Scholarly Link Exchange
Scholix
RDA 10th plenary Montreal

research data sharing without barriers
rd-alliance.org

Agenda

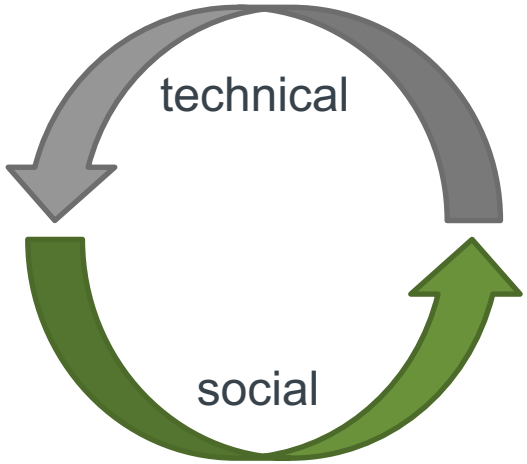
- Quick recap of Scholix
- Learning through real implementations
  10x examples & feedback
- Community materials ‘how to’
- Finalizing the guidelines: the last 10% takes 90% of our time
- Joint assessment: where do we stand

Recap: what's the problem?

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

What is the problem?

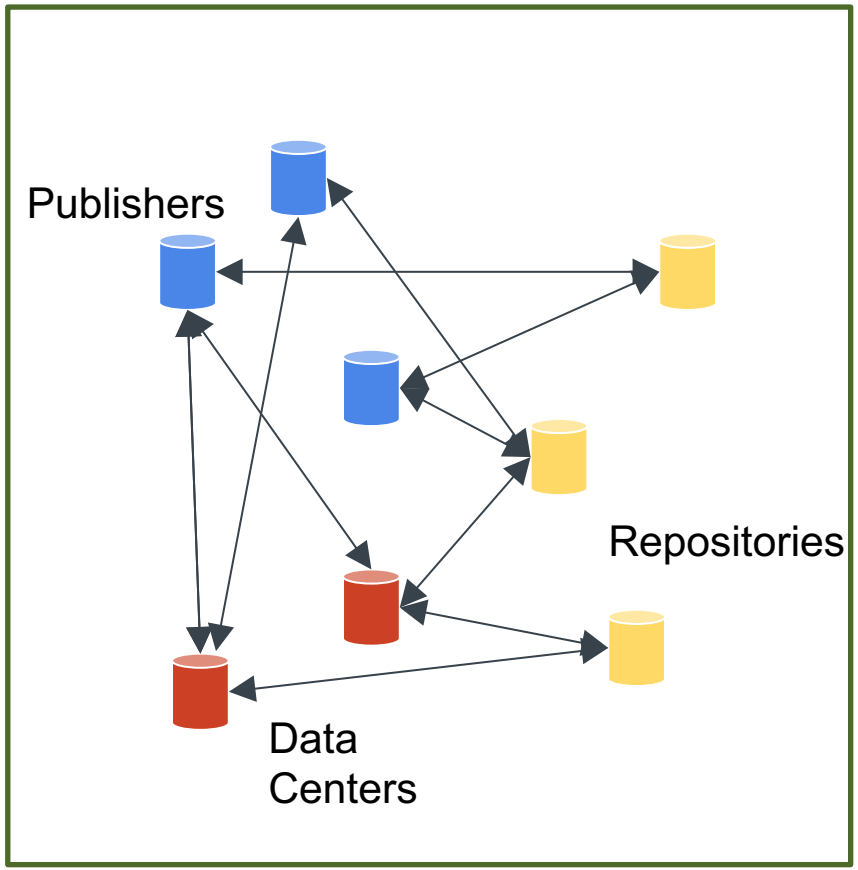
1. Many disconnected sources (publishers, data centers, repositories, infrastructure providers, …)
2. Heterogeneity of practices, for example:
   - Different PID systems (DOI, accession numbers)
   - Different ways of referencing data (formal citations, in-text references, …)
   - Different moments of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

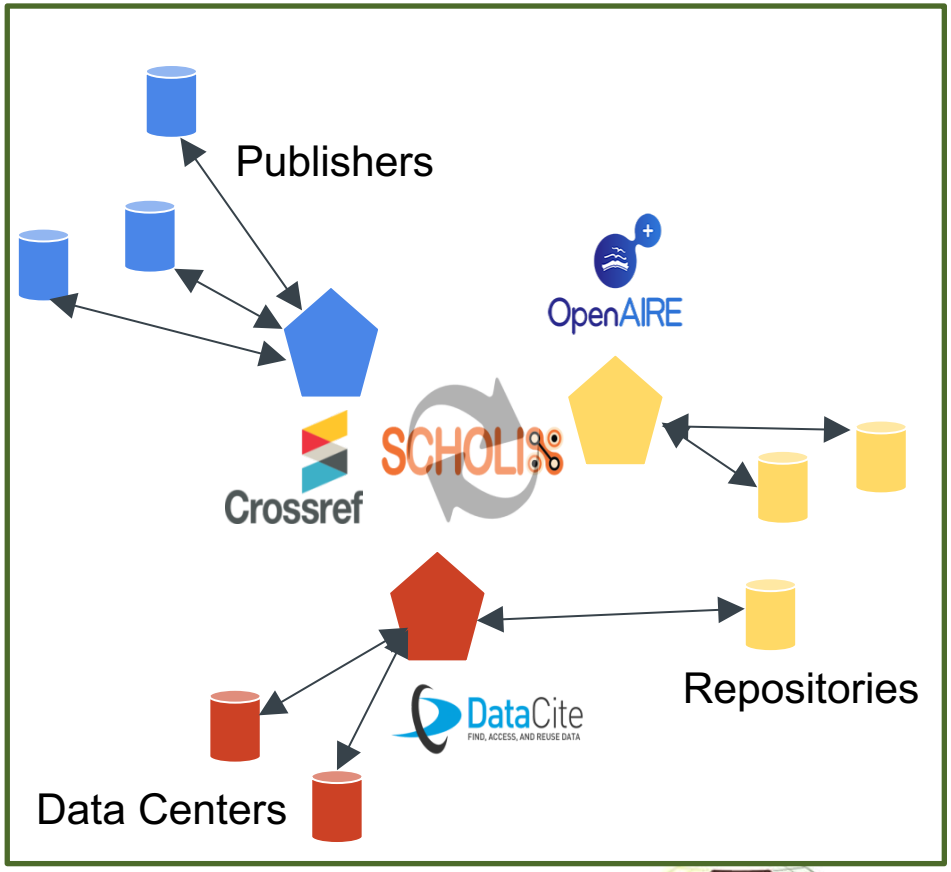


Recap Scholix: connecting the dots

Past: disconnected sources using heterogeneity of practices

Future: standard set of guidelines for exposing and consuming links, supported by hubs





Recap: 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

SCHOLIX

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

Recap of Scholix benefits see: www.Scholix.org

For data repositories and **journal publishers**

- linking data and the literature will increase their visibility and usage
- can support additional services to improve the user experience on online platforms
- More scalable and robust due to a global standard rather than bilateral agreements

For **research institutes, bibliographic service providers, and funding bodies**

- the infrastructure will enable advanced bibliographic services and productivity assessment tools
- track datasets and journal publications within a common and comprehensive framework

For **researchers**:

- Easier finding and accessing relevant articles and data sets
- track long-term impact of their data (and publications)
- thereby providing additional incentives to share data.

Example: Scholix at Europe PMC

Florian Graef

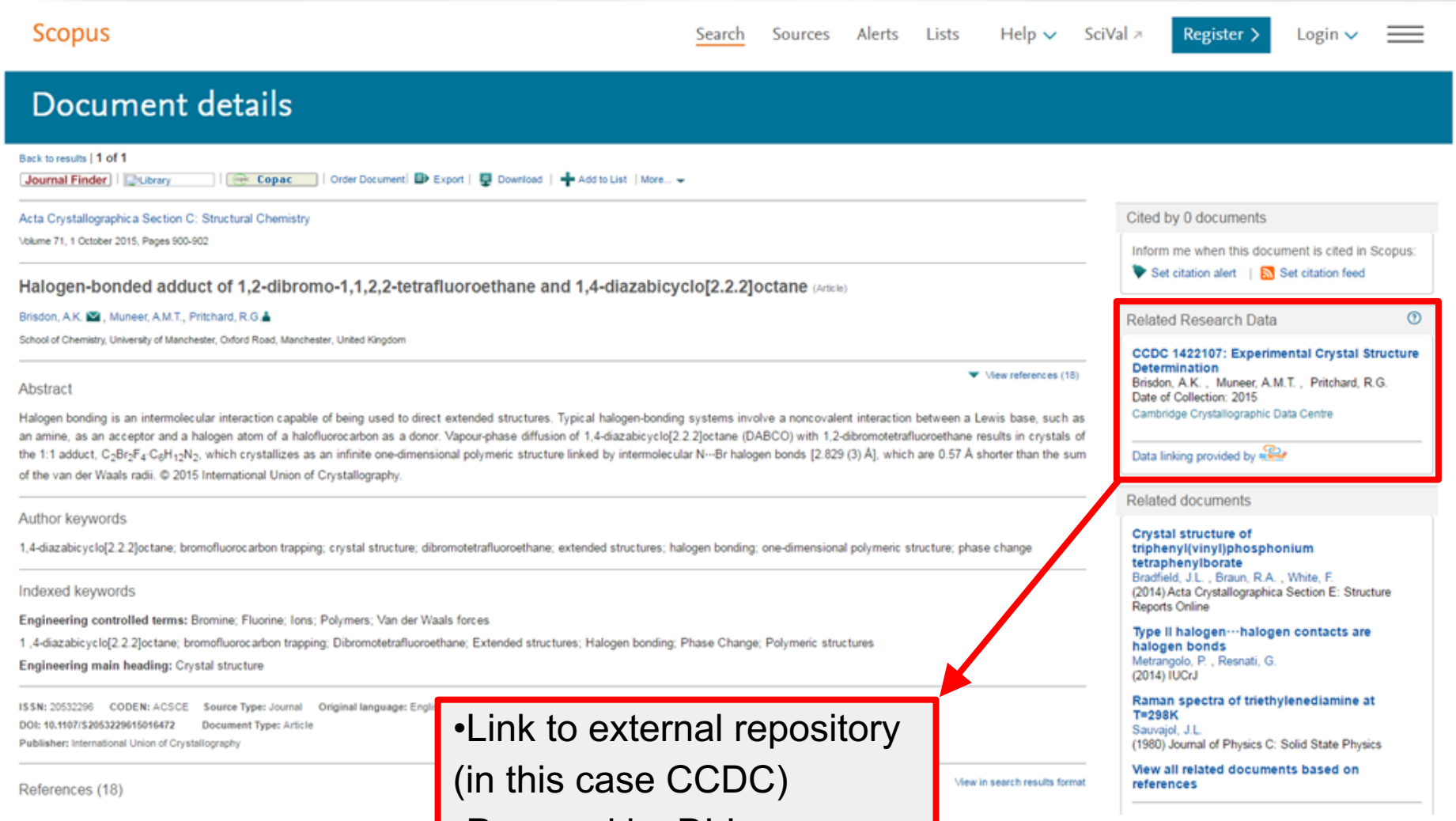
- Different origin Data-Literature links in various places in API and User Interface
- Consolidation into one API method providing links in Scholix format

EuropePMC obtains data-literature links in a few ways:
-DB-Crossreferences (external data records cite a publication and tell us about it (e.g. PDBe)
-Text mined accessions – extracted by our text mining pipeline in publications (PDBe, ENA)
-External links – various entities provide us with links to resources which are related to individual publications (e.g. Altmetrics, Wikipedia, Publons)
There are differences in directionality/ the way we obtain them but all are data-literature links
-> Consolidation of the API to provide all data in one response (Scholix format)
-> Will gradually replace current API methods starting with a single Tab summing up data cited/produced in a study
Internal testing is ongoing

Questions?
graf@ebi.ac.uk

Example: Scopus.com (articles/citations) => dataset

Eleonora Presani



- Link to external repository (in this case CCDC)
- Powered by DLI

Document details

Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms
Volume 346, 1 March 2015, Pages 26-44

Activation cross-sections of proton induced reactions on ^{nat}Sm up to 65 MeV (Article)

Tárkányi, F.[a], Hermanne, A.[b], Takács, S.[a], Ditrói, F.[a], Ignatyuk, A.V.[c]

[a]Institute for Nuclear Research, Hungarian Academy of Sciences (ATOMKI), Debrecen, Hungary
[b]Cyclotron Laboratory, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, Brussels, Belgium
[c]Institute of Physics and Power Engineering (IPPE), Obninsk, Russian Federation

Abstract

View references (24)

Activation cross sections for proton induced reactions on Sm are presented for the first time for $^{nat}Sm(p,xn)^{154,152m2,152m1,152g,150m,150g,149,148,147,146,145}Eu$, $^{nat}Sm(p,x)^{153,145}Sm$, $^{nat}Sm(p,x)^{151,150,149,148g,148m,146,144,143}Pm$ and $^{nat}Sm(p,x)^{141}Nd$ up to 65 MeV. The cross sections were measured via activation method by using a stacked-foil irradiation technique and high resolution gamma ray spectroscopy. The results were compared with results of the nuclear reaction codes ALICE, EMPIRE and TALYS (results taken from TENDL libraries). Integral yields of the activation products were calculated from the excitation functions. © 2015 Elsevier B.V.

Author keywords

Cross-section measurement | Europium samarium promethium and neodymium radio-isotopes | Proton irradiation | Samarium target | Yield calculation

Indexed keywords

Engineering controlled terms: Chemical activation | Protons | Samarium

Activation method
Activation products
Cross-section measurement
Excitation function
Gamma-ray spectroscopy
High resolution | Integral yields
Proton induced reaction

Engineering main heading: Nuclear reactions

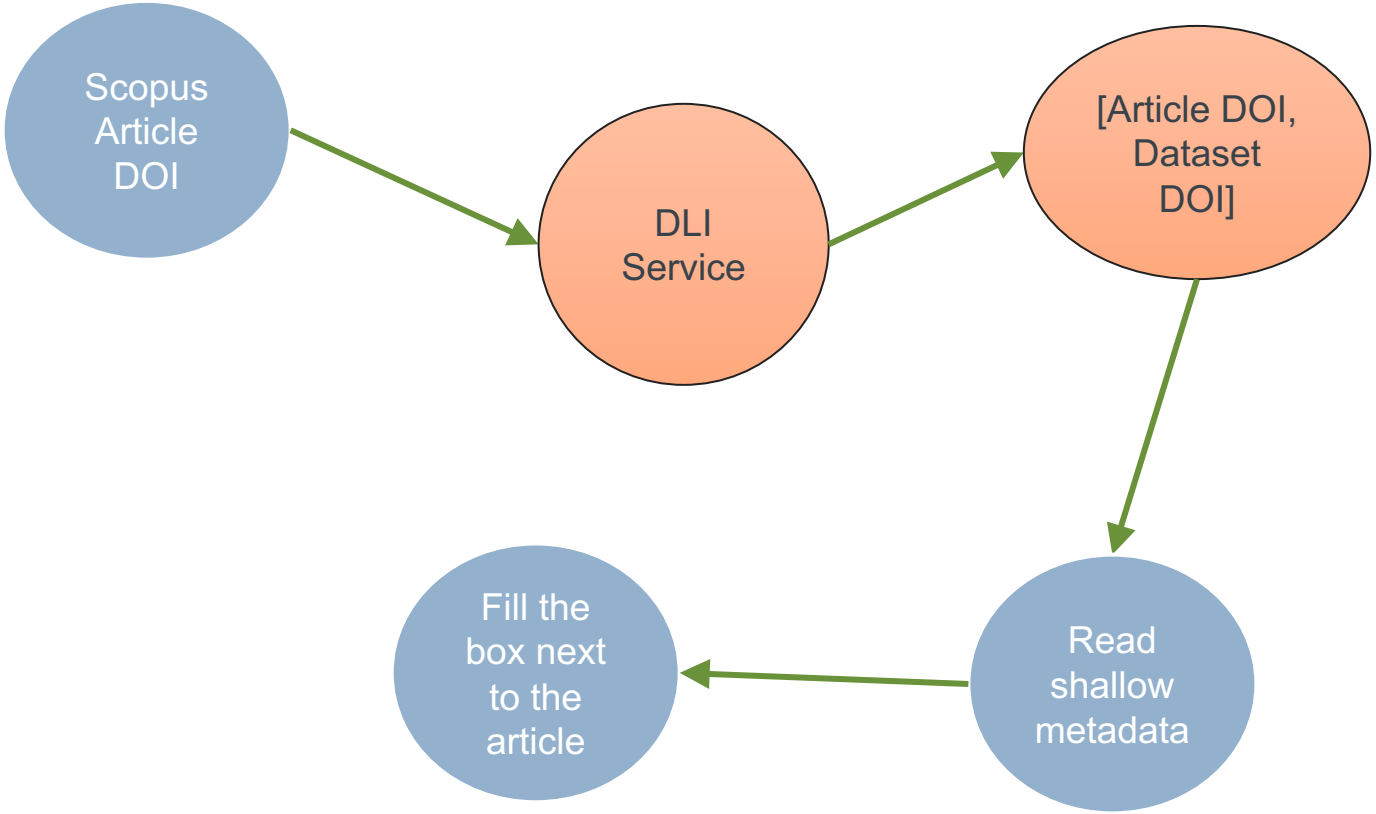
- Example2: Scopus => Mendeley Data repository
- also powered by DLI

- https://www.scopus.com/record/display.uri?eid=2-s2.0-84922265195&origin=inward

Scopus and DLI Service within Scholix

Eleonora Presani

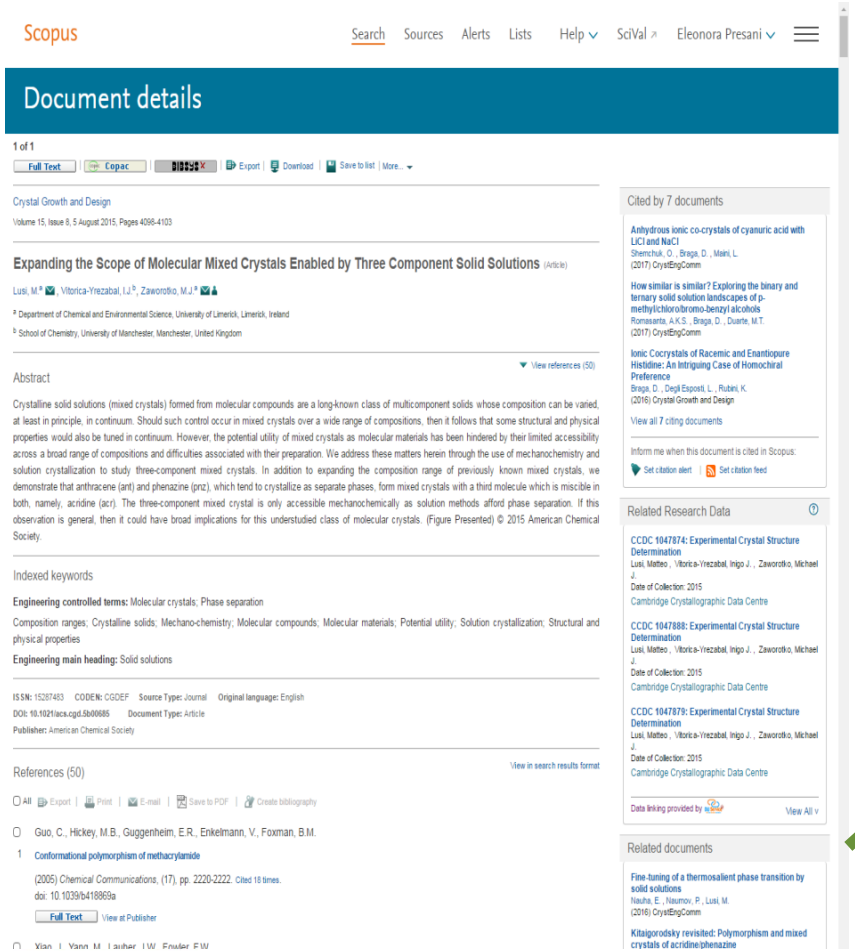
- For each document entry in Scopus (with DOI) DLI receives a query

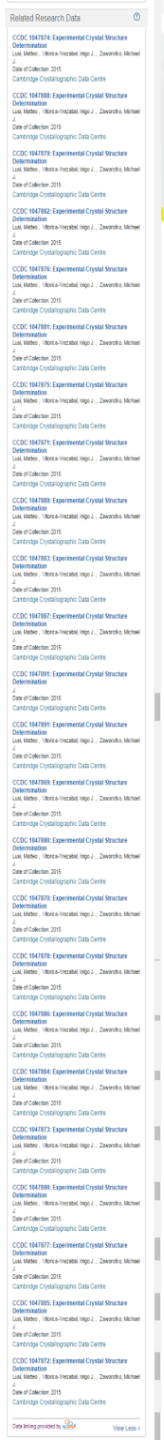


Scopus example: many datasets

Eleonora Presani

- Scopus shows only the first three datasets, and then the user can expand the view
- https://www.scopus.com/record/display.uri?eid=2-s2.0-84938632232&origin=inward





RESEARCH DATA ALLIANCE

Scopus suggestions for Scholix & DLI

Eleonora Presani

- **A clearer information on where the data is actually stored (URL resolution)**
- **A recipe to retrieve deeper metadata from the data repository**
  - Dataset description / abstract
  - Version
  - References (if any)
  - Keywords
  - Data type
  - ….
- **Maybe contributing repositories can offer a key to retrieve those from them?**


```
- {
    facet: "Content Provider",
    query_name: "provenance",
  - values: [
    - {
        key: "CrossRef",
        value: 1
      },
    - {
        key: "Mendeley Data and published articles",
        value: 1
      }
    ]
  }
,
```

Example: Dryad participation in the Scholix initiative

Elizabeth Hull & Fiona Murphy

Dryad's vision is to promote a world where research data is openly available, integrated with the scholarly literature, and routinely re-used to create knowledge.

- As a repository specifically for data associated with publications, recognize the value of a common, global approach to these linkages, and pleased to be an early implementer
- Launched in July 2017 and applied to all existing data packages - **18,000 article links and counting**

Work in progress / for discussion:

- Traditional use of / preference for "IsReferencedBy" field for article DOIs over "IsSupplementTo," but currently reporting to DataCite in both fields

SCHOLIX

research data sharing without barriers
rd-alliance.org

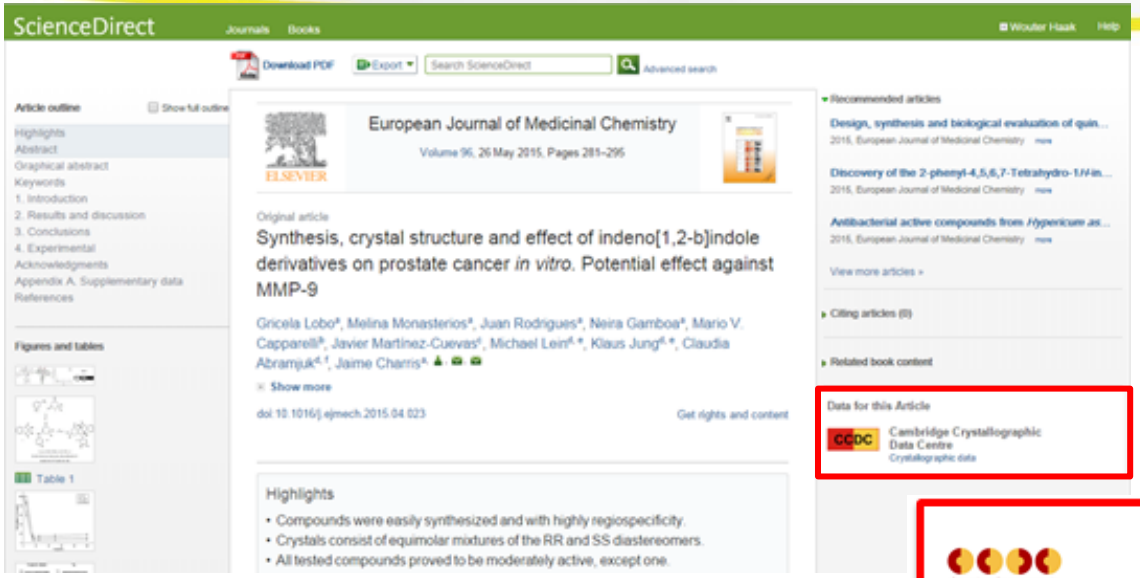
DRYAD

http://datadryad.org

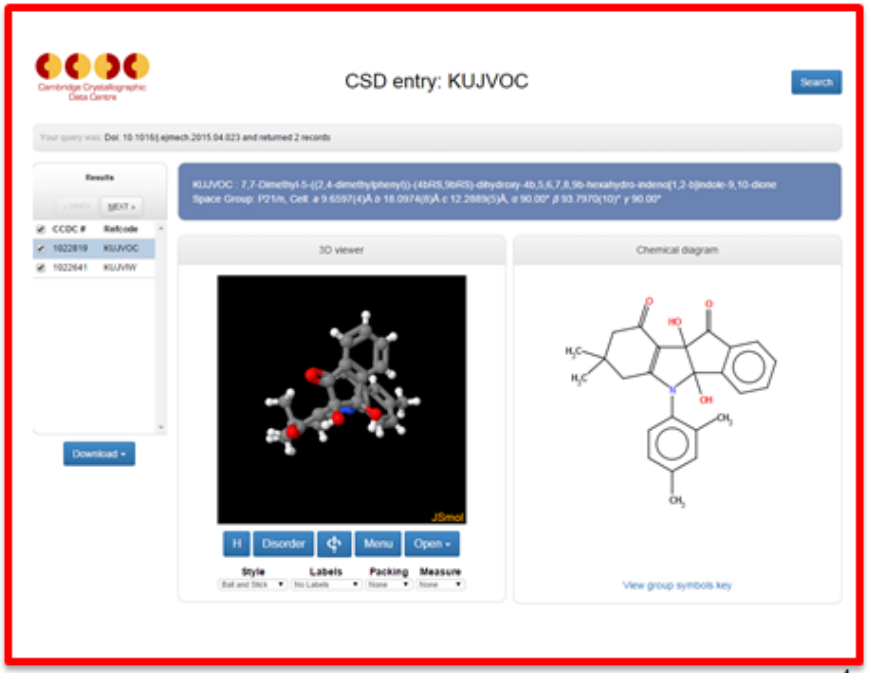
RDA
RESEARCH DATA ALLIANCE

ScienceDirect/CCDC Example

Ian Bruno, Helena Cousijn



Currently linking between article and dataset is facilitated by bespoke Elsevier and CCDC services



Plan is to pilot linking using the DLI service

- Concerns: timeliness of link being available relative to article being published - more systems need to update compared to current mechanisms
- Benefit: CCDC will no longer need to maintain services provided to specifically support Elsevier linking

ScienceDirect/CCDC Example

Ian Bruno, Helena Cousijn

Query:

https://api-dliservice-prototype-dli.d4science.org/v1/linksFromPid?pid=10.1016/j.poly.2007.03.016&pidType=doi

Response:

{"schema":"doi","identifier":"http://doi.org/10.5517/ccnc0qq","repoAcronym":"CCDC","publisher":"Cambridge Crystallographic Data Center","title":"CCDC 606413: Experimental Crystal Structure Determination","repoDescr":"Crystallographic data"}],"timestamp":1503930494052}

Extract:

'Publisher' , 'title', 'identifier' and combine with 'repository name', 'descriptor' and logo from existing database

Display:

Hub example: DataCite

Martin Fenner

- DataCite is a non-profit organization based in Germany with currently 60 member organizations across the globe
- DataCite provides persistent identifier services to its users, focussing on DOIs and associated metadata for research data
- DataCite services link research data to journal articles, software, people, funding and samples and helps locate, identify, and cite research data

Hub example: Data Literature Interlinking Service (ScholExplorer) *- Paolo Manghi*

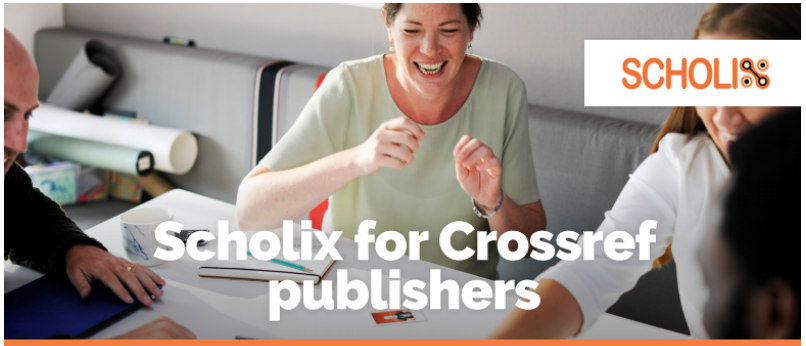
- **Beta system** at https://dliservice-prototype-dli.d4science.org
- Content from publishers and data archives
  - Datasets+Pubs: 4,200,000
  - Dataset-Pub links: 38,000,000
- In sync with DataCite every 5-6 hours
- Scholix compatible APIs to resolve DOIs
  - Scopus is a consumer

- **Production system** @OpenAIRE Nov 2017

How to take part

- Community materials ‘how to’
- Scholix for repository managers (Natasha Simons)
- Scholix for DataCite members (Catherine Brady)
- Scholix for CrossRef members (Joe Wass)
- Frequently Asked Questions (FAQs) (Natasha Simons)
- Scholix Scopus story (Natasha Simons & Eleonora Presani)

How to take part



HOW LEODEN IS USING SCHOLIX TO ENHANCE DATA-LITERATURE LINKS

Leoden is a publisher member of Crossref. The member publishes journals and registers their content with Crossref.

Crossref, as one of the Scholix hubs, assists in the global aggregation of literature-data link information for their publisher members. Leoden already provides a wealth of metadata through the DOI registration process. Upon hearing about the Scholix initiative, she decides to include links to data as part of the citation metadata records she provides to Crossref. In doing this, Leoden makes sure that globally unique persistent identifiers - DOIs wherever possible - are used for both the dataset and piece of literature. As a result, the data-literature links provided by Leoden are made available via

HOW ANNE IS USING SCHOLIX TO FACILITATE RESEARCH DISCOVERY

In doing this, Anne makes sure that globally unique persistent identifiers - DOIs wherever possible - are used for both the dataset and piece of literature. As a result, the data-literature links provided by Anne are made available via Scholix aggregators such as the DLI service.

Publishers and service providers such as Scopus can query the DLI Scholix aggregation to find datasets related to journal articles they hold. Articles on the journal website can now display a link back to the datasets held in Anne's data centre (via the DOI link), driving web traffic to the data centre and facilitating research discovery.

HOW PEDRO IS USING SCHOLIX TO MAKE DATASETS MORE VISIBLE

Pedro manages the institutional research data repository at a university. The repository holds and provides access to datasets from his institution. Information about journal publications related to each dataset are included in the repository metadata records. However, when Pedro visits the journal websites there are no links back to the datasets held in his repository.

Pedro decides to take advantage of the Scholix interoperability framework to get the message out about the literature that is related to data in his repository.

Pedro already provides descriptions of his data through to the international OpenAIRE aggregation. He now simply includes any links to literature in the metadata records harvested by OpenAire (making sure that persistent identifiers - DOIs wherever possible - are used for both the dataset and piece of literature). OpenAire, as one of the Scholix hubs, assists in the global aggregation of data-literature link information. The data-literature links from Pedro's repository are now available in Scholix aggregators such as the DLI service.

Publishers and service providers such as Scopus can then query the DLI Scholix aggregation to find datasets related to journal articles they hold. Articles on the journal website now display a link back to the datasets held in Pedro's repository (for example via the DOI link), driving web traffic to the repository and facilitating research discovery.

Next steps - guidelines

- Guidelines. Almost final, but there are still loose ends. Focus was on testing it (e.g. DLI, Scopus, CCDC, Pangaea).
- We will now put it into a final version. Uncertainties remaining:
  - Scope: focus on data-article right now. Software, labs, and other entities more in the future. Agreed to stay open for these kind of relations
  - “Cited by” and “Supplement to” are enough as relations. We use the subset of the DataCite list for now. Out of scope right now is the standardized terms.
  - “Link Provider” = place where you got it (and -tbd- not the original source)
  - “Publisher” = mandatory (but discussion still open)
  - Does a PID have to be URL? (e.g. does the http part need to be part of it; does the means of resolving it be separate from the PID) => ask advice from the PID group
- Work on documenting schema, support materials and validation tool will kick off after the Barcelona meeting

Timeline: where do we stand

RDA-WDS Data Publishing Services WG

Deliverables

- Inventory of interlinking, cross-referencing, and other tools and processes relevant to data publication currently in place.
- An analysis of pros and cons, with an emphasis on scalability and doability.
- Gap analysis, including an analysis of needs & use cases for key stakeholders (data repositories, journal publishers, providers of bibliographic services, funding bodies, research institutions, researchers)
- Recommendations for a one-to-all cross-resolving service. These recommendations will include technical, organizational, governance, and cost aspects
- An operational and publicly available service for cross-referencing datasets and articles

RDA-WDS Scholix WG

Deliverables:

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

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

Thank you!

Questions?

Email info@scholix.org

RDA WG site & Scholix web site: http://www.scholix.org/

Co-chairs:

- Adrian Burton adrian.burton@ands.org.au
- Wouter Haak W.Haak@Elsevier.com
- Paolo Manghi paolo.manghi@isti.cnr.it
- Martin Fenner Fenner@datacite.org