http://www.slideshare.net/hvdsomp/the-oaiore-interoperability-framework

The OAI-ORE Interoperability Framework in the current Scholarly Communication Context

<http://www.openarchives.org/ore/toc>

Herbert Van de Sompel
Los Alamos National Laboratory, USA

OAI-ORE was funded
by the Andrew W. Mellon Foundation, the National Science Foundation, JISC, and Microsoft

The ORE Editors are: Carl Lagoze (Cornell U.), Herbert Van de Sompel (LANL), Pete Johnston (Eduserv Found.), Michael Nelson (Old Dominion University), Robert Sanderson (U. of Liverpool), Simeon Warner (Cornell U.)

In This Talk

Some thoughts about scholarly communication based on the forthcoming paper:

Herbert Van de Sompel & Carl Lagoze (2009) All Aboard: Towards a Machine-Friendly Scholarly Communication System.

Written as a contribution to a book compiled in honor of Jim Gray.

Herbert's slogan used for effect in many talks

The current scholarly communication system is nothing but a scanned copy of the paper-based system.

Systemic issues
Economic strains
Technical aspects

Working Towards Change

About 10 years ago, the OAI & OA started out of dissatisfaction with the existing scholarly communication system.

And many efforts since then were launched to try and introduce real change.

Working Towards Change is Starting to Pay Off

Rapid changes are occurring along various dimensions in scholarly communication.

Technical trends:

1. Augmentation of the scholarly record with a machine-actionable substrate.
2. Integration of datasets into the scholarly record.
3. Exposure of process and its integration into the scholarly record.

Working Towards Change is Starting to Pay Off

Rapid changes are occurring along various dimensions in scholarly communication.

Technical trends:

1. Augmentation of the scholarly record with a machine-actionable substrate.
2. Integration of datasets into the scholarly record.
3. Exposure of process and its integration into the scholarly record.

Augmentation with a machine-actionable substrate

- There is just too much literature for researchers to be able to keep up [1].
- There is a need to absorb literature across disciplines, to connect the dots, to combine existing disparate findings to arrive at new insights.
  - E.g. Life sciences: many interconnected disciplines & the lack of uniformly structured data across disciplines is a barrier for moving discoveries in basic research to clinical applications [2].
- There is a need for machine-agents that filter, *read* and reason on behalf of researchers.

[1] Palmer, C.L., Cragin, M.H. and Hogan, T.P. (2006) Weak information work in scientific discovery. Information Processing & Management, 43 (3). 808-820. doi:10.1016/j.ipm.2006.06.003

[2] Ruttenberg, A., Clark, T., Bug, W., et al. (2007) Advancing translational research with the Semantic Web. BMC Bioinformatics, 8 Suppl 3. S2. doi:10.1186/1471-2105-8-S3-S2

Augmentation with a machine-actionable substrate

- There is a strong push towards a machine-actionable representation of knowledge that is embedded in scientific literature.
  - Post publication: Entity/Relation extraction via text mining.
  - At publication time: Semantic publications [1].
  - Shared ontologies across disciplines for uniform knowledge representation [2].

- The Elsevier Grand Challenge [3] and the Concept Web Alliance [4] are strong indications that both the information industry and the research community are taking this seriously.
  - Knowlets as snippets of semantically expressed knowledge that can be shared and … traded. The emergence of A&I databases for machine consumption …

[1] Shotton, D. (2009) Semantic Publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, *22* (2), pp 85-94. doi:10.1087/2009202

[2] The Open Biomedical Ontologies. http://www.obofoundry.org/

[3] Elsevier Grand Challenge. http://www.elseviergrandchallenge.com

[4] The Concept Web Alliance. http://conceptweblog.wordpress.com/declaration/

Working Towards Change is Starting to Pay Off

Rapid changes are occurring along various dimensions in scholarly communication.

Technical trends:

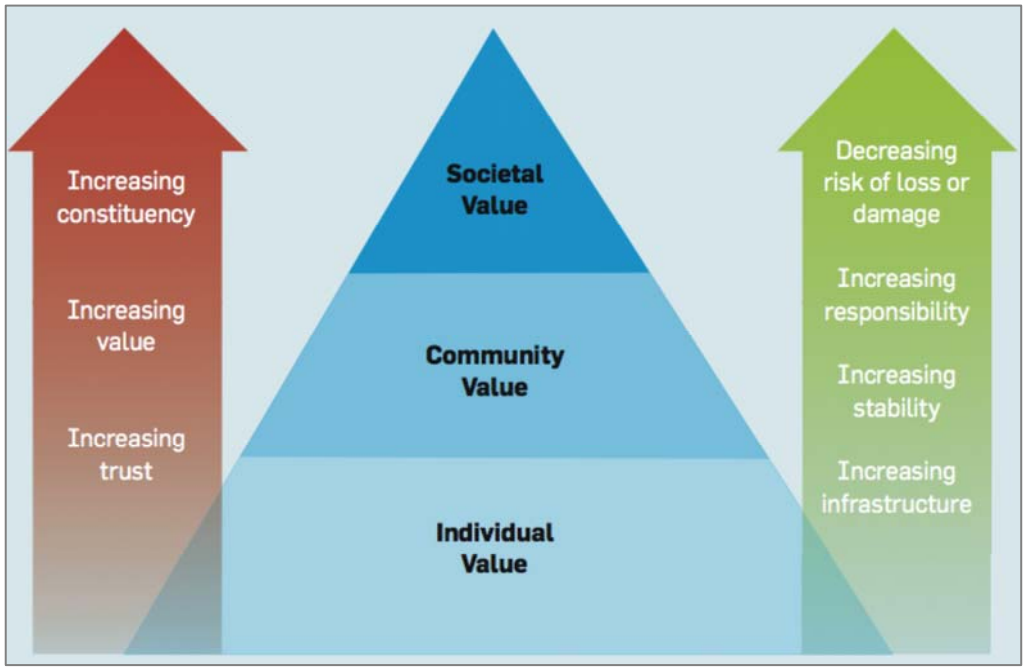
1. Augmentation of the scholarly record with a machine-actionable substrate.
2. Integration of datasets into the scholarly record.
3. Exposure of process and its integration into the scholarly record.

Integration of datasets in the scholarly record

- Data has always been important to scholarship.
- But until recently data was not treated as a first-class scholarly object.

Integration of datasets in the scholarly record

- Now all functions of scholarly communication – *registration*, *certification*, *awareness*, *archiving*, *rewarding* – are being discussed/implemented for datasets:
  - Data Pyramid [1] indicates how trust – *certification* – and preservation – *archiving* – are crucial as the applicability of data reaches beyond the personal realm.



[1] Berman, F. (2008) Got data?: a guide to data preservation in the information age. *Communications of the ACM*, *51* (12). 50-56. doi:10.1145/1409360.1409376

Integration of datasets in the scholarly record

- Now all functions of scholarly communication – *registration*, *certification*, *awareness*, *archiving*, *rewarding* – are being discussed/implemented for datasets:
  - Data Pyramid [1] indicates how trust – *certification* – and preservation – *archiving* – are crucial as the applicability of data reaches beyond the personal realm.
  - General recognition of the necessity of capabilities to discover datasets – *awareness* [2].
  - Efforts aimed at identifying and referencing datasets in a standard way [3] take their position as primary artifacts for granted (and are partly motivated by allowing researchers to gain credit – *rewarding* – for sharing data).

[1] Berman, F. (2008) Got data?: a guide to data preservation in the information age. *Communications of the ACM*, *51* (12). 50-56. doi:10.1145/1409360.1409376

[2] Ruusalepp, R. (2008) Infrastructure Planning And Data Curation: A Comparative Study Of International Approaches To Enabling The Sharing Of Research Datahttp://www.dcc.ac.uk/docs/publications/reports/Data_Sharing_Report.pdf.

[3] Altman, M. and King, G. (2007) A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib Magazine*, *13* (3/4). doi:10.1045/march2007-altman

Working Towards Change is Starting to Pay Off

Rapid changes are occurring along various dimensions in scholarly communication.

Technical trends:

1. Augmentation of the scholarly record with a machine-actionable substrate.
2. Integration of datasets into the scholarly record.
3. Exposure of process and its integration into the scholarly record.

Exposure of process and its integration into the scholarly record

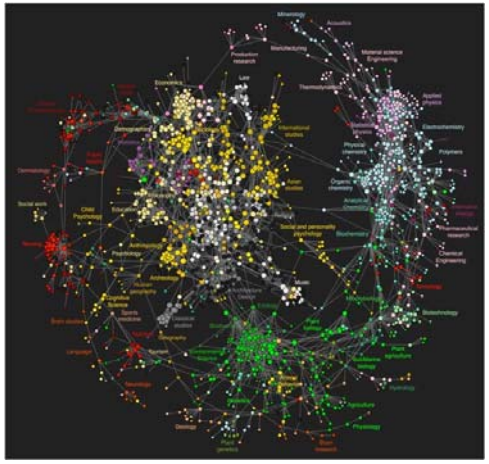
- The process of using of prior knowledge to generate new insights has been exposed for a long time via citations in the literature.
- The citation graph reveals aspects of scholarly dynamics and is heavily studied for that reason.
- But the citation graph is problematic (imperfect citation extraction; challenging author disambiguation; partial coverage; proprietary nature).

- A more meaningful graph can emerge:
  - Web-oriented citation approach, e.g. CLADDIER [1].
  - Author identifiers [2].
  - Citation semantics, e.g. Citation Typing Ontology [3;4].

[1] The CLADDIER Project. http://claddier.badc.ac.uk/trac
[2] Enserink, M. (2009) Are you ready to become a number? Science, 323(5922) pp 1662-1664. doi:10.1126/science.323.5922.1662
[3] Citation Typing Ontology. http://purl.org/net/cito/
[4] Shotton, D. (2009) Semantic publishing: the coming revolution in scientific journal publishing. Learned Publishing, 22(2) pp 85-94(10). doi:10.1087/2009202

Exposure of process and its integration into the scholarly record

- New: Exposing the process of consuming or paying attention to scholarly information (by sharing usage data) as revealed by COUNTER [1], MESUR [2], bX [3] projects.



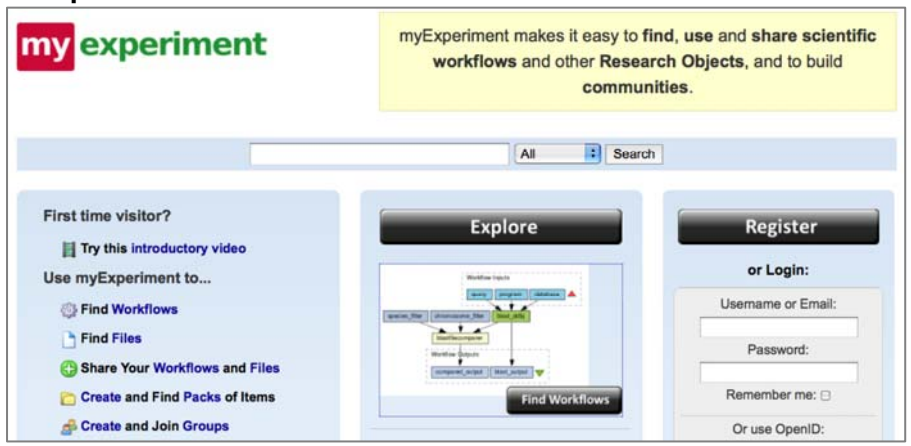
- Applications in:
  - Collection management.
  - Novel metrics of scholarly impact [4].
  - Study of scholarly dynamics/trends [5].
  - Discovery - Recommender systems [6].

[1] Project COUNTER. http://www.projectcounter.org/
[2] MESUR. http://www.mesur.org/
[3] Ex Libris bX. http://www.exlibrisgroup.com/category/bXOverview
[4] Bollen, J., Van de Sompel, H., Hagberg A., Chute, R. (2009) A principal component analysis of 39 scientific impact measures. http://arxiv.org/abs/0902.2183 . Accepted by PLoS ONE
[5] Bollen J., Van de Sompel, H., Hagberg, A. et al. (2009) Clickstream Data Yields High-Resolution Maps of Science. PLoS ONE 4(3): e4803. doi:10.1371/journal.pone.0004803
[6] Bollen, J., and Van de Sompel, H. (2006) An architecture for the aggregation and analysis of scholarly usage data. Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 298-307. doi:10.1145/1141753.1141821

Exposure of process and its integration into the scholarly record.

- New: Exposing scientific workflows (Taverna, Kepler, …) , experiments, bundles of resources used in workflows as revealed by myExperiment [1;2].

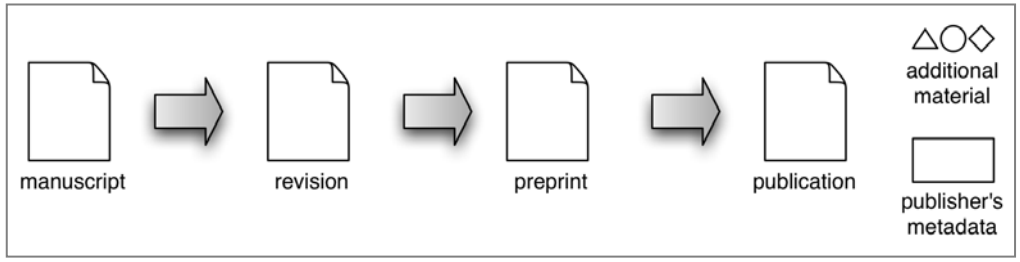
- Applications in:
  - Discovery and reuse of data and workflows.
  - Novel metrics of scholarly impact.

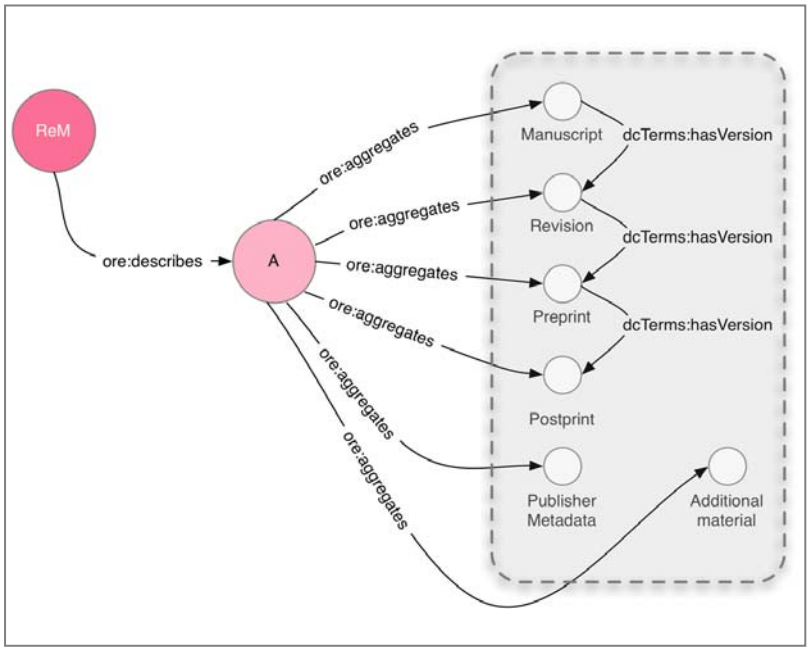


[1] myExperiment. http://www.myexperiment.org/
[2] De Roure, D.C. and Goble, C. (2007) myExperiment – A Web 2.0 Virtual Research Environment, in International Workshop on Virtual Research Environments and Collaborative Work Environments, (Edinburgh, UK, 2007). Available at http://eprints.ecs.soton.ac.uk/13961/.

Exposure of process and its integration into the scholarly record.

- New: Exposing scholarly value chains, and scientific life cycles [1]

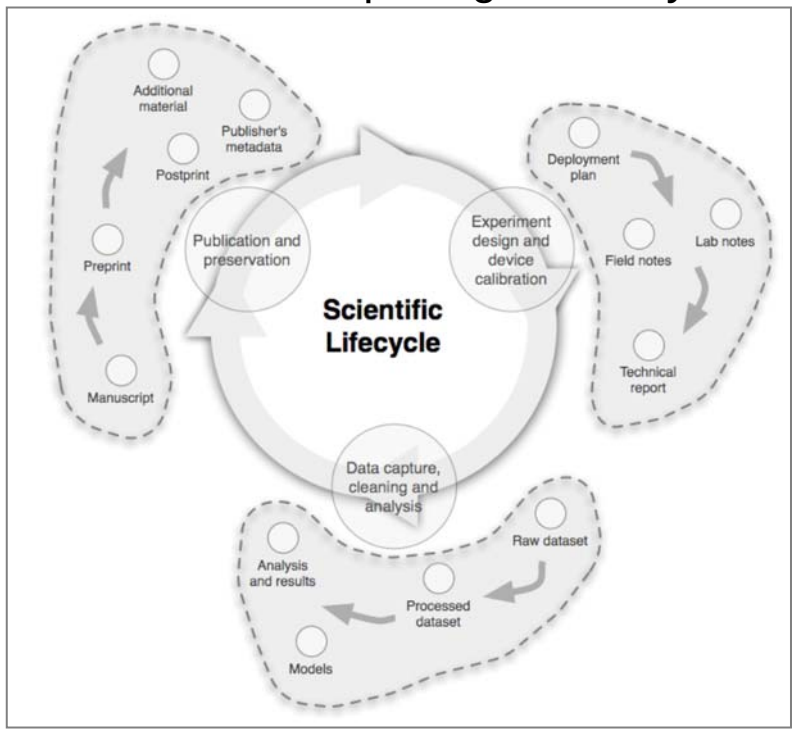


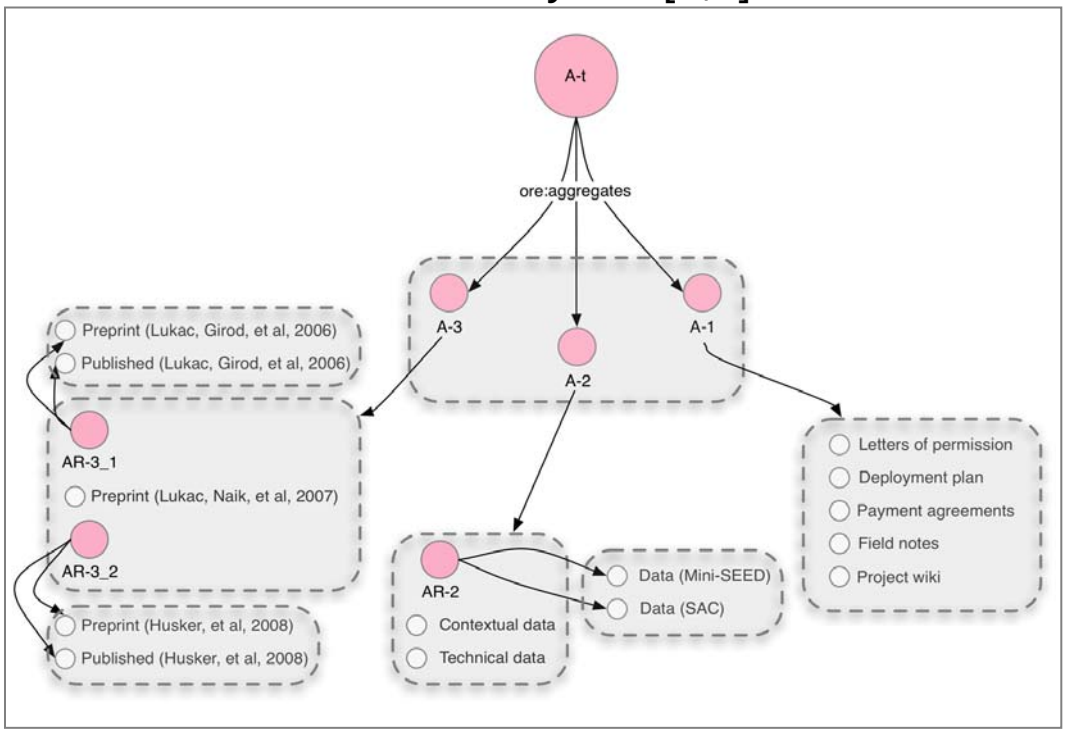


[1] Pepe, A., Mayernik, M., Borgman, C., Van de Sompel, H. (2009) Technology to Represent Scientific Practice: Data, Life Cycles, and Value Chains. arXiv preprint arXiv:0906.2549v1 ; submitted to JASIST.

Exposure of process and its integration into the scholarly record.

- New: Exposing scholarly value chains, and scientific life cycles [1;2]





[1] Pepe, A., Mayernik, M., Borgman, C., Van de Sompel, H. (2009) Technology to Represent Scientific Practice: Data, Life Cycles, and Value Chains. arXiv preprint arXiv:0906.2549v1; submitted to JASIST.

Working Towards Change is Starting to Pay Off

These trends have a Web-orientation in the development of scholarly interoperability in common.

URIs, HTTP URIs to identify scholarly artifacts, concepts, researchers, institutions, etc.

Use of Web Architecture and associated standards (XML, RDF, RDFS, OWL, RSS, Atom, …) to support the representation and communication of scholarly information and knowledge.

Working Towards Change is Starting to Pay Off

These trends have a Web-orientation in the development of scholarly interoperability in common.

The OAI-ORE interoperability framework fits right into this wave: based on Web Architecture, and Linked Data (the Semantic Web made real) guidelines.

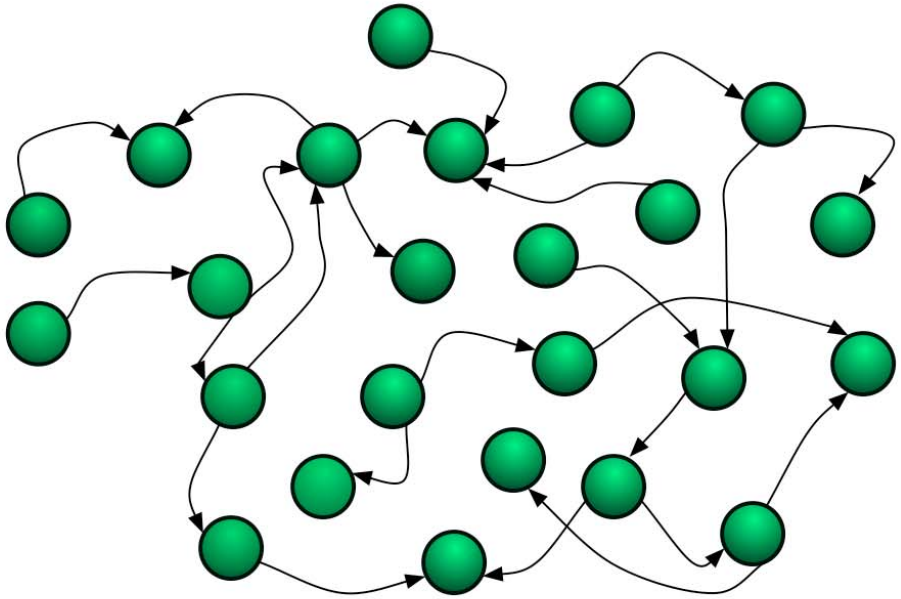
OAI Object Reuse and Exchange

Subject: How to handle an **Aggregation** of Web resources?

Approach: Publish **Resource Maps** to the Web that Instantiate, Describe, and provide an Identity for the **Aggregation**

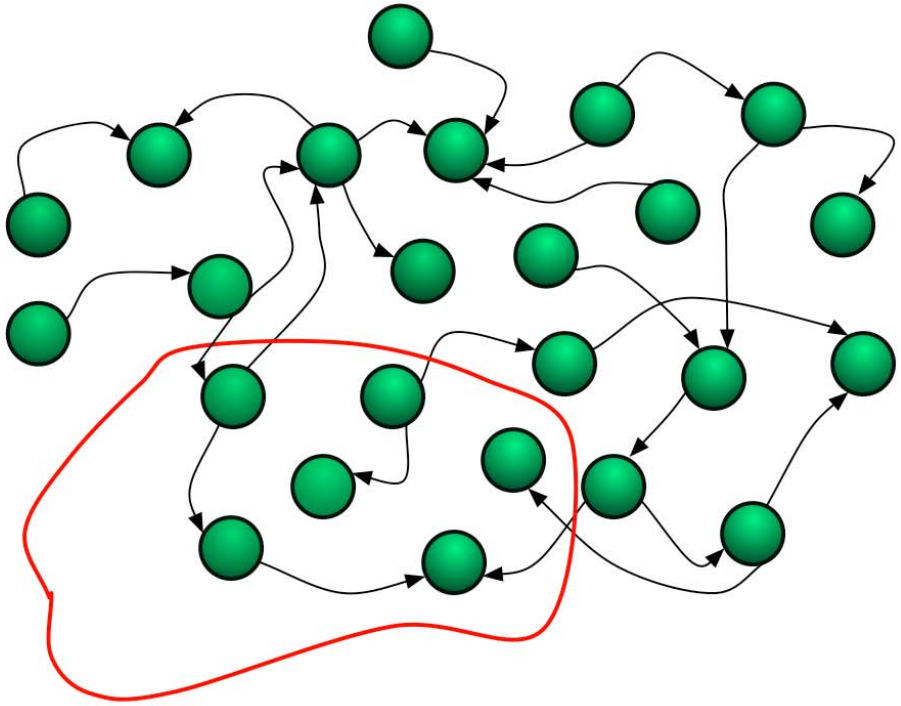
Aggregation: a *non-information* resource
Resource Map: an information resource that describes an Aggregation

The Web

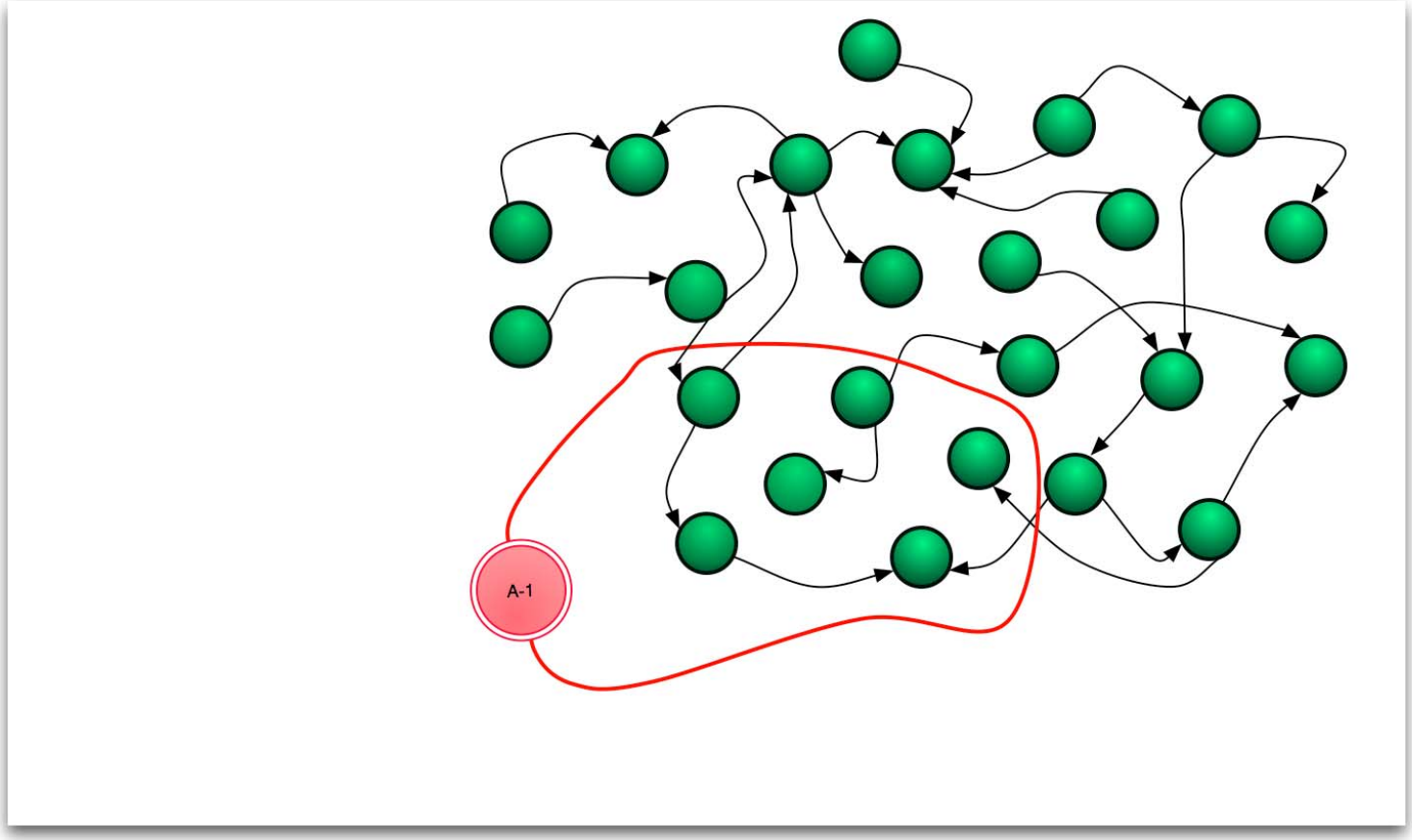


An Aggregation and the Web

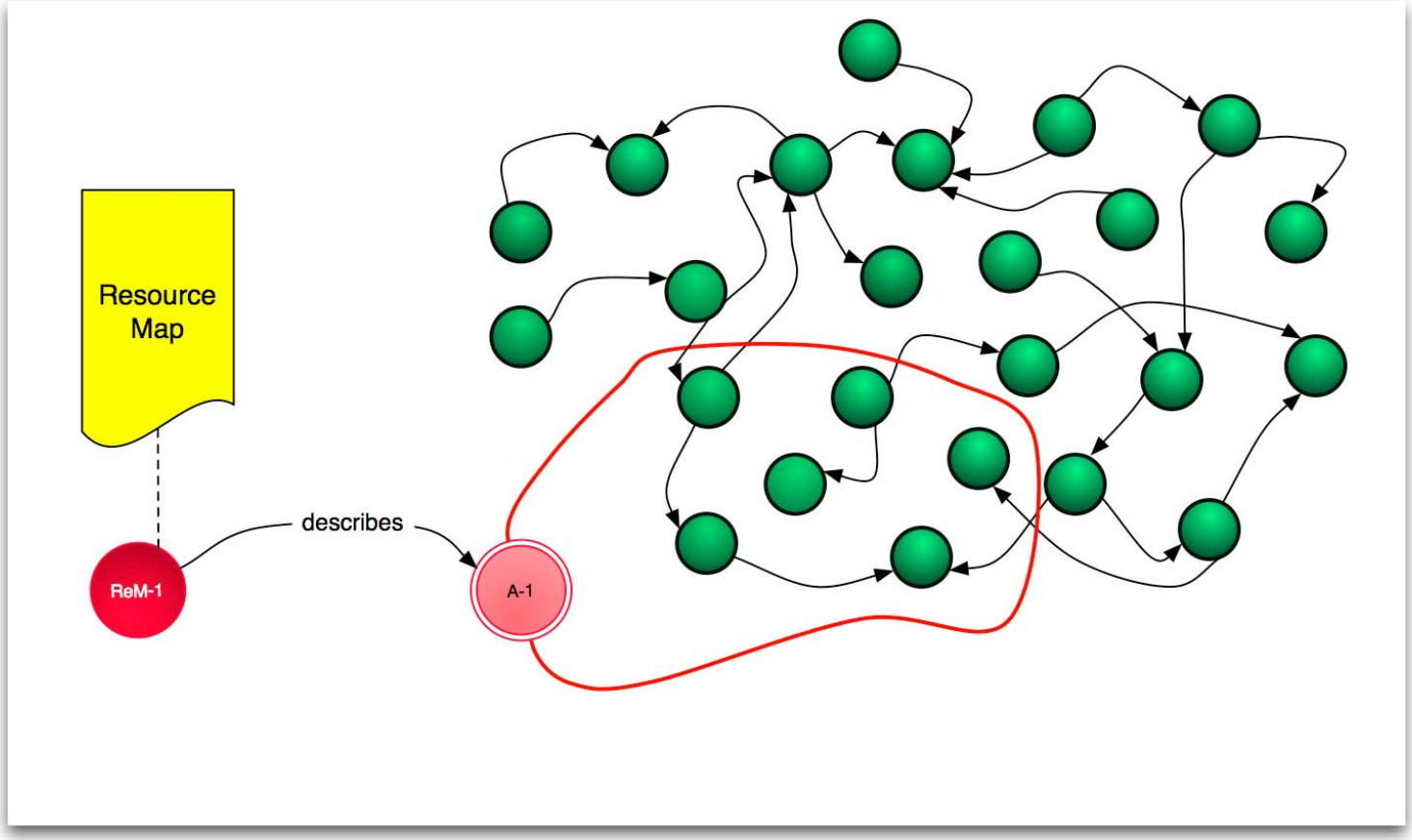
- Resources of an aggregation are distinct URI-identified Web resources
- To handle aggregations, missing are:
  - The boundary that delineates the aggregation in the Web
  - An identity (URI) for the aggregation



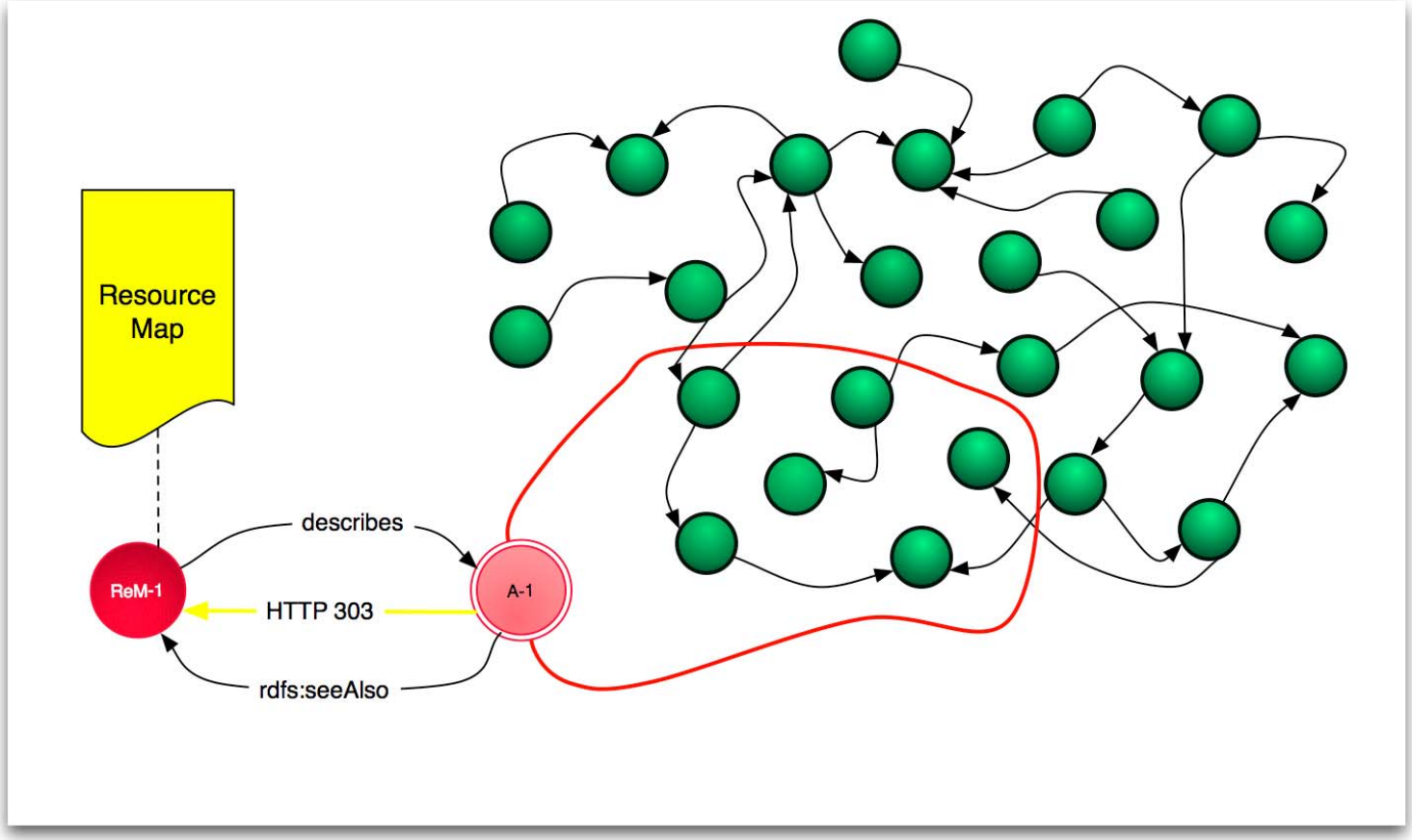
Introduce a Resource that stands for the Aggregation



Publish a Resource Map that describes the Aggregation



Discover the Resource Map via the Aggregation



All the OAI-ORE details at …

ORE Specs – http://www-openarchives.org/ore/toc

ORE Tutorial - http://www.slideshare.net/hvdsomp/an-overview-of-the-oai-object-reuse-and-exchange-interoperability-framework

And now …

Robert Sanderson. Visualizing JSTOR: Exploring OAI-ORE for Information Topology Navigation.

Maarten Hoogerwerf. A demonstrator of enhanced publications using OAI-ORE.

Tim DiLauro. Using OAI-ORE to Simplify Data Publishing Workflow.