Infrastructure, relationships, trust, and RDA

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

e-Infrastructures & RDA for data intensive science
Paris, France
22 September 2015
research infrastructure vs. e-infrastructure
“Research infrastructure is stuff like telescopes”
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.
Conceptions of Infrastructure

- Infrastructure — French 1875; English 1927 (originally a military context)
- Spatial data infrastructure — c. 1990 (or earlier)
- Information infrastructure — c. 1993
- Cyberinfrastructure — USA c. 2003
- e-Infrastructure — UK c. 2006 (really took off c. 2009-10 with eIRG and ESFRI)
- Hybrid data infrastructure and infrastructure-as-a-service — c. 2011
- Data infrastructure — ??
Infrastructure is hard to conceive and describe because when it works, it’s transparent, ubiquitous, and embedded in our daily work.
• 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)
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

Traffic Image: Mike Gonzalez

Training, Education, Workforce

Fran Berman, Research Data Alliance
Shared Principles

- Openness
- Consensus
- Balance
- Harmonization
- Community Driven
- Non-profit
The Research Data Alliance Community Today

Total RDA Community Members: 3243 from 103 countries

56 Working and Interest Groups
RDA Organisational Members and Affiliates
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-facetted 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
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.
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
See you tomorrow!

CNAM, Paris, France
23 - 25 September 2015
Info:

enquiries@rd-alliance.org

@resdatall