Open Data is Not Enough
Making Data Sharing Work

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

Trusted Data Services to Support Climate Change Research
Paris, France
6 July 2015
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

- 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)

NO SMOKING

Adopted Policy

Sustainable Economics

Systems Interoperability

Common Types, Standards, Metadata

Adopted Community Practice

Training, Education, Workforce

Traffic Image: Mike Gonzalez

Fran Berman, Research Data Alliance
Shared Principles

• Openness
• Consensus
• Balance
• Harmonization
• Community Driven
• Non-profit
Implementing the Components of a “Data Engine”

- **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
The Research Data Alliance Community Today

from 102 countries


Academia: 1942
Government/Public service: 441
Other: 169
SMEs: 160
IT Consultancy/Development: 97
Large Enterprise: 64
Policy/Funding Agency: 44
Press & Media: 19

North America: 38%
Europe: 48%
Australasia: 4%
Asia: 6%
Africa: 3%
South America: 1%
RDA Organisational Members and Affiliates

[Logos of various organisations]
Initial Impact

• Data are having their day! RDA is both cause and effect.

• Collaborative value
  • Accelerating harmonization—the citation story.
  • Discovering shared themes—PIDs to “platform”
  • 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 from outputs book (see webpage)

• Money
  • funding success is advancing the field
  • measurable return on investment

• More stories of data success at http://datastories.jiscinvolve.org/wp/
Preservation of LHC data
100PB growing to ~5EB for decades

<table>
<thead>
<tr>
<th>Veni</th>
<th>... to RDA plenaries, WG &amp; IG meetings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vidi</td>
<td>... how other disciplines attacked the problem</td>
</tr>
<tr>
<td>Vici</td>
<td>... developed, refined and now implementing a strategy, including cost model and business case</td>
</tr>
</tbody>
</table>

→ A better solution, more sustainable and advanced by years

slide courtesy Jamie Shiers, CERN
Plenary 6
Enterprise engagement
Special focus on Research Data for Climate Change

Plenary 6 and Data Challenge!
CNAM, Paris, France
23 - 25 September 2015