The RMap Project: Linking the Products of Research and Scholarly Communication
Motivation

• Compound objects fast becoming the norm for outputs of scholarly communication.

• In many circumstances, the traditional article is not the object of long-term interest for at least some segment of the community.

• Components may reside in different repositories, maintained by different institutions, employing different technologies.
  – Some of these components and their repositories are not part of the traditional scholarly communication ecosystem.

• Acknowledgement that these objects do not stand alone -- and of a broad need to understand their context.
Research Partnership

- Data Conservancy: Expertise in management of large data archives from multiple disciplines
- IEEE: Expertise in management of data-intensive scholarly journal publications
- Portico: Expertise in digital preservation, publisher workflow requirements, and existing relationships with 275 publishers

- Funding from the Alfred P. Sloan Foundation
Some High-Level Goals

- RMap tool working prototype
- Collaborative partnerships with the community
- System that supports emerging forms of digital scholarship and publishing
- Plan for sustainability of the project
Work Plan

• Year One—Planning Phase: Gather requirements, create use cases, hold workshop with stakeholders, refine use scenarios based on community feedback

• [You are here] Year Two—Prototype Development: Create system to identify, store, update, and retrieve relationships among publications and other forms of scholarly output, including data and software
TECHNOLOGY
The RMap Project
Key Objectives

• Support assertions from broad set of contributors
• Integrate with Linked Data
• Leverage data from existing scholarly publishing stakeholders (publishers, identifier providers, data and software repositories)
• Provide support for agents and other resources without identifiers (authors, textual citations)
Data Model (simplified)
Data Model - Resource

- Things (abstract or concrete) that can have an identifier
- Basic building block of the WWW
- Key entity for description and retrieval within RMap
- Other core entities in the data model are also Resources
Data Model - Agent

- A person or thing (or group of these) responsible for some action
- Distinction between scholarly (e.g., author, funder, publisher, data processing program) and system (RMap component, user, etc.)
Capture provenance within RMap system
An action or activity involving System Agents and other resources
Provenance of Scholarly Resources can be captured separately by registering it in RMap via DiSCOs.
Data Model – RDF Statement (triple)

- Building blocks of the semantic web
- Conceptually of the form: `<subject> <predicate> <object>`
- Like subject-verb-object in English
Data Model - DiSCO

- Distributed Scholarly Compound Object
- Primary unit of registration within RMap
- Basically a set of resources and related RDF description.
- Similar to OAI-ORE
Data Model - DiSCO

DiSCO

C-2

A-2

D-2

creator

cites

D-1

S-1

source

outputOf

Compound Object
Create dataset D-1
Create software S-1
Generate dataset D-2
Article related to D-2

Software Repositories

Data Repositories

Publishers

RMap Linked Data Graph

Incoming Discord

Creator

Article

Dataset

Software
Creation of software S-2

Software Repositories

Data Repositories

Publishers

RMap Linked Data Graph

Incoming DiSCO
Generation of dataset D-3
Article A-2 related to D-3
Correct article identifier

RMap Linked Data Graph

Incoming DiSCO

Creators
Articles
Datasets
Software
Repositories
Data
Repositories
Publishers

C-1
C-2
D-1
D-2
D-3
S-1
S-2
C-1
C-2
A-1
A-3
Dataset D-1 connections

Diagram showing connections between Software Repositories, Data Repositories, and Publishers. The diagram includes nodes labeled S-1, S-2, C-1, C-2, D-1, D-2, D-3, A-1, A-3, and C-1. Arrows indicate relationships such as creator, source, and outputOf. The diagram is titled "RMap Linked Data Graph."
Creator C-1 connections

```
Creator C-1 connections
```

```
Creator C-1 connections
```

```
Creator C-1 connections
```
Associate resources with C-1 identity
Associate resources with more identities
RESTful APIs

- Programming language independent
- Easy to test with web tools (curl, wget)
- Abstraction away from underlying implementations and models, which we expect to change more often
## REST APIs (subset)

<table>
<thead>
<tr>
<th>Function</th>
<th>HTTP verb</th>
<th>API rel path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieve related triples</td>
<td>GET</td>
<td><code>{resourceURI}/stmts</code></td>
</tr>
<tr>
<td>Retrieve related events</td>
<td>GET</td>
<td><code>{resourceURI}/events</code></td>
</tr>
<tr>
<td>Retrieve related DiSCO</td>
<td>GET</td>
<td><code>{resourceURI}/discos</code></td>
</tr>
<tr>
<td>Create DiSCO</td>
<td>POST</td>
<td>/disco</td>
</tr>
<tr>
<td>Retrieve DiSCO</td>
<td>GET</td>
<td>/disco/{discoId}</td>
</tr>
<tr>
<td>Update DiSCO</td>
<td>POST</td>
<td>/disco/{discoId}/update</td>
</tr>
<tr>
<td>Delete a DiSCO</td>
<td>DELETE</td>
<td>/disco/{discoId}/delete</td>
</tr>
<tr>
<td>Retrieve an Event</td>
<td>GET</td>
<td>/event/{eventId}</td>
</tr>
<tr>
<td>Get DiSCOs related to event</td>
<td>GET</td>
<td>/event/{eventId}/discos</td>
</tr>
<tr>
<td>Perform SPARQL query</td>
<td>POST</td>
<td>/sparql</td>
</tr>
</tbody>
</table>
API Specification and Documentation

• Behaviors
• API paths
• Data Models
• Serializations (media types, content negotiation)
• Implementations
Function: Update DiSCO

• Behavior within RMap
  – Failed requests will be rolled back, so as not to require manual cleanup (transaction)
  – Insufficient authorization will result in failed transaction and offer to authenticate with other credentials.
  – A new DiSCO will be instantiated; the previous (old) DiSCO will be marked “inactive”
  – Add triple <new-DiSCO-URI> <hasVersion> <old-DiSCO-URI>
  – Resources will be instantiated for objects without identifiers (e.g., citation as string)
  – Scholarly Agents will be instantiated for agents lacking URIs (e.g., creator as string)
  – Event(s) created capture activity

• Request
  – Verb/relative path: POST /disco/{id}/update
  – Path parameters: {id} - URI of existing (old) DiSCO
  – Model: Resources + relationships (like OAI-ORE)
  – Serializations: RDF/XML, Turtle, or JSON-LD

• Response
  – Model: (custom)
  – Serializations: JSON, HTML
  – New DISCO URI in header: Location: <new-DiSCO-URI>
  – Old DISCO URI in header: Link <old-DiSCO-URI>;rel="predecessor-version"
  – Event URI(s) in header: Link <event-URI>;rel="http://www.w3.org/ns/prov#wasGeneratedBy"
  – [Enumerate response codes, labels, and their meanings]
API Coverage

• Current focus on APIs to populate and access the graph

• Future focus
  – Authentication
  – Administration
  – Composition & normalization
  – Inferencing
  – System operability
Technical Team Activity

• Developed and captured initial set of use cases
• Developed and documented initial data model
• Specified API behaviors
• Developed and documented API methods, including REST paths, request and response formats, models, and serializations (media types)
  – Still a couple issues to sort out
• Prototype platform implementation
• Participation in RDA Data Publishing groups
• Actively working on harvesting relationship data to push into RMap
Harvesting links and proxy registration
Community Engagement
The RMap Project
Workshop: Key Feedback

• RMap Project should be a clearinghouse or meta-service that captures information about various data-linking services

• Important to add value to the publication & data linking work already underway in the community

• Having an established publisher as a research partner is a comparative advantage for the RMap Project
Workshop Feedback (continued)

• One approach would be to focus on the “input” side of the process (with special attention for software and research workflows) in order to create a generalizable approach to gathering content.

• The challenge of “secondary data”, such as the inferred connections between publications and data or software remains unaddressed and important.
Some of the things you can do for RMap

• Feedback
  – Do the articulated use cases, approach, goals, and proposed offerings align with your interests. Where they don’t, how could we better align?

• Share Your Data
  – As we populate our prototype, we need to gather a broad swath of test data, covering a variety of resource types (e.g., journals, repositories, funders, creators, articles, data, software, instruments, samples) and the relationships that connect them.

• Use
  – Consider using RMap capabilities to register, discover connections to, and augment your own content, once those capabilities become available.
Some of the things RMap will do for you

• Aggregate and offer an inclusive and normalized view of distributed scholarly compound objects and associated resource relationships, including those from sources without membership in existing identity services (e.g., source code management platforms, institutional repositories).
  – Reduce cost and complexity of transforming information from multiple systems.

• Provide a single mechanism to discover context (e.g., relationships and related resources) for scholarly objects in which you are interested.
  – Reduce cost and complexity of developing and managing multiple interfaces for multiple systems.

• Expose records of a particular statement (e.g., who has asserted that Resource X was created by Agent Y?) or the history of assertions associated with a particular resource (i.e., what has been said about Resource X?).

• Capture sufficient provenance information to allow evaluation of assertions by their source and content.
  – Streamline logic for automatic integration of citation and reference to objects of interest.
Team Members and Acknowledgements

• Sayeed Choudhury, Tim DiLauro: Data Conservancy, Johns Hopkins

• Mark Donoghue, Gerry Grenier, Renny Guida, Ken Rawson: IEEE

• Vinay Cheruku, Karen Hanson, Amy Kirchhoff, John Meyer, Sheila Morrissey, Stephanie Orphan, Jabin White, Kate Wittenberg: Portico

This research project is made possible through generous support from the Alfred P. Sloan Foundation

We thank our workshop participants for their valuable feedback
Q&A

• For more information, please visit:
  – http://rmap-project.info/rmap/