Research Objects: Towards Exchange and Reuse of Digital Knowledge

Sean Bechhofer, David De Roure, Matthew Gamble, Carole Goble, Iain Buchan

University of Manchester, University of Southampton

FWCS Workshop, WWW 2010

Raleigh NC

25th April, 2010

MANCHESTER 1824

UNIVERSITY OF Southampton
School of Electronics
and Computer Science

Research Objects

Semantically rich aggregations of resources, supporting a research objective

- Publications
  - Argumentation: *Convince* the reader [Mesirov]
  - Results reinforced by *Reproducability* [De Roure]
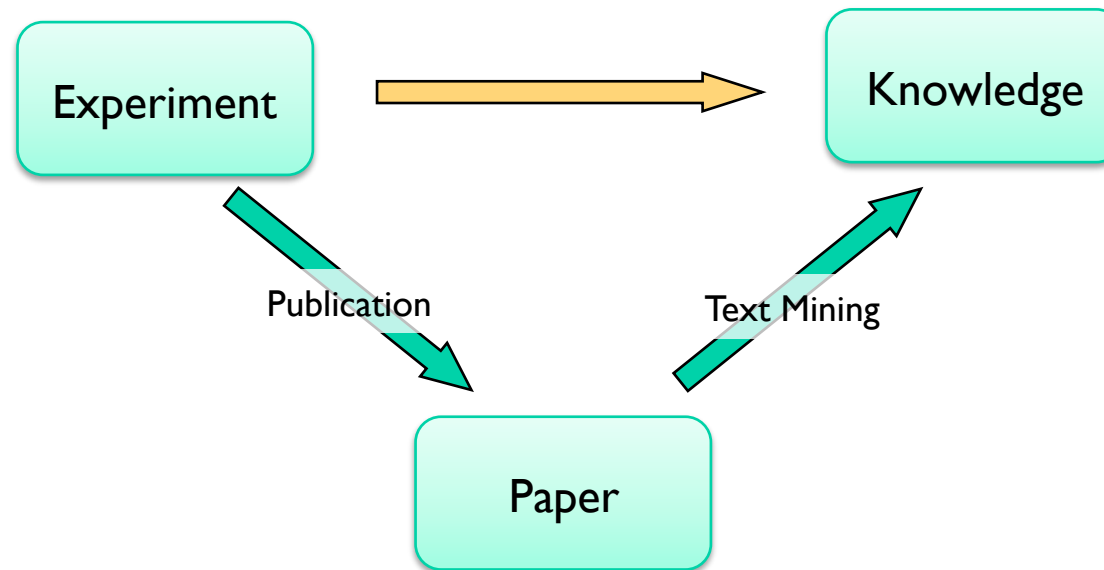
Disclaimer

- This is a *position paper* or "*manifesto*".
- A collection of desiderata.
- A work in progress

vote **Research Object!**

Knowledge Burying [Mons]

- Publishing/mining cycle results in loss of knowledge
  - $\geq$ 40% of information lost
- RIP – *Rest in Paper*
- ROs as a mechanism for publication of knowledge, preserving information about the *process*.

4

What is an e-Laboratory?

- A **laboratory** is a facility that provides *controlled conditions* in which scientific research, experiments and measurements may be performed, offering a work space for researchers.

- An **e-Laboratory** is a set of integrated components that, used together, form a distributed and collaborative space for e-Science, enabling the planning and execution of *in silico* experiments **--** processes that combine data with computational activities to yield experimental results
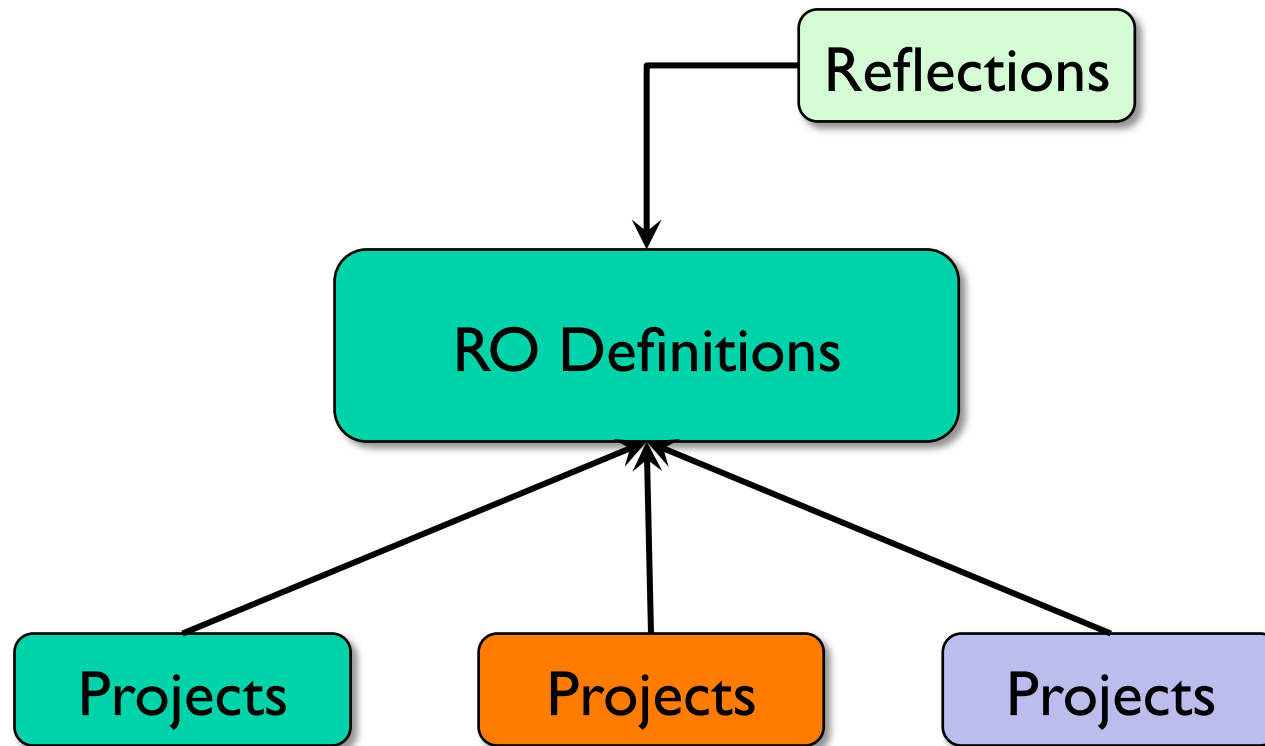
e-Labs

- An e-Lab consists of:
  1. a community;
  2. work objects;
  3. generic resources for building and transforming work objects.

| People | + | Data | + | Methods |
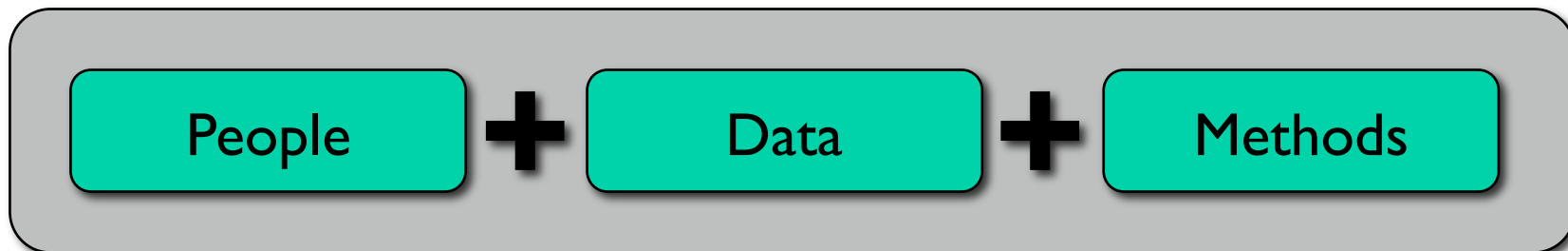
- e-Labs built from a collection of services, consuming and producing ROs, a common currency for communication between e-Lab services

Top Down, Bottom Up

Motivating Projects

- myExperiment
  - Workflow sharing
  - Packs supporting (simple) aggregations.
  - *Links* not just references
  - Packs as nascent ROs
- Sysmo-DB
  - Assets catalogue supporting exchange of data, models, SOPs
- Obesity e-Lab/MethodBox
  - Sharing survey data/analysis scripts
  - "Research Objects Inside"
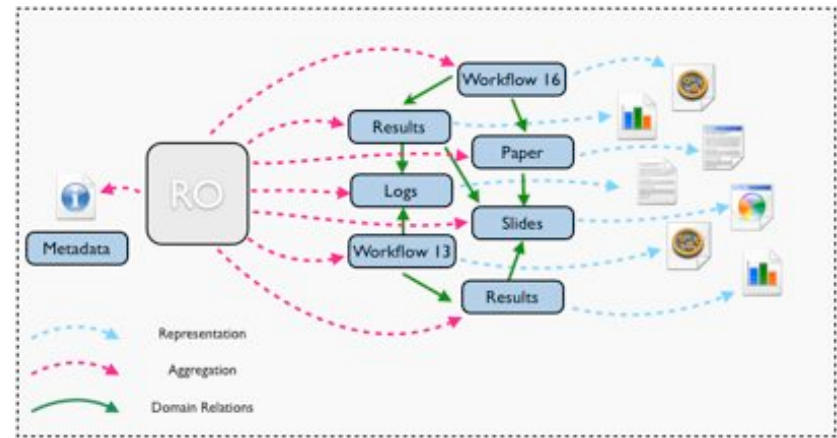
Principles

- Reusability
- Repurposeable
- Repeatable
- Reproducible
- Replayable
- Traceable

- Aggregation
- Identity
- Metadata
- Lifecycle
- Versioning

Graceful Degradation

Research Object services are able to consume Research Objects without necessarily understanding or processing all of their content

ROs and Linked Data

- Linked Data: Collection of best practices for publishing and connecting structured data on the web.

- ROs should be independent of mechanisms for representation and delivery

- ROs as non-information resources
  - "Named Graphs for LD"

LD Cloud

RO
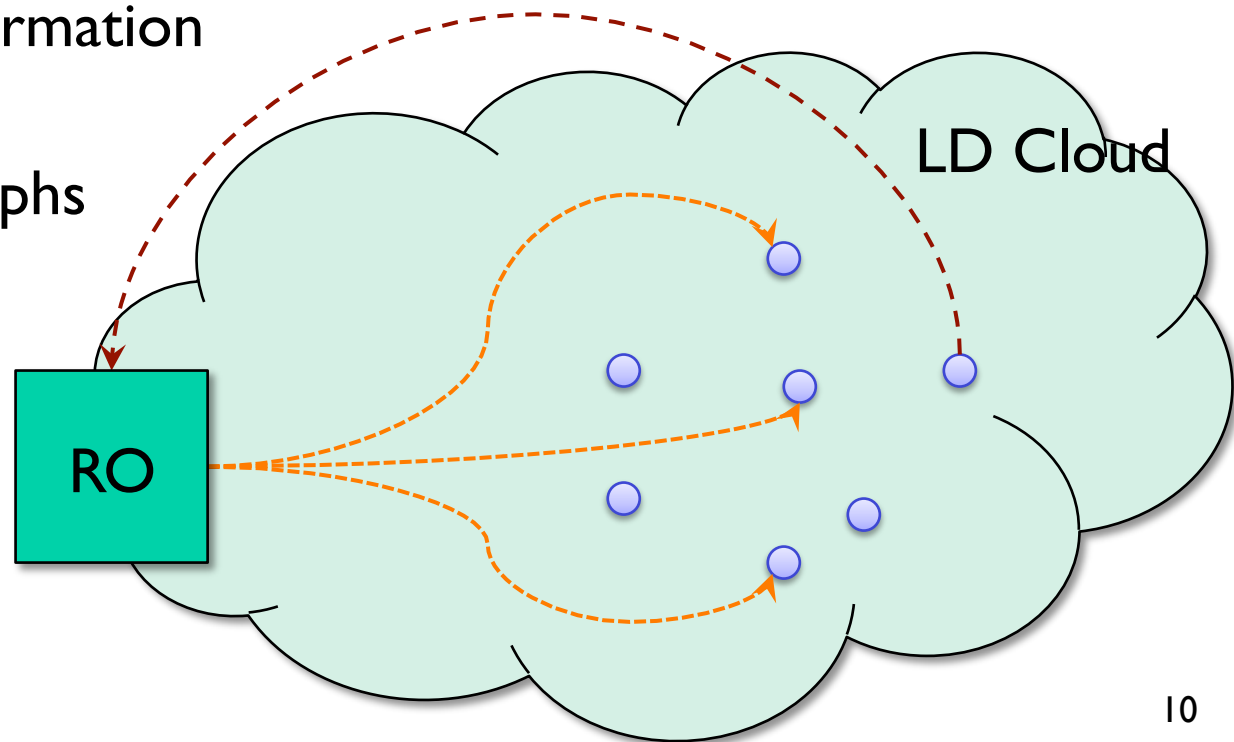
Stereotypes

- **Publication Object**
  - Record of Activity
  - Credit/attribution
- **Live Object**
  - RO as work in progress
  - Up to date references to appropriate resource
- **View Object**
  - Named Graphs for LD

- **Archived Object**
  - RO as a record of what happened
  - Curated, "fossilised", immutable aggregation
- **Exposing Object**
  - Standardised wrapper around data sources
- **Method Object**
  - RO as protocol

Next Steps

- Prototype implementations of core Research Object functionality and services

- Building on myExperiment experiences/codebase.

- OAI-ORE to support RDF-based aggregation structures.

- Exploring lifecycle

- Stereotypes

- Reuse of existing vocabularies (SWAN-SIOC)
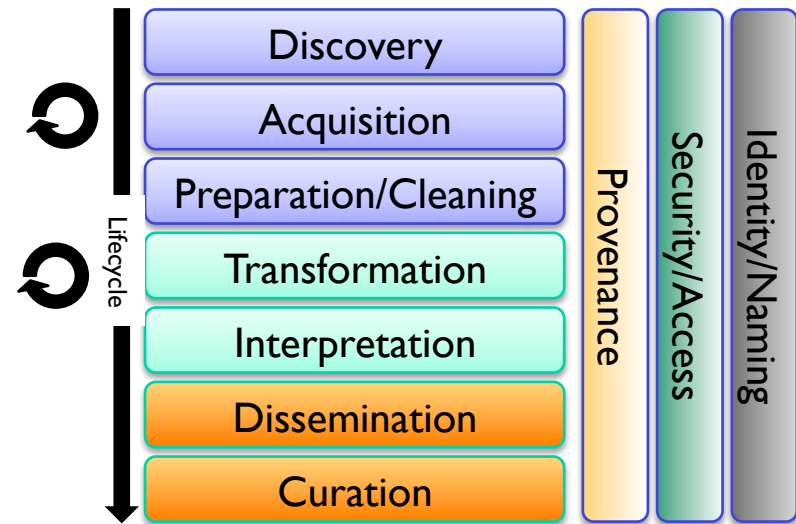
The Vision

- Research Objects will allow us to conduct research in ways that are
    - *Efficient*: cheaper to borrow than recreate;
    - *Effective*: larger scale through reuse;
    - *Ethical*: Benefiting wider communities, not just individuals.

- Paper publication is an archaic human-readable form of *Research Object*
- *Could I have a copy of your Research Object please?*

Thanks!

- Thanks to all who contributed, including John Ainsworth, Jiten Bhagat, Phillip Couch, Doc Cruickshank, Mark Delderfield, Ian Dunlop, Danius Michaelides, Paolo Missier, Stuart Owen, David Newman, Shoaib Sufi.

vote

Research Object!