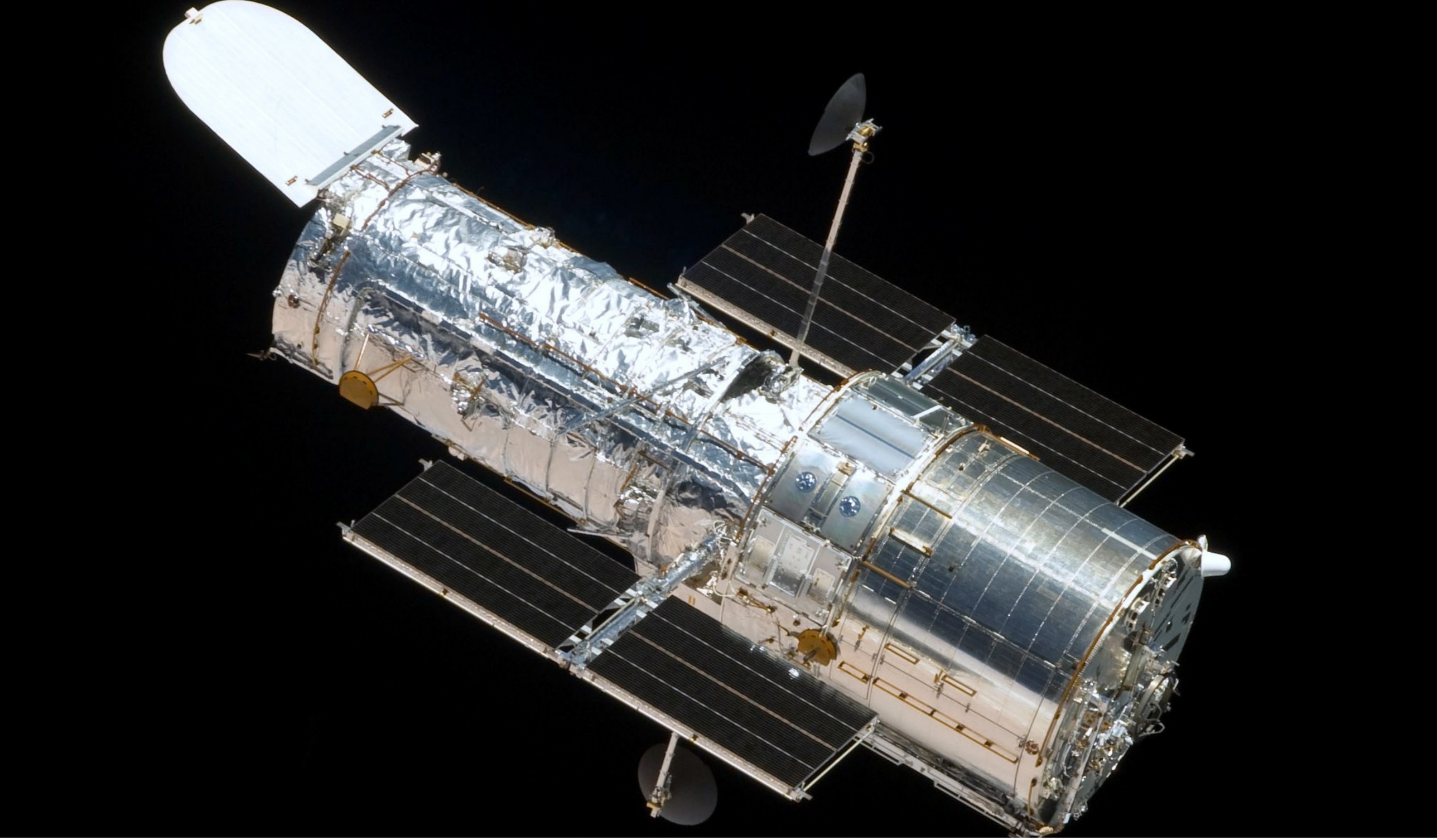
Some reflections on (e-)Infrastructure and RDA

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

e-Infrastructures Reflection Group
Bratislava, Slovakia
15 November 2016

research infrastructure vs. e-infrastructure



"Research infrastructure is stuff like telescopes"

"HST-SM4" by Ruffnax (Crew of STS-125)

Overwhelming data volume

Artist rendition courtesy SKA Organisation

research infrastructure vs. e-infrastructure

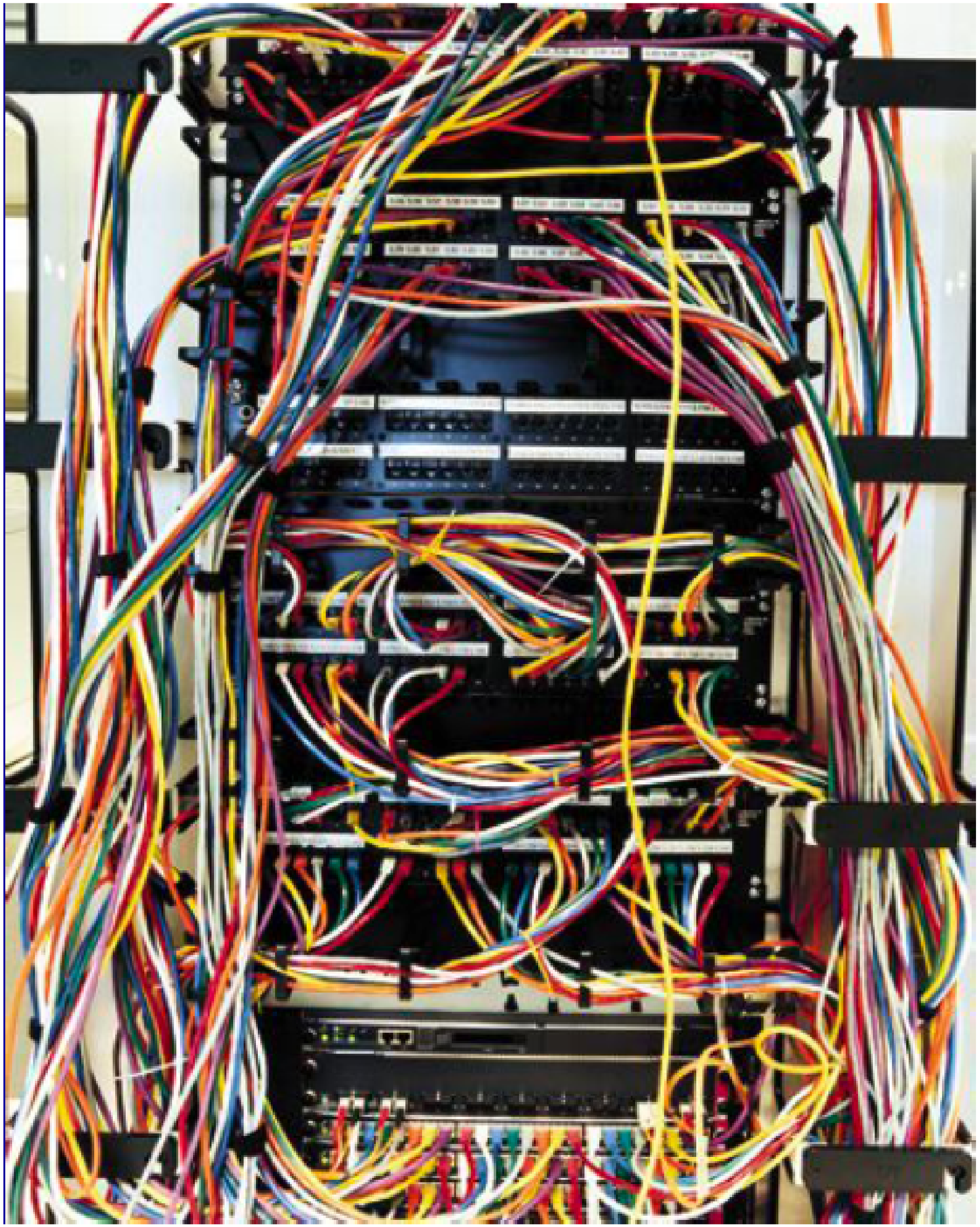
a false dichotomy

e-Infrastructure *is* research infrastructure.

Modern research infrastructure *is* (or at least requires) e-Infrastructure.

It's about the *data*

Infrastructure is hard to conceive and describe because when it works, it’s transparent, ubiquitous, and embedded in our daily work.



Dynamics of Infrastructure

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

RESEARCH DATA ALLIANCE

- 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

(that helps data flow and be useful)

THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

building the social and technical bridges that enable open sharing of data

15 FLAGSHIP OUTPUTS

of which 4 ICT Technical Specifications

75 ADOPTION CASES

across multiple disciplines, organisations & countries

72 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

of which 27 WORKING GROUPS & 45 INTEREST GROUPS

4,527 INDIVIDUAL MEMBERS FROM 115 COUNTRIES

66% Academia & Research
16% Public Administration
11% Enterprise & Industry

46 ORGANISATIONAL MEMBERS

RDA
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.

“Create - Adopt - Use”
(in 12-18 months)

Adopted Policy

Systems Interoperability

Common Types, Standards, Metadata

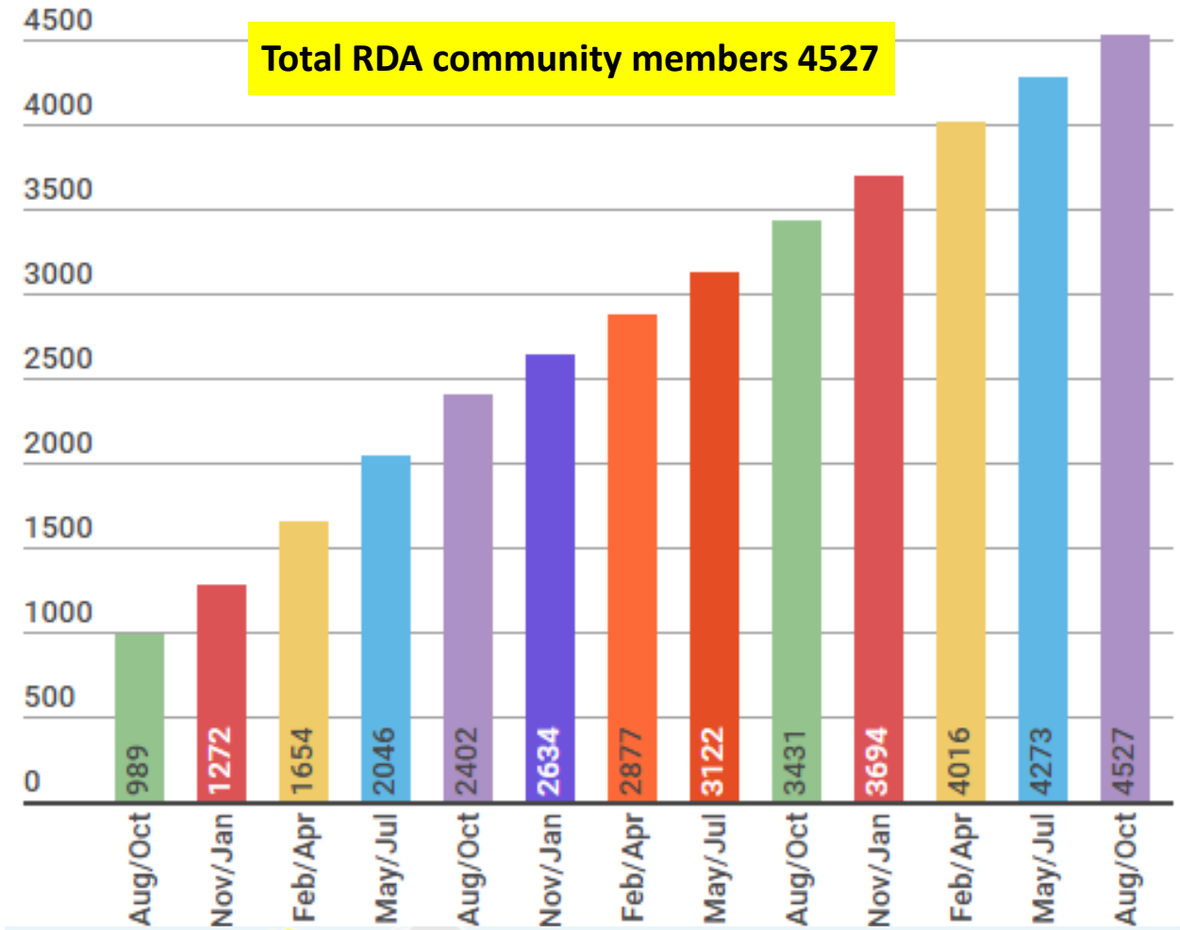
Sustainable Economics

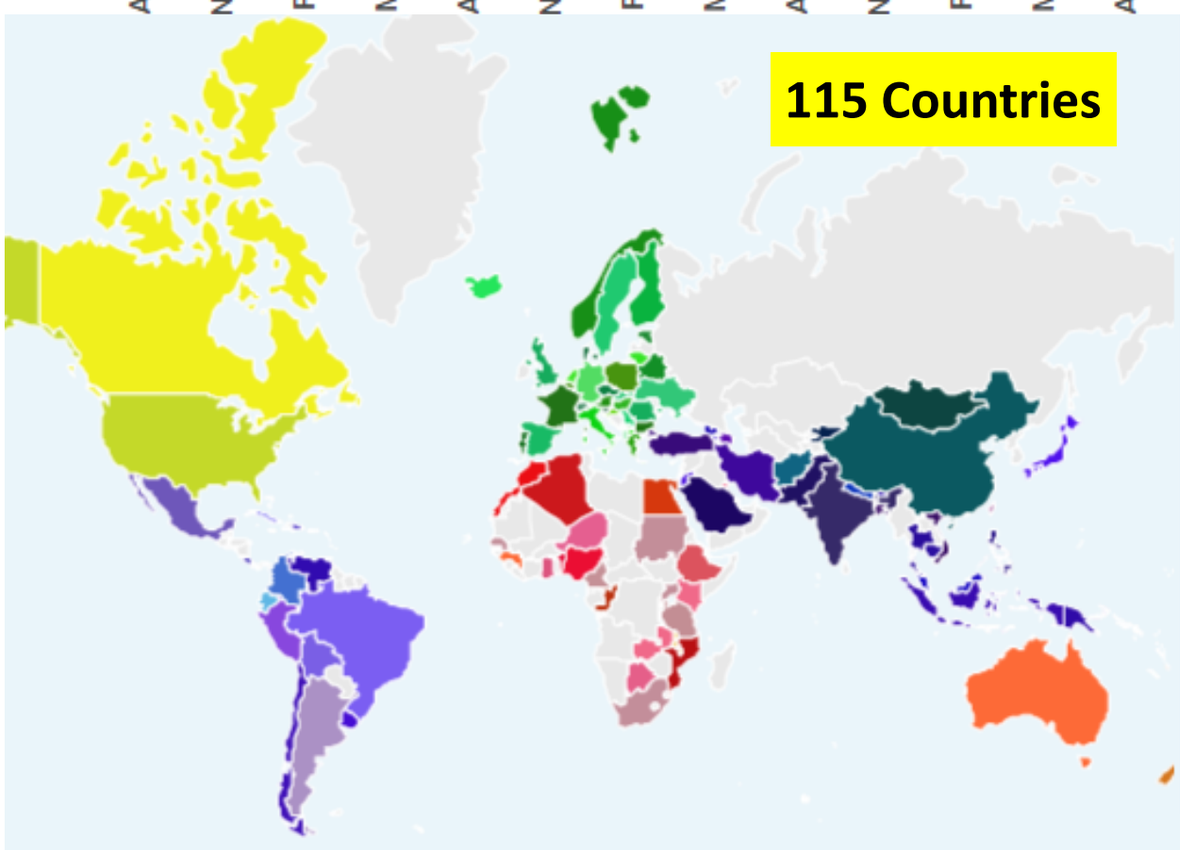
Adopted Community Practice

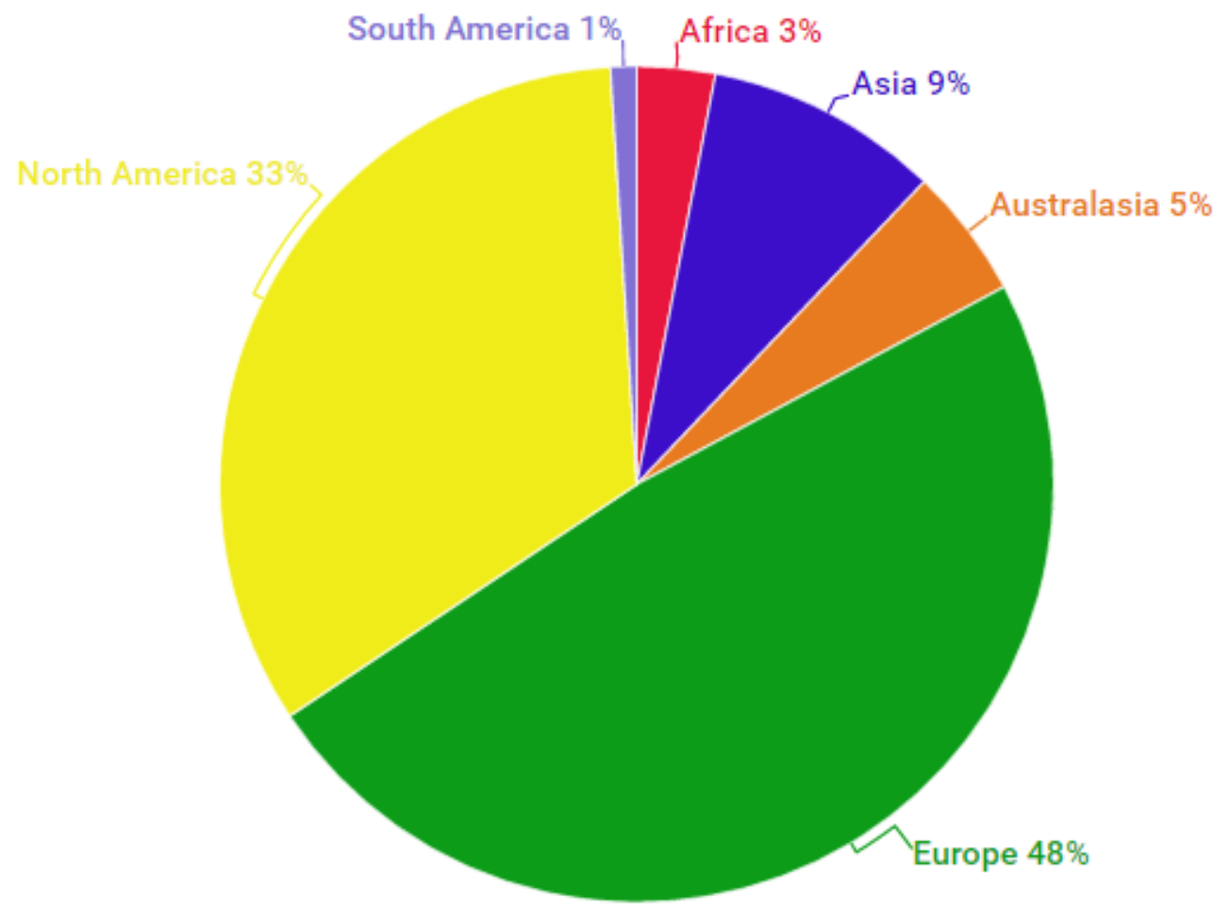
Training, Education, Workforce

Traffic Image: Mike Gonzalez

RDA worldwide growth







RDA
RESEARCH DATA ALLIANCE

Organisational & Affiliate Members

46 RDA Organisational Members

6 RDA Affiliate Members

The Association of Commonwealth Universities
ands
BLACKFYNN
BSC
Caltech
CAICYT CONICET
canarie
cap-digital
ASU CENTER FOR BIOLOGY + SOCIETY ARIZONA STATE UNIVERSITY
COLUMBIA UNIVERSITY
CSC
HELMHOLTZ ASSOCIATION
Corporation for National Research Initiatives®
DANS Data Archiving and Networked Services
DataONE Data Observation Network for Earth
DCC
egi
ELSEVIER
EUDAT
EIROforum
eurisy ACTING COLLECTIVELY TO BRIDGE SPACE AND SOCIETY
RatSWD German Data Forum
Griffith UNIVERSITY
icm interdyscyplinarne centrum modelowania matematycznego i komputerowego Uniwersytet Warszawski
National Institute of Advanced Industrial Science and Technology AIST
stm
Pitt
INTERNET2
LIBER
casrai Connecting Research
CODATA
DataCite International Data Citation
MPCDF MAX PLANCK COMPUTING & DATA FACILITY
NCSA
netherlands eScience center
NeSI New Zealand eScience Infrastructure
PURDUE UNIVERSITY LIBRARIES
GODAN Global Open Data for Agriculture & Nutrition
QDR QUALITATIVE DATA REPOSITORY
QIVIVO Fournisseur d'économies d'énergie
RDC DRC Research Data Canada – Données de Recherche Canada
RDS www.rds.edu.au
SPARC
ORCID
ICSU WORLD DATA SYSTEM
Science & Technology Facilities Council
TERN Terrestrial Ecosystem Research Network
WASHINGTON UNIVERSITY IN ST. LOUIS 1853
webscience TRUST
WILEY

rd-alliance.org/get-involved/organisational-membership

WWW.RD-ALLIANCE.ORG
@RESDATALL

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

Common Types, Standards, Metadata

Systems Interoperability

Policy and Practice

Sustainable Economics

Training, Education, Workforce

Fran Berman, Research Data Alliance

“e-Infrastructures are in a peculiar situation which is characterised by both a constant intention of collaborating and a position of debating with their users, the research communities and the domain specific (discipline oriented) research infrastructures. It is a huge challenge for an e-infrastructure to fulfill the **diverse and sometimes conflicting requirements** of many research environments.”

—e-IRG Roadmap 2016

Friction is inevitable and necessary in collaboration

"A wheel turns because of its encounter with the surface of the road; spinning in the air it goes nowhere."

cover notes for
Friction—An Ethnography of Global Connection
by Anna Lowenhaupt Tsing

RESEARCH DATA ALLIANCE

Marcel DuChamp's *Bicycle Wheel*
photo © nuzz—www.flickr.com/photos/nuzz/

RESEARCH DATA ALLIANCE

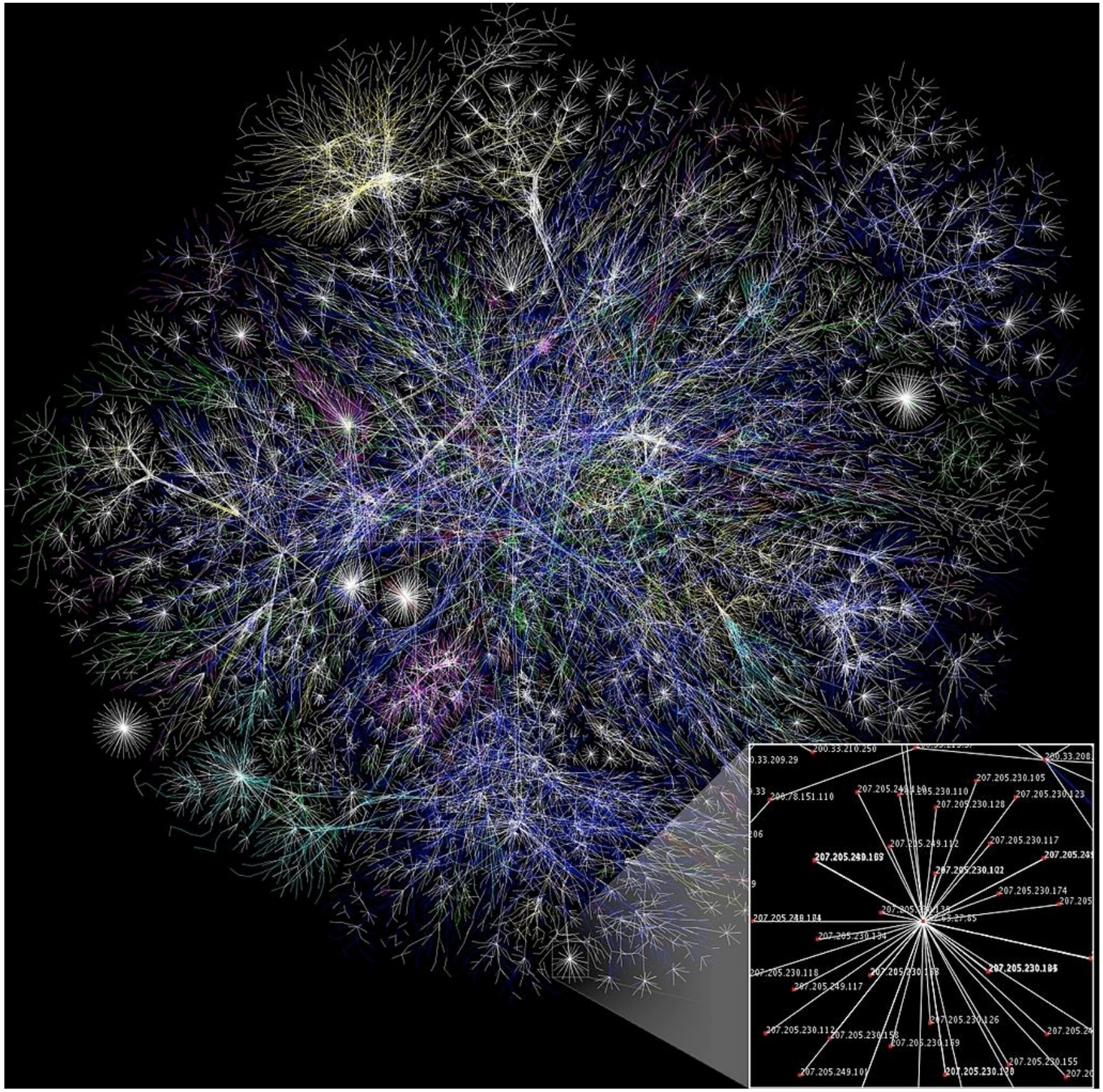
Coalition Politics

"There is no reason to think collaborators have common goals."

- A. Tsing, *Friction*

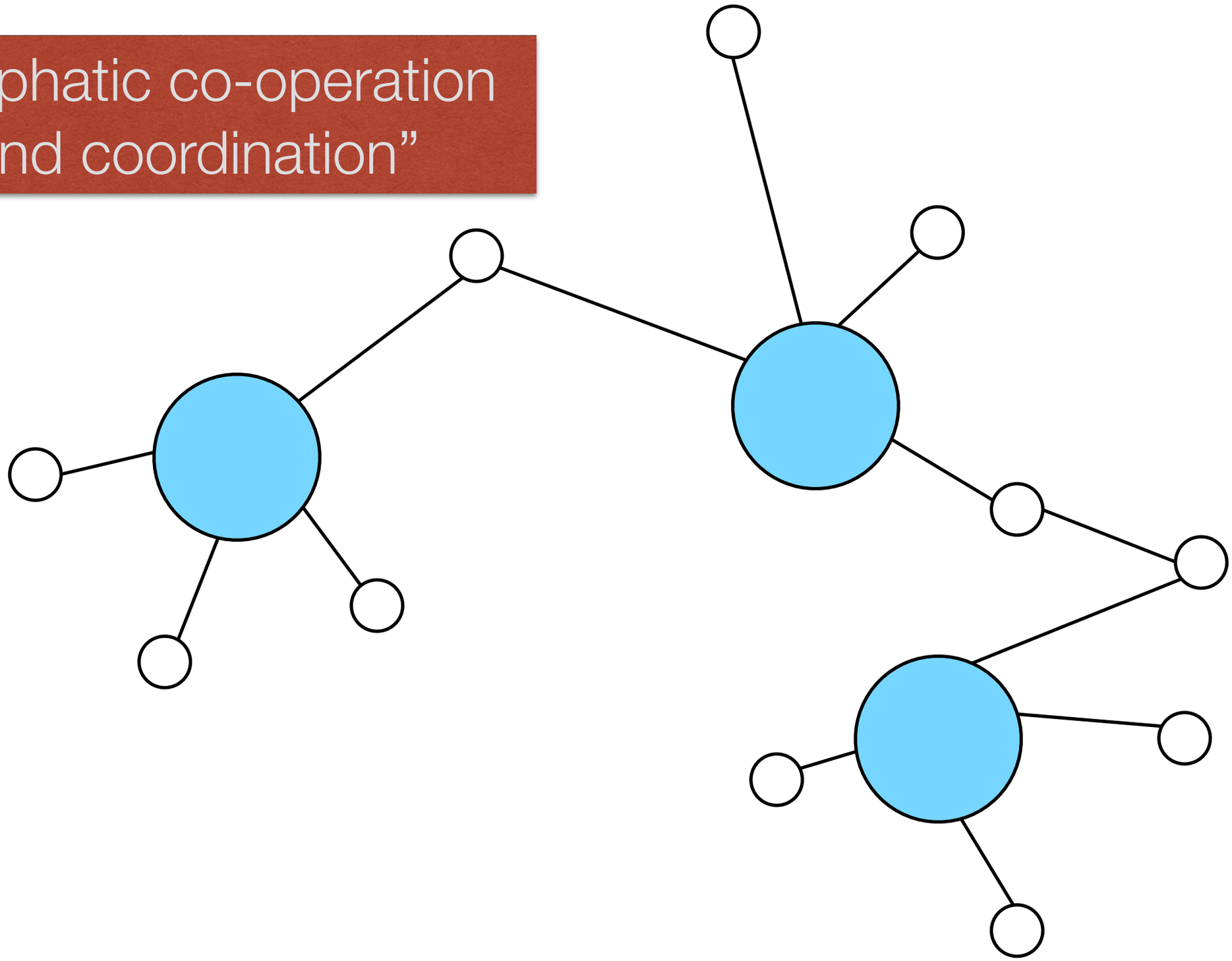
Ashby's Law of Requisite Variety

Only variety absorbs variety



Map of the internet by the Opte Project [CC-BY] via Wikimedia Commons

“emphatic co-operation and coordination”



Networks or ecosystems often rely on “weak” links, so partner and build relationships. (See Barabási A-L and R Albert. 1999 and others)

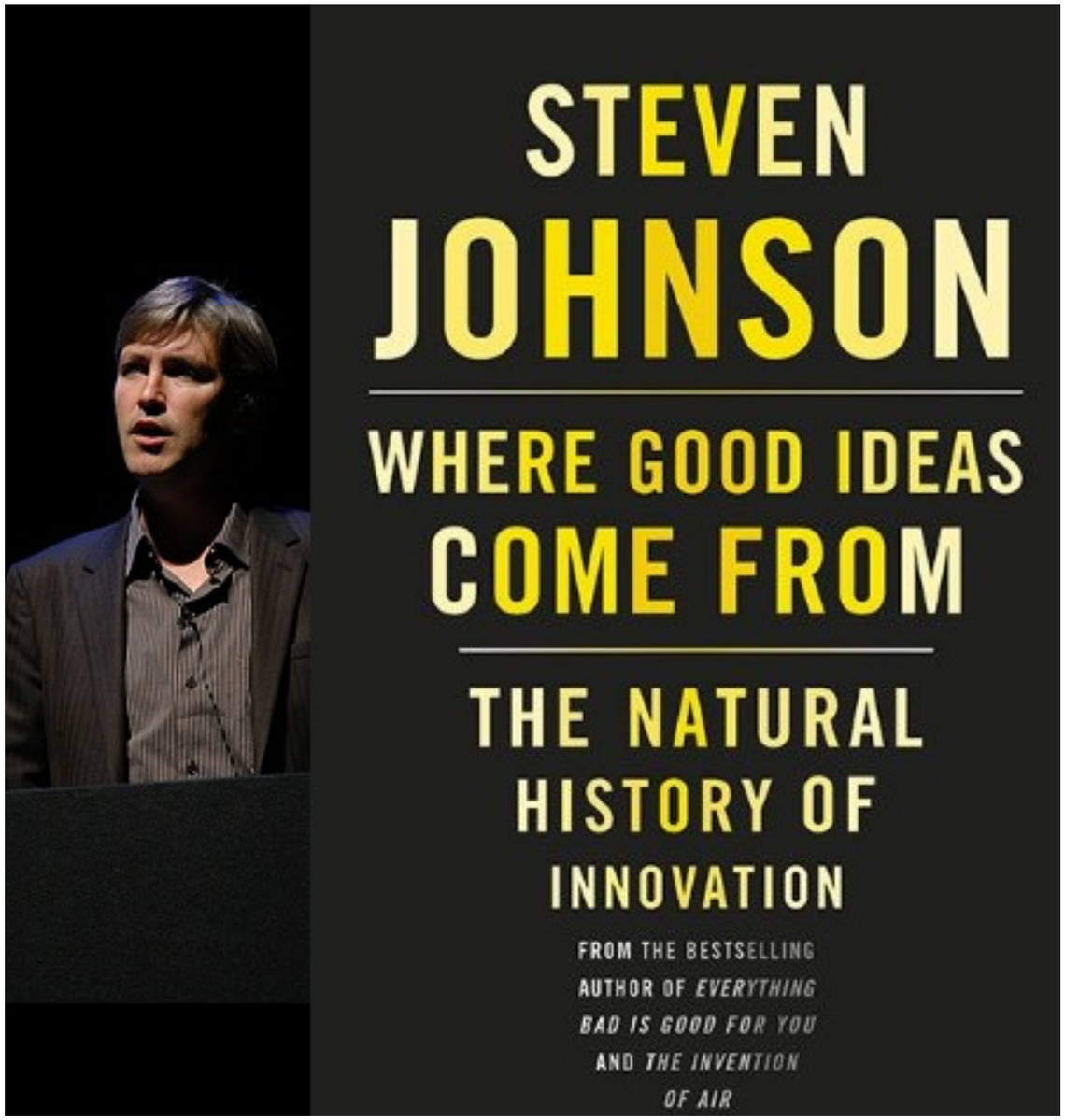
One stop shop?

“without promoting monopolies or implying the need of creating of a single integrated provider”

Or Grand Bazaar!

“Where Good Ideas Come From”

- The Adjacent Possible—the importance of local
  - Often not “Eureka!” but rather a slow hunch fading in to view over time.
  - Hunches need to collide with other hunches so create that environment. Don’t protect IP share it. Connecting vs. protecting
  - Sharing of failures as well.
- Create spaces for that to happen—virtual and real coffee shops
- “Chance favors the connected mind.”



Solving the problem must include adopters in the process.

Image courtesy bigthink.com

Open problem solving is key.

Figure courtesy webbirdmedia.com

No defined architecture.

RDA Principles

Openness
Consensus
Balance
Harmonization
Community Driven
Non-profit

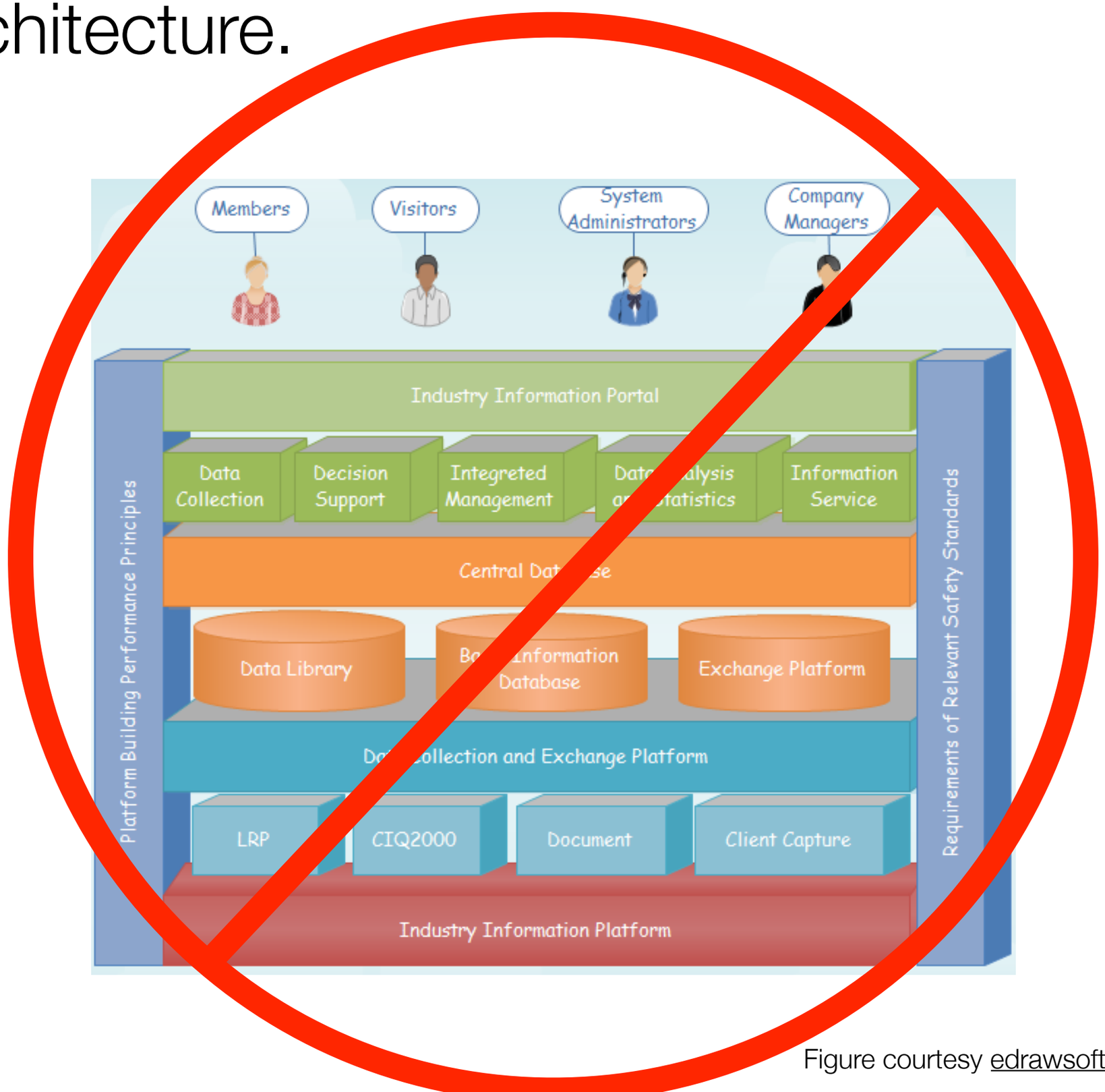


Figure courtesy edrawsoft.com

RDA Recommendations & Outputs

Data Foundation & Terminology: a model for data in the registered domain.

PID Information Types: a common protocol for providers and users of persistent ID services worldwide.

Data Type Registries: allowing humans and machines to act on unknown, but registered, data types.

Practical Policy: defining best practices of how to deal with data automatically and in a documented way with computer actionable policy.

Metadata standards directory: Community curated standards catalogue for metadata interoperability

RDA Recommendations & Outputs

Data Citation: defining mechanisms to reliably cite dynamic data

Data Description Registry Interoperability solutions enabling cross platform discovery based on existing open protocols and standards

Wheat Data Interoperability impacting the discoverability, reusability and interoperability of wheat data by building a common framework for describing, representing linking and publishing wheat data

Brokering Governance WG: Sustainable Business Models for Brokering Middleware to support Research Interoperability

RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG: A framework to run a series of Summer Schools in Data Science and data sharing in low and middle income countries (LMICs)

RDA Recommendations & Outputs

Repository Audit and Certification DSA–WDS: A convergent DSA-WDS certification standard to help eliminate duplication of effort, increase certification procedure coherence and compatibility thus benefitting researchers, data managers, librarians and scientific communities.

RDA/WDS Publishing Data Bibliometrics: improved research data metrics and corresponding services, with the final goal of increasing the overall availability and quality of citations and research data itself.

RDA/WDS Publishing Data Services: A universal interlinking service between data and the scientific literature.

RDA/WDS Publishing Data Workflows: enhance the possibilities for greater discoverability and a more efficient and reliable reuse of research data benefitting other stakeholders like publishers, libraries and data centres.

RDA Recommendations & Outputs

23 Things: Libraries For Research Data An overview of practical, free, online resources and tools that users can immediately take advantage of to incorporate research data management into the practice of librarianship.

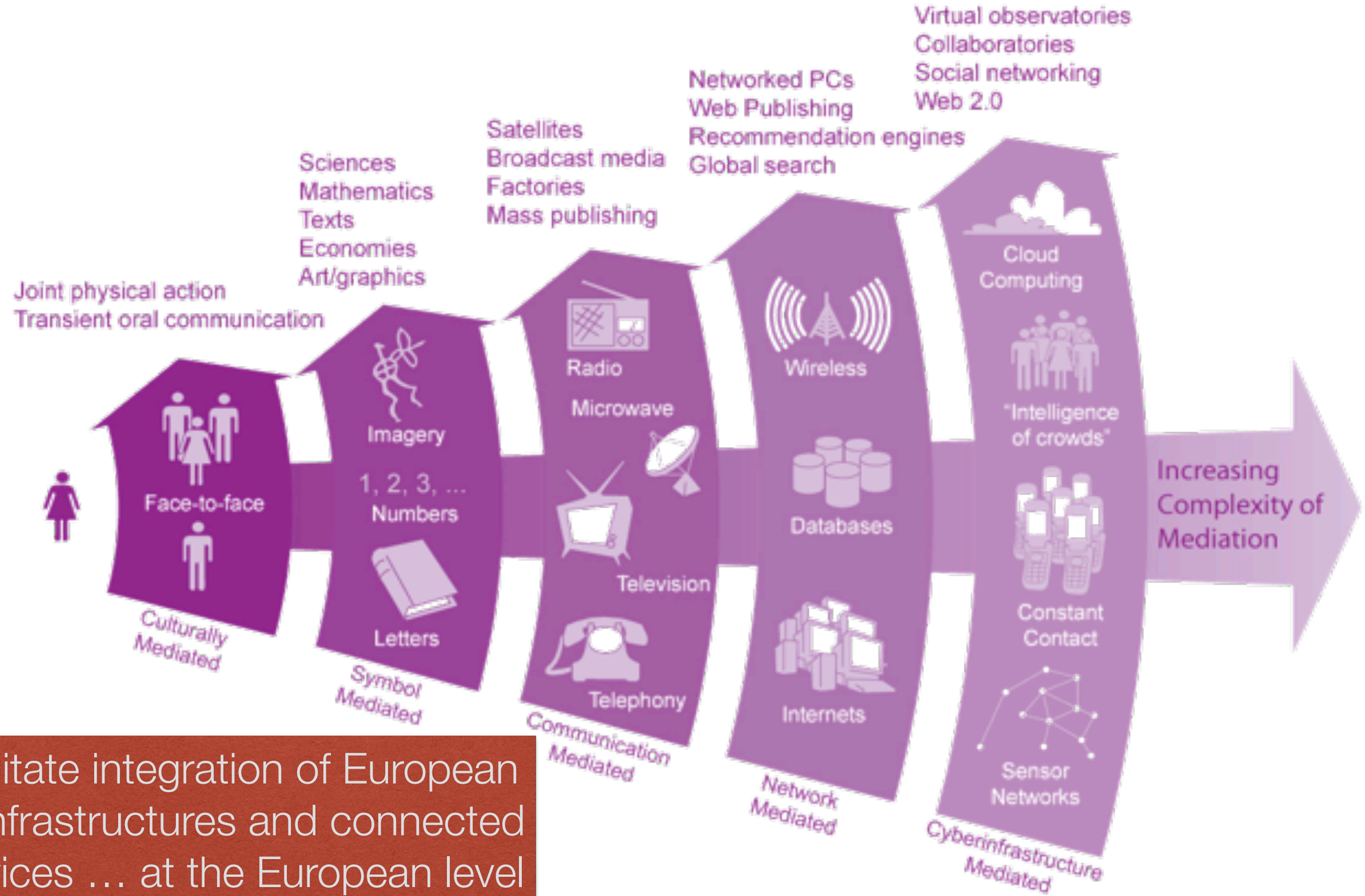
Legal Interoperability of Research Data Principles and Implementation Guidelines: a set of principles and practical implementation guidelines offered as high-level guidance to all members of the research community — the funders, managers of data centers, librarians, archivists, publishers, policymakers, university administrators, individual researchers, and their legal counsel.

The Scholix initiative a high level interoperability framework for exchanging information about the links between scholarly literature and data. It aims to build an open information ecosystem to understand systematically what data underpins literature and what literature references data.

THE RESEARCH DATA ALLIANCE
RECOMMENDATIONS & OUTPUTS

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.



facilitate integration of European e- Infrastructures and connected services … at the European level and globally.

Increasing Complexity of Mediation

From: C. Borgman, 2008, NSF Cyberlearning Report

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 reflections on trust and sharing and infrastructure

RESEARCH DATA ALLIANCE

- When or do we need to certify trust? Do we?
- We must preserve the freedom to tinker. That is challenging in large-scale data facilties/RIs
- Build in decentralization where possible. Any centralization must be community governed. A role for e-IRG?
- 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 — an area where e-IRG, ESFRI, EOSC, RDA, etc. must find common purpose

RESEARCH DATA ALLIANCE

RDA 9th Plenary Meeting

Data Infrastructures for Open Science

5-7 April 2017, Barcelo Sants Hotel, Barcelona, Spain

Organised by Barcelona Supercomputing Center (BSC) with the support of RDA Europe

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

RESEARCH DATA ALLIANCE EUROPE

Host a **collocated or associated** event on April 3rd, 4th or 7th (PM) – Applications open until **27 Nov 2016 17:00 CET** **https://www.rd-alliance.org/plenaries/ninth-plenary/rda-9th-plenary-collocated-associated-events**

https://www.rd-alliance.org/plenaries/rda-ninth-plenary-meeting-barcelona

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

RESEARCH DATA ALLIANCE

Info:

enquiries@rd-alliance.org

@resdatall

research data sharing without barriers

rd-alliance.org