Introduction to Scholix – Infrastructure to Collect and Report Data Citations
# The State of Data Citation

<table>
<thead>
<tr>
<th>Policy</th>
<th>Data as first-class scholarly output</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>Support for data citation workflow</td>
<td>Evolving</td>
</tr>
<tr>
<td>Bibliometrics</td>
<td>Understand the impact of data citations</td>
<td>Beginning</td>
</tr>
<tr>
<td>Culture Change</td>
<td>Changes in researcher behaviour</td>
<td>Very early</td>
</tr>
</tbody>
</table>
Data Citation Workflow

Journal Data Policies
RDA Interest Group

Finding appropriate repository
re3data
FAIRsharing
AGU FAIR Data
FAIRsFAIR

Coordinated submission workflow
THOR
AGU FAIR Data

Linking data and researcher
THOR
FREYA

Understand data citations and credit
Make Data Count

Aggregate and show data usage and citations
Make Data Count

Collect and distribute data citations
RDA Scholix Event Data

Publish article and data
DCIP
Data Citation Workflow

Journal Data Policies
RDA Interest Group

Finding appropriate repository
re3data
FAIRsharing
AGU FAIR Data
FAIRsFAIR

Coordinated submission workflow
THOR
AGU FAIR Data

Linking data and researcher
THOR
FREYA

Understand data citations and credit
Make Data Count

Aggregate and show data usage and citations
Make Data Count

Collect and distribute data citations
RDA Scholix
Event Data

Publish article and data
DCIP
RDA/WDS Scholarly Link Exchange Working Group
What is the problem?

1. Many disconnected sources (publishers, data centers, repositories, infrastructure providers, …)
2. Heterogeneity of practices, for example:
   • Different PID systems (DOI, accession numbers)
   • Different ways of referencing data (formal citations, in-text references, …)
   • Different moments of citing data (at publication, post publication, …)
Scholix is a schema

- A schema for standardizing the exchange of scholarly link information between scholarly infrastructure providers
  - Information Model for scholarly links representation
  - Recommendation and provision of exchange formats and protocols

See also http://www.scholix.org/guidelines
Scholix is a community

• Scholix helps with the information exchange between infrastructure providers.
• Scholix does not change how data repositories working with DataCite or publishers working with Crossref provide data citations via DOI metadata.
• The Scholix community encourages data repositories and publishers to submit data citations to DataCite and Crossref.
Recap Scholix: connecting the dots

**Past:** disconnected sources using heterogeneity of practices

**Future:** standard set of guidelines for exposing and consuming links, supported by hubs
Scholix benefits
see: www.Scholix.org

For data repositories and journal publishers
• increase their visibility and usage
• improve the user experience
• More scalable and robust

For research institutes, bibliographic service providers, and funding bodies
• Make data count
• track datasets and publications within common and comprehensive framework

For researchers:
• Easier finding and accessing
• track long-term impact of their data
• additional incentives to share data.
Scholix Timeline

2014
RDA-WDS Data Publishing Services WG

2015

2016
Version 1.0 Scholix Interoperability Framework

2017

Version 2.0 Draft Schema

2018

2019
Version 3.0 of the Scholix schema; in production
http://doi.org/10.5281/zenodo.1120265
## Information model: properties

### Link Information Package

- Link Publication Date (1)
- Link Provider (1..N)
- Relationship Type (1)
- License URL (0..1)

<table>
<thead>
<tr>
<th>Source Object</th>
<th>Target Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Identifier (1)</td>
<td>Object Identifier (1)</td>
</tr>
<tr>
<td>Object Type (1)</td>
<td>Object Type (1)</td>
</tr>
<tr>
<td>Object Title (0..1)</td>
<td>Object Title (0..1)</td>
</tr>
<tr>
<td>Object Publisher (0..1)</td>
<td>Object Publisher (0..1)</td>
</tr>
<tr>
<td>Object Creator (0..N)</td>
<td>Object Creator (0..N)</td>
</tr>
<tr>
<td>Object Publication Date (0..1)</td>
<td>Object Publication Date (0..1)</td>
</tr>
</tbody>
</table>
Scholix in practice: the Hubs

Event Data
https://support.datacite.org/docs/eventdata-guide

Scholexplorer
https://scholexplorer.openaire.eu/
The joint Crossref/DataCite Event Data service holds the following information regarding literature/data links as of 17 June 2019:

- Crossref -> DataCite: 5295
- DataCite -> Crossref: 1342975
- DataCite -> DataCite: 44396
- IsSupplementTo: 760081
- IsCitedBy: 7891
- references: 585223
- isReferencedBy: 9901
- Circles not at scale.
Next steps

- Further adoption with focus on publishers
- Support and training
- Coordination between hubs
- User interfaces
### User Interfaces

Display of usage stats in DataCite Search, work on display of citations has started.

https://search.datacite.org/works/10.7272/q63r0qr4
Display of citations and other connections in PID Graph using a Jupyter notebook and the DataCite GraphQL API.

https://github.com/datacite/notebooks/blob/master/pid-graph/r-grant-publications.ipynb