# Scholix Framework: Building a Bridge Between Research Data and Publications

**Amir Aryani,** Australian National Data Service, Melbourne, Australia, orcid.org/0000-0002-4259-9774, amir.aryani@ands.org.au

**Adrian Burton,** Australian National Data Service, Melbourne, Australia, orcid.org/0000-0002-8099-7538, adrian.burton@ands.org.au

**Paolo Manghi,** Institute of Information Science and Technology - CNR, Pisa, Italy, orcid.org/0000-0001-7291-3210, paolo.manghi@isti.cnr.it

**Sandro La Bruzzo**, Institute of Information Science and Technology - CNR, Pisa, Italy, orcid.org/0000-0003-2855-1245, sandro.labruzzo@isti.cnr.it

**Markus Stocker**, PANGAEA, Bremen, Germany, orcid.org/0000-0001-5492-3212, mstocker@marum.de

**Uwe Schindler**, PANGAEA, Bremen, Germany, orcid.org/0000-0002-1900-4162, uschindler@pangaea.de

**Michael Diepenbroek**, PANGAEA, Bremen, Germany, orcid.org/0000-0003-3096-6829, mdiepenbroek@pangaea.de

**Martin Fenner**, DataCite, Hannover, Germany, orcid.org/0000-0003-1419-2405, martin.fenner@datacite.org

**Hylke Koers**, Elsevier, Amsterdam, The Netherlands, orcid.org/0000-0001-6538-7590, h.koers@elsevier.com

## Session Type:

- Presentation

## Abstract

Identifying the connections between datasets and publications has been a known challenge for scholarly communications and research repositories. In the last three years, there has been a significant development in promoting these connections among data centres and publishers. The major force behind identifying the data-literature connections is emerging new funders' policies that encourage (in some cases enforce) reproducibility of science. In this talk, we will present the Scholix (Scholarly Link exchange) framework as a high-level interoperability approach toward exchanging information about the links between scholarly literature and data. This work was initiated by the Research Data Alliance working on Publishing Data Services.

Over the past decade, publishers and data centres, have agreed on and implemented numerous bilateral agreements to establish bidirectional links between research data and the scholarly literature. However, because of the considerable differences inherent to these many agreements, there is very limited interoperability between the various solutions. This talk will present the vision of a universal interlinking service and proposes the technical guidelines of a multi-hub interoperability framework.

## Conference Themes:

- Supporting Open Scholarship, Open Data, and Open Science
- Managing Research Data, Software, and Workflows
- Integrating with the Wider Web and External Systems


## Keywords

scholarly communication, data article interlinking, interoperability, open science, research data


## Audience

Data managers, librarians, repository managers and research administrators, project leaders

## Background

At the time of writing this paper, for most repositories, there is no trivial solution to establish the connections between articles and the data that supports the findings of the article. For some time, publishers and data centres have established bilateral agreements to connect their research data and the scholarly literature. However, these bilateral agreements between two (or at most few) organisations limits the availability of this information to open access repositories. For example, the data-literature linking between PANGAEA, a Data Publisher for Earth and Environmental Science, and selected Elsevier journals are not easily available to DSPACE repositories. In addition to being bilateral, such linking lacks an industrial standard.

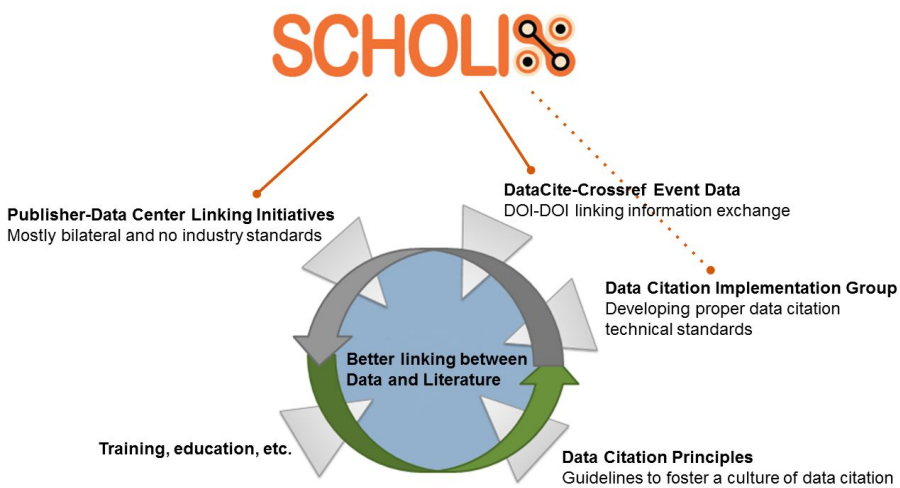



**Figure 1:** Scholix: An overarching framework for existing technical initiatives that individually address parts of the overall problem that is hindering better linking between data and the literature.

Scholix (Scholarly Link exchange) is a community and multi-stakeholder driven effort to address the current unsatisfactory situation in data-literature information exchange. Scholix is a high-level interoperability framework that facilitates the exchange of information about the links between data

and scholarly literature, as well as between data. The framework proposes an overall cohesive vision and approach to bring together the existing initiatives. At present, such aspirations are hampered by the lack of agreed ways of expressing links between data and literature and common ways of exposing those links to a potentially comprehensive global information system, as known for literature to literature links. The Scholix framework targets the latter by establishing an agreed interoperability framework and thus creating a pull factor to incentivise the former.

## Presentation content

Scholix is not a product, service, or infrastructure. It is a conceptual framework for interoperability. Scholix maintains an evolving lightweight set of guidelines rather than a normative standard. The framework is developed on consensus from various stakeholder groups in the research data landscape, including data centres, publishers, Crossref, DataCite, OpenAIRE (Manghi et al. 2012), and many others. The complete list of participants are available at http://www.scholix.org/about.

These stakeholders are the main hubs (Figure 2) of literature-data and data-data link information. In principal, hubs are (existing) services that collect and aggregate information about links from their respective communities. Scholix proposes a standard information exchange model between these hubs and any third party data provider or consumer.

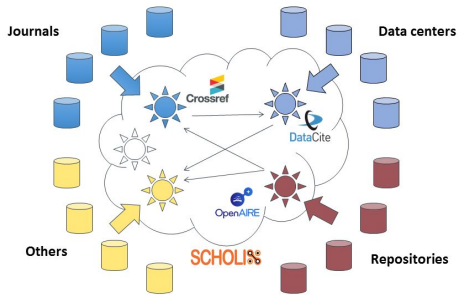
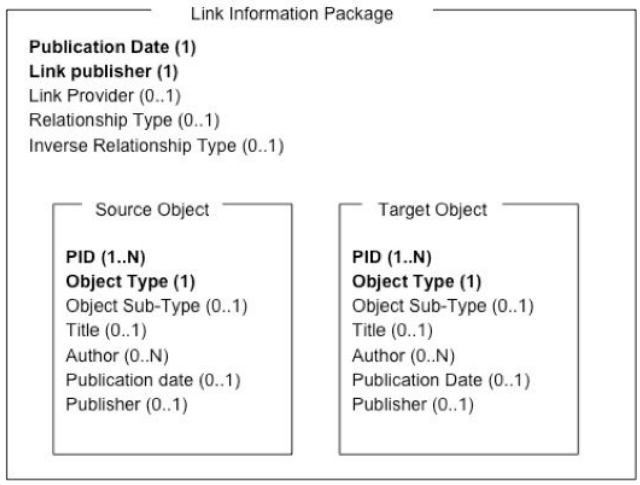



**Figure 2:** Schematic overview of the proposed multi-hub interoperability framework.

Interoperability among the hubs is achieved by a common information model and open exchange methods described earlier. This exchange of information does not affect existing community specific practices. There is no requirement for "many to many" interoperability between data centres, journals and repositories. In contrast to the uniformity among the hubs, heterogeneity is a natural part of the outer layer and is addressed by the hubs with their communities.




The information model of SCHOLIX is quite abstract and simple to adopt. It is tailored toward adaptability and it can support multi-discipline information exchange without affecting the

domain-specific metadata.

The three main elements of this model are the `link` and the objects involved, named `source` and `target` to enable determining the link's direction. The `link publisher` is the party that makes and publishes the assertion that two research objects are linked (such as a journal publisher).

The Scholix interoperability framework does not mandate how to format and exchange information. Data can be formatted and exchanged using a range of models and protocols such as JSON, XML, or RDF formats and RESTful, OAI-PMH or SPARQL protocols.

The partners of this project are in the progress of integrating the SCHOLIX framework in the following information hubs: Crossref and DataCite Event Data (eventdata.datacite.org), ANDS Research Data Australia (researchdata.ands.org.au), Research Data Switchboard and the Graph Graph clusters (researchgraph.org). The DLI service (Burton et al. 2015) is today accessible in BETA[1], features more than 7M links. The authors will provide an update on the progress of this initiative at the Open Repository conference.

## Conclusion

Scholix is a proposed conceptual framework to drive interoperability between providers of links between research data and the literature. Currently, the framework is a set of informal guidelines and high level models. Nevertheless, it represents a significant first step toward a shared vision, standardisation, and the realisation of alluring benefits of a global information commons around data and literature. A further ICSU-WDS / RDA Working Group is being organised to elaborate the Scholix framework, coordinate the development of hubs, and support community adoption. The working group will also work actively with: service providers for benefit realisation; hubs and their communities for buy-in; and with international advocacy and peak bodies for culture change.

## References

**[Burton et al. 2015]** Burton, A., Koers, H., Manghi, P., La Bruzzo, S., Aryani, A., Diepenbroek, M., & Schindler, U. (2015, September). On bridging data centers and publishers: the data-literature interlinking service. In Research Conference on Metadata and Semantics Research (pp. 324-335). Springer International Publishing.

**[Manghi et al. 2012]** Manghi, P., Bolikowski, L., Manola, N., Schirrwagen, J., and Smith, T. (2012). OpenAIREplus: the european scholarly communication data infrastructure. D-Lib Magazine, 18(9). https://doi.org/10.1045/september2012-manghi

[1] http://dliservice.research-infrastructures.eu