



# Datenobjekte in zuverlässigen Repositorien

Peter Wittenburg, Thomas Zastrow  
Max Planck Computing and Data Facility

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

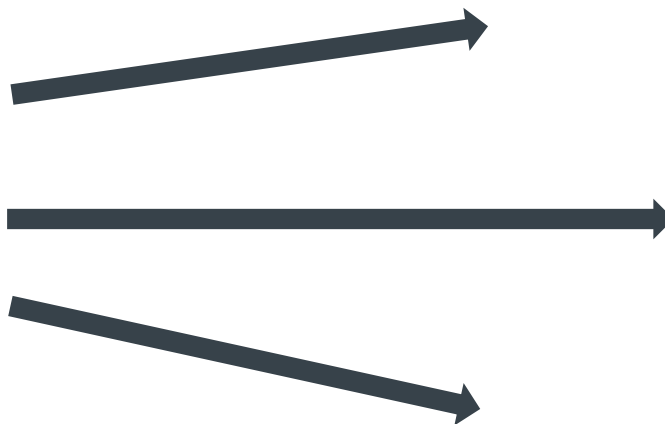
... not at the end of the development



**75%-80% of Data Scientist time for DM&A**  
**~ 50 Billions of devices creating a Data Monster!**  
**Are we ready to fight the monster – NO**

# Hope is on Stable Common Components

1000  
flowers



Scientific  
Analytics

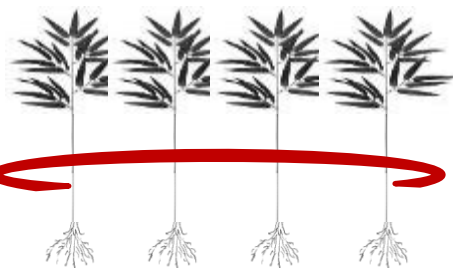
Management  
Curation  
Access

Scientific  
Creation

PID, AAI, MD, WF,  
Registries,  
Repositories,  
meta-semantics,  
etc.

# Hope is on Stable Common Components

1000  
flowers

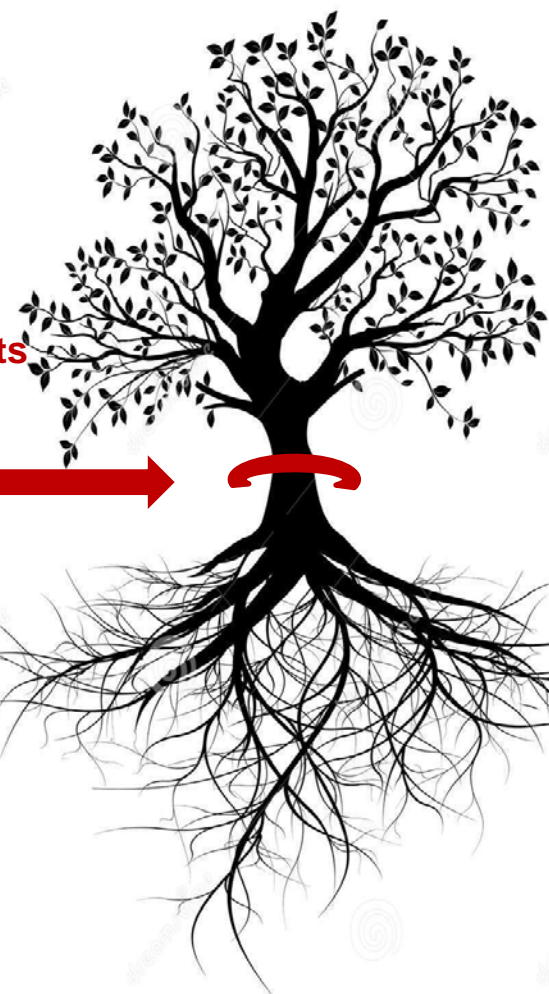


Leave flexibility  
Even more opportunities

Reduce heterogeneity & costs  
Make solutions stronger  
Achieve sustainability

Leave flexibility  
Even more opportunities

**PID, AAI, MD, WF,  
Registries,  
Repositories,  
meta-semantics,  
etc.**



Scientific  
Analytics

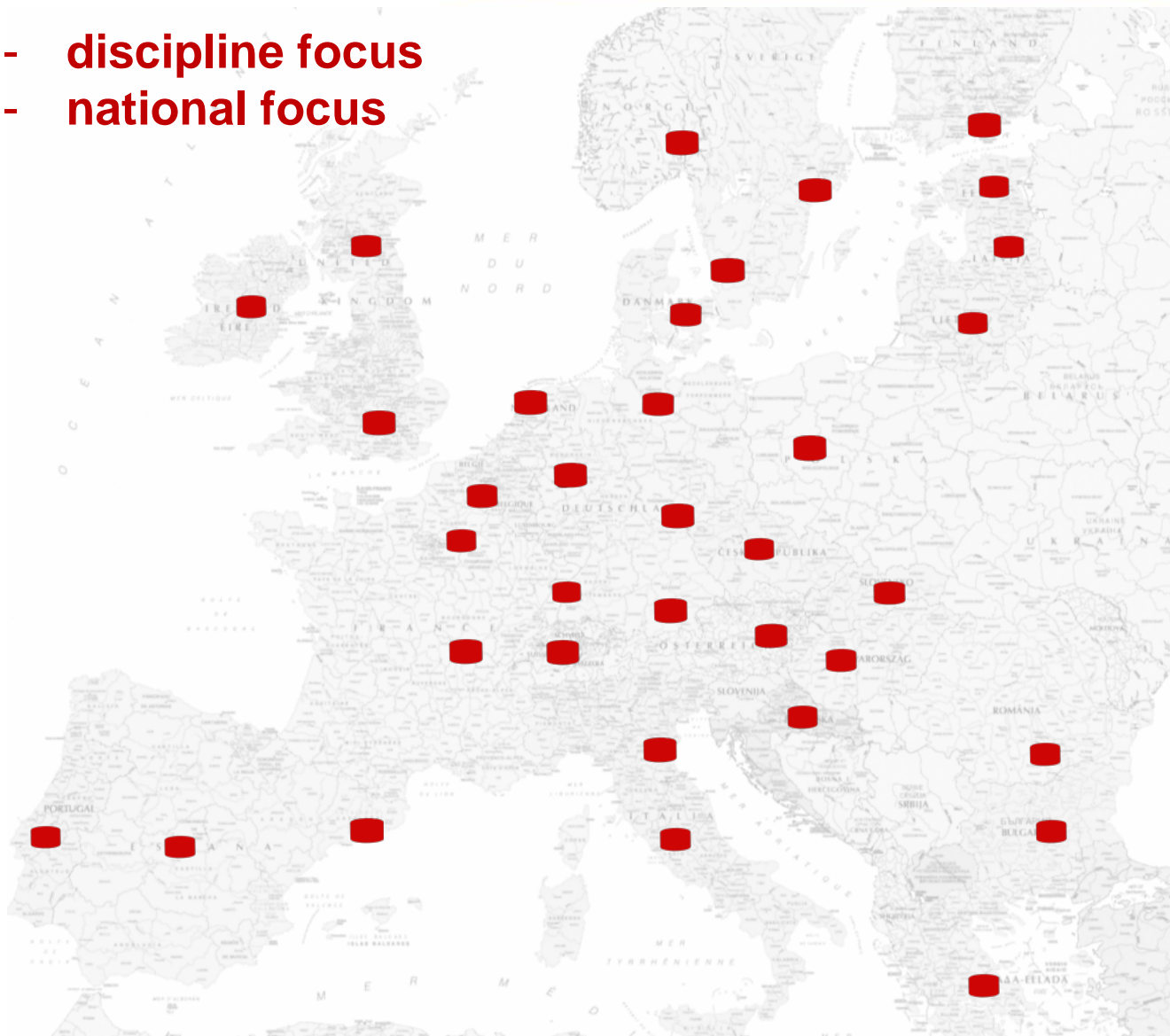
**Management  
Curation  
Access**

Scientific  
Creation



# Federations of Trustful Repositories

- discipline focus
- national focus

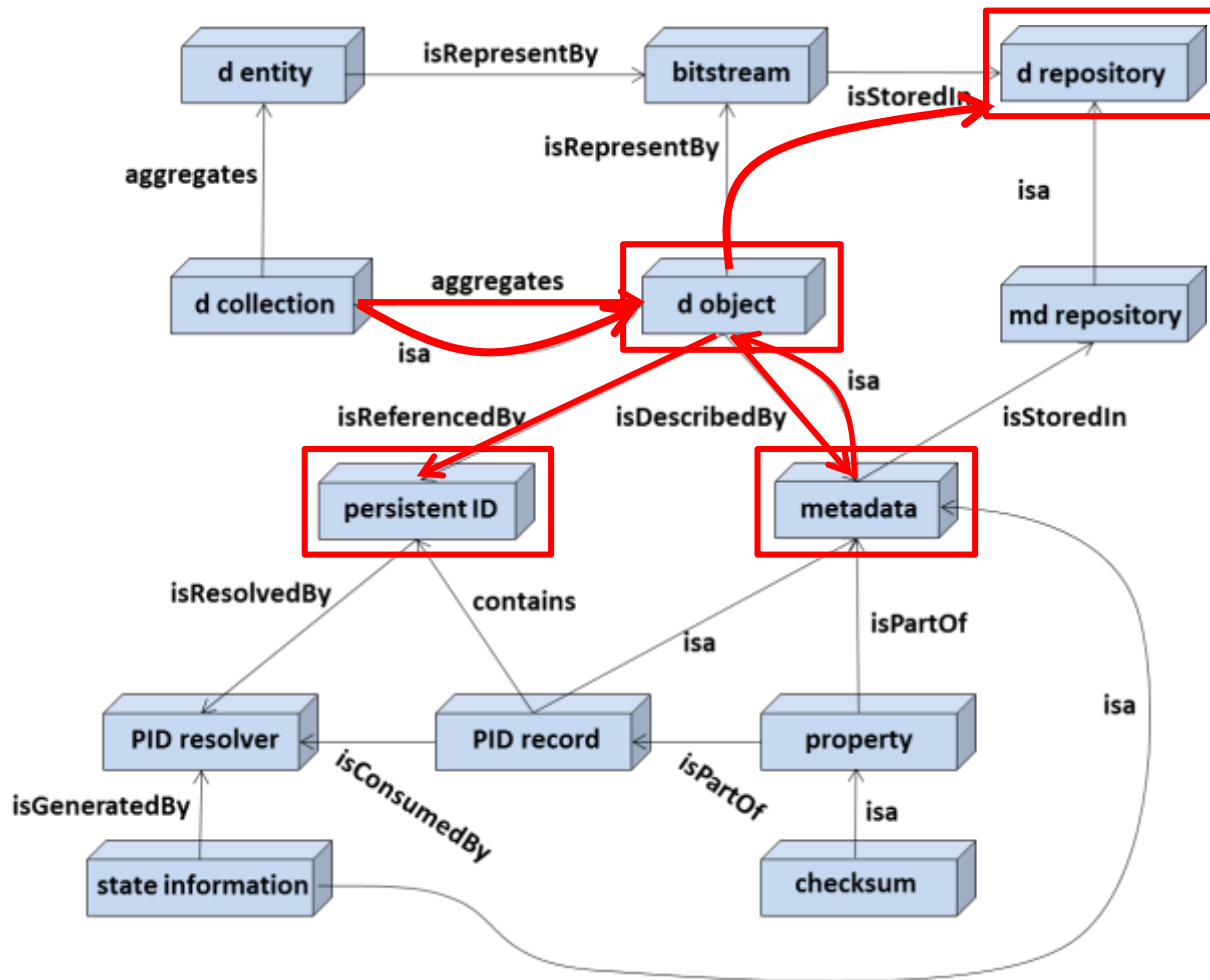


Social contracts to enable integration, seamless access and built on **trust**

at the core trustful centres being assessed by Data Seal of Approval / World Data Systems



# RDA DFT – Simple powerful data model



DFT core model is simple.

Messages are on purpose very simple.

Need to learn to speak with one voice.

If all SW builders would adhere to it, we would have gained a lot!

- **Data should be Findable**
- **Data should be Accessible**
- **Data should be Interoperable**
- **Data should be Re-usable.**

## To be Findable:

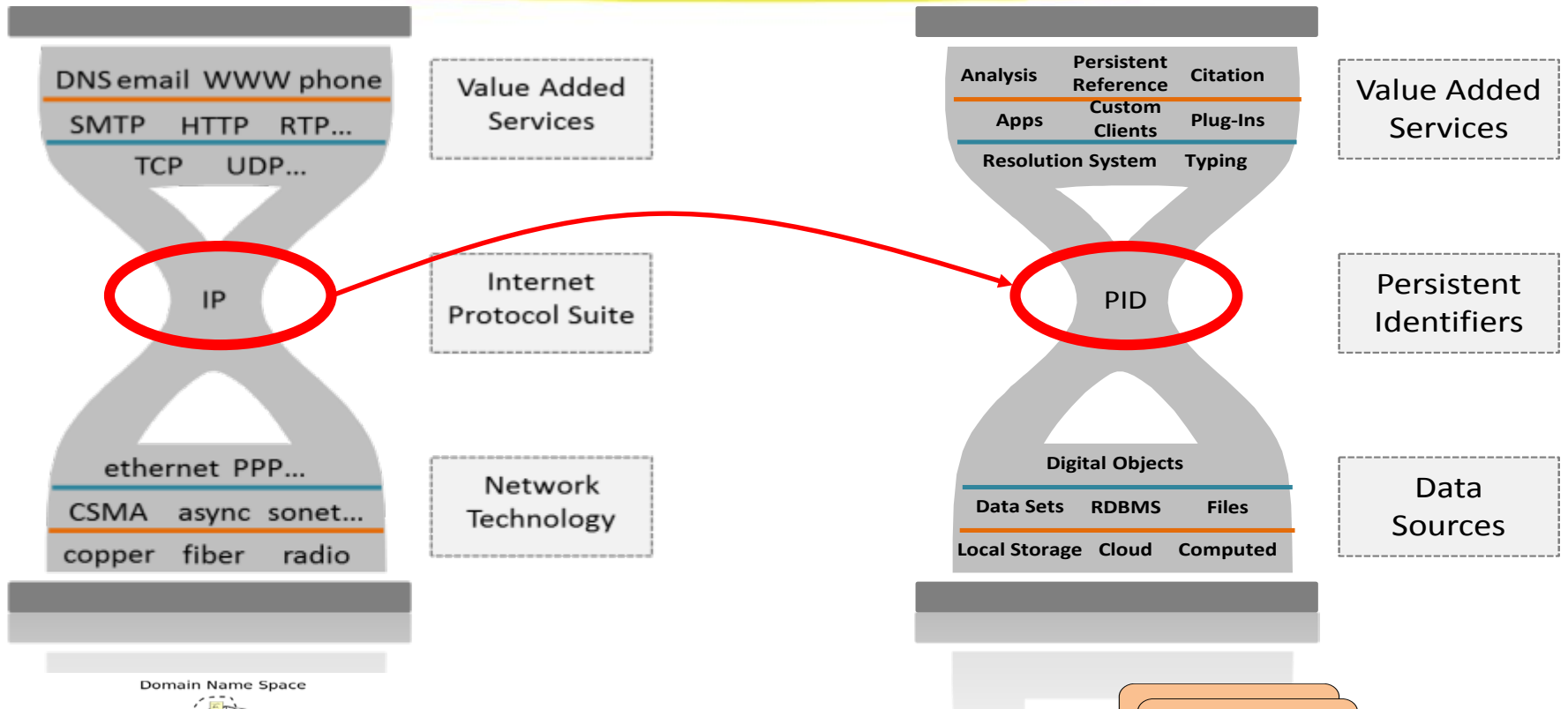
- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

## To be Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
  - A1.1 the protocol is open, free, and universally implementable.
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.



# PIDs to initiate Momentum



**PID System is just a catalyser for new layered services and new businesses. Need to change – need a momentum.**

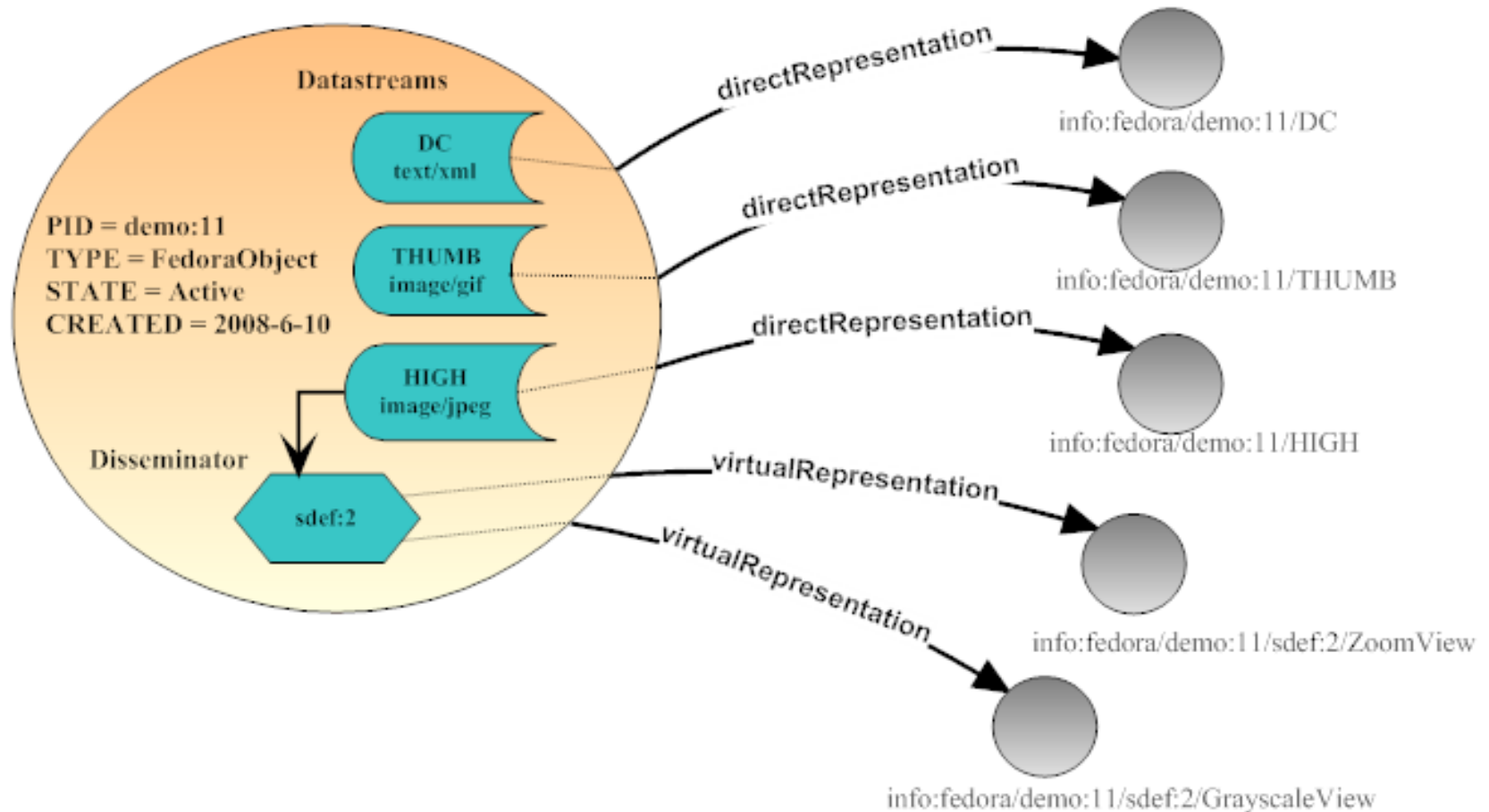


point to each other

metadata attributes



# Former Fedora C Object Model



- DO has a PID
- binding is done within the object

# Worldwide Handle system as Reference



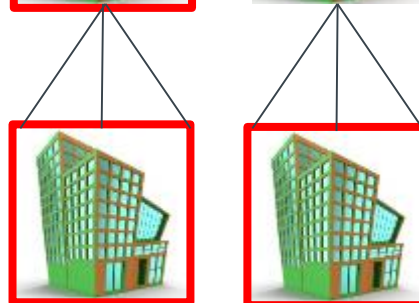
Independent Swiss Foundation



DONA Board of International Experts

Redundant network of root nodes

Contracts



- industry: need of a global identification solution  
- China: towards a national identification solution



**Alles klar oder Fragen?**