

PUBLISHING AND PRESERVING DATA

STM Annual U.S. Spring Conference 2014



Enabling New Forms of Scholarly Publishing

- Scholars continue to expand the role of data in their research
- Professionals in data curation, publishing, and preservation must develop expertise to support this work as it moves from research to formal publication
- The goal of this project is to enable innovation in scholarly communication and digital publishing

Changing Role of Data in Publications

- Increasing expectations for connections between publications and underlying data
- Importance of collaboration among publishers, data centers, and preservation services to build tools to serve this need
- Goal is to preserve not just publications and data, but the relationships among them

Changing Definition of “Article”

- Primary unit of scholarly communication is becoming a multi-part, distributed object
- Need to maintain the connections among the various components and locations
- Increasingly elaborate relationship information will be necessary going forward

Tool Prototype

- Graph-based view of connections among publications, data, agents, and their properties
- Many-to-many relationships rather than one-to-one view of current systems
- Tracking and preservation of these connections through the scholarly communications cycle

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 240 publishers

Work Plan

- Year One—Planning Phase: Gather requirements, create use cases, hold workshop with stakeholders, refine use scenarios based on community feedback
- Year Two—Prototype Development: Create system to identify, store, update, and retrieve relationships among publications and their data

Example Use Case: Publisher

- Publisher wants to know if there are reference links to an article, its author, or a funding agency that it can include in its publication
- Publisher submits article metadata and identifiers to the Matchmaker tool
- Matchmaker consults its database and returns any relationship maps it finds

Working with the Community

- Project PIs will consult with and share project plans and findings with the publishing, research, and data curation communities
- Project workshop in Year One will test prototype plans with stakeholders and use feedback to refine Year Two work plan for prototype development

Outcomes and Deliverables

- Matchmaker tool working prototype
- Plan for long-term sustainability of the service
- Collaborative partnership with the community
- System that supports emerging forms of digital scholarship and publishing

Team Members and Acknowledgements

- Sayeed Choudhury, Tim DiLauro: Data Conservancy, Johns Hopkins
- Gerry Grenier, Renny Guida: IEEE
- Kate Wittenberg, Sheila Morrissey: Portico

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