

Trusted Data + Trusted Cyberinfrastructure + Trusted Connections = Trusted Research

Andrew Treloar

ardc

australian research data commons

Built from   

Outline

- Importance of data
- Role of software and other cyber-infrastructure
- Trusted research needs trusted connections
- Open persistent identifier infrastructure
- Changing research practice

It's all about the data...

Centrality of data

- Research enterprise? Yes
- Research infrastructure? Convince me





StorageTek



Photo CC-BY www.flickr.com/photos/torkildr/3462607995/



But not just the data...

We also need software, because

- Data volumes
- Data complexity
- Growth in Machine Learning

- All require greater use of software

The importance of trust

Why? A local example



Trusted data

- Need trust in both the container and the contents
- Container:
 - Trusted data repositories (CoreTrustSeal has momentum)
- Contents:
 - Reliance on existing discipline practices
 - Plus provenance information, including software processing steps

Trusted software

- Much harder problem
- Formal proofs don't scale (yet)
- Alternatives
 - many eyes?
 - wide adoption?
 - convergence on smaller set of solutions?
- Perhaps settle for reproducibility in the short term

Trusted connections

- Need reliable, future-proof connections between trusted elements
- Infrastructure supporting these elements is brittle
- Identifiers can engineer out some of the brittleness

Importance of open persistent
identifier infrastructure

Need persistent identifiers for...

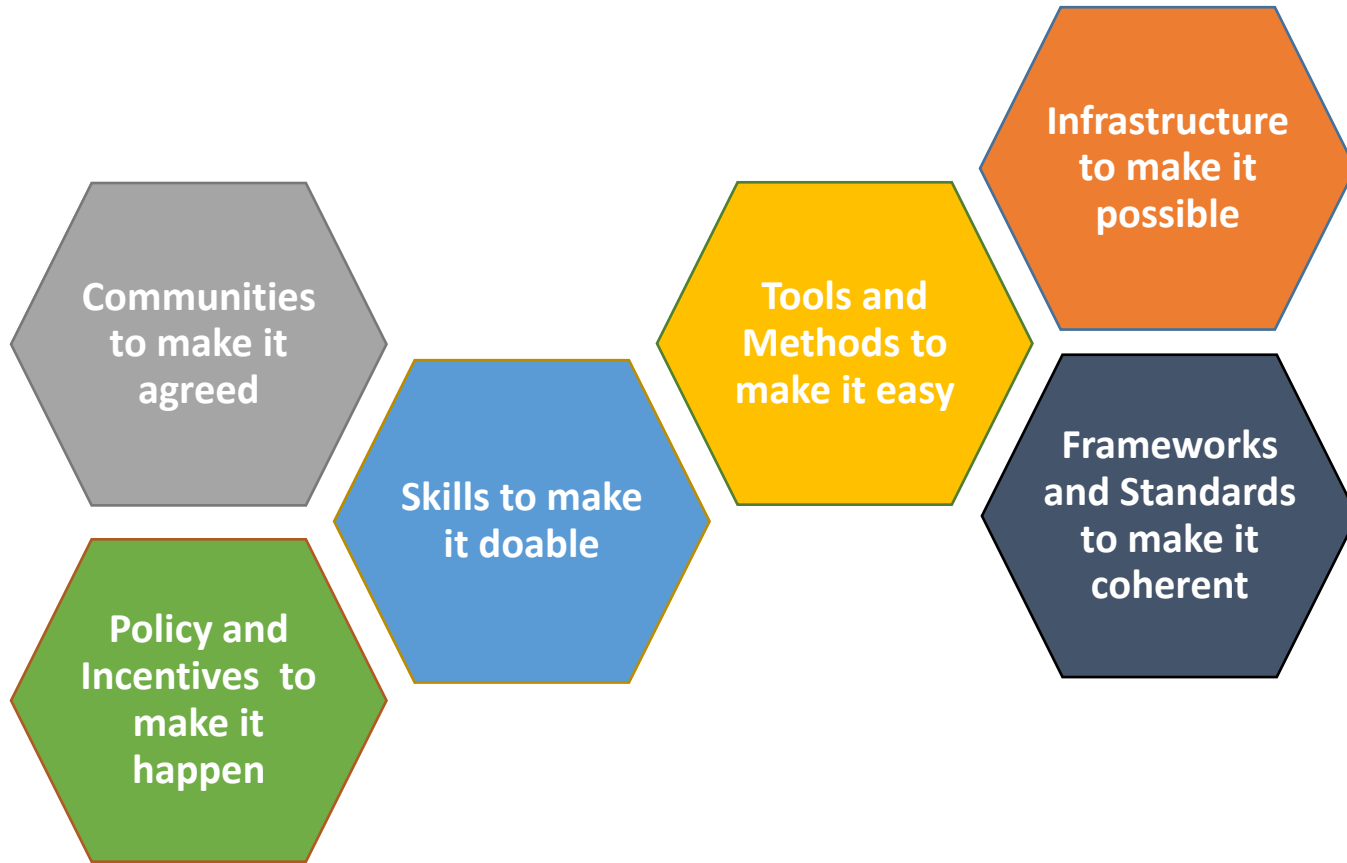
- Agents: individuals (such as researchers or curators); organisations (such as funders, research institutions, data centres, publishers and archival institutions); and other legal entities (such as consortia)
- Resources: publications, data, and other research artefacts, such as lab notebooks, software, equipment, or physical specimens
- Rights statements: grant agreements, licenses, patents
- Events: processes that are relevant to the provenance of resources, such as creation, curation, access, claiming, updates, citation, review
- Derived entities: such as projects

Source: Dappert, A et al 2017 Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. Data Science Journal, 16: 28, pp. 1–16, DOI: <https://doi.org/10.5334/dsj-2017-028>

Why open?

- Because if we are building identifiers into our connections, we don't want to be held to ransom by identifier providers
- Need trust across each of the following areas:
 - running the infrastructure (governance)
 - funding it (sustainability)
 - preserving community ownership of it (insurance).
 - *Bilder G, Lin J, Neylon C (2015) Principles for Open Scholarly Infrastructure-v1, retrieved 10 Sept 2018, <http://dx.doi.org/10.6084/m9.figshare.1314859>*

Changing research practice to
bring all of this about



Derived from [COS pyramid](#)

Trusted Data + Trusted
Cyberinfrastructure +
Trusted Connections =
Trusted Research

Andrew Treloar

E: andrew.treloar@ardc.edu.au

T: @atreloar



With the exception of third party images or where otherwise indicated, this work is licensed under the Creative Commons 4.0 International Attribution Licence



The Australian Research Data
Commons is supported by the
Australian Government
through the National
Collaborative Research
Infrastructure Strategy
Program