



The Scholix Framework for Interoperability in Data-Literature Information Exchange

*Adrian Burton, Hylke Koers, Paolo Manghi,
Martin Fenner, Sandro La Bruzzo, Amir Aryani,
Michael Diepenbroek, Uwe Schindler, Markus Stocker*

research data sharing without barriers
rd-alliance.org

Linking Datasets and the Literature: why? ²

Linking Research Data with the Literature is of great value, yet current solutions are not realizing the potential

Why linking?

1. Increase visibility & discoverability of research data (and articles)
2. Support credit attribution mechanisms
3. Facilitate reproducibility, i.e. place research data in the right context to enable proper re-use.

Examples

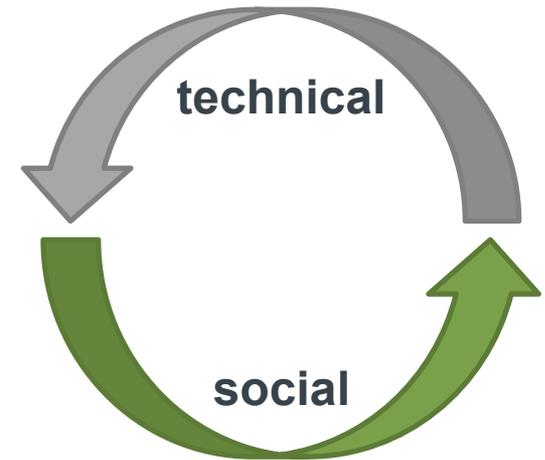
- Some data repositories keep track of articles that cite, or refer to, their data
- Some publishers have applications to link articles with data hosted externally
- Providers of bibliographic information and infrastructure providers are taking efforts to “connect the dots”

So.. what's the problem?

Linking Dataset with the Literature is of great value, yet current solutions are not realizing the potential

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, ...)



There's several initiatives that are addressing (parts of) the problem in different ways...

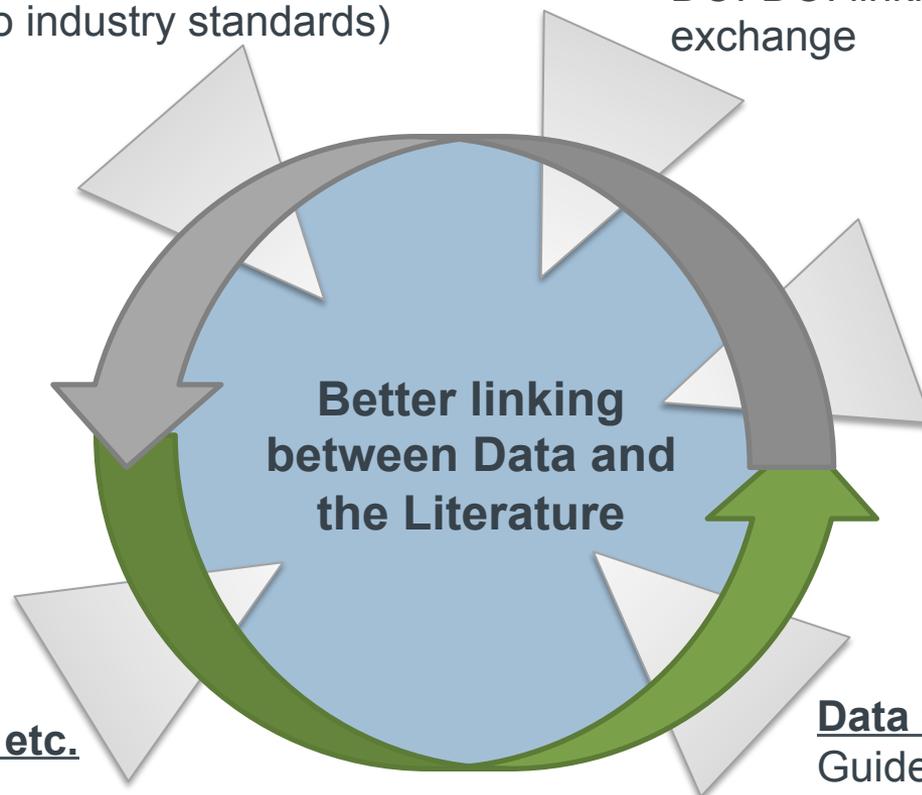
Publisher – Data Center linking initiatives
(Mostly bilateral, no industry standards)

DataCite-CrossRef eventdata:
DOI-DOI linking information exchange

Data Citation Implementation Group:
Developing proper data citation technical standards

Training, education, etc.

Data Citation Principles:
Guidelines to foster a culture of data citation

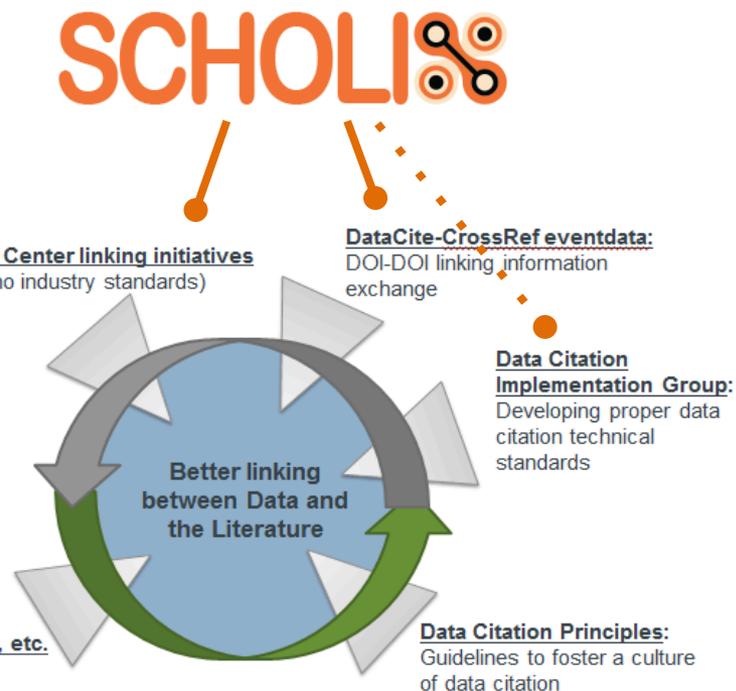


... what is lacking is an overall, cohesive vision and approach to bring these together

Scholix: a high-level interoperability framework for enabling exchange of scholarly links between literature and data

Connecting technical linking solutions by providing an overarching framework with practical guidelines, underpinned by a common vision

- An enabler of services e.g.
 - Global aggregation
 - Impact evaluation



See also <http://www.scholix.org/about>

Scholix is a framework supporting a vision enabling exchange of scholarly links between data and literature using a common, global approach

- Maintains an evolving lightweight set of Guidelines to increase interoperability rather than a normative standard.
- Based on consensus from various stakeholder groups in the research data landscape
 - data centers, publishers, Crossref, DataCite, OpenAIRE, many others
- A product of the ICSU-WDS / RDA working group “Data Publication Services” - endorsed by ICSU-WDS, RDA endorsement pending

Scholix is NOT a product, service, or infrastructure



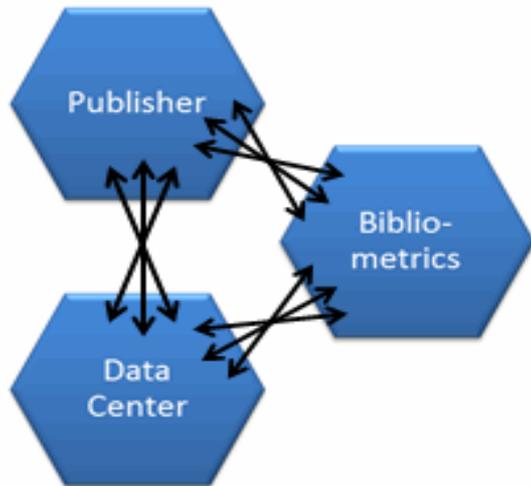
See also <http://www.scholix.org/about>

research data sharing without barriers
rd-alliance.org

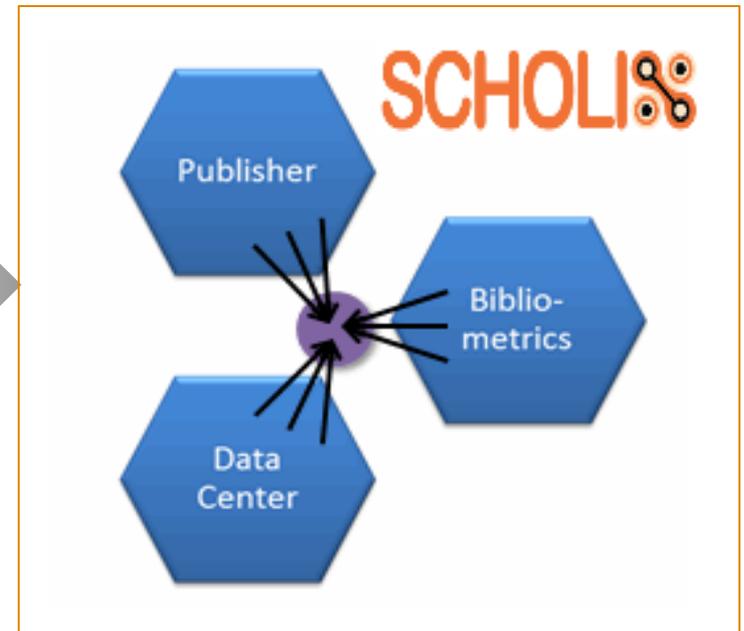


Objective: move from **to**

a plethora of (mostly) bilateral arrangements between the different players...



.. a one-for-all cross-referencing framework for articles and data



Scholix Guidelines

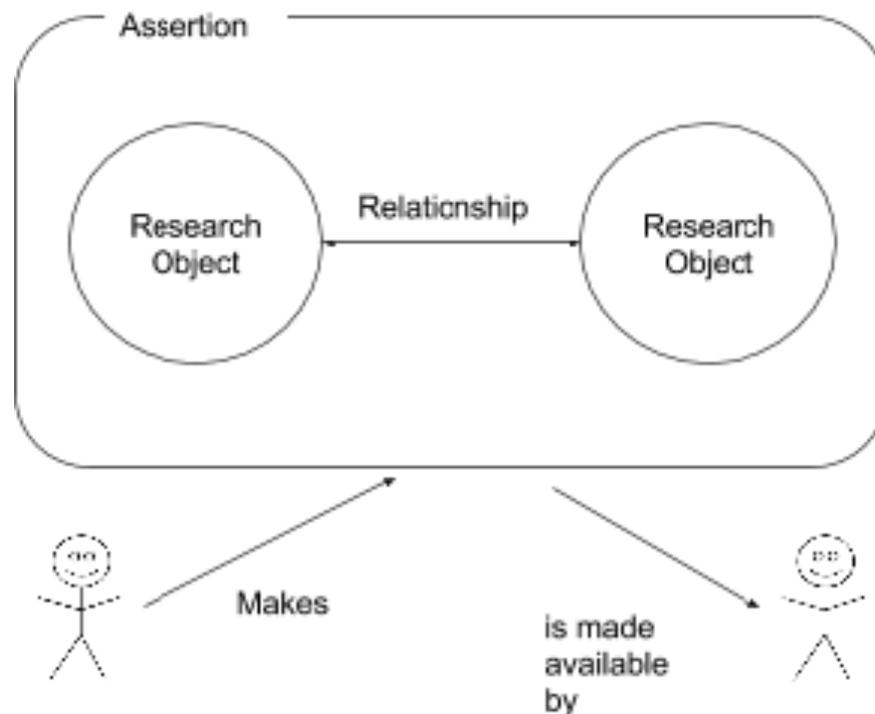
- A framework for standardizing the exchange of *scholarly link* information between scholarly infrastructure providers
 - Information Model for scholarly links representation
 - Recommendation and provision of exchange formats and protocols



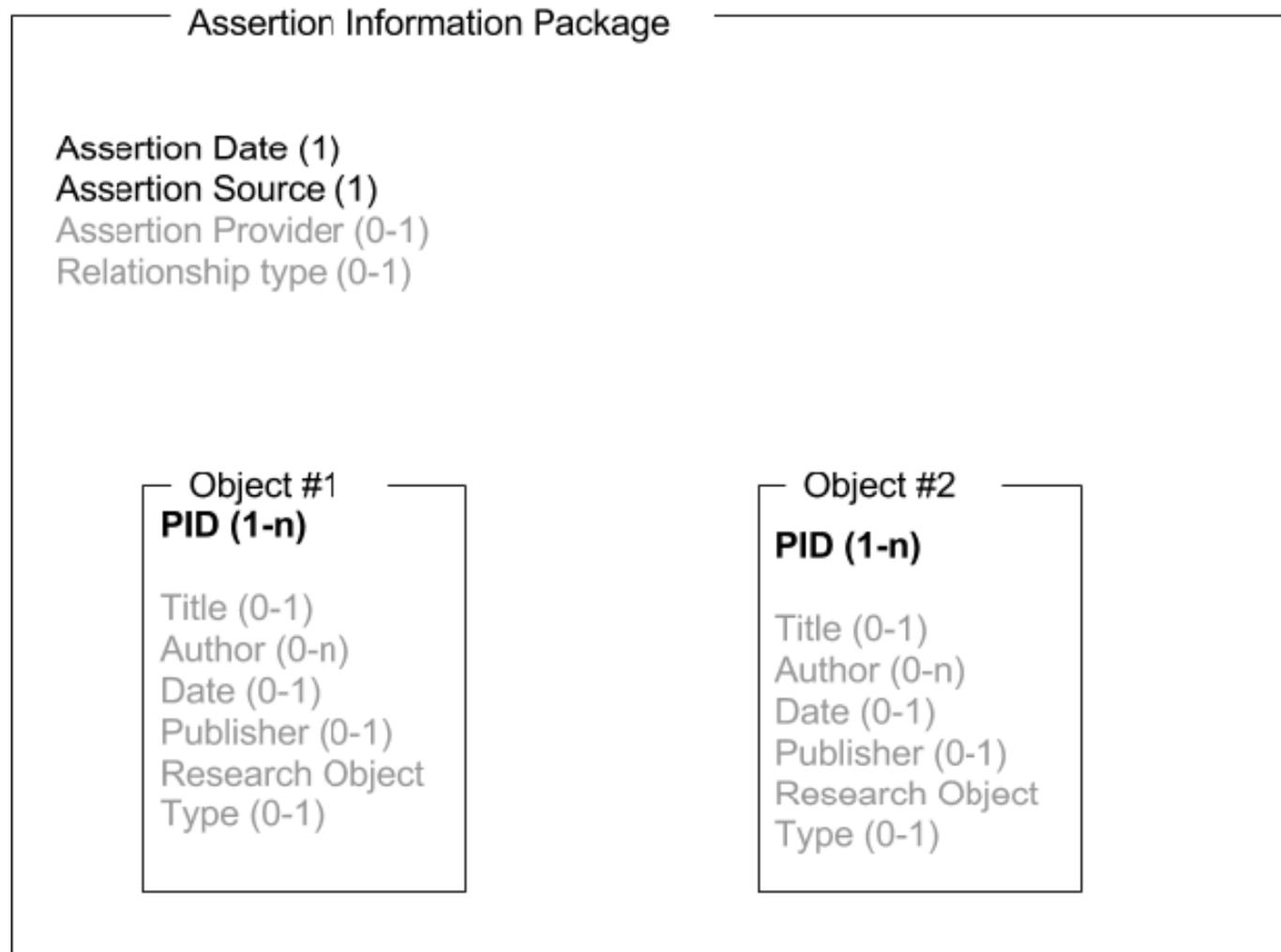
See also <http://www.scholix.org/guidelines>

Information model 1/3

- Research objects include data and literature (and other types out of scope for this particular framework)
- Research objects have relationships with other research objects
- Parties in the research system assert that two research objects are related
- Parties in the research system collect and/or make those assertions available



Information model 2/3



Information model 3/3

Information Element	Examples of potentially applicable standards
Assertion Date Research Object Date	ISO8601, XML Schema "dateTime" format, W3CDTF, EDTF - to be explored
Relationship Type	DataCite Metadata Kernel: Relationship
Research Object Type	Citation Styles - Types? - to be explored further.
Assertion Source Assertion Provider Research Object Publisher	ISNI, Ringgold, Digital Science GRID, PROV? - to be explored further.
Author	ORCID

- **Formats**

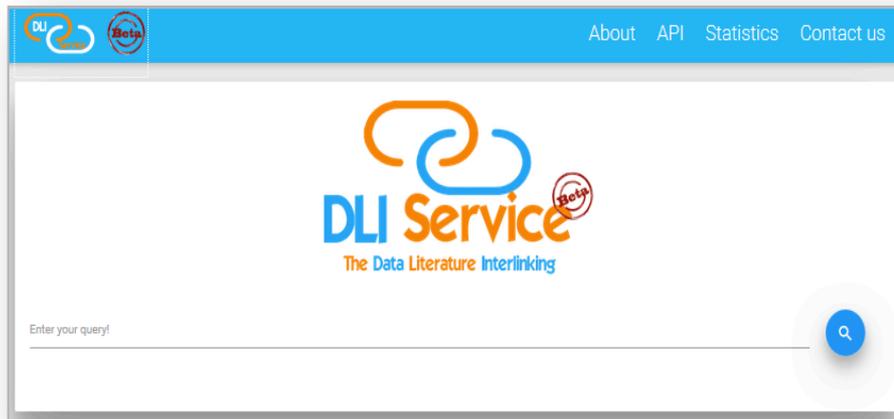
- Profiles of the information model based on XML, JSON, DISCO-RDF, etc.

- **Protocols:**

- OAI-PMH, Open API, etc.

Organizations are already starting to develop services that follow the Scholix framework:

1. [DataCite Event Data](#)
2. [Crossref Event Data](#) and [Linked Clinical Trials](#)
3. OpenAIRE and PANGAEA Data-Literature Interlinking ([DLI](#)) Service

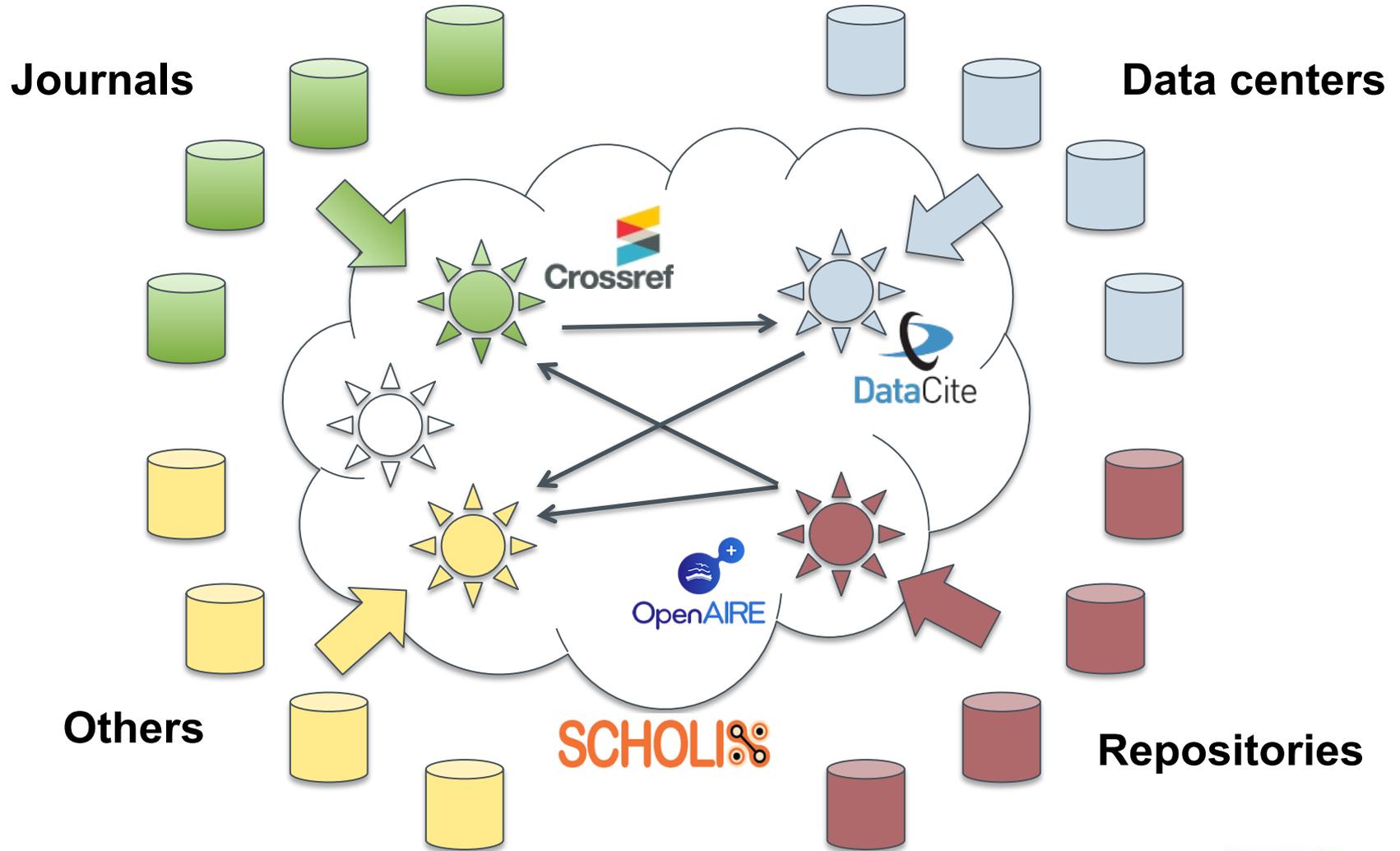


DLI: a prototype / demonstrator service developed by OpenAIRE with support from PANGAEA and the Data Publishing Services WG.

Give it a spin:

<http://dliservice.research-infrastructures.eu>

Multi-Hub Vision



How to join in?

Option 1: Feed data-literature info to an existing Scholix hub

Extend your existing metadata feed to e.g.,

- DataCite
- CrossRef
- OpenAIRE

Option 2: Become a hub

- Expose a new feed using the [Scholix Guidelines](#)
- Register as an [implementor](#) to be included in the DLI aggregation and other hub exchanges

How to join in?

Option 3: Develop a third-party service

- Use the information from the Scholix hubs in your own service

Option 4: Help develop the guidelines

- Help expand and document the [Scholix Guidelines](#) or to take them into new areas.

Questions?

Email info@scholix.org
Join the the RDA-WDS [Working Group!](#)

- *Web references*
 - Scholix web site: <http://www.scholix.org/>
 - Data-Literature Interlinking Service:
<http://dliservice.research-infrastructures.eu>

Assertion metadata

- **Assertion** about the relationship (link) between two research objects, i.e. literature or datasets
 - Workflows, software, etc. are not in the scope of the framework
- **Assertion source (mandatory)**, i.e. the party that makes the assertion that two research objects are related (e.g. a journal publisher), as well as the date on which the assertion was made
- **Assertion provider (optional, highly desirable)**, i.e. the party that aggregates information from sources (e.g. Crossref) and makes information available - The provider should be asserted if it is different from the source
- **Relationship type (optional, highly desirable)**, e.g. isCitedBy or isDerivedFrom

- **PID (optional)**, Resource objects should be identified by resolvable and globally unique persistent identifiers (PID), in which case information about research objects should be obtained by resolving the PID
- **Title, Author, Date, Publisher, and Type of research object (optional, but highly desirable if PID is missing)**

- Another WG is working on the creation of profiles of the information model appropriate to a number of protocols and formats, e.g. OAI-PMH, with XML, Open API with JSON, DISCO RDF, Linkback, etc.