



RESEARCH DATA ALLIANCE

Globally Interoperable IoT Identification and Data Processing

- joint RDA and IoT Forum workshop -

Peter Wittenburg

Max Planck Society, Max Planck Computing & Data Facility

RDA Europe Director



Welcome to the Joint Workshop

1. Why this workshop
 - do we have a data problem in science and other sectors?
 - let's have a look to actual cross-sector discussions and work
2. Does IoT change the Game?
3. What does RDA do (brief)?
4. Are there concepts to overcome barriers?
5. Agenda

Do we have a problem?

- data intensive projects and large data aggregations are a fact
- however, facts are also:
 - data is hardly visible and not accessible (only 18% of data in registered repositories is accessible)
 - 80% of created data is not accessible any longer after short periods
 - data domain is fragmented – data integration is a costly job (identification, organisation and description of data, etc.)
 - 75-80% (RDA EU, MIT) of data scientists time is lost with data integration/management work
 - 60% of costs of data intensive projects is spent for pure integration tasks
 - sorting out rights is a never ending story
 - data volumes and complexity are continuously increasing
- **our methods do not scale and many are excluded**

Will IoT change the game?

- › 50 billion smart devices (Intel) will create true data monsters



- › continuous streams with high-granularity
- › optimisations and real-time decisions required
- › much more re-purposing of data in various contexts

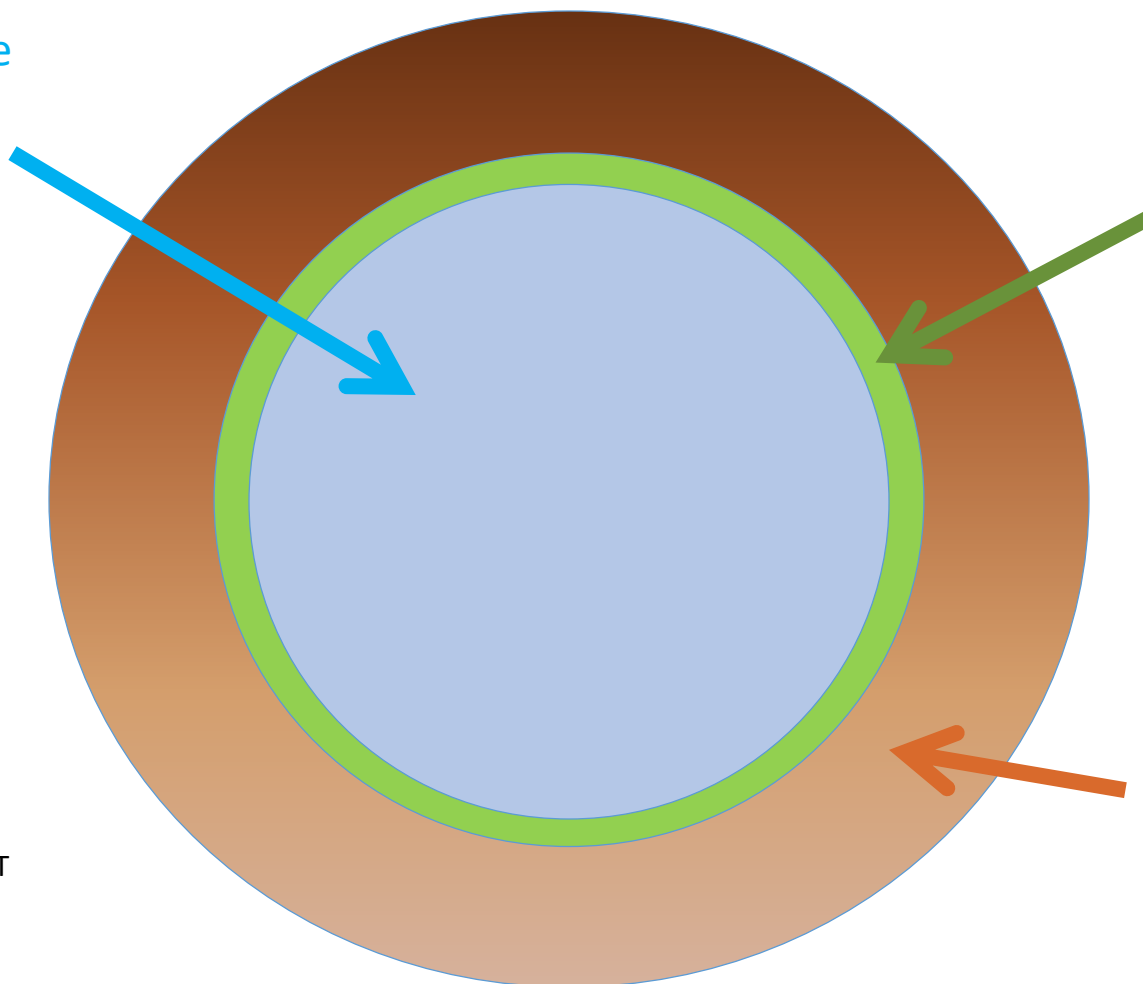
Are we fit for these challenges?

No our methods are not scalable - we need a change!

Fundamental Change Through IoT I

Cyber
infrastructure

Humans



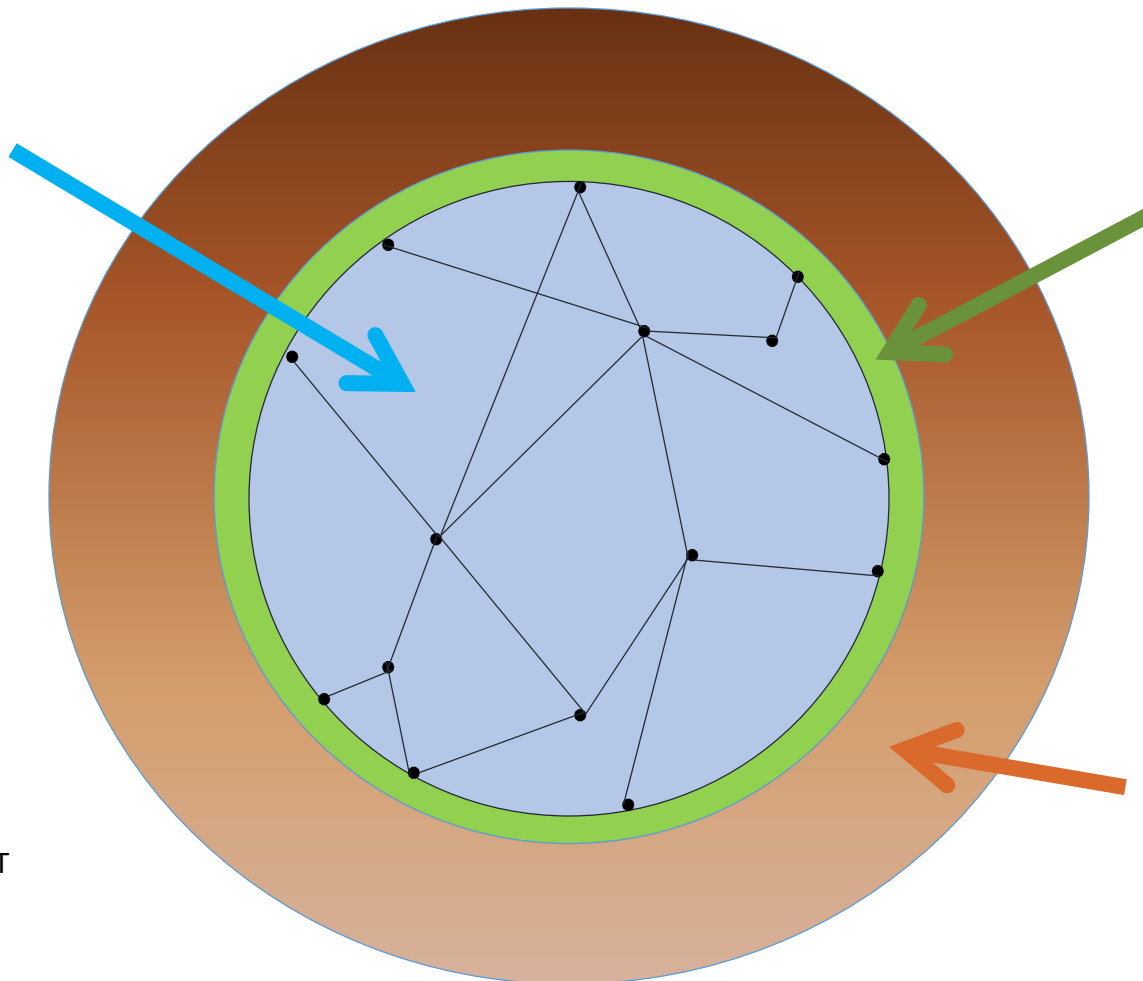
Physical
Objects

adapted from
Chris Greer, NIST

Fundamental Change Through IoT II

Internet
WWW
etc.

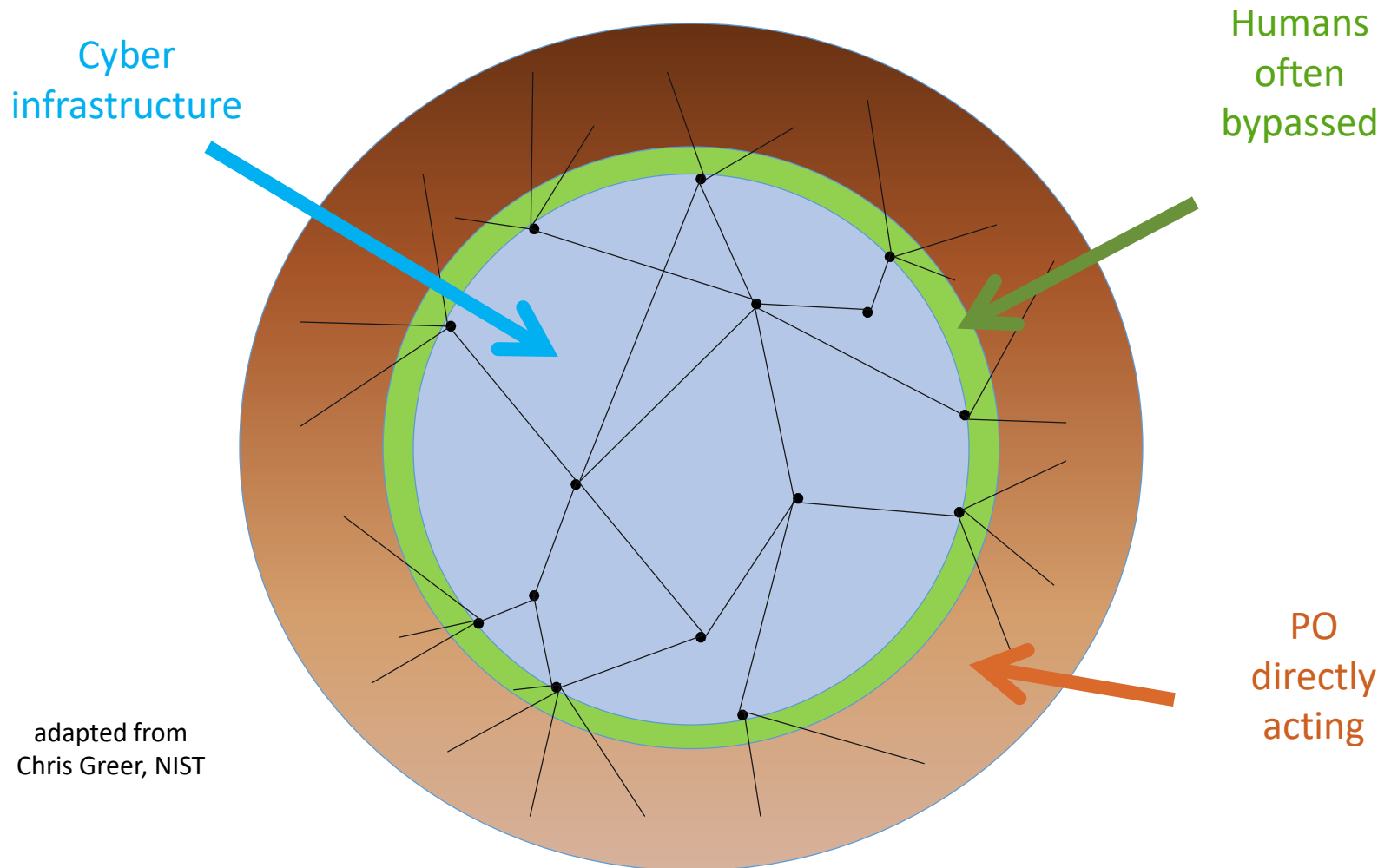
Humans
Actors
Mediators



Physical
Objects

adapted from
Chris Greer, NIST

Fundamental Change Through IoT III

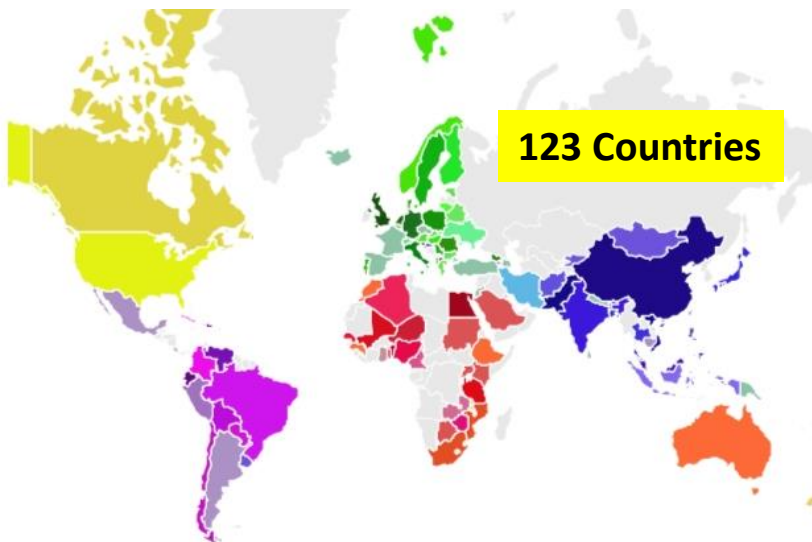
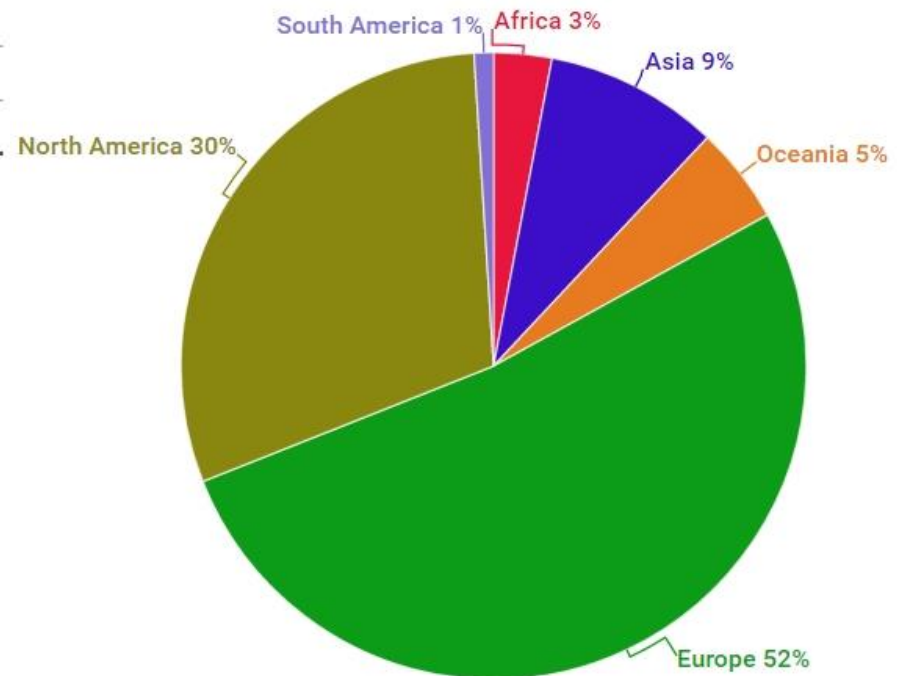
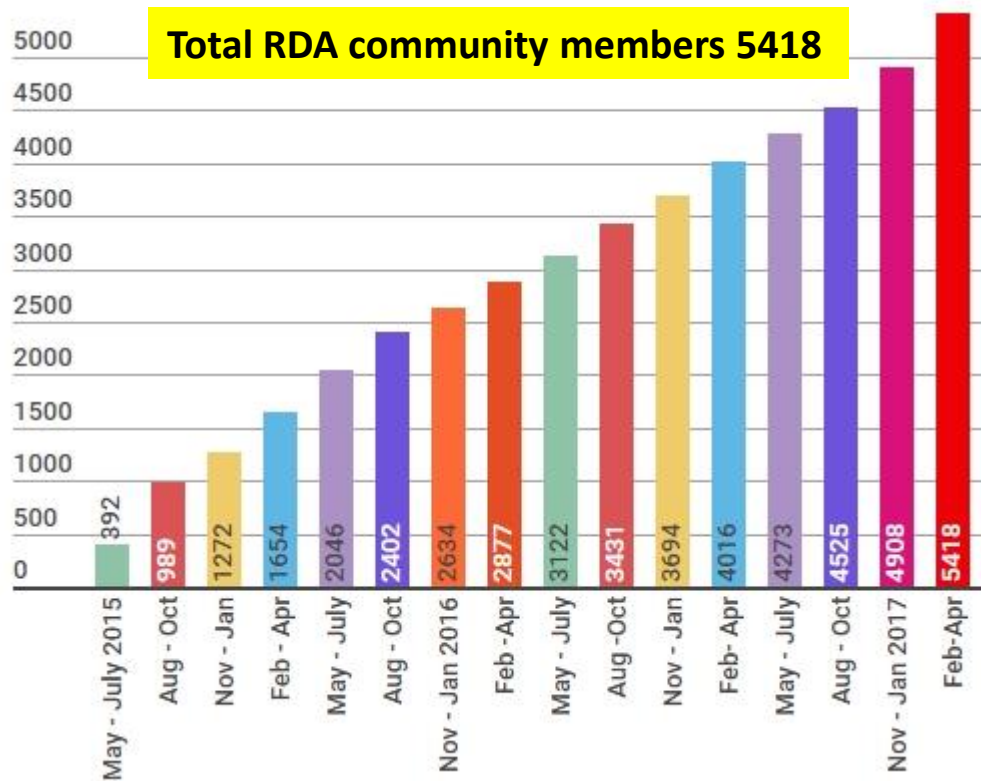


How to come to global solutions?

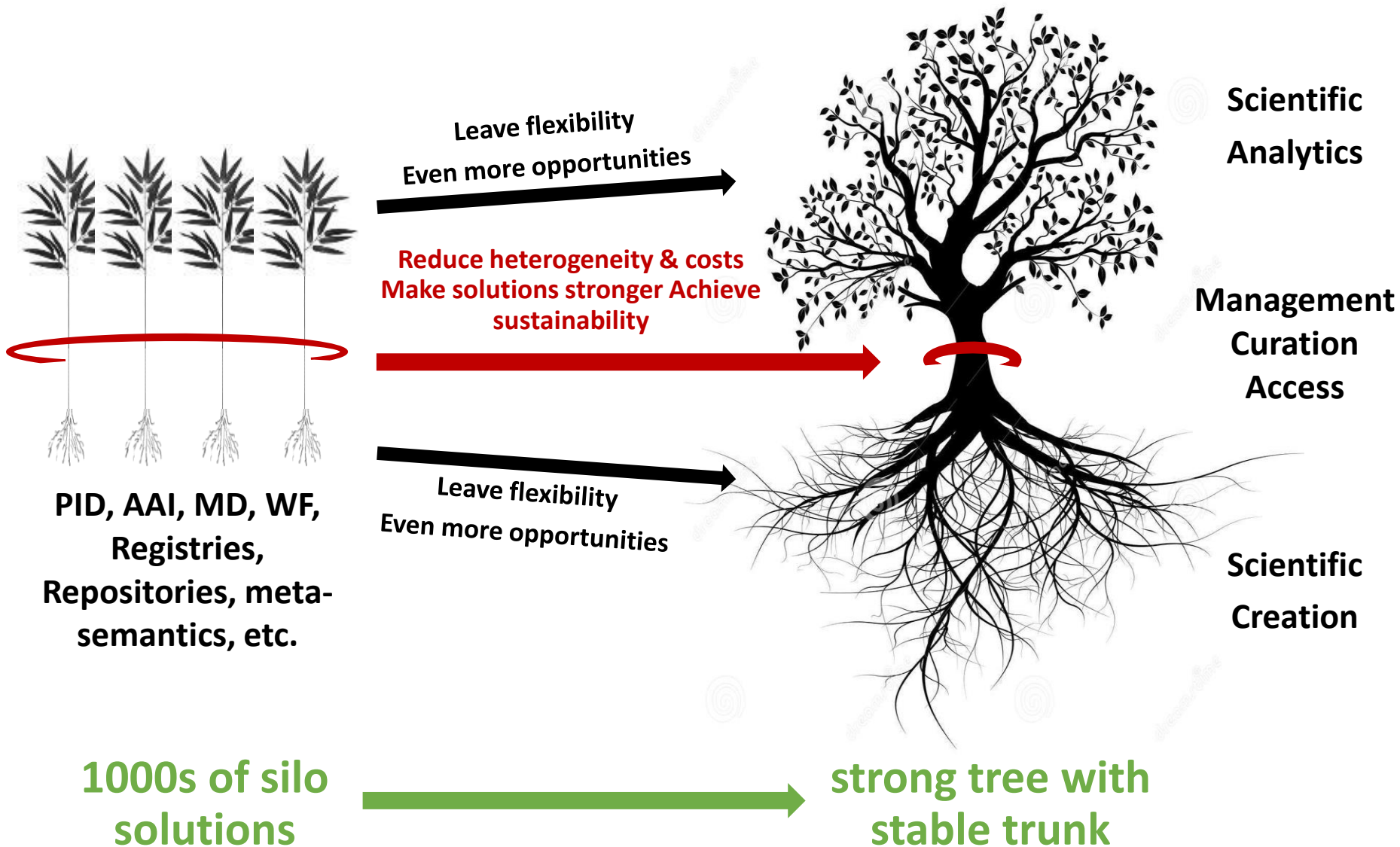
- in RDA ~ 5000 experts from ~123 countries join forces to improve conditions for data sharing and re-using
- it's all about interoperability based on open BP/standards
 - but global interoperability is expensive – where to settle?
- the actual work is
 - done in working and interest groups of experts which create specifications of policies, procedures, interfaces, etc.
 - bottom-up – people who want to overcome a barrier join
 - based on actual use cases
 - based on rough consensus – not everyone needs to agree
 - based on adoption as major success criteria
- is it familiar to you – no surprise: IETF model
- RDA 1st plenary March 2013 in Göteborg
- RDA 9th plenary April 2017 in Barcelona
- from 4 to 75 RDA groups
- 17 endorsed recommendations by working groups
- RDA is known in science and needs to open up

Total RDA community members 5418

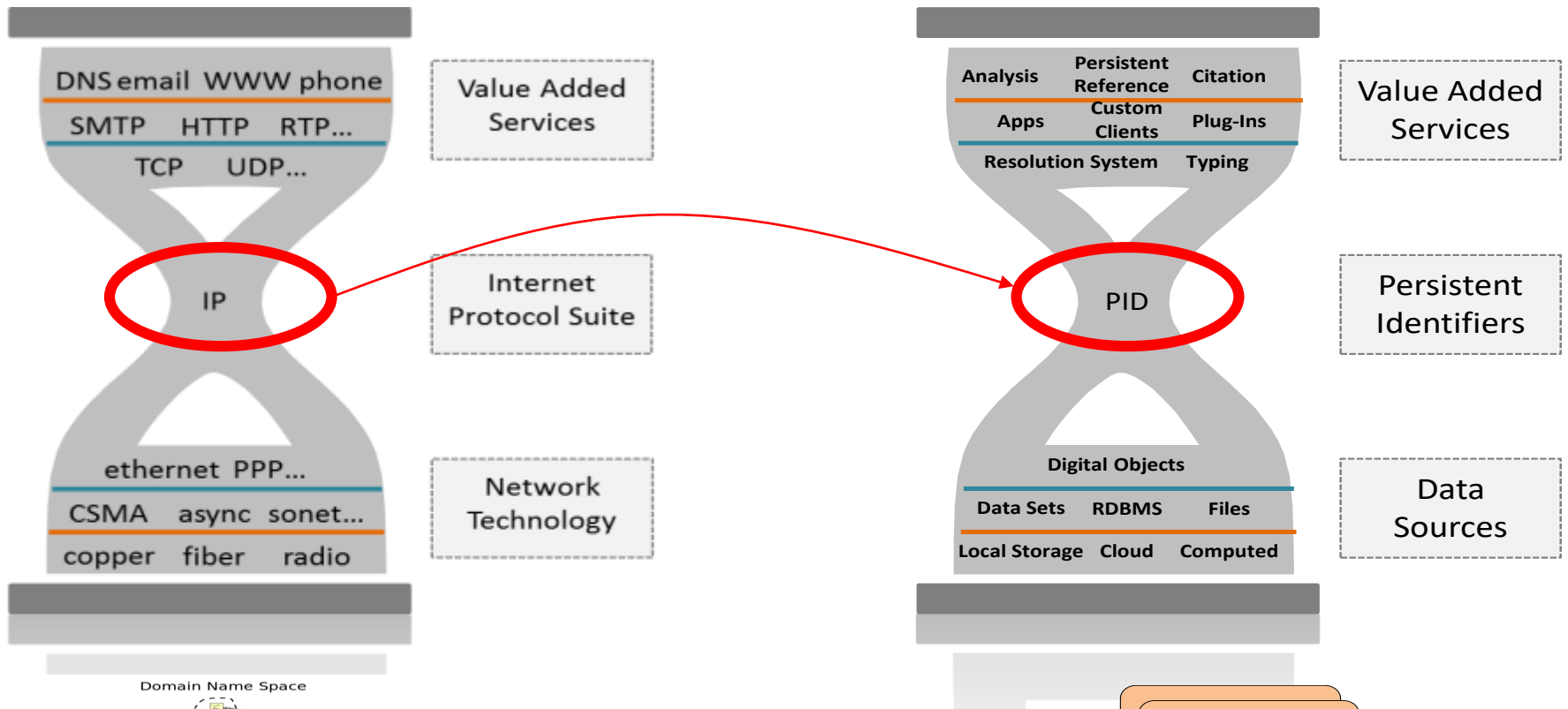
RDA worldwide growth



Fundamental Observation

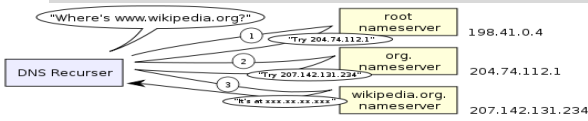


Global und Persistent IDs as Anchors



PID System is a catalyzer to define a new basis and to come to new services

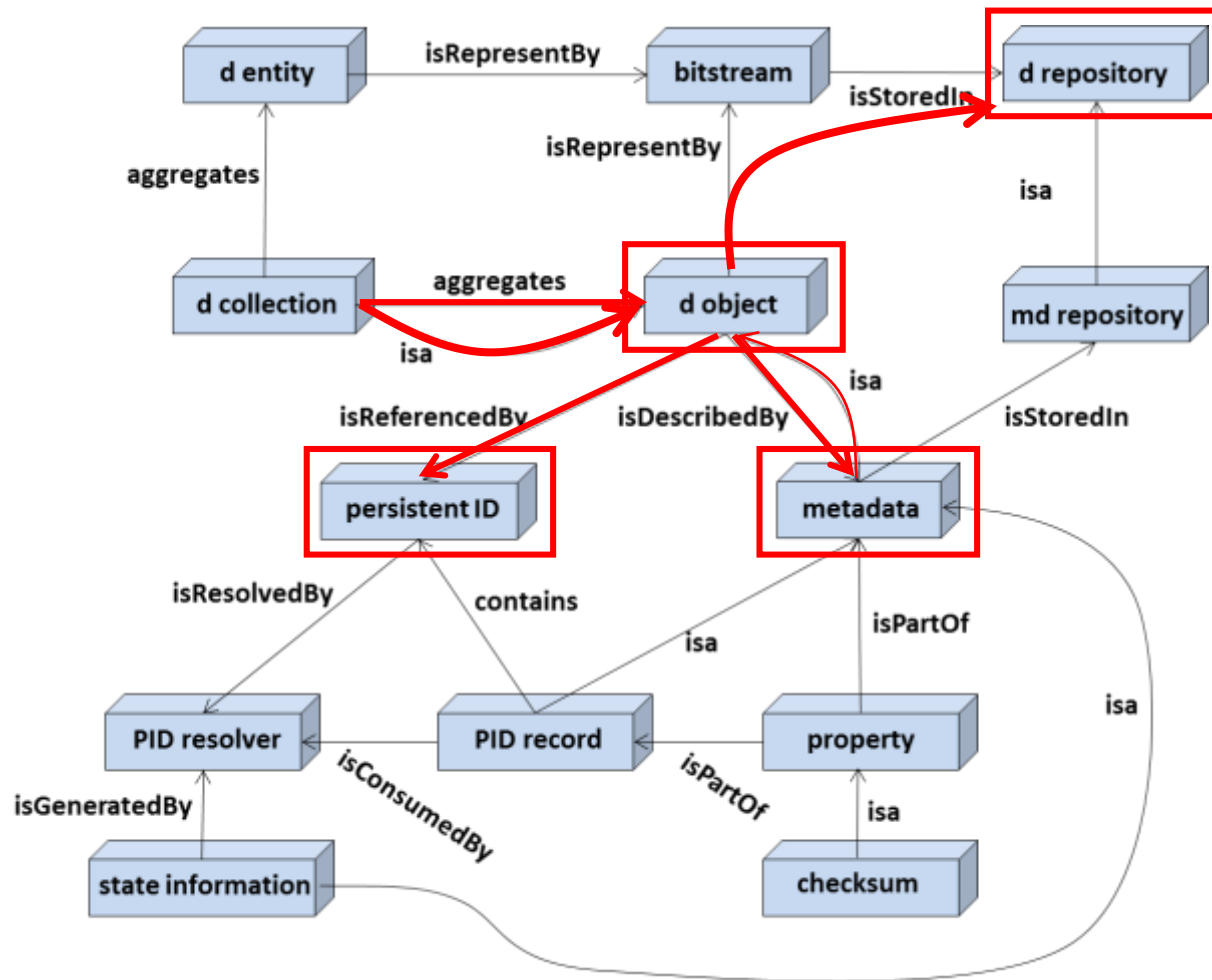
this is now excepted almost everywhere in science!



point to each other

metadata attributes

RDA DFT – Simple powerful data model



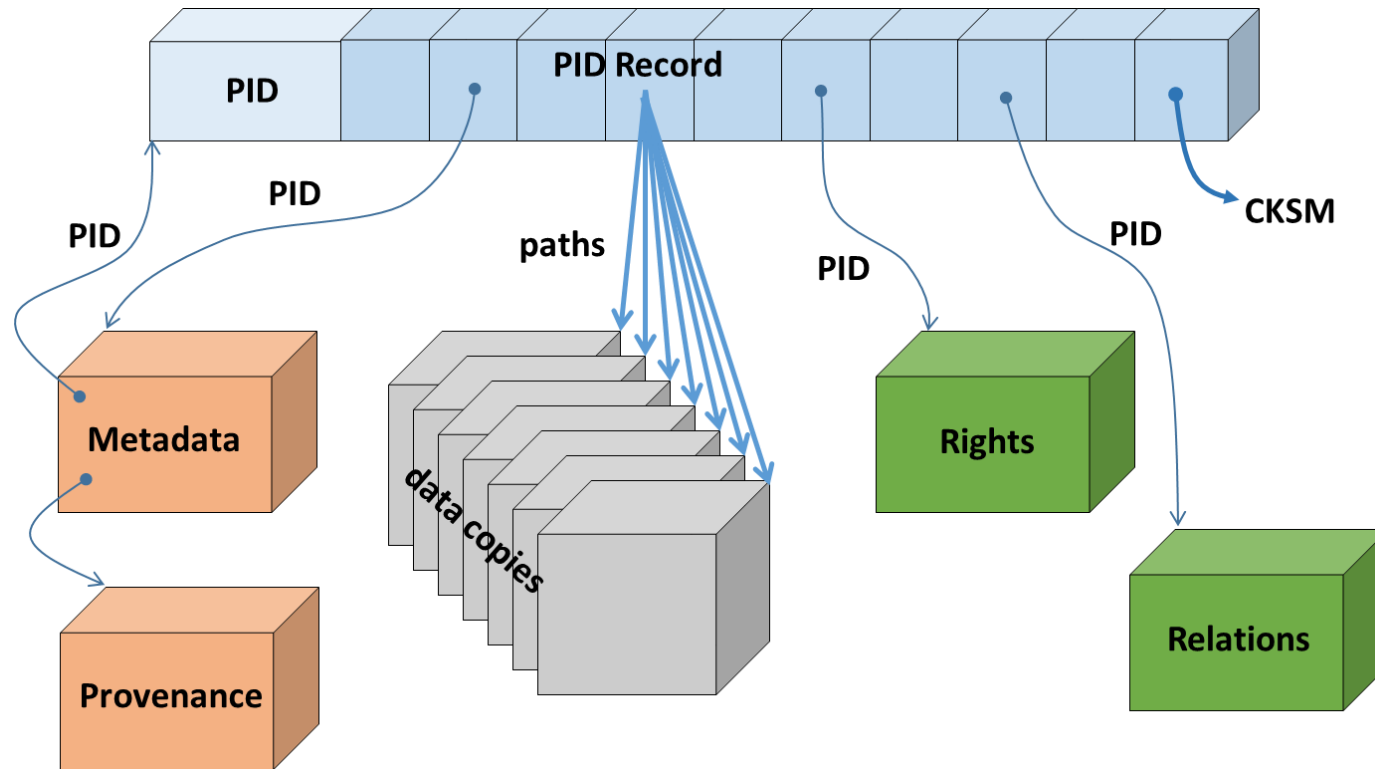
Data Foundation and Terminology WG

Core model is very simple.

If all software developers would implement this model, we would get an enormous increase in efficiency.

Deviations can become very expensive.

PID resolution to state information



PID Information Type WG & PID Kernel Information WG

- specify principles of interoperability
- specify core types such as „checksum“
- be compliant to ITU X.1255

Worldwide Handle System



**Independent Swiss
Foundation**

**DONA Board of
International Experts**

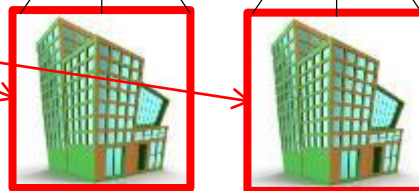


**Redundant network
of root nodes**

Contracts



**Services in
Germany**



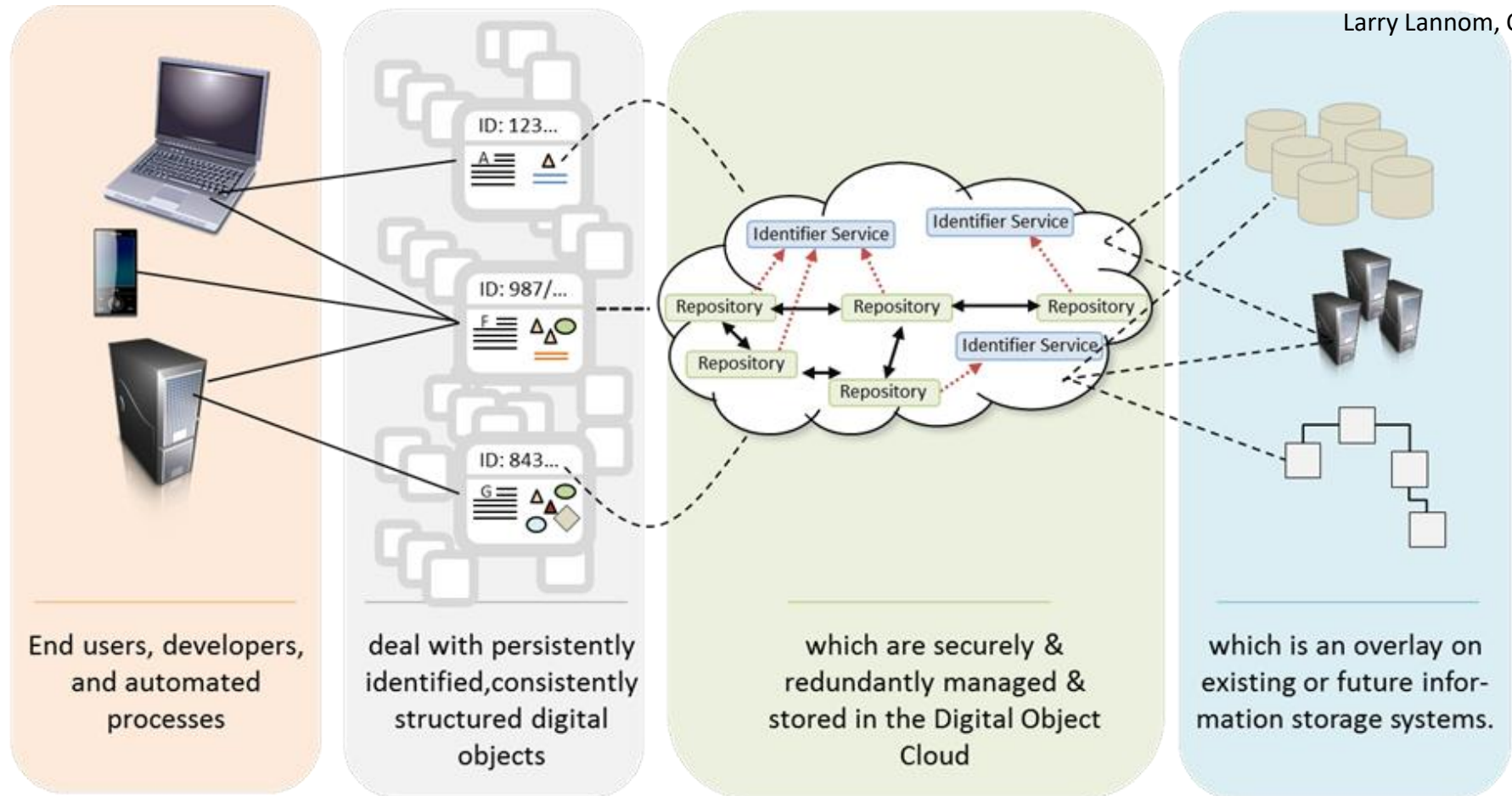
EPIC

DataCite

**increase transparency
get recognition by bodies
(ITU, etc.)**

Global Digital Object Cloud

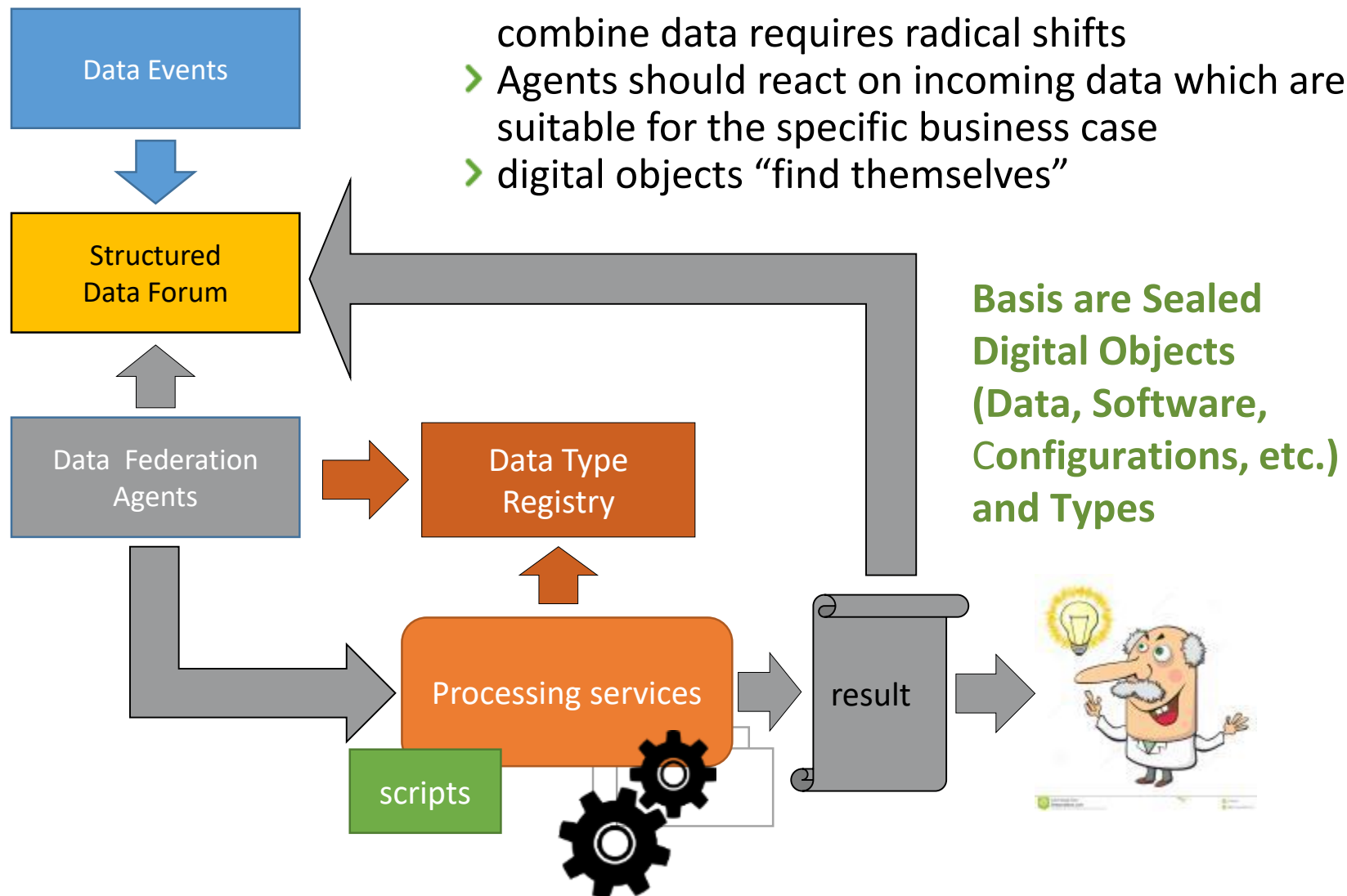
taken from
Larry Lannom, CNRI



- implementing PIDs systematically
- Global Climate Community implementing GDOC already

Type-Triggered Automatic Processing

- massiveness of data streams and wish to recombine data requires radical shifts
- Agents should react on incoming data which are suitable for the specific business case
- digital objects “find themselves”



RDA has more to offer, but not today ...



The 10th RDA Plenary Meeting will take place from 19 to 21 September 2017 in Montreal, Canada. The meeting is co-organised by RDA, the University of Montreal and Research Data Canada, Canada.

<https://www.rd-alliance.org/plenaries/rda-tenth-plenary-meeting-montreal-canada>

The 11th RDA Plenary Meeting will take place in March 2018 in Berlin.

Agenda

9.00	Peter Wittenburg (RDA)	Welcome and RDA Concepts for efficient data practices
9.30	Robert Kahn (CNRI)	From IP addresses to Persistent Identifiers – principles of Interlinking
10.15		Discussion
10.30	Coffee Break	
	Session 1 – Towards Interoperable Solutions	
11.00	Tobias Weigel (WDCC)	Infrastructure for the Global Climate Reporting
11.30	Ye Tian (CNIC-CAS)	Application of Digital Object Architecture in China
12.00	Peggy Irelan (Intel)	IoT Standards drive Frictionless Analytics
12.30	André Zwanziger (T-Systems)	MISP - Multi-IoT-Service-Plattform
13.00	Lunch Break	
	Session 2 – Towards Interoperable Solutions	
14.00	Alex Ntoko (ITU)	State of discussions in ITU
14.30	Jürgen Heiles (Siemens)	The role of Identifiers in IoT
15.00	Juanjo Hierro (FIWARE)	Supporting Researchers' Data Interoperability: FIWARE for the development of the IoT Cloud for RDA Europe
15.30	Coffee Break	
16.00	Leif Laaksonen (moderator): Robert Kahn, Tobias Weigel, Ye Tian, Peggy Irelan, André Zwanziger, Alex Ntoko, Jürgen Heiles, Juanjo Hierro, Peter Wittenburg	Expert Panel on Globally Interoperable IoT Identification and Data Processing – which ways to go?
17.00	End	

**RDA is ready to
leave the scientific
domain and open
up to industrial
domain!**

RDA Global

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - [@resdatall](https://twitter.com/resdatall)

LinkedIn -

www.linkedin.com/in/ResearchDataAlliance

Slideshare -

<http://www.slideshare.net/ResearchDataAlliance>

RDA Europe

Email - info@europe.rd-alliance.org

Twitter - [@RDA_Europe](https://twitter.com/RDA_Europe)

Email – peter.wittenburg@mpcdf.mpg.de



RDA Interest (IG) & Working Groups (WG) by Focus (1)

Total 81 groups:
30 Working Groups & 51 Interest Groups

Domain Science - focused

- ☐ **Agrisemantics WG**
- ☐ **BioSharing Registry WG**
- ☐ **Fisheries Data Interoperability WG**
- ☐ **On-Farm Data Sharing (OFDS) WG**
- ☐ **Rice Data Interoperability WG**
- ☐ **Wheat Data Interoperability WG**
- ☐ Agricultural Data IG (IGAD)
- ☐ Biodiversity Data Integration IG
- ☐ Chemistry Research Data IG
- ☐ Digital Practices in History and Ethnography IG

- ☐ Geospatial IG

- ☐ Global Water Information IG
- ☐ Linguistics Data Interest Group
- ☐ Health Data IG
- ☐ Mapping the Landscape IG
- ☐ Marine Data Harmonization IG
- ☐ Quality of Urban Life IG
- ☐ RDA/CODATA Materials Data, Infrastructure & Interoperability IG
- ☐ Research data needs of the Photon and Neutron Science community IG
- ☐ Small Unmanned Aircraft Systems' Data IG
- ☐ Structural Biology IG
- ☐ Weather, Climate and air quality IG

Community Needs - focused

- ☐ **Certification and Accreditation for Data Science Training and Education WG**
- ☐ **RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG**
- ☐ **Teaching TDM on Education and Skill Development WG**
- ☐ Archives & Records Professionals for Research Data IG

- ☐ Data for Development IG
- ☐ Development of Cloud Computing Capacity and Education in Developing World Research IG
- ☐ Education and Training on handling of research data IG
- ☐ Ethics and Social Aspects of Data IG
- ☐ International Indigenous Data Sovereignty IG

RDA Interest (IG) & Working Groups (WG) by Focus (2)

Total 81 groups:
30 Working Groups & 51 Interest Groups

Reference and Sharing - focused

- ❑ Data Citation WG
- ❑ Data Description Registry Interoperability WG
- ❑ Data Security and Trust WG
- ❑ Empirical Humanities Metadata WG
- ❑ Provenance Patterns WG
- ❑ RDA / WDS Publishing Data Bibliometrics WG
- ❑ Research Data Collections WG
- ❑ QoS-DataLC Definitions WG
- ❑ International Materials Resource Registries WG
- ❑ National Data Services IG
- ❑ RDA/CODATA Legal Interoperability IG
- ❑ Reproducibility IG
- ❑ Data Discovery Paradigms IG
- ❑ Repository Core Description WG
- ❑ Research Data Repository Interoperability WG

Partnership Groups

- ❑ RDA / TDWG Metadata Standards for attribution of physical and digital collections stewardship WG
- ❑ RDA/NISO Privacy Implications of Research Data Sets IG
- ❑ RDA/WDS Scholarly Link Exchange Working Group
- ❑ RDA/WDS Publishing Data IG
- ❑ ELIXIR Bridging Force IG

RDA Interest (IG) & Working Groups (WG) by Focus (3)

Total 81 groups:
30 Working Groups & 51 Interest Groups

Data Stewardship and Services – focused

- ❑ **Brokering Framework WG**
- ❑ **WDS/RDA Assessment of Data Fitness for Use WG**
- ❑ **RDA / WDS Publishing Data Workflows WG**
- ❑ Active Data Management Plans IG
- ❑ Data in Context IG
- ❑ Data Rescue IG
- ❑ Data Versioning IG
- ❑ Domain Repositories IG
- ❑ Libraries for Research Data IG

- ❑ Long tail of research data IG
- ❑ Preservation e-Infrastructure IG
- ❑ Preservation Tools, Techniques, and Policies IG
- ❑ RDA/WDS Certification of Digital Repositories IG
- ❑ RDA/WDS Publishing Data Cost Recovery for Data Centres IG
- ❑ Repository Platforms for Research Data IG
- ❑ Research Data Provenance IG
- ❑ Virtual Research Environments IG

Base Infrastructure – focused

- ❑ **Array Database Assessment WG**
- ❑ **Data Type Registries WG**
- ❑ **Metadata Standards Catalog WG**
- ❑ **Metadata Standards Directory WG**
- ❑ **PID Kernel Information WG**
- ❑ Data Fabric IG
- ❑ Data Foundations and Terminology IG
- ❑ Big Data IG

- ❑ Brokering IG
- ❑ Federated Identity Management IG
- ❑ Metadata IG
- ❑ PID IG
- ❑ Vocabulary Services IG

Towards an Open Forum of Sealed Data

Sealing means:

- unique identifier for each DO (Handle)
and associating crucial information with it such as
- unique identifier for owner (person, organisation) (personal ID)
 - unique identifier for devices (IPv6)
- a fingerprint of the DO (checksum, etc.)
 - a type ID (Handle)

It's all there, let's use it systematically

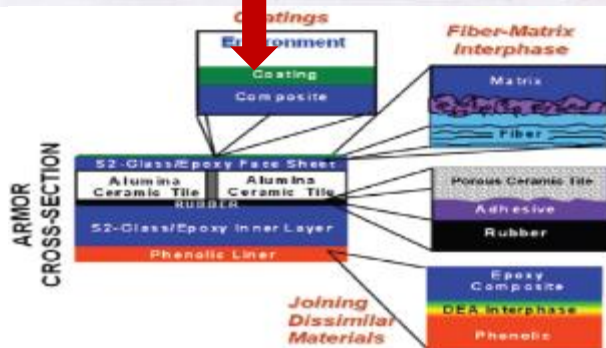
NoMaD - Material Science



- many Labs create data about materials and compounds (experiments + simulations)
 - space of chemical compounds is infinite
 - let's categorise this space to quickly find useful compound materials?
 - > 3 Mio aggregated entries now



The periodic table is color-coded by groups: Group 1 (purple), Group 2 (blue), Groups 3-10 (green), Group 11 (yellow), Group 12 (orange), Groups 13-18 (red), and Groups 19-20 (pink). The table includes element symbols, atomic numbers, and names.



- categorisation via Machine Learning etc.
- *Revolution: writing paper is not the only scientific goal anymore – it's repurposing data*

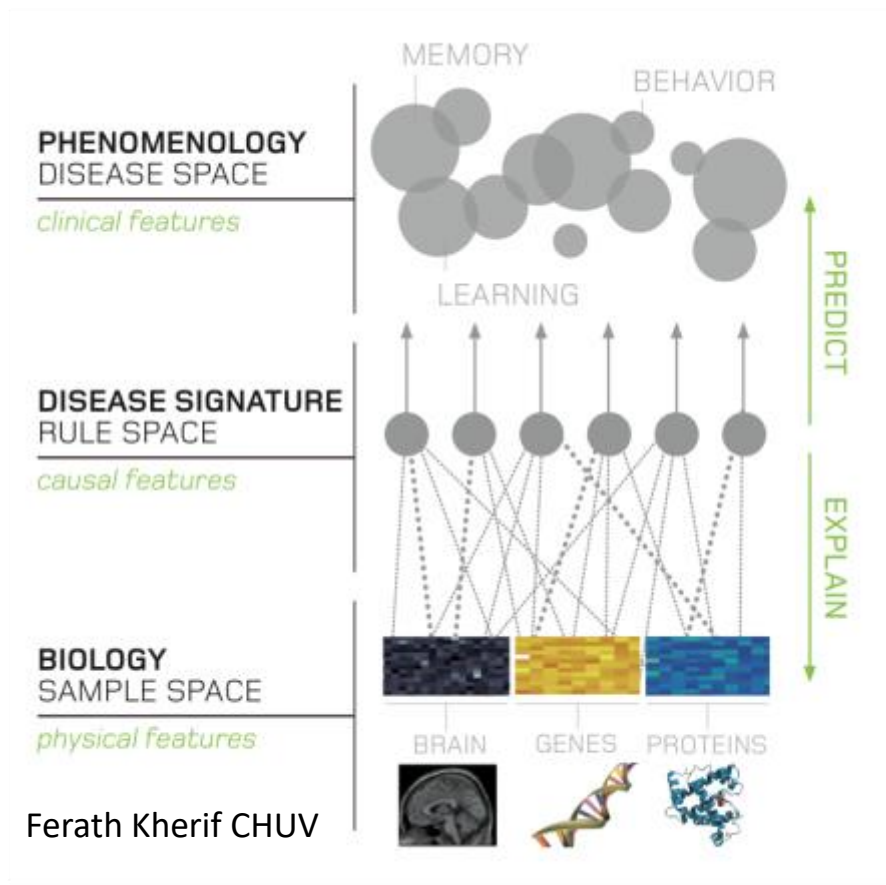
DOBES – Humanities/Languages



- ~70 global teams
- ~80 TB in online archive
- 4 dynamic external copies
- remote archives

- how can one use data to validate theories about the evolution of languages (and cultures) over thousands of years
- how to understand which languages are more "economic" than others
- *Revolution in humanities: scientific paper is not only goal anymore – it's about repurposing data*

Brain Research – Detect Disease Patterns



- early detection of causal basis of brain diseases
- machine learning to correlate patterns in data with phenomena
- much data from various specialized labs and hospitals is required

- **Revolution in medical world:**
 - *sharing data outside of the hospitals for new purposes*
 - *solving rights & ethical problems*

Organisational & Affiliate Members

43 Organisational Members

8 Affiliate Members

