



Open Data is Not Enough

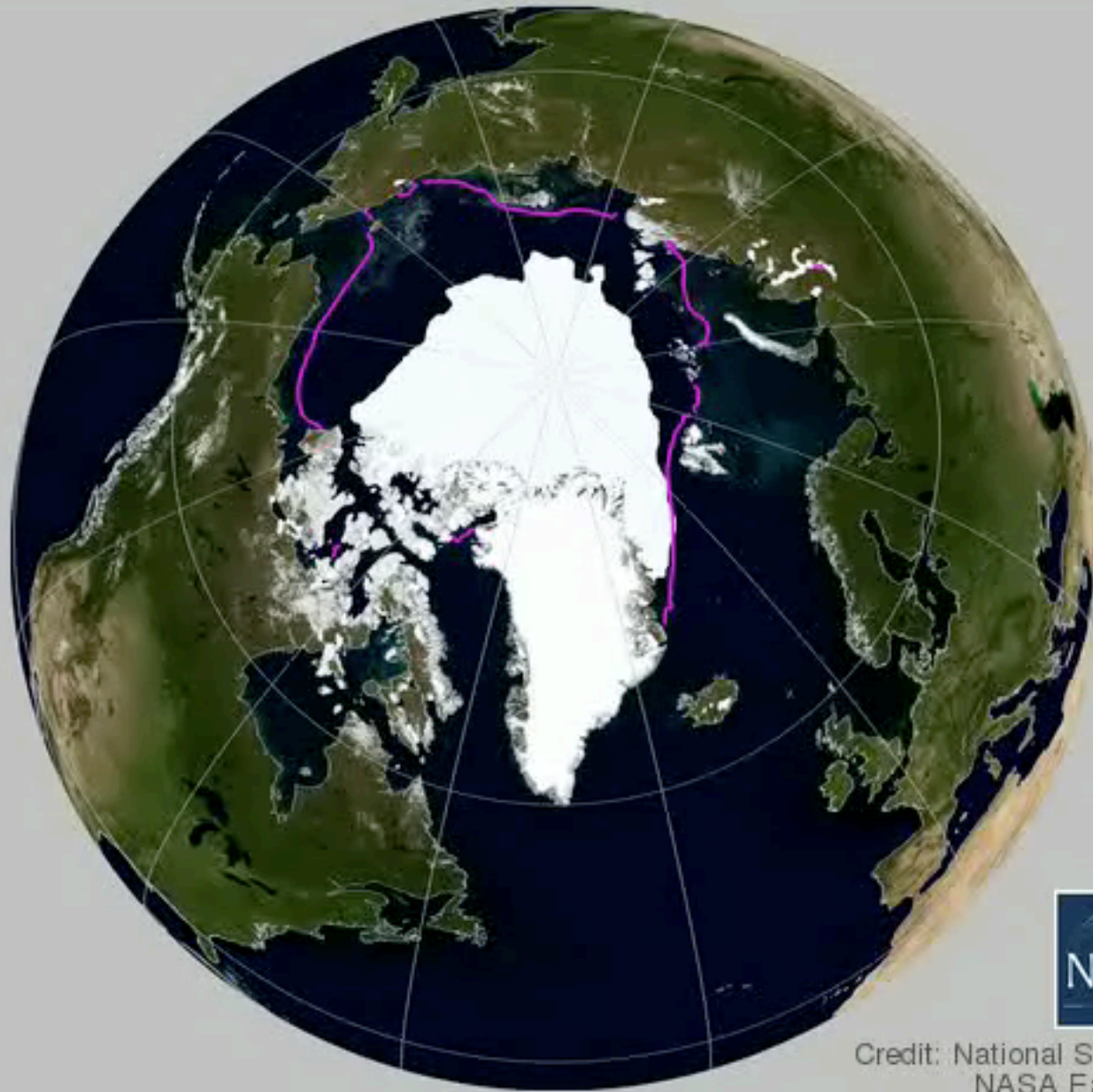
Making Data Sharing Work

Mark A. Parsons
0000-0002-7723-0950
Secretary General

The 2nd SPARC Japan Seminar 2015
Tokyo, Japan
21 October 2015



September Arctic Sea Ice Extent, 1979 to 2012



September 2012

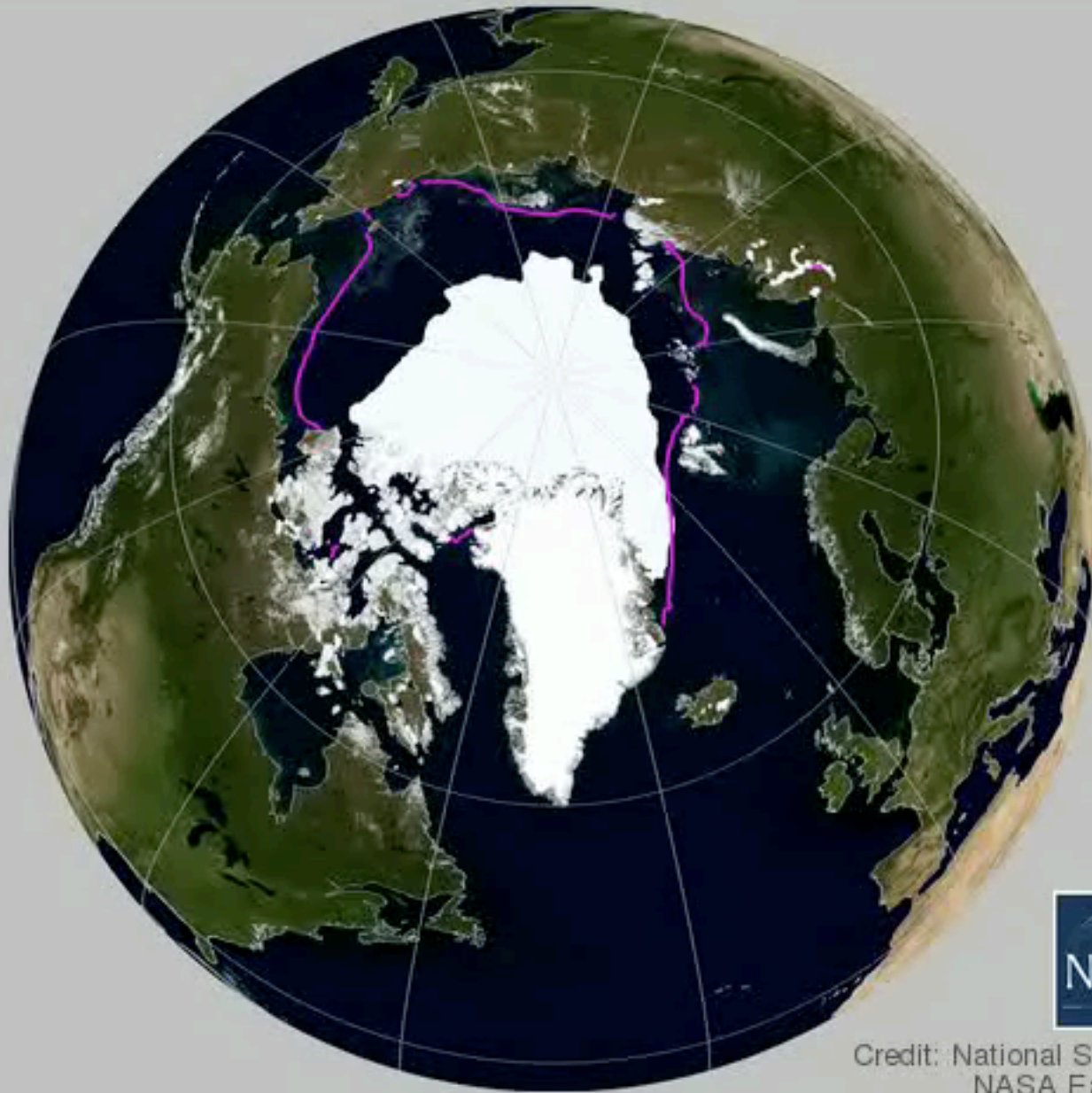


September 1979



Credit: National Snow and Ice Data Center,
NASA Earth Observatory

September Arctic Sea Ice Extent, 1979 to 2012



September 2012



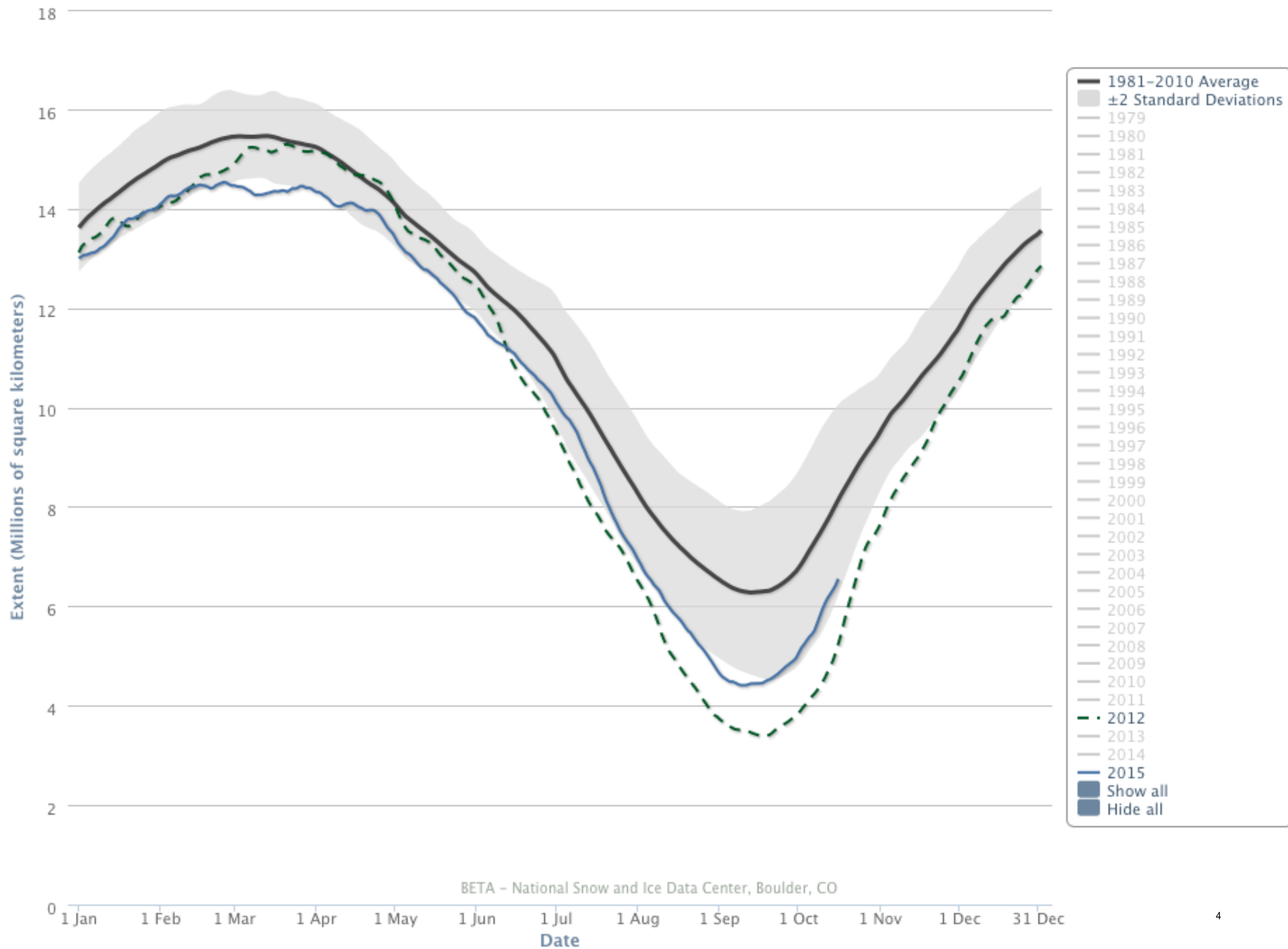
September 1979



Credit: National Snow and Ice Data Center,
NASA Earth Observatory

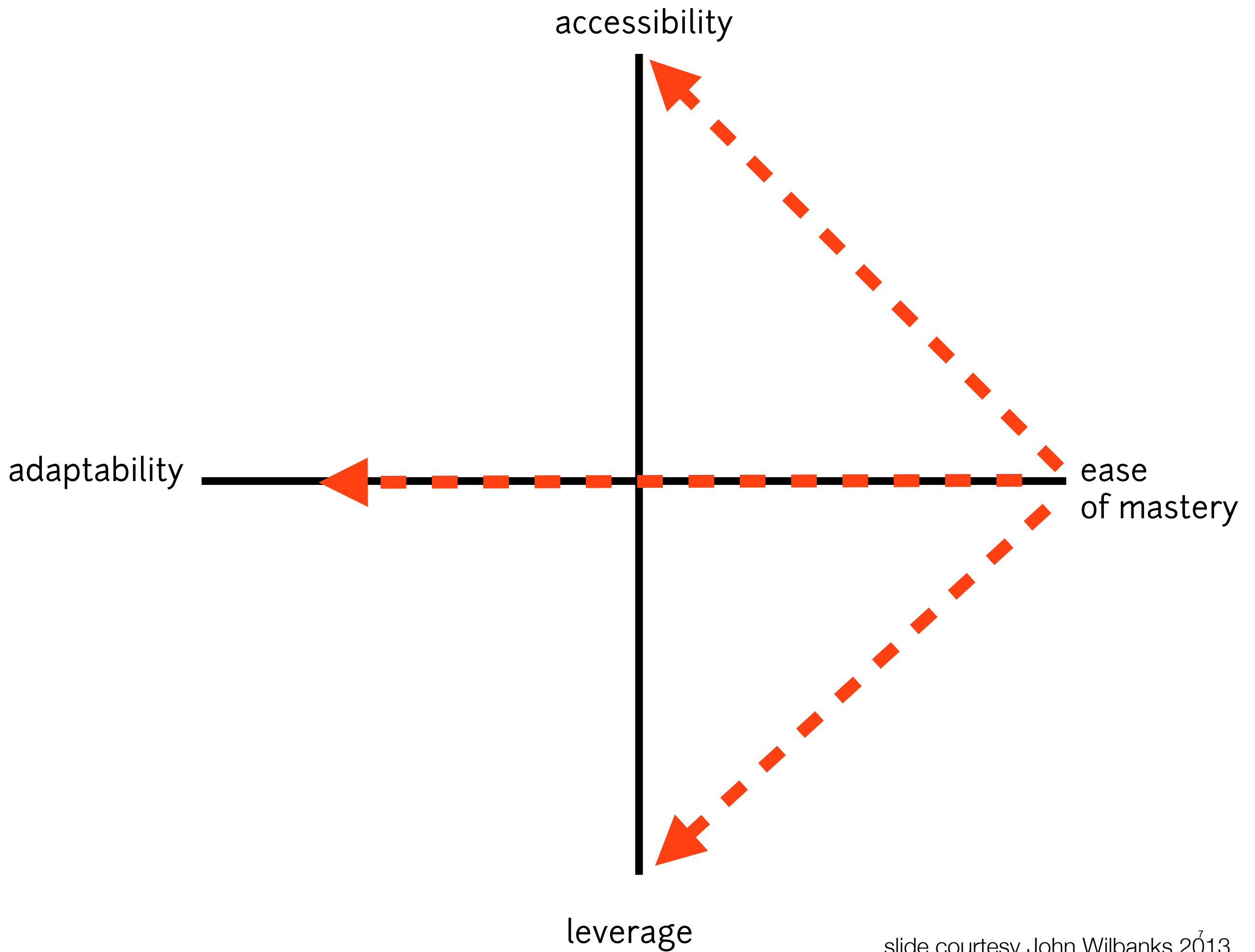
Arctic Sea Ice Extent

(Area of Ocean with at least 15% sea ice)



The generative value of data

- Generative value per Jonathan Zittrain (2008) as interpreted and extended to data by John Wilbanks:
“the capacity to produce **unanticipated change** through **unfiltered** contributions from **broad and varied audiences**.” —J. Zittrain
- Data become more generative by being more adaptable, more easily mastered, more accessible, and more connected and influential.
- Not net *present* value but net *potential* value.



To make open work we need

- **Curation**—increasing the (generative) value of the data
- **Context**—of both data and application of provider and user
- **Trust**—of data, information, organisations, institutions....
- Interfaces, connections, relationships, mediation—**Bridges**
- **People**

Research Data Alliance

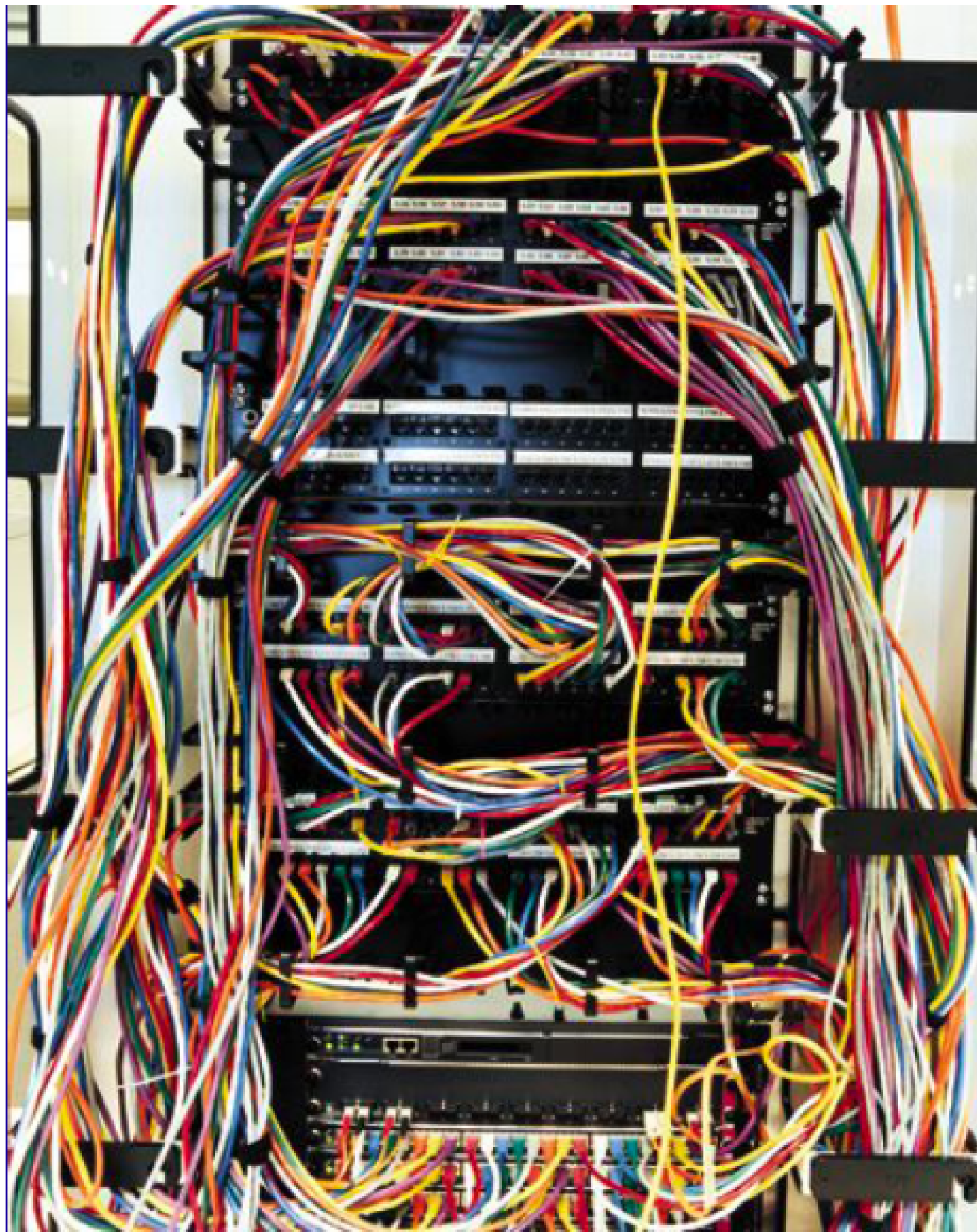


Vision

Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the **social and technical bridges** that enable open sharing of data.











Dynamics of Infrastructure

Edwards, et al. 2007 Understanding Infrastructure: Dynamics, Tensions, and Design.



- Infrastructures become “ubiquitous, accessible, reliable, and transparent” as they mature.
- Systems  Networks  Inter-networks
 - “system-building, characterized by the deliberate and successful design of technology-based services.”
 - “technology transfer across domains and locations results in variations on the original design, as well as the emergence of competing systems.”
 - Finally, “a process of **consolidation characterized by gateways** that allow dissimilar systems to be linked into **networks**.”

Not what, but

When is infrastructure?

Not what, but

When and

Who is infrastructure?

Bridges and Gateways

Gateways are often wrongly understood as “technologies,” i.e. hardware or software alone. A more accurate approach conceives them as combining **a technical solution with a social choice**, i.e. a standard, both of which must be integrated into existing users’ communities of practice. Because of this, gateways rarely perform perfectly.

— Edwards et al. 2007



Infrastructure is

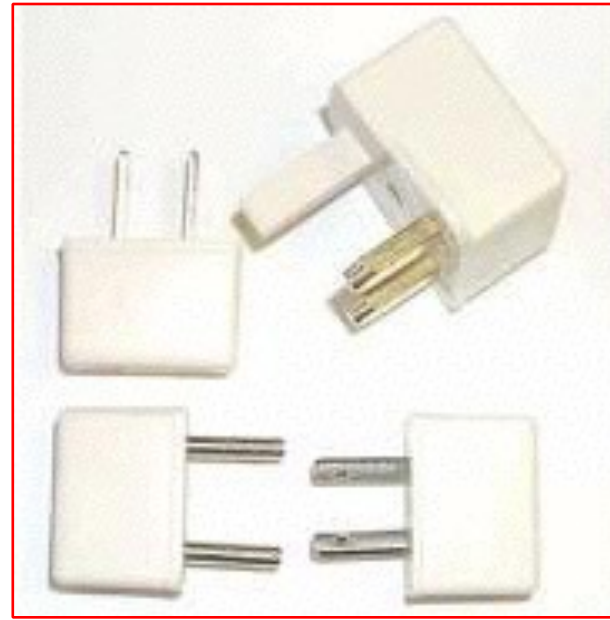
Relationships, interactions, and connections
between people, technologies, and institutions

“Create - Adopt - Use”

(in 12-18 months)



Adopted Policy



Systems
Interoperability



Common Types,
Standards, Metadata



Sustainable Economics



Adopted Community
Practice



Training, Education,
Workforce

*Traffic Image:
Mike Gonzalez*

Shared Principles



- *Openness*
- *Consensus*
- *Balance*
- *Harmonization*
- *Community Driven*
- *Non-profit*

RDA: Accelerate Data Sharing and Interoperability Across Cultures, Communities, Scales, Technologies

■ **Technical parts of the data engine:**

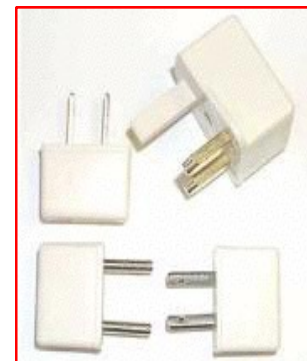
- Data type registries reference model
- Wheat data interoperability framework

■ **Rules of the road:**

- Common agreement on data citation
- Common practice for data repositories
- Principles of legal interoperability

■ **Better drivers**

- Summer schools in data science and cloud computing in the developing world (with CODATA)
- Active data management plan development and monitoring



Systems
Interoperability



Common Types,
Standards, Metadata



Policy and Practice



Sustainable
Economics



Training, Education,
Workforce

Initial Products—adopt one today!

- A basic vocabulary of **foundational terminology** and query tool to make sure we know what we're talking about.
- A **data type model and registry** ("MIME-types" for data) to help tools interpret, display, and process data.
- A **persistent identifier type registry** to help search engines understand what they are pointing to and retrieving.
- A basic set of **machine actionable rules** to enhance trust

New Products — adopt one today!

- A **metadata standards directory** so we can describe similar things consistently
- A **dynamic-data citation methodology** so we can reference precise subsets of changing data.
- Semantically linked **terms describing wheat data** so we can share harvest and related information around the world
- **Services and methods for finding data across multiple registries**, to help cross disciplinary and multi-faceted discovery.

Next Products—coming next Plenary!



- A **unified repository certification scheme** to reduce confusion and improve trust.
- A suite of **data publishing-related services** for
 - measuring bibliometrics
 - managing data workflows
 - interconnecting articles and data

The Research Data Alliance Community Today



Total RDA Community Members: 3243

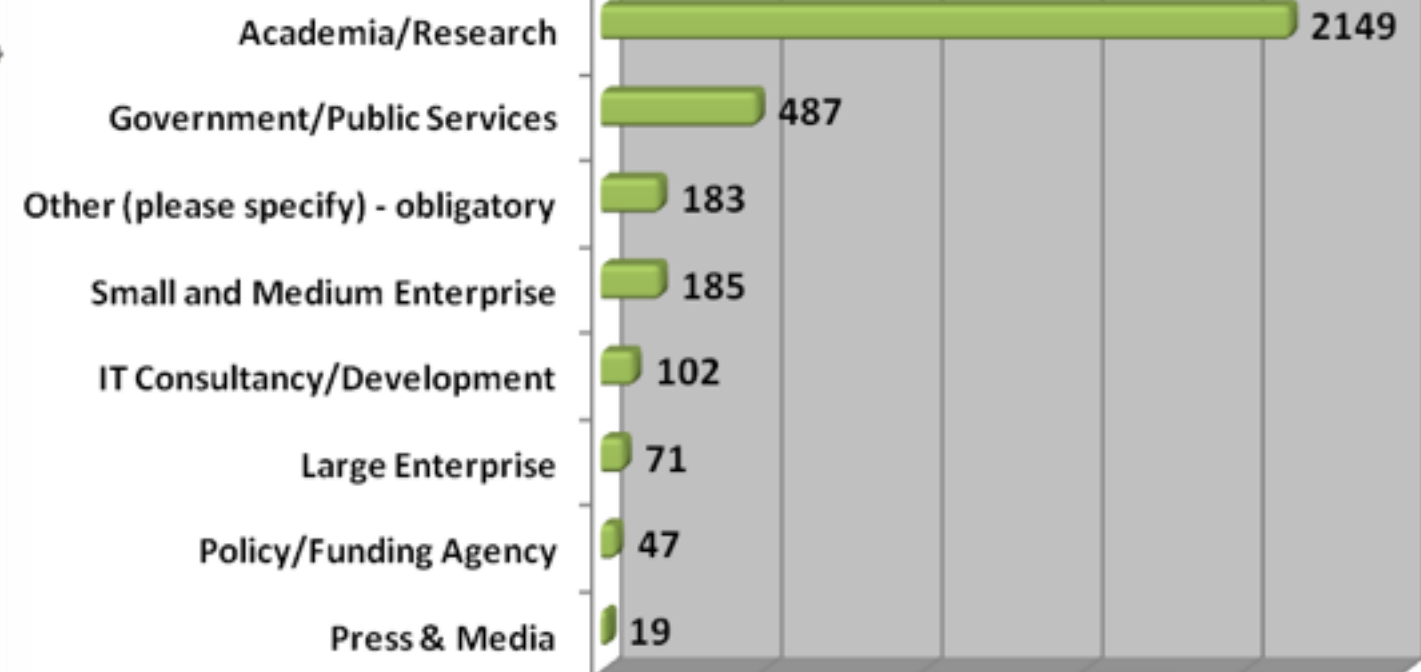
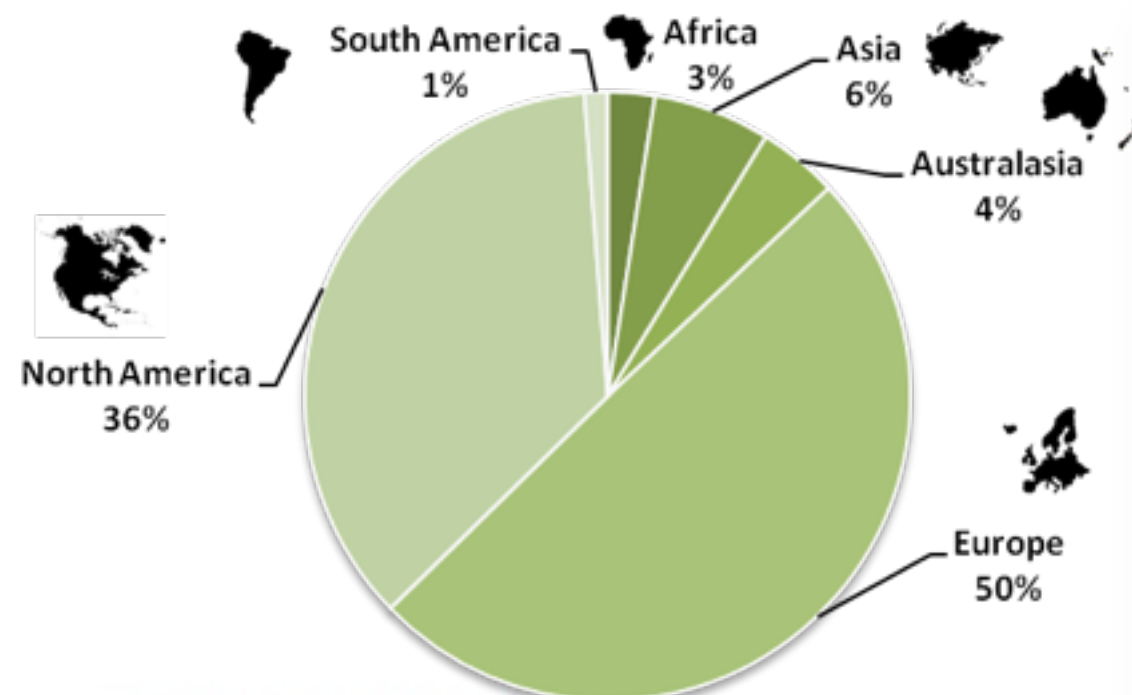


from 103 countries



56 Working and Interest Groups

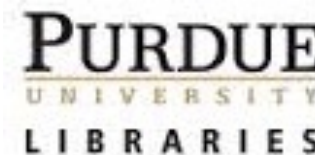
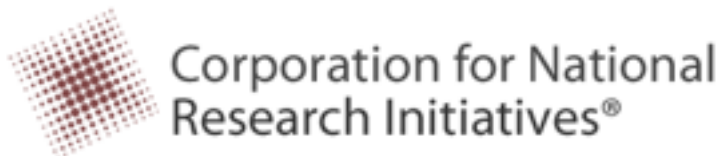
May - July Aug - Oct Nov - Jan Feb - Apr May - July Aug - Oct Nov - Jan Feb - Apr May - July Aug - Sept



RDA Organisational Members and Affiliates



canarie

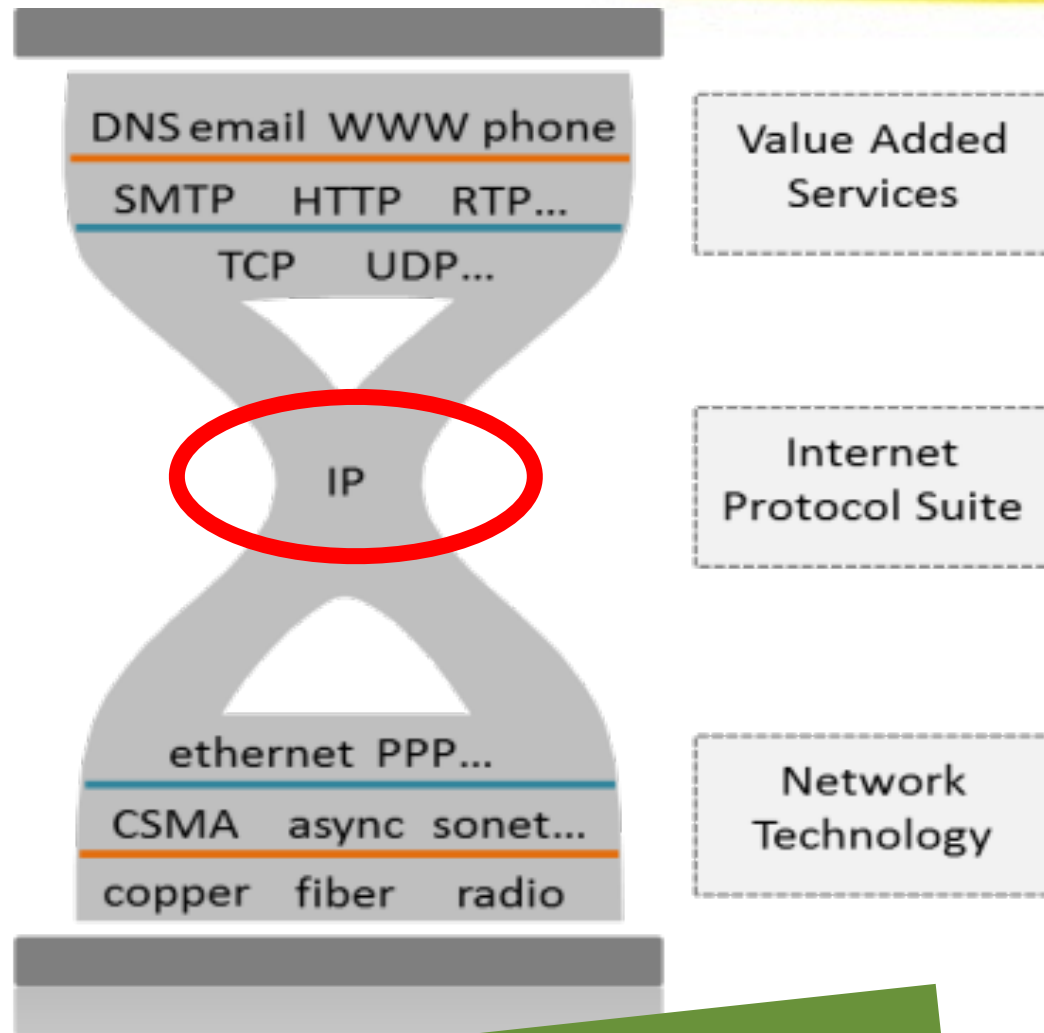


Some themes amidst the difference

1. **Persistent Identifiers** for data, documents, people, organisations, instruments — Everything!
2. **Certifying Trust** in assertions, evidence, organisations, processes...
3. The value of **Conversations, Relationships, and Mediation** — an agile network effect.

An Area of Convergence and Agreement

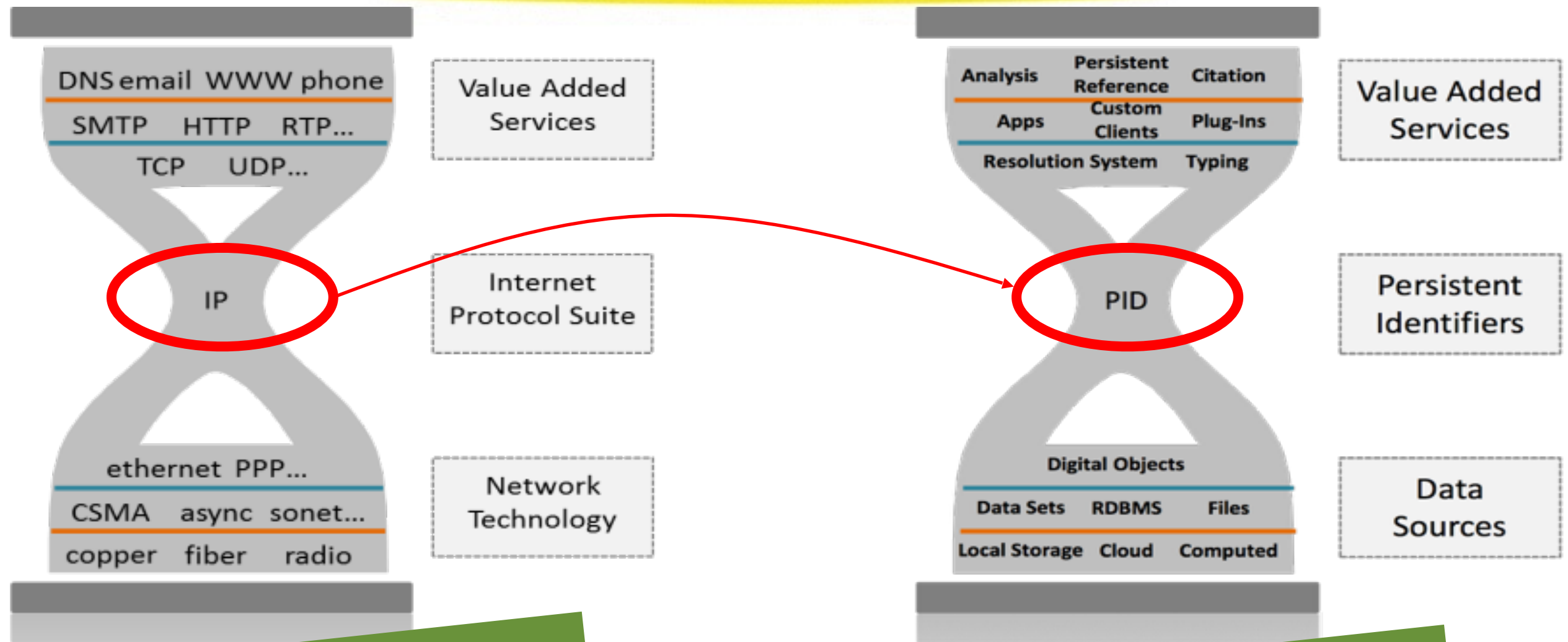
<#>



Internet Domain
nodes with IP numbers
packages being exchanged
standardized protocols

An Area of Convergence and Agreement

<#>

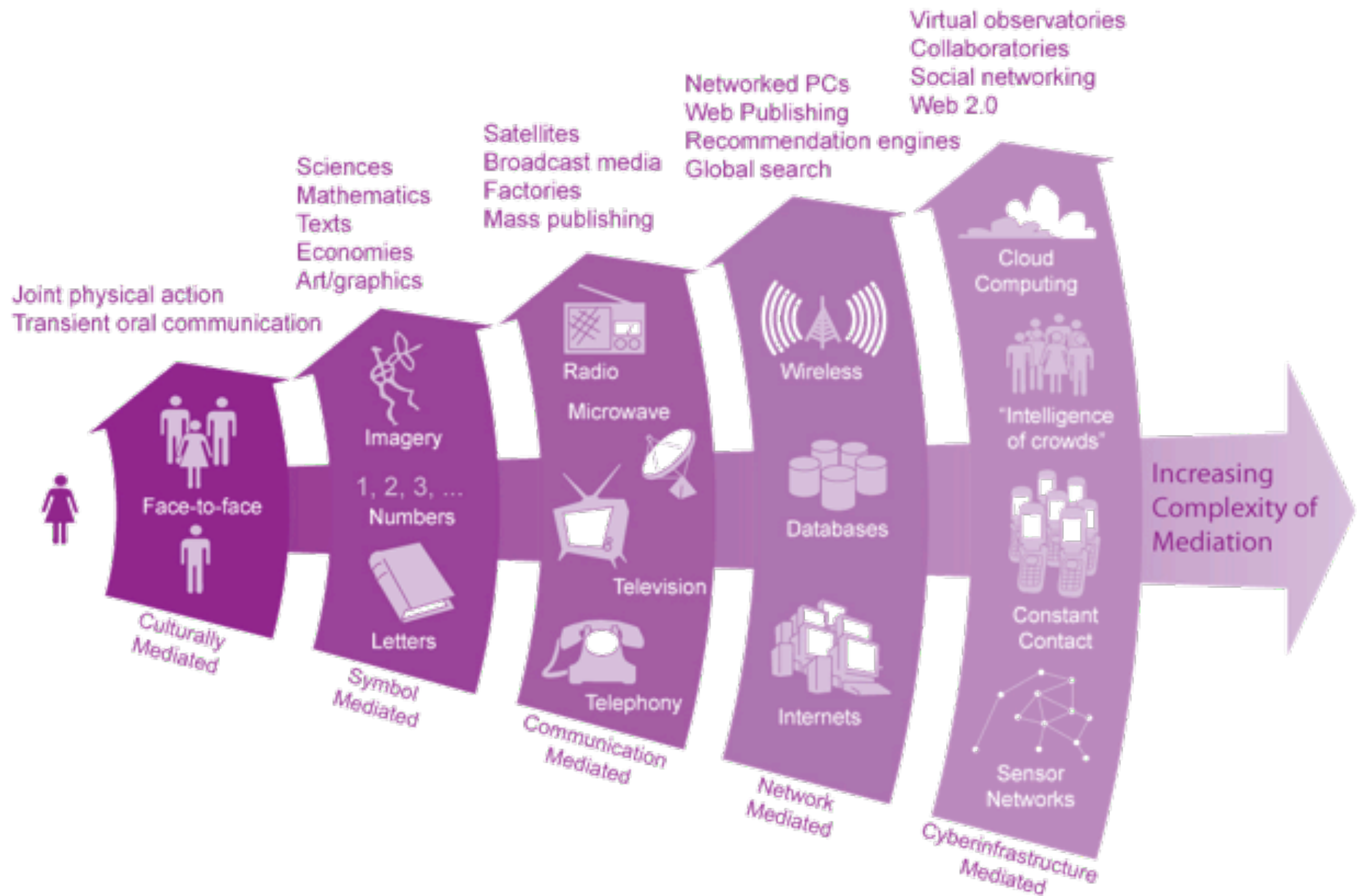


Internet Domain
nodes with IP numbers
packages being exchanged
standardized protocols

Data Domain
objects with PID numbers
objects being exchanged
standardized protocols

Some themes amidst the difference

1. **Persistent Identifiers** for data, documents, people, organisations, instruments — Everything!
2. **Certifying Trust** in assertions, evidence, organisations, processes...
3. The value of **Conversations, Relationships, and Mediation** — an agile network effect.



Increasing Complexity of Mediation

From: C. Borgman, 2008, NSF Cyberlearning Report

An Agile Manifesto for Organisations

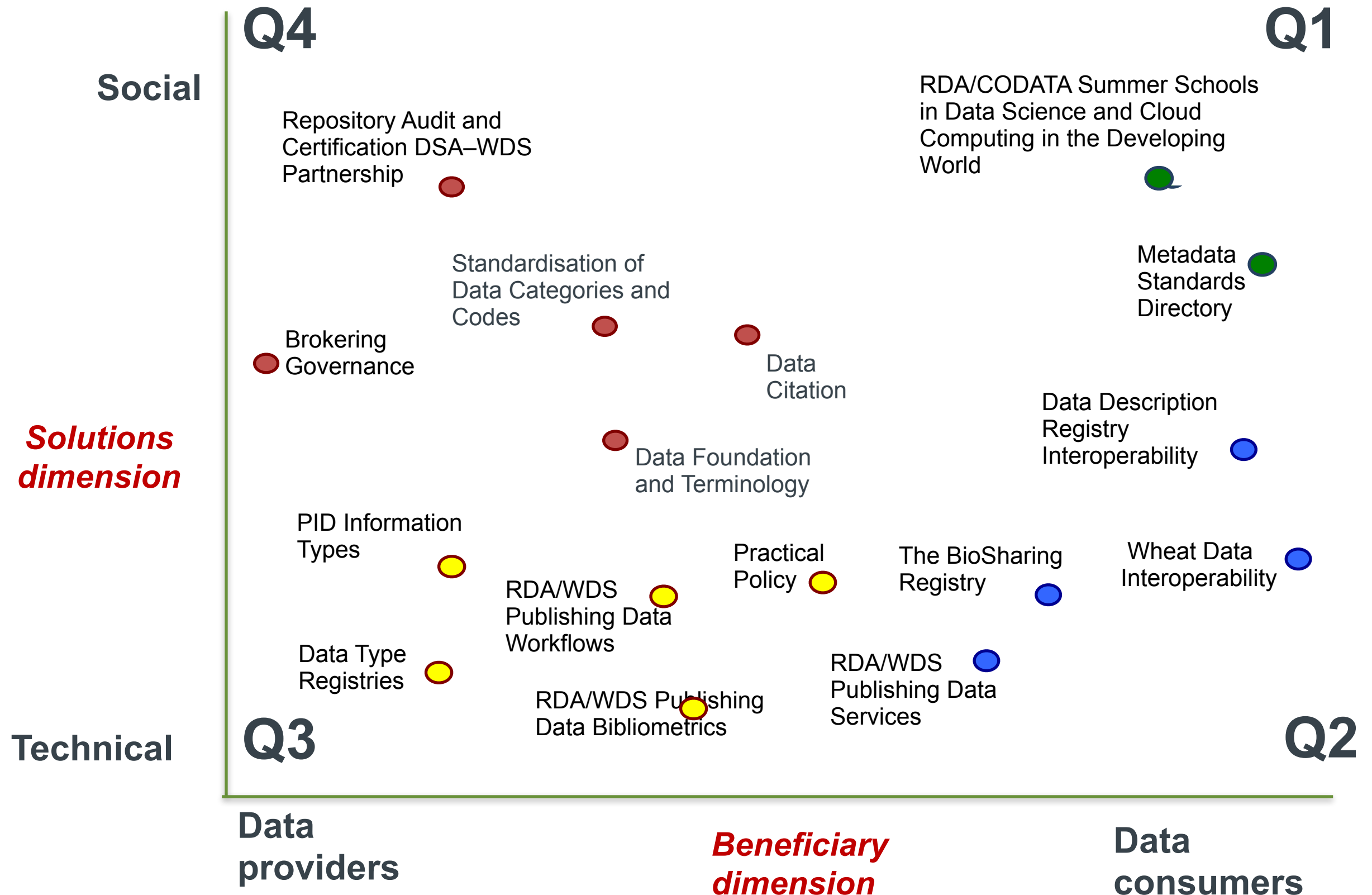
(courtesy Bruce Caron)



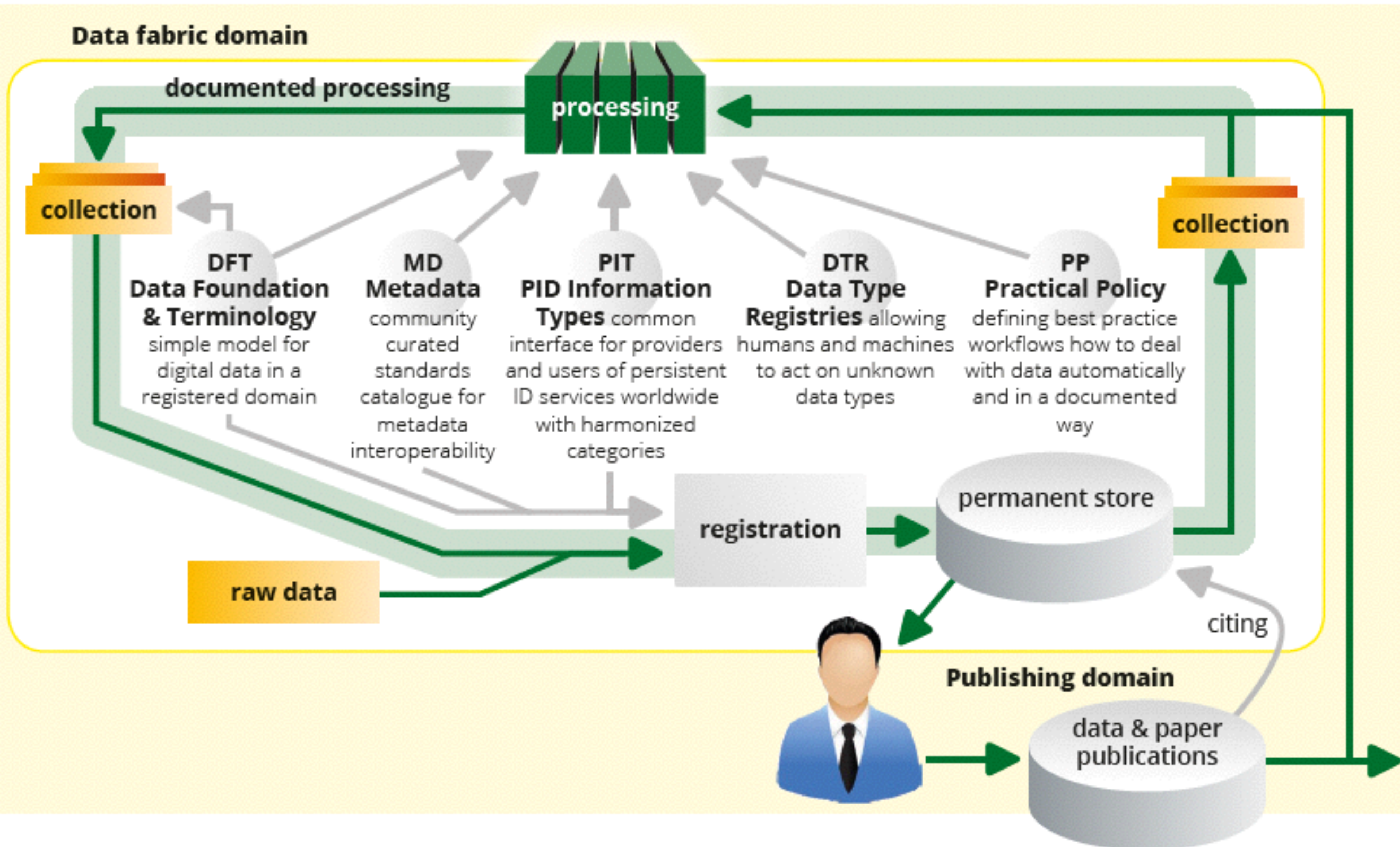
We value:

- **Individuals and interactions** over processes and tools
- **Working volunteers** over comprehensive documentation
- **Member collaboration** over contract negotiation
- **Responding to change** over following a plan.

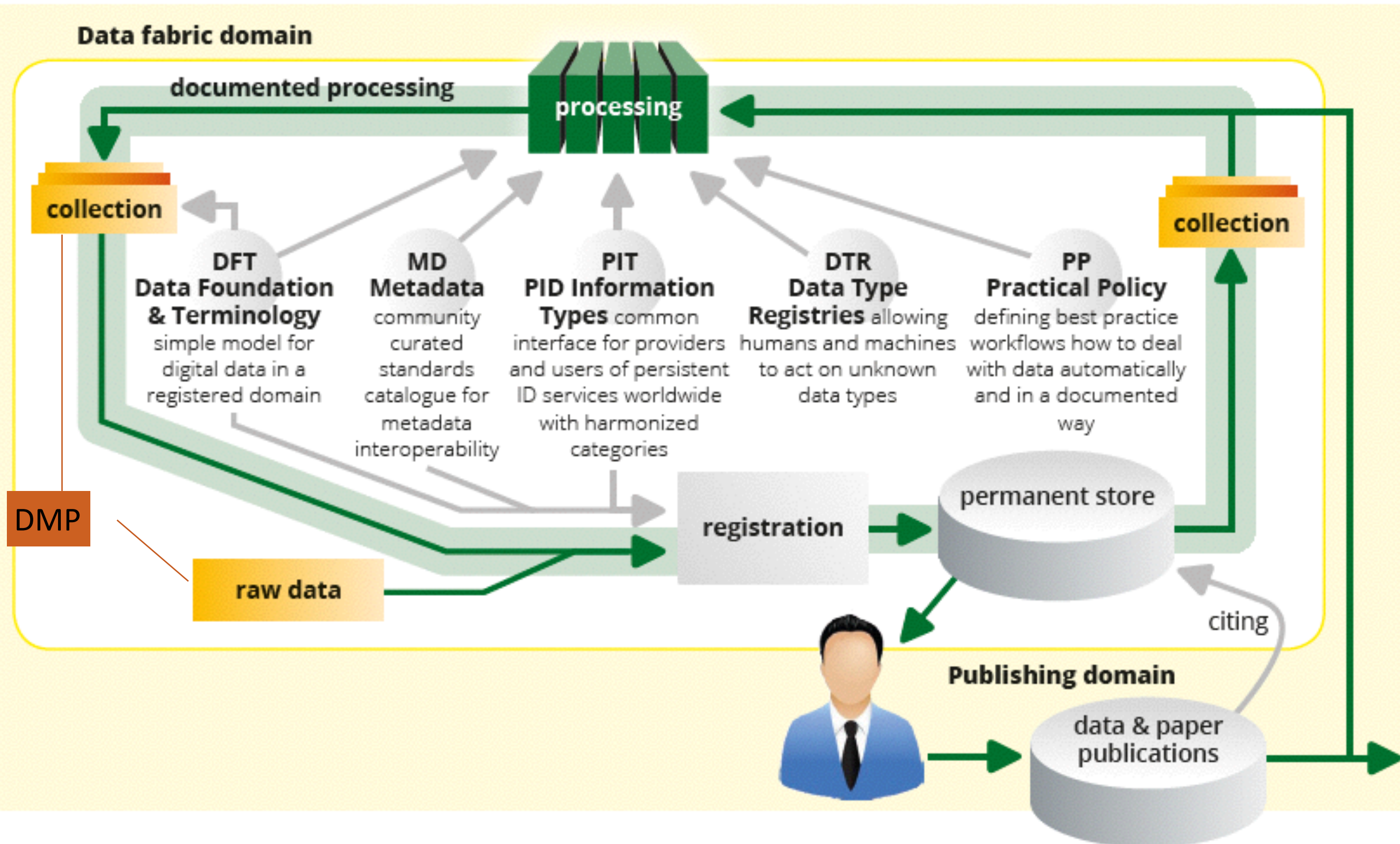
Working Groups Clusters



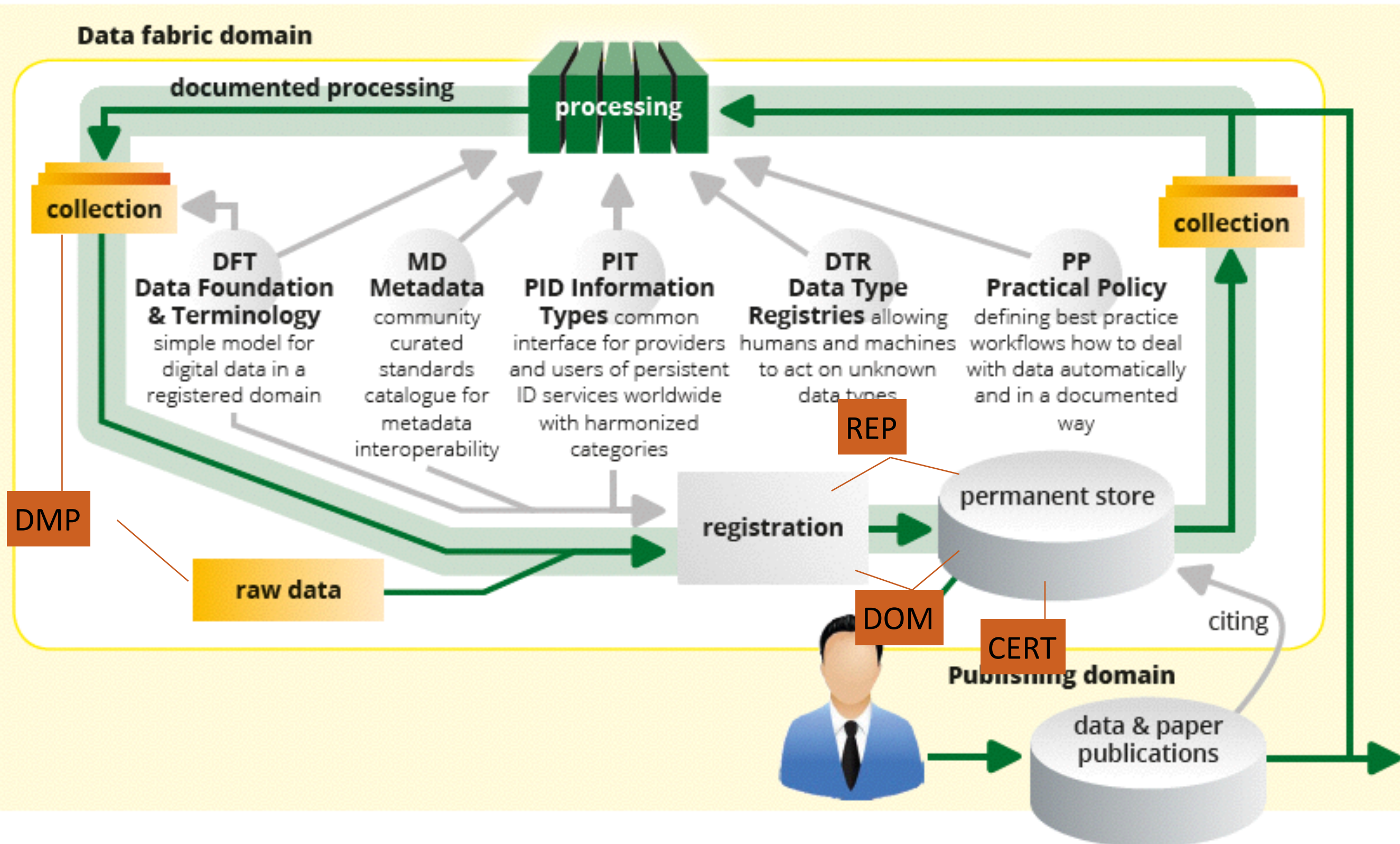
The “Data Fabric”



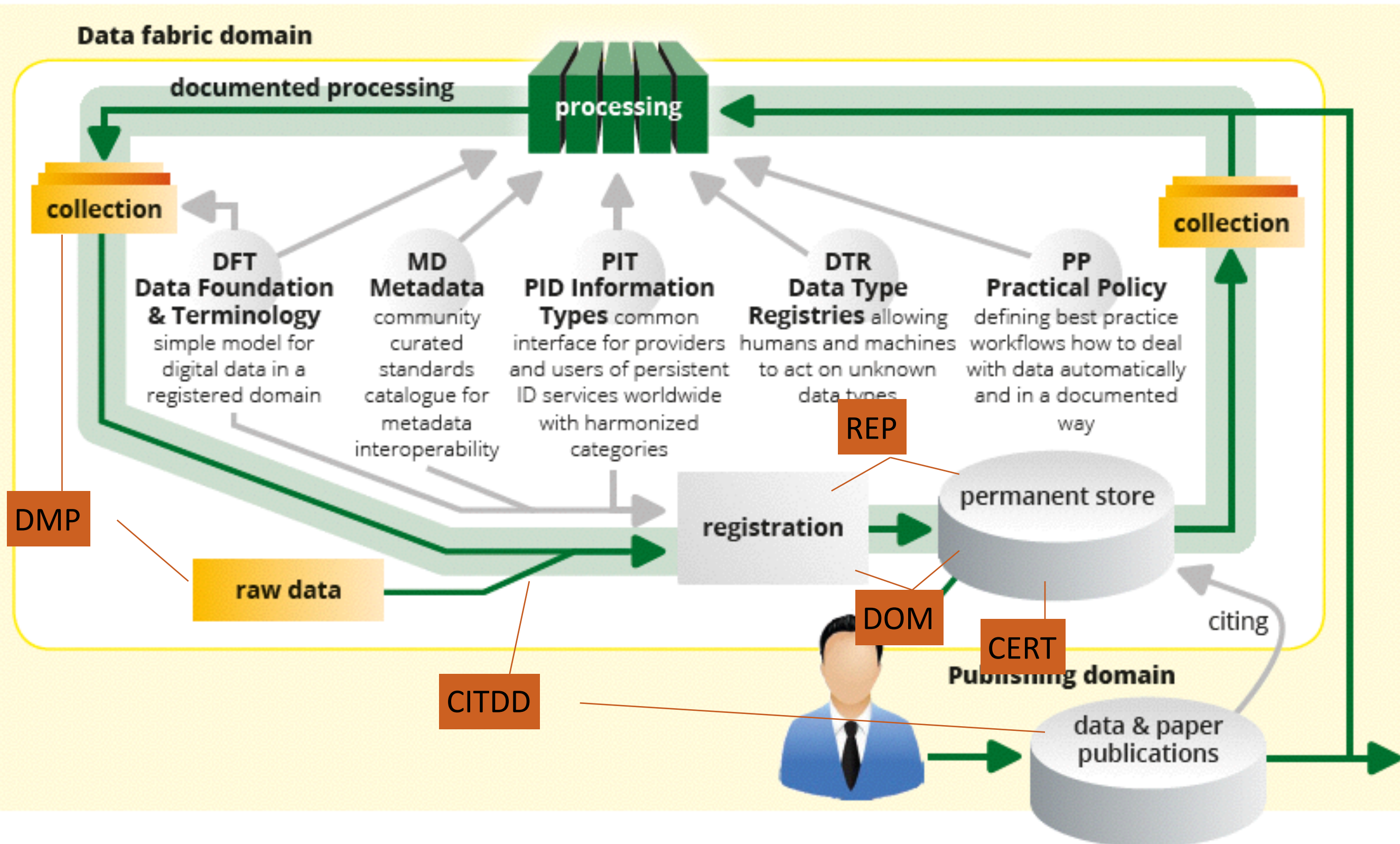
The “Data Fabric”



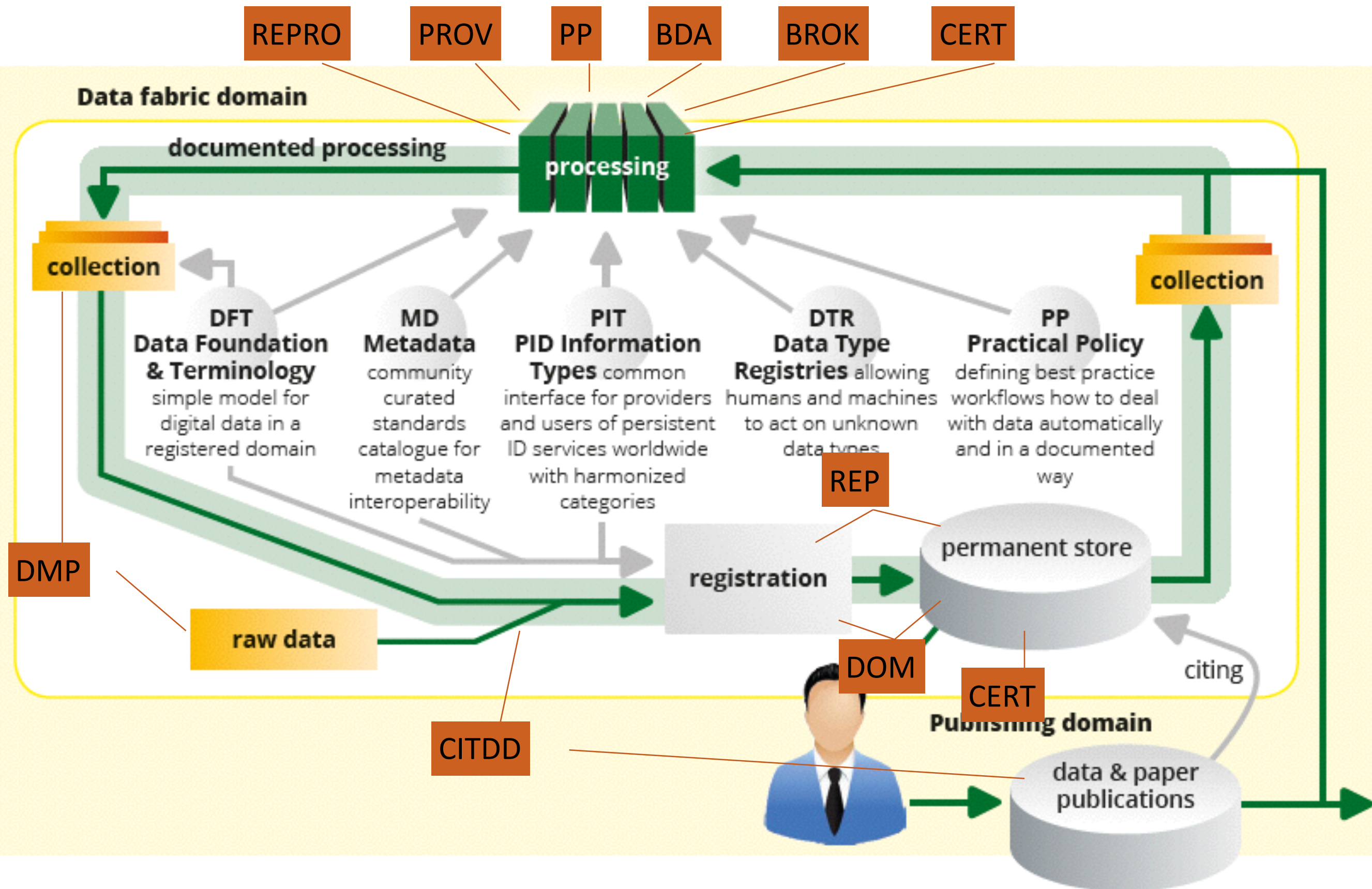
The “Data Fabric”



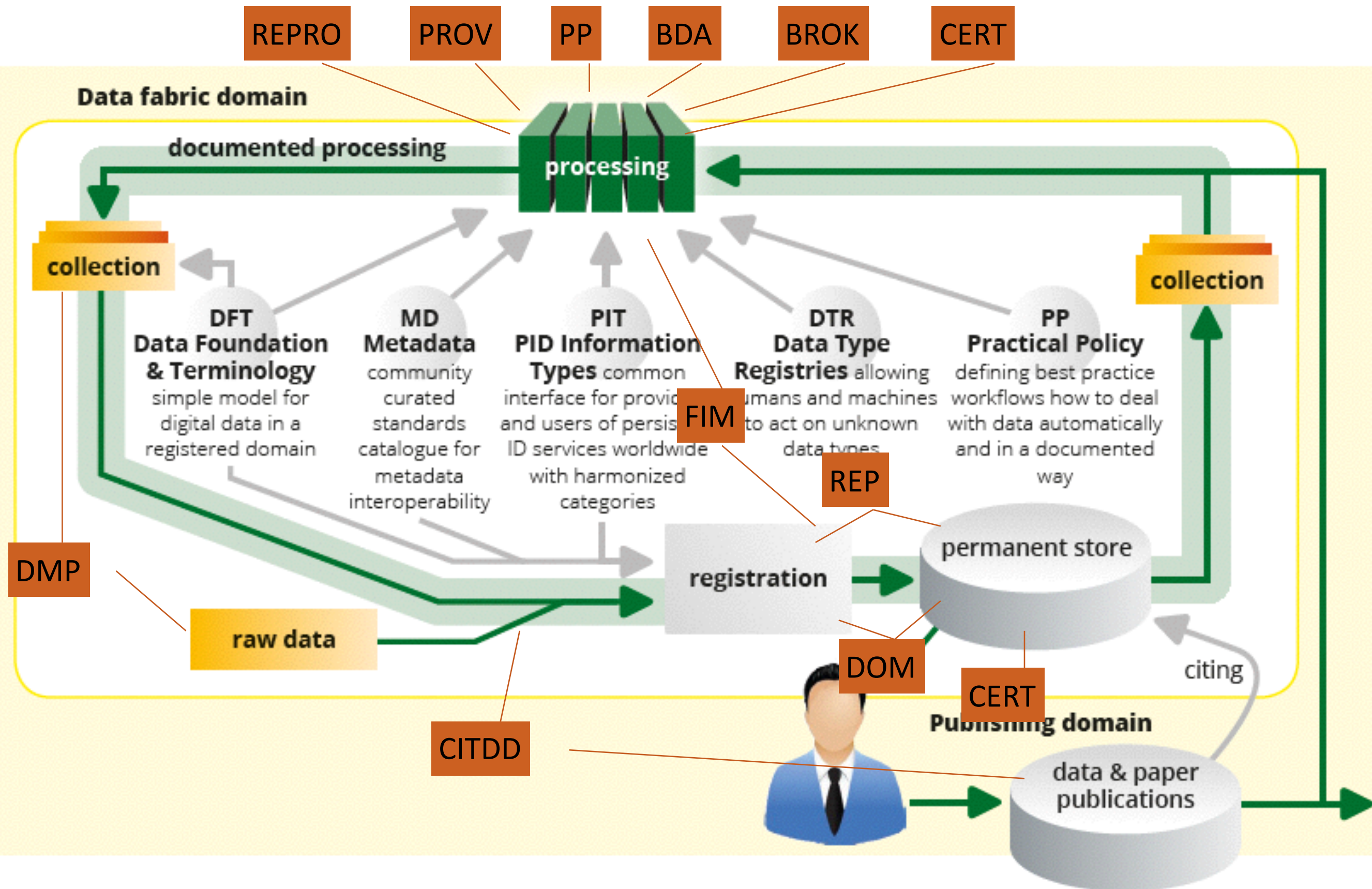
The “Data Fabric”



The “Data Fabric”



The “Data Fabric”





1st: Save the data. This is hard.



2nd: Share the data. This is harder.

Some themes amidst the difference

1. **Persistent Identifiers** for data, documents, people, organisations, instruments—Everything!
2. **Certifying Trust** in assertions, evidence, organisations, processes...
3. The value of **Conversations, Relationships, and Mediation**— an agile network effect.

Some themes amidst the difference

1. **Persistent Identifiers** for data, documents, people, organisations, instruments — Everything!
2. **Certifying Trust** in assertions, evidence, organisations, processes
3. The value of **Conversations, Relationships, and Mediation** — an agile network effect.

Trust

Some amateur thoughts on trust and sharing and infrastructure



- When or do we need to certify trust? Do we?
- We must preserve the freedom to tinker.
- Build in decentralization where possible. Any centralization must be community governed.
- Trust is built through
 - shared experience— e.g., RDA Plenaries
 - shared perspectives — RDA is a forum for engagement and constructive disagreement
 - actual reuse and adoption — in RDA consensus is defined through use.
 - sustained performance — RDA seeks to build a broad coalition of international support

Regional RDAs



- Australian National Data Service, RDA/United States, RDA/Europe,
- Implement RDA deliverables locally and enhance adoption.
- Ensure regional or national issues are addressed globally.
- Support plenaries and support attendance at plenaries.



Initial Impact



- Data are having their day! RDA is both cause and effect.
- Collaborative value
 - Accelerating harmonization—the citation story.
 - Discovering shared themes—PIDs, the “platform” story
 - New insight from rethinking old paradigms—roadmaps, architectures, and lifecycles are passé. Reuse, agility and bridging are hip.
- Real deliverables in 2 years.
 - Learning fast through openness and 18 months.
 - demonstration of delivery
 - interconnection between components already occurring
- Money
 - funding success is advancing the field
 - measurable return on investment

Preservation of LHC data

100PB growing to ~5EB for decades

| | |
|--------------------|---|
| <i>Veni</i> | ... to RDA plenaries, WG & IG meetings |
| <i>Vidi</i> | <i>... how other disciplines attacked the problem</i> |
| <i>Vici</i> | ... developed, refined and now implementing a strategy, including cost model and business case → A better solution, more sustainable and advanced by years |

Next Steps for RDA: Stay Pragmatic, Focus on Impact

More Infrastructure

Continuing pipeline of infrastructure deliverables adopted, used, coordinated and amplified to accelerate data sharing

More effective Community

Increasing coordination and collaboration between domains, sectors, organizations, communities. Effective advocacy for national and international data issues and communities.

Impact-focused Outreach

Stronger partnerships with industry, governments, domains, organizations.

Substantive engagement of students and early career professionals, greater spectrum of international cultures.

- **Next Plenaries** (Plenaries are both community and working meetings. Meetings held twice yearly around the world.):
 - March, 2016: **Tokyo, Japan (P7)**
 - September, 2016: ~ **Washington, DC (P8)**
 - March, 2017: **Barcelona, Spain (P9)**

Joining RDA:

Go to rd-alliance.org and register

- Must agree to RDA principles (openness, community-driven, etc.)
- Free for individuals

Plenary 7

1-3 March 2016

Tokyo, Japan





Info:

enquiries@rd-alliance.org

@resdatall

research data sharing without barriers
rd-alliance.org

RDA Membership

Working Groups

implementable, impactful outcomes

Interest Groups

domain coordination, idea generation, maintenance, ...

Technical
Advisory Board
*socio-technical vision
and strategy*

Secretariat
*administration and
operations*

Organisational
Advisory Board
*needs, adoption,
business advice*

Council

organisational vision and strategy

RDA
Foundation

Funders Forum