RDA-WDS Scholarly Link Exchange WG - WDS Working Group Charter

Background

The Working Group (WG) follows on from the RDA/WDS Publishing Data Services WG. The original working group established demonstrator services enabling infrastructure. The follow on working group will support the "hardening" of that infrastructure and services as well as an increase in the number of participating hubs and services. The original working group established an interoperability framework. The follow on group will provide further specification, documentation and profiling of that framework to support adoption by link contributors and consumers. The original working group established a consensus among large infrastructure providers and early adopters; the follow up group will extend that consensus to the next stage of adopters and to a more diverse set of infrastructure providers. The original working group harnessed the energy and interest of specialists; the follow up group will provide support for a number of communities and services as they implement and adopt the framework and vision established in the original group.

www.scholix.org

https://www.icsu-wds.org/community/working-groups/data-publication/services

https://www.rd-alliance.org/groups/rdawds-publishing-data-services-wg.html

Terms of Reference

The WG will enables a comprehensive global view of the links between scholarly literature and data. It will leverage existing work and international initiatives to work towards a global information common. To achieve this, the WG will have multiple roles:

- Oversee and guide the technical development of Hubs and Services as well as elaborate the Scholix framework
- Create coordination and governance materials
- Assist with the engagement and community buy-in and adoption.
- Communicate broadly on the initiative

Deliverables and Milestones

The Scholarly Link Exchange Working group aims to enable a comprehensive global view of the links between scholarly literature and data. The working group will leverage existing work and international initiatives to work towards a global information common by establishing:
- Pathfinder services and enabling infrastructure
- An interoperability framework with guidelines and standards (see also www.scholix.org)
- A significant consensus
- Support for communities of practice and implementation

- Better understanding of current data-literature interlinking landscape viewed from the perspective of e.g. disciplines, publishers, repositories etc.

Schedule

By the end of this 18 month WG period there will be:
- A critical mass of Scholix conformant hubs providing the enabling infrastructure for a global view of data-literature links
- Pathfinder services providing aggregations, query services, and analyses
- Beneficiaries of these services accessing data-literature link information to add value to scholarly journal sites, data centre portals, research impact services, research discovery services, research management software, etc.
- Operational workflows to populate the infrastructure with data-literature links

Membership

The Chair/Co-chairs and members of the WG are appointed by the WDS-SC. Its membership will be reviewed every two years by the WDS-SC during one its biannual face-to-face meetings.

Co-chairs:

Initial workstream leads and **Co-chairs:**
1. Technical
    - technical specs and docs (**Paolo Manghi**)
    - hub development and interoperability (**Martin Fenner**)
    - Scholix service development (Jeff Grethe)
2. Community
    - publisher (Iain Hrynaszkiewicz)
    - repository (Ian Bruno)
    - general outreach (Fiona Murphy)
3. WG Coordination
    - WG program oversight (**Wouter Haak**)
    - WG component integration (**Adrian Burton**)

Members:

Initial members are coming from the existing working group on an opt-out basis; they will be asked again if they want to join this newly formed working group by e-mail following the RDA Plenary 8th

- https://rd-alliance.org/groups/rdawds-publishing-data-services-wg.html

Adrian Burton	George Mbevi	Kathrin Beck	Paul Dlug
Amir Aryani	Håkan Grudd	Kerstin Helbig	Peter Rose
Amye Kenall	Haralambos Marmanis	Kerstin Lehnert*	Peter Fox
Aris Gkoulalas-Divanis	Howard Ratner	Lars Vilhuber	Rabia Khan
Arnold Rots	Hua Xu	Laura Rueda*	Rainer Stotzka
Arthur Smith	Hylke Koers	Laurel Haak	Richard Kidd
Bernard Avril	Iain Hrynaszkiewicz	Leonardo Candela	Rick Johnson
Carly Strasser	Ian Bruno*	Luiz Olavo Bonino da Silva Santos	Robert Arko
Carole Goble	Ingrid Dillo*	Lyubomir Penev	Rorie Edmunds*
Caroline Martin	Jamus Collier	Mark Donoghue	Sarah Callaghan*
Claire Austin	Jeffrey Grethe	Martin Fenner*	Sheila Morrissey
Claudio Atzori	Jingbo Wang	Martina Stockhause*	Siddeswara Guru
Dan Valen	Jo McEntyre	Michael Diepenbroek*	Simon Hodson*
David Martinsen	Joachim Wackerow	Mohan Ramamurthy	Suenje Dallmeier-Tiessen
David Arctur	Johanna Schwarz	Mustapha Mokrane*	Tim DiLauro
Donatella Castelli	John Helly	Natalia Manola	Timea Biro
Eefke Smit*	Jonathan Tedds	Niclas Jareborg	Tom Demeranville
Elise Dunham	Juanle Wang*	Nigel Robinson	Ui Ikeuchi
Elizabeth Moss	Kate Roberts	Paolo Manghi	William Mischo
Francis ANDRE	Katerina Iatropoulou	Patricia Cruse*	Wouter Haak*
			Xiaoli Chen
			Yolanda Meleco

* Representatives of a WDS member

Finances

Participation of members of the WDS-WG is voluntary. Administrative support is provided by the WDS-IPO. This includes organizing online meetings (GoToMeeting) and providing an online collaborative space. The WDS-IPO cannot provide financial support for travelling nor for any other costs. Face-to-face meetings will be organized as the opportunity arises at international conferences, and participants will be expected to fund their travel and expenses through their own organizations. Additional grants might be secured by the WDS-IPO or its partners to support the activities of the WG.

Procedures

The Chair/Co-chairs will be responsible for the agenda of WG meetings agenda and for calling those meetings. They will report the WDS-SC Executive Committee regularly; at least twice a year on the occasion of the biannual WDS-SC face-to-face meetings.

Term and Review

This WG activity and its continued operation will be reviewed in Oct 2018
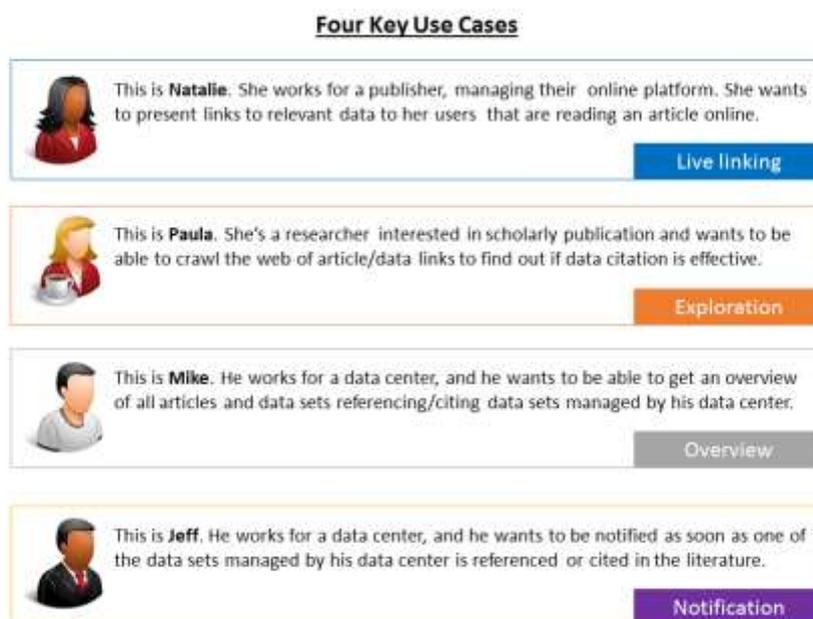
ANNEX

Value Proposition

The WG aims to oversee and guide the maturation of a distributed global system to collect, normalize, aggregate, store, and share links between research data and the literature. This will build upon the output of the preceding Data Publishing Services Working Group, which delivered a consensus vision and set of guidelines called the Scholix Framework, together with an operational system called the Data-Literature Interlinking (DLI) System, which puts these guidelines into practice as a pathfinder implementation. The WG proposed here will build out these assets into an operational infrastructure and service layer that is to become the de facto go-to place for organizations to deposit or retrieve links between research data and the literature.

The value of such a system ultimately rests on the value of links between research data and the literature. The utility of such links is threefold (see also the Case Statement of the Data Publishing Services WG):
1. They improve the <u>visibility and discoverability</u> of research data (and relevant literature), so that researchers can find relevant material more easily.

2. They help place research data in the right context, so that researchers can re-use data more effectively.
3. They support credit attribution mechanisms, which incentivize researchers to share their data in the first place.

These value elements are illustrated below, and in more detail in Annex A.

Four Key Use Cases

This is **Natalie**. She works for a publisher, managing their online platform. She wants to present links to relevant data to her users that are reading an article online.

Live linking

This is **Paula**. She's a researcher interested in scholarly publication and wants to be able to crawl the web of article/data links to find out if data citation is effective.

Exploration

This is **Mike**. He works for a data center, and he wants to be able to get an overview of all articles and data sets referencing/citing data sets managed by his data center.

Overview

This is **Jeff**. He works for a data center, and he wants to be notified as soon as one of the data sets managed by his data center is referenced or cited in the literature.

Notification

While there is broad support for the value and utility of data-literature links amongst the various stakeholders in research data publishing (including researchers as the ultimate end-users of this information), organizing the associated information space is not an easy feat: there are many disconnected sources with overlapping information, and there is a wide heterogeneity in practices today - both at a technical level (different PID systems, storage systems, etc.) and at a social level (different ways of referencing a data set in the literature, different moments in time to assert a link, etc.). As a consequence, the landscape today is incomplete and patchy, characterized by independent, many-to-many non-standard solutions - for example a bilateral arrangement between a journal publisher and a data center. This is both inefficient and limiting in the value that can be delivered to researchers.

The universal linking infrastructure which this WG strives to put in place represents a systemic change. It will offer an overarching, cohesive structure that binds together many of today's practices into a common interoperability framework - which will ensure that links between research data and the literature can be easily shared, aggregated, and used on a global scale. This will drive a network effect, where the value in the system as a whole is greater than the sum of individual parts: for researchers as end-users, this value lies in the comprehensiveness and quality of link information; for service providers and infrastructure providers (including journal publishers and data centers), the value also lies in simplicity, efficiency, and reduction of friction in the process by being able to work with a single interface to deposit and retrieve links (and, potentially, the possibility to benefit from additional services developed on top of the core infrastructure).

Who will benefit and Impact

Mapping the value proposition as described in the above to the various stakeholders and actors in research data publishing (copied largely from the Data Publishing Services WG Summary & Recommendations), benefits and impact may be summarized as follows:

- For **data repositories** and **journal publishers**: linking data and the literature will increase their visibility and usage, and can support additional services to improve the user experience on online platforms (for example, offering links to relevant data sets with articles, or offering links to the literature that will help place data in context). In contrast to the bilateral arrangements that we often see today between data centers and journal publishers, the global linking infrastructure will make the process of linking data sets and research literature a more robust, comprehensive, and scalable enterprise.
- For **research institutes, bibliographic service providers, and funding bodies**: the infrastructure will enable advanced bibliographic services and productivity assessment tools that track datasets and journal publications within a common and comprehensive framework.
- For **researchers**: firstly, the infrastructure will make the processes of finding and accessing relevant articles and data sets easier and more effective. Secondly it will make it possible for researchers to track long-term impact of their data (and publications), thereby providing additional incentives to share data.

Engagement with existing work in the area:

1. Building upon previous work of the RDA/WDS Publishing Data Services WG
2. RDA/WDS Publishing Data IG
    a. RDA/WDS Publishing Data Bibliometrics WG
    b. RDA/WDS Publishing Data Workflows WG
3. Infrastructure providers
    a. Crossref: http://eventdata.crossref.org/guide/
    b. DataCite: https://blog.datacite.org/its-all-about-relations/
    c. ICSU WDS: https://www.icsu-wds.org/
    d. ANDS: http://rd-switchboard.net/
    e. Pangaea: https://pangaea.de/
4. Infrastructure projects
    a. OpenAIRE: https://www.openaire.eu/
    b. THOR: https://project-thor.eu/
5. Related projects
    a. RMAP: http://rmap-project.info/rmap/
    b. Making Data Count (MDC): http://mdc.lagotto.io/
    c. Force11 Data Citation Implementation Pilot (DCIP): https://www.force11.org/group/dcip
    d. FAIR data (http://www.nature.com/articles/sdata201618)
6. Data Center Community

        a. ICSU WDS
        b. DataCite
7. Publisher Community
        a. Crossref
        b. STM Association: http://www.stm-assoc.org/
        c. CHORUS
8. Institutional Repository Community
        a. OpenAIRE
        b. SHARE
9. Discipline-specific Communities
        a. Pangaea (Earth and Environmental Science)
        b. EBI-EMBL (Life Sciences)
        c. ICPSR (Social Sciences)
        d. CERN (High Energy Physics)

Adoption Plan:

The Adoption Plan for this Working Group is quite mature since it builds on a previous working group, includes adopter work packages, includes outreach and documentation work packages, targets new hubs, and focuses on benefit realisation.

Previous Working Group: The proposed working group builds directly on the Data Publishing Services Working Group which has a considerable membership with an active core of contributors. The WG is representative of publishers, data centres, research organisations and research information infrastructure services who are the key stakeholder and adopter communities. The existing momentum and buy-in of this group will be leveraged for adoption.

Technical Development of Hubs: In a similar vein, the WG activity plan includes targeted activity to extend existing hubs (CrossRef, DataCite, OpenAire, RMap) and establish new hubs in new community areas (such as Astronomy, Life Sciences).

Implementation Sub Projects: The working group case statement "Activities" section provides details of a number of adoption sub projects. The Scholix framework that underpins the WG approach involves content publishers (eg journal publishers or data centres) communicating with natural hubs (eg CrossRef and DataCite). This WG activity plan includes implementation projects from publisher to hub.

Documentation and Support Materials: The WG activity plan includes an extension of the Scholix framework by providing documentation of instantiation of the abstract Scholix information model in various technologies or formats (such as xml, rdf, json) and using a number of common protocols (such as open api calls, sparql, oai-pmh, resourceSync). These specification and implementation materials will also be the product of the development and adoption projects described above.

Outreach, Liaison, Collaboration: This Working Group focuses on a technical solution to the exchange and aggregation of data-literature link information. Other peak bodies and

advocacy groups focus on changing practice and integrating data citation as part of scientific practice. The WG work plan includes collaboration with those organisations to leverage their established agendas. Current members of the WG include leaders in these organisations and further such activity is slated in that area of the work plan.

Benefit realisation: The sustainable driver of adoption is benefit for the adopter. The overall work plan is underpinned by the objective of delivering benefits to end users, as outlined in the use cases of the Annex A.

Work Plan:

The work plan will be implemented through a set of interconnected activities outlined below. Categories exist only for planning and pragmatic purposes; they are not at all independent and activities will not be siloed. Cross-category contributions by working group members will be the norm.

Stream 1.Technical Development.

The objective of this stream to put the Scholix framework into practice such that both hubs and services develop operational functionality.

A. Develop Hubs
1. OpenAIRE
   - Make OpenAIRE APIs compatible with Scholix to export and import links to and from DLI Service
2. DataCite
   - Further develop standardised interfaces for query and export
3. CrossRef
   - Further develop standardised interfaces for query and export
4. New domain-specific hubs, e.g. EMBL/EBI (TBC by opportunity)
5. Interim hubs (direct feed to DLI): standardisation (using Scholix framework) of feeds from previous working group and improvement of dynamic currency of feeds
   - ANDS to DLI direct (only non-DOI content)
   - ...
6. Further interoperation of the hubs (extensions to the Scholix conceptual framework during the course of the working group)

B. Develop Services (in relation to the user scenarios defined in previous WG)
1. DLI aggregation service https://dliservice.research-infrastructures.eu/#/
   a. Transition to production at OpenAIRE data centre and infrastructure
   b. APIs for PID resolution (Scholix conformant) - Pangaea
   c. Improving quality: e.g. de-duplication of objects (datasets and literature)
   d. Improving service level: live updates of links
2. Use of the Scholix framework to access and expose links between articles and data in exemplar end-user services
   a. OpenAIRE APIs compatible with Scholix to export and import links to and from DLI Service
   b. Data centre/ publisher exemplar projects using DLI as per user scenarios

C. Elaborate the Scholix framework
1. Create profiles of the inf model for use in different technologies
   a. XML for oai-pmh
   b. JSON for RESTful api
2. Investigate how best to apply
   a. DISCO (through cooperation with RMAP)
   b. ResourceSync
   c. Others?…(RDF for Sparql)
3. Provide documentation and support materials for the above

2. Community buy-in stream

This stream supports buy-in from different communities such that exchange of scholarly link information is implemented and accepted as standard practice.

D. Support Community Adoption:
1. Create strategies for community adoption:
   a. Publishers
   b. Data centres
   c. Repositories
   d. ….
2. Implement these strategies through:
   ■ Early adopter groups (eg CrossRef early adopters; e.g. Force11 DCIP project; eg via the THOR project; with COAR)
   ■ Implementation projects
   ■ Webinars
   ■ Presentations
   ■ Support materials and activities

E. Communicate Broadly
1. Create communications plans
2. Implement communications plans

F. Create Coordination and Governance Materials. Investigate and document issues such as:
   ○ Quality of data links
   ○ Requirements to be a hub
   ○ Access
   ○ Benefits for contributors
   ○ Measures of success

Key Stakeholder Groups
The above Activity Plan will be delivered with involvement of the following groups who bring complementary resources, approaches, focus, and expertise.
   A. Advocacy and Peak Bodies
   ● Force11 (application data citation standards & advise on implementation standards)
   ● CODATA (application data citation standards & advise on implementation standards)

- ICSU World Data System (e.g. get more citations into DataCite)
- STM (outreach, training, Crossref early adopter project)
- ESIP / COPDESS
- FAIR Data

B. Other data literature linkage projects
- National Data Service
- RMAP (application of DISCO)
- SHARE
- RDA Working Groups (Publishing Data IG, ….)

C. Prospective Hubs
- BIOCaddie (DataMed)
- EMBL-EBI/ELIXIR
- NASA ADS