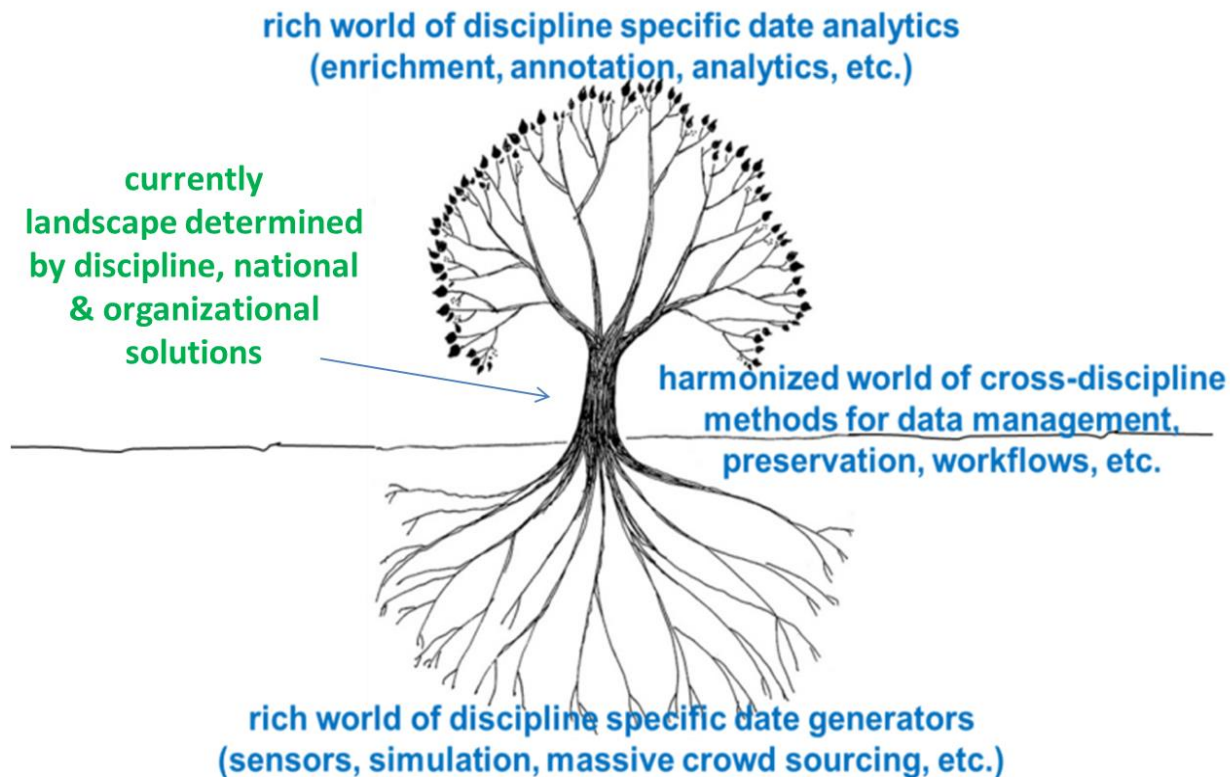
- Agriculture / **Wheat Interop**
- Biodiversity
- Structural Biology
- Biosharing Registry
- ELIXIR
- Toxicogenomics
- Metabolomics
- Geospatial
- Materials Data
- Photon&Neuron
- Marine
- History&Ethnography
- Urban Life

Social Groups

- Community Capability
- Data Re-use
- Data Life Cycle
- Engagement
- Ethical Aspects
- Legal Interoperability
- Data Rescue
- Data Handling Training
- Data for Development
- Cloud Worldwide Training

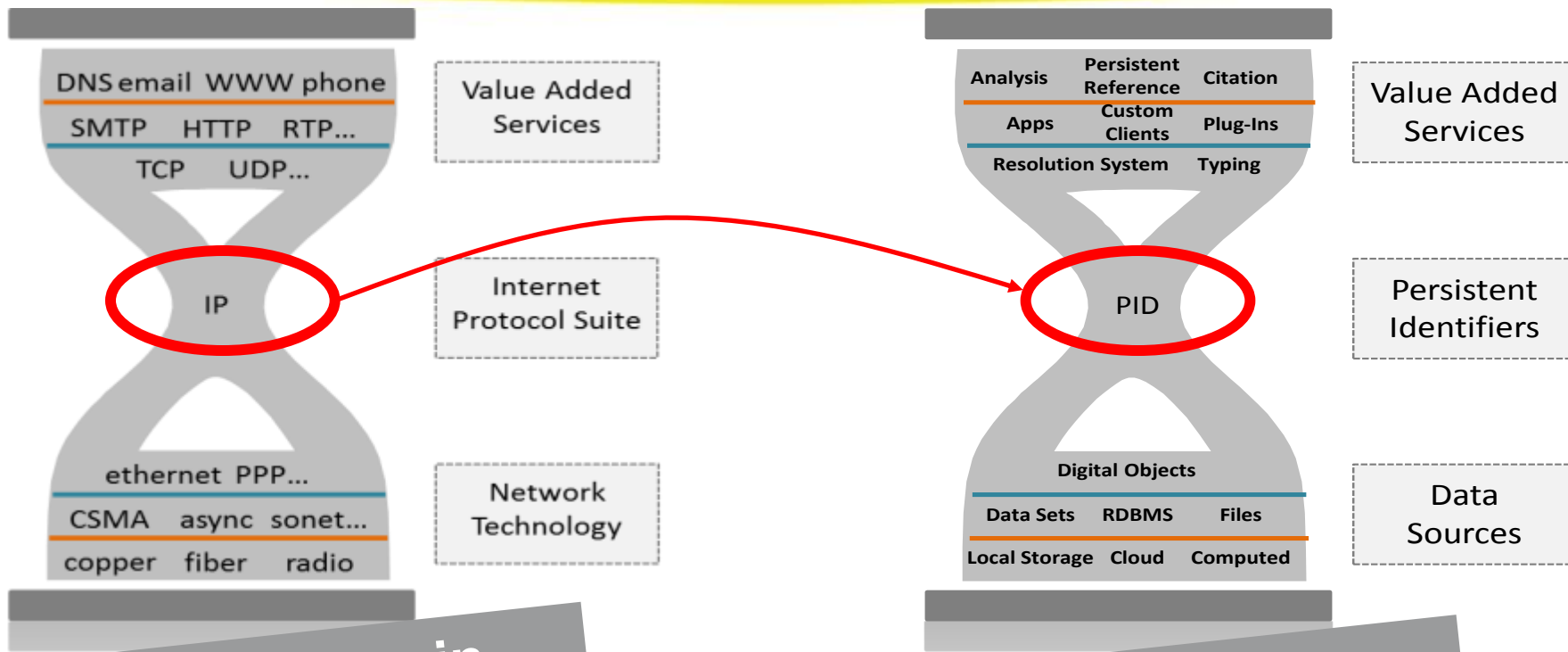
- Dynamic Development of IGs/WGs
- TAB Clustering of IGs/WGs
- Data Fabric Cluster
- Other Clusters
- **Wide Agreements**
- Results in Detail (-> Raphael)

Widely agreed in RDA I



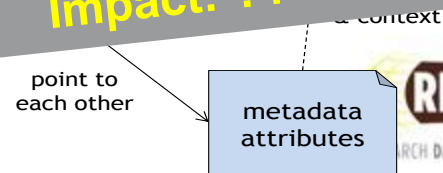
- ❑ management of data objects is widely type and discipline independent
- ❑ still every project defines its own strategies leading to huge stack of software that will not be maintainable

Widely agreed in RDA II



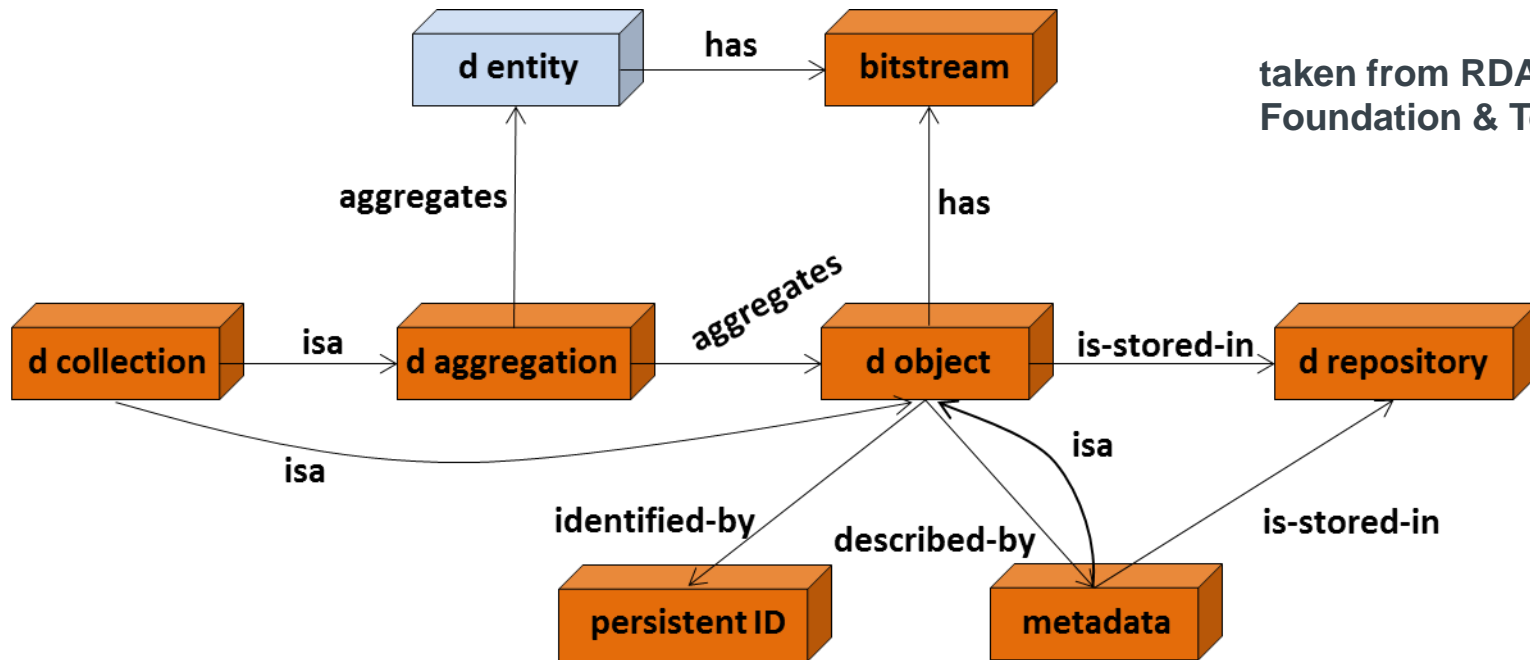
Internet Domain
 nodes with IP numbers
 packages being exchanged
 standardized protocols
Impact: worldwide boost in connectivity and information

Data Domain
 objects with PID numbers
 objects being exchanged
 standardized protocols
Impact: ??



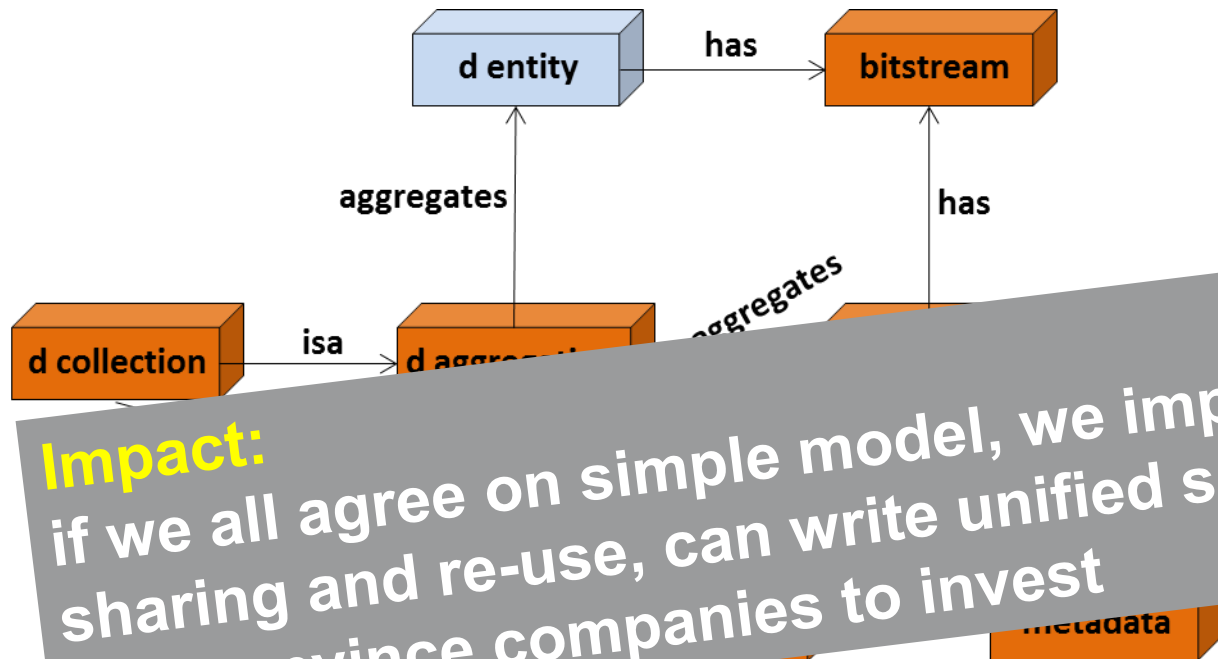
- Dynamic Development of IGs/WGs
- TAB Clustering of IGs/WGs
- Data Fabric Cluster
- Other Clusters
- Wide Agreements
- Results in Detail (-> Raphael)

RDA Results I: common data model



taken from RDA WG Data Foundation & Terminology

- PIDs at the beginning of trust chain
- need a worldwide, independent and robust PID system worldwide
- metadata are essential in anonymous data world



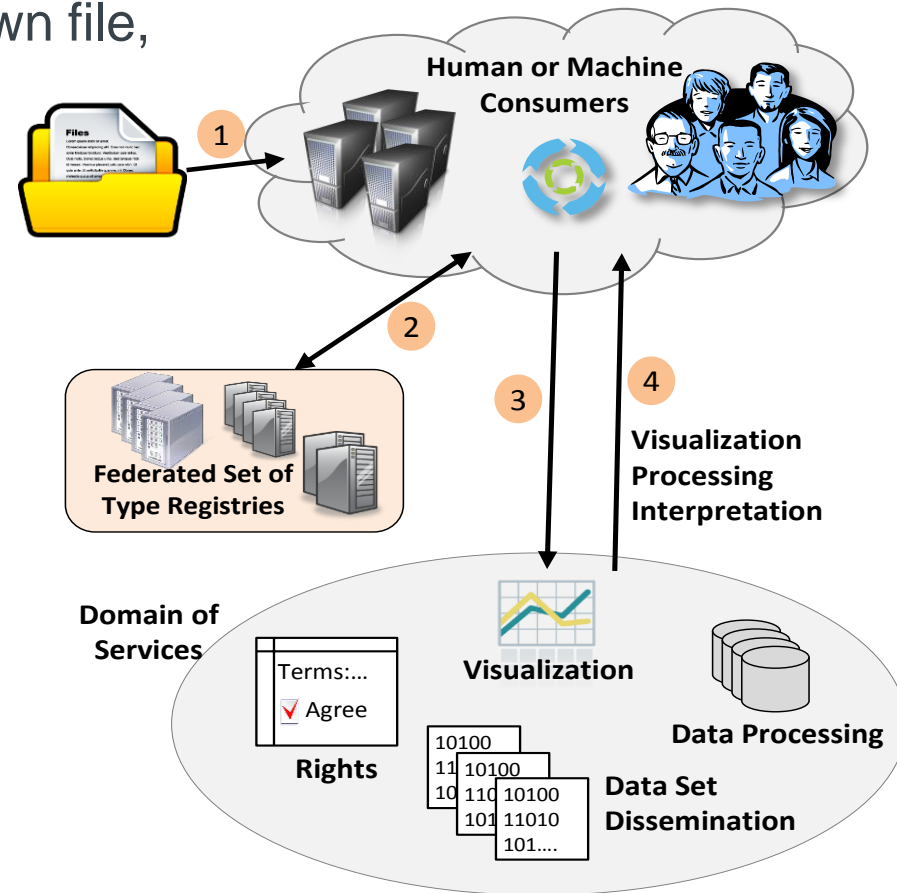
taken from RDA WG Data Foundation & Terminology

Impact:
if we all agree on simple model, we improve sharing and re-use, can write unified software and convince companies to invest

- PIDs at the beginning of trust chain
- need a worldwide, independent and robust PID system worldwide
- metadata are essential in anonymous data world

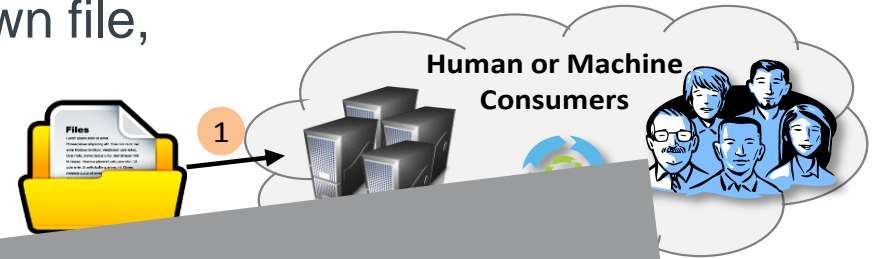
RDA Results II: Data Type Registry

- result: a **registry for data types**
- **simple example**: you get an unknown file, pull it on DTR and content is being visualized
- DTR can also be used to describe and re-use semantic content
- **no free lunch**: someone needs to register and define type
- **PIT Demo already working with DTR** and NIST is working with communities

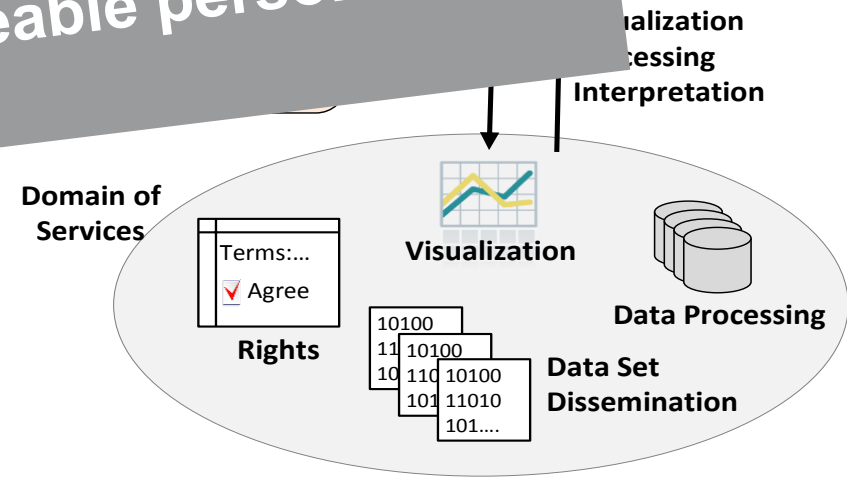


RDA Results II: Data Type Registry

- result: a registry for data types
- simple example: you get an unknown file, pull it on DTR and content is being visualized
- DTR can also be used to describe and register



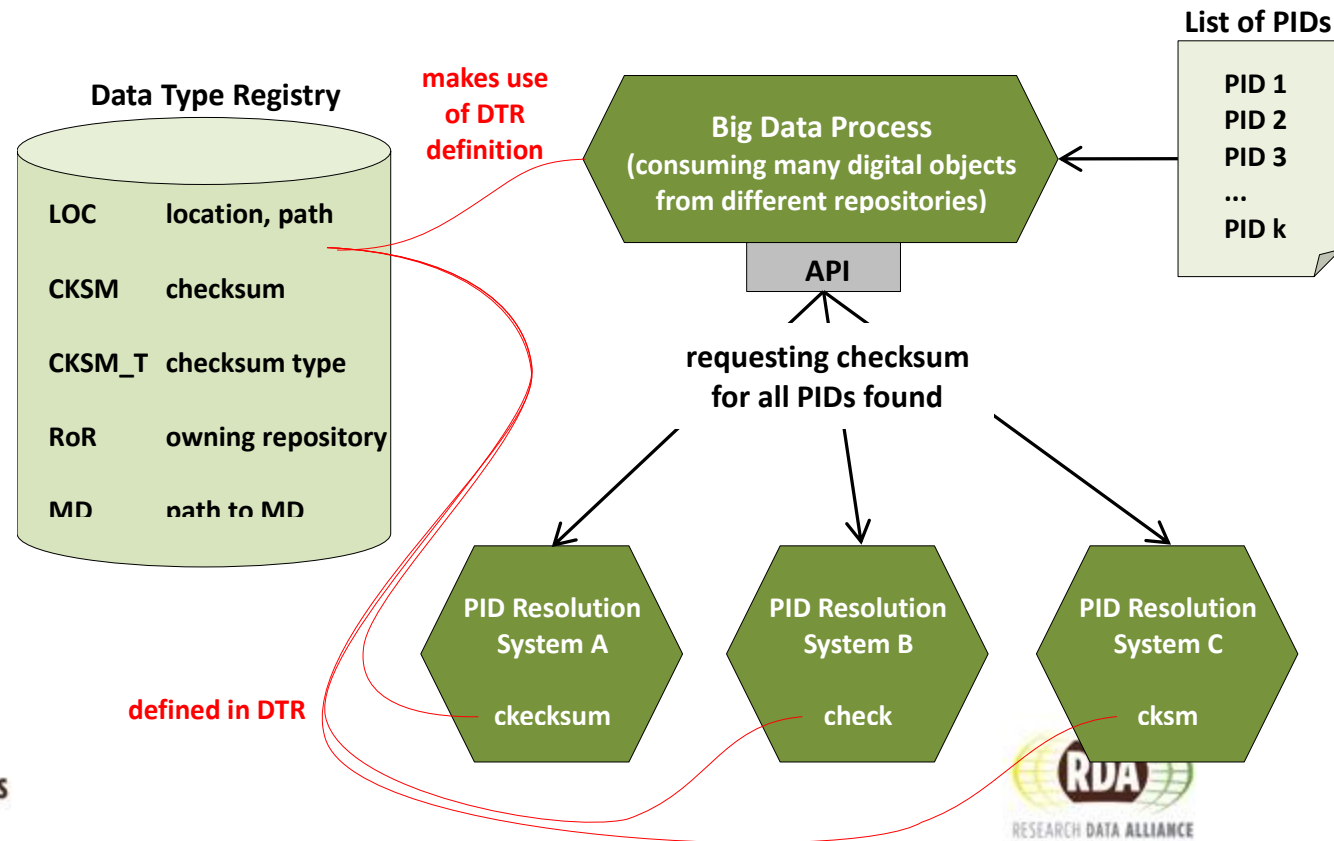
Impact: automatic processing of any data type by machines and not-knowledgeable persons comes into reach



RDA Results III: PID Information Types

- result: a **generic API** and a set of **basic attributes**
- a **PID Record** is like a Passport (Number, Photo, Exp-Date, etc.)
- if all **PID Service-Provider** agree on one API and talk the same language (registered terms) SW development will become easy
- **Test-Installation**

in operation
together with
DTR

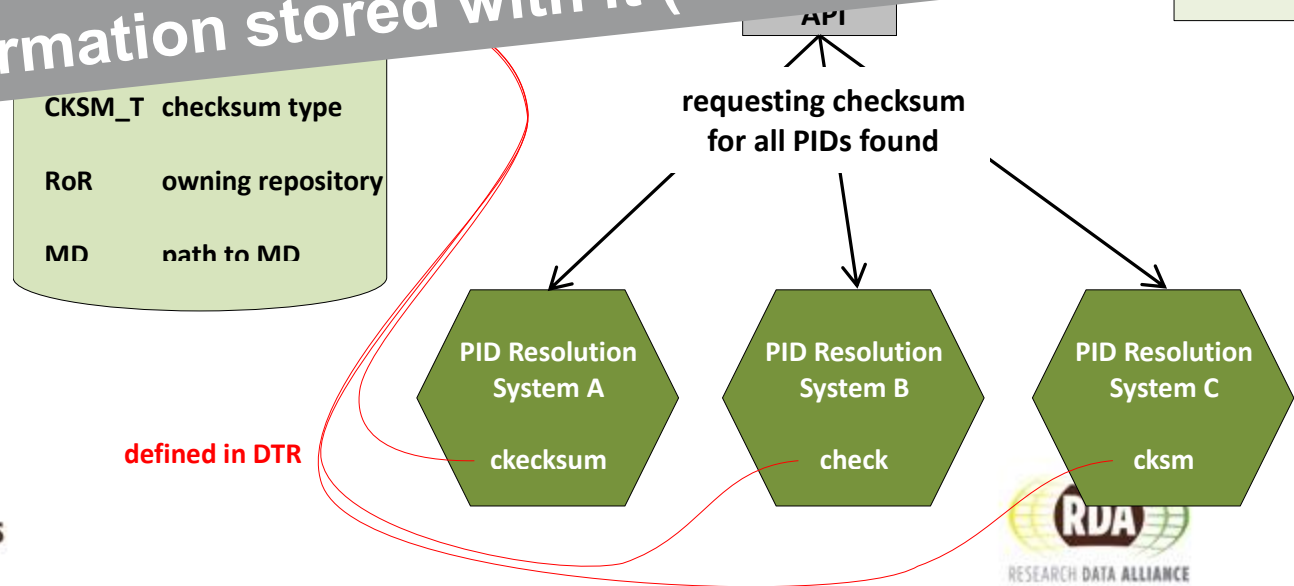


RDA Results III: PID Information Types

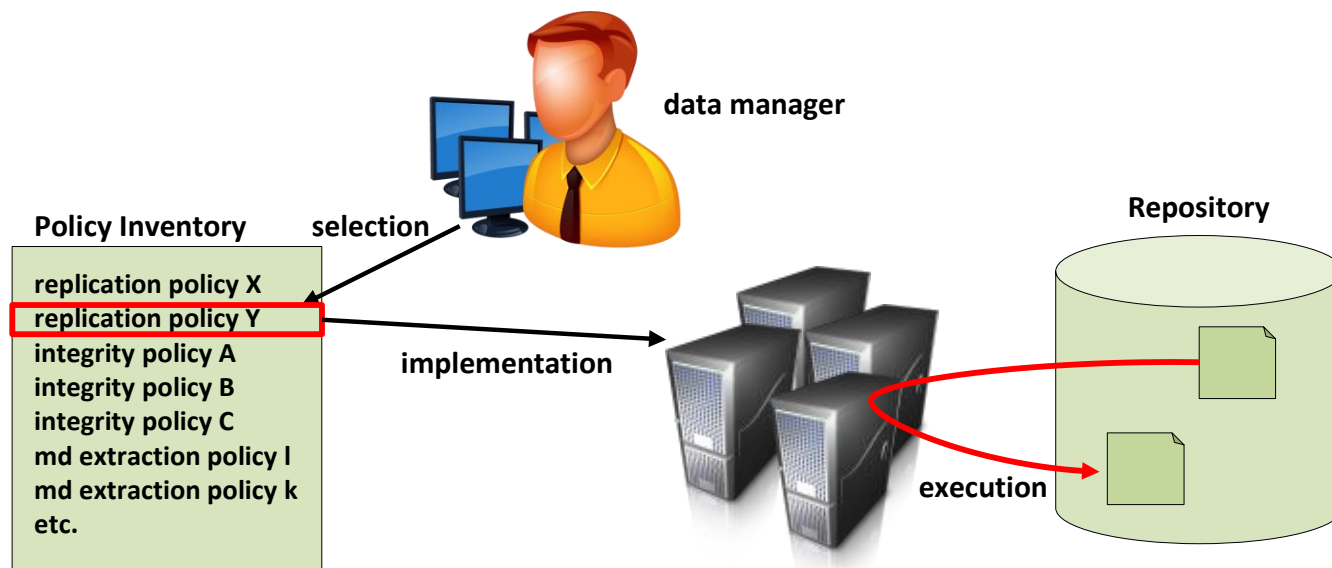
- result: a **generic API** and a set of **basic attributes**
- a **PID Record** is like a Passport (Number, Photo, Exp-Date, etc.)
- if all **PID Service-Provider** agree on one API and talk the same language (registered terms) SW development will become easier
- **Test-Installation**

Impact:

- working with PIDs and service providers much easier
- worldwide interoperability in accessing PIDs and relevant information stored with it (like a passport)



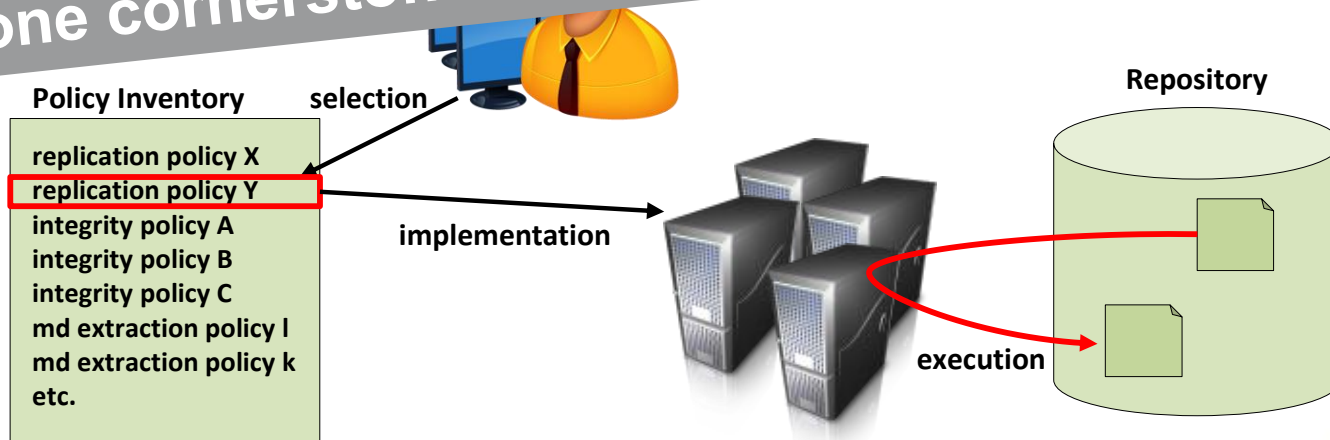
- due to unforeseen circumstances work until P5
- **Practical Policies = executable Workflow Statements**
- result at P5: a set of **Best Practice PPs** for a number of typical DM/DP tasks (Integrity Check, Replication, etc.)
- currently a large **collection of PPs**, currently being evaluated
- you could add your policies



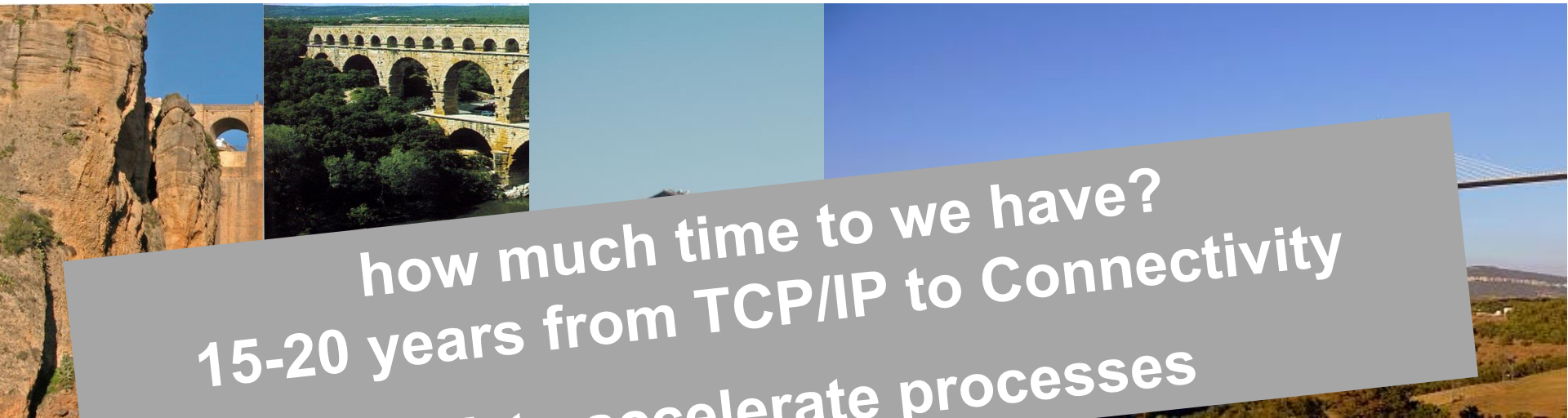
- due to unforeseen circumstances work until P5
- Practical Policies = executable Workflow Statements
- result at P5: a set of Best Practices for data tasks (4)

Impact:

- huge simplification for data stewards
- finally feasible quality checks and certification
- huge step in improving trust
- one cornerstone towards reproducible data



- organizing Uptake and Engagement is crucial now – also: need to remain sensitive enough to listen & adapt
- heard about concrete adoptions at Sunday (all 4 results)
- joint collaboration calls with e-Infrastructures will come (EUDAT, etc.)
- many information transmission & training meetings are scheduled
 - large European infrastructures in May 2015 (ESFRI, HBP, etc.)
 - National infrastructure meetings (FI February, DE May & November, FR several also around P6, GR planned, others to come)
- also outreach to SMEs and startups is planned
 - discussions with local networks going on in FR, DE, ??
- national RDA related adoptions plans by funders in progress
- EU Work programme 16/17: evaluation of opportunities
- supporting this will cost a lot of time
 - therefore from September on RDA EU support/help team (seniors/juniors)
 - visit of many infrastructure projects in Europe incl. adoption discussions/help



how much time to we have?
15-20 years from TCP/IP to Connectivity
RDA to accelerate processes

about building the **social and technical bridges** that enable global open sharing and re-use of data. **Researchers, scientists, data practitioners** from around the world are invited to work together to achieve the vision

Funders: NSF, EC, AU, Japan, Brazil, DE?, UK?, ZA?, FI?, etc.

Thanks for your attention.

<http://www.rd-alliance.org>

<http://europe.rd-alliance.org>

Next RDA Plenary P6:

23-25. September Paris