

Paolo Manghi
Institute of Information Science and Technologies (CNR)

Iryna Kuchma
EIFL

scholexplorer.openaire.eu

SCHOLIX



ScholarXplorer
The Data Literature Interlinking Service



 Search this site

- Home
- Participate
- Scholix schema
- Implementors
- About
- FAQ

Scholix: A Framework for Scholarly Link eXchange

The goal of the Scholix initiative is to establish a high level interoperability framework for exchanging information about the links between scholarly literature and data. It aims to enable an open information ecosystem to understand systematically what data underpins literature and what literature references data. [Scholexplorer](#) is the first exemplar aggregation and query service fed by the Scholix open information ecosystem. The Scholix framework together with the Scholexplorer aggregation are designed to enable other 3rd party services (domain-specific aggregations, integrations with other global services, discovery tools, impact assessments etc).

Scholix is an evolving lightweight set of guidelines to increase interoperability. It consists of:

- A consensus among a growing group of publishers, datacentres, and global/ domain service providers to work collaboratively and systematically to improve exchange of data-literature link information
- Information model: conceptual definition of what is a Scholix scholarly link
- Link metadata schema: metadata representation of a Scholix link
- Options for exchange protocols (forthcoming)

Scholix is the “wholesaler to wholesaler” exchange framework, to be implemented by existing hubs or global aggregators of data-literature link information such as DataCite, CrossRef, OpenAIRE, or EMBL-EBI. These hubs in turn work with their natural communities of data centres or literature publishers to collect the information through existing community-specific workflows and standards. Scholix thus enables interoperability between a smaller number of large hubs and leverages the existing exchange arrangements between those hubs and their natural communities (eg between CrossRef and journal publishers).

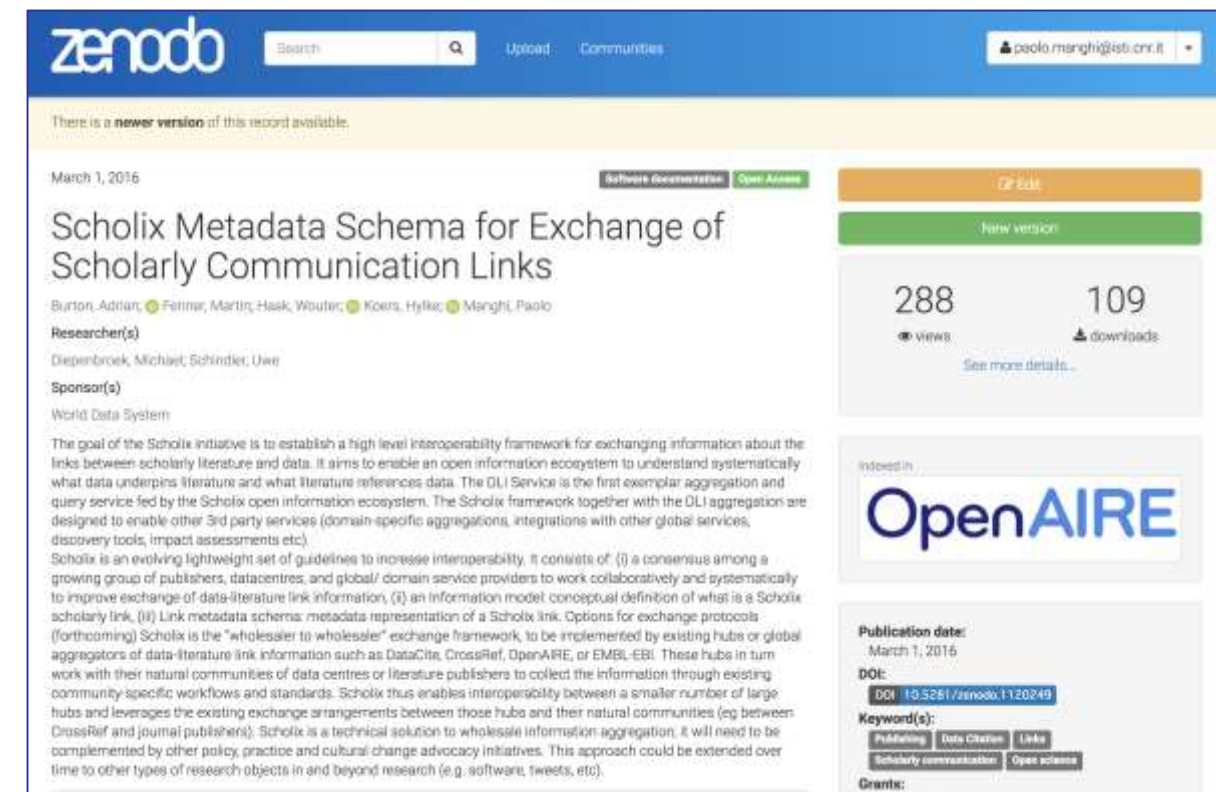
Scholix is a technical solution to wholesale information aggregation; it will need to be complemented by other policy, practice and cultural change advocacy initiatives. This approach could be extended over time to other types of research objects in and beyond research (e.g. software, tweets, etc).

Scholix is supported by the following organizations:



SCHOLIX RDA/WDS Working Group

- 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 (JSON/XML-based)



The screenshot shows a Zenodo record page for the 'Scholix Metadata Schema for Exchange of Scholarly Communication Links'. The page includes a search bar, user profile, and a notification about a newer version. The main content area displays the title, authors (Burton, Adrian; Ferrer, Martin; Hask, Wouter; Koers, Hylke; Manghi, Paolo), researcher(s) (Deenbroek, Michael; Schindler, Uwe), and sponsor(s) (World Data System). The abstract describes the goal of the Scholix initiative to establish a high-level interoperability framework for exchanging information about links between scholarly literature and data. It also mentions that Scholix is an evolving lightweight set of guidelines to increase interoperability, consisting of a consensus among publishers, datacentres, and global/domain service providers. The page shows 288 views and 109 downloads, and is indexed in OpenAIRE. The publication date is March 1, 2016, and the DOI is 10.5281/zenodo.1120249. Keywords include Publishing, Data Citation, Links, Scholarly communication, and Open science. Grants are also listed.

Scholix data model: from the moon

Link Information Package

Link Publication Date (1)

Link Provider (1..N)

Relationship Type (1)

License URL (0..1)

Source Object

Object Identifier (1)

Object Type (1)

Object Title (0..1)

Object Publisher (0..1)

Object Creator (0..N)

Object Publication Date (0..1)

Target Object

Object Identifier (1)

Object Type (1)

Object Title (0..1)

Object Publisher (0..1)

Object Creator (0..N)

Object Publication Date (0..1)



Scholix: Object Types

Value	Definition	Examples
literature	Literature, of an academic or scientific nature, published in human-readable, text-based form. This includes non-peer-reviewed literature.	journal articles theses monographs book chapters scientific reports standards
dataset	Distinct pieces of information, usually organised in a structured way, that have been published permanently on a repository or data center.	database table spreadsheet

Scholix: Relationship types

Value	Definition (Derived from DataCite Metadata Schema v4.1)	Usage and Examples
IsSupplementTo	Indicates that A is a supplement to B when both are published together	A dataset "IsSupplementTo" an article
IsSupplementedBy	Indicates that B is a supplement to A when both are published together	An article "IsSupplementedBy" a dataset
References	Indicates B is used as a source of information for A	An article "References" a dataset
IsReferencedBy	Indicates A is used as a source of information by B	A dataset "IsReferencedBy" an article
IsRelatedTo	Indicates a generic relation between A and B.	An article "IsRelatedTo" a dataset A dataset "IsRelatedTo" an article Use only where a more specific relationType does not apply.

Scholix in practice: the Hubs

Data repositories



<https://www.datacite.org/eventdata.html>

Publishers



<https://www.crossref.org/services/event-data>

Institutional repositories



<http://scholexplorer.openaire.eu/>



ScholeXplorer

The Data Literature Interlinking Service

Type your search!

Search

Search by persistent Identifier

Search PID

Search and browse in **1332853** literature objects, **8239642** datasets, **56223198** bi-directional scholix links,
from **13000** publishers, **10** data centers, CrossRef, DataCite, and OpenAIRE.

Query: oral history ukraine Search Total Items: 391634 Total pages: 39164

< 1 2 3 4 5 >

Typology	
dataset	373899
publication	17700
unknown	35

Link Provider	
Datasets in Datacite	369236
Crossref	24800
European Nucleotide Archive	9498
PubMed	9151
OpenAIRE	5882
	5423
ICPSR	2454

dataset

Identity Formation and Social Problems in Estonia, Ukraine, and Uzbekistan: Focus Group and Oral History Transcripts, 1996-1998

Anderson, Barbara A. , Kennedy, Michael D. , Hopf, Ted , Kamp, Marianne , Malanchuk, Oksana

This study sought to understand how various forms of social identity structured the articulation of social problems that accompanied the post-Soviet transition, and how the articulation of these social issues related to the formation of identities in post-Soviet society. These questions underlie the structure of the interviews conducted in Estonia, Ukraine, and Uzbekistan between 1996 and 1998. Interviews were conducted in two formats: either a focus group consisting of 6-8 individuals or an individual oral history interview. Thirty-six focus groups, evenly divided between the three countries, were conducted in 14 locations across the three countries. Groups were divided by gender and ethnicity/nationality. During the focus groups, participants were asked to describe positive and negative events of the previous 10 years and whether these changes specific groups of people more than others. Depending on the site, focus groups were sometimes asked to comment on specific issues. For example, the three locations chosen for their proximity to ecological crises (Sillamäe, Ivankiv, and Moynak) were asked about environmental problems. Focus group participants were asked to provide the names of prominent local people who could comment on the social issues discussed during the session. A total of 35 oral history interviews were conducted with prominent citizens in each location: 10 each in Estonia and Uzbekistan and 15 in Ukraine. The 2-hour interviews were designed to record information about the interviewee's personal background as well as their involvement in the changes occurring in the past 10 years, assessment of the current situation, and thoughts on the future directions of their countries.

Sources: [Datasets in Datacite](#)

Publisher: [Inter-University Consortium for Political and Social Research](#)

[Link to Datasets: 1](#) [Link to Literature: 9](#) [Link to other Objects: 1](#)

dataset

Oral History as Process-generated Data

Freund, Alexander

Historical Social Research/Historische Sozialforschung

Sources:

dataset

Identity Formation and Social Problems in Estonia, Ukraine, and Uzbekistan: Focus Group and Oral History Transcripts, 1996-1998

Anderson, Barbara A., Kennedy, Michael D., Hopf, Ted, Kamp, Marianne, Malanchuk, Oksana
1996

This study sought to understand how various forms of social identity structured the articulation of social problems that accompanied the post-Soviet transition, and how the articulation of these social issues related to the formation of identities in post-Soviet society. These questions underlie the structure of the interviews conducted in Estonia, Ukraine, and Uzbekistan between 1996 and 1998. Interviews were conducted in two formats: either a focus group consisting of 6-8 individuals or an individual oral history interview. Thirty-six focus groups, evenly divided between the three countries, were conducted in 14 locations across the three countries. Groups were divided by gender and ethnicity/nationality. During the focus groups, participants were asked to describe positive and negative events of the previous 10 years and whether these changes specific groups of people more than others. Depending on the site, focus groups were sometimes asked to comment on specific issues. For example, the three locations chosen for their proximity to ecological crises (Sillamäe, Ivankiv, and Moynak) were asked about environmental problems. Focus group participants were asked to provide the names of prominent local people who could comment on the social issues discussed during the session. A total of 35 oral history interviews were conducted with prominent citizens in each location: 10 each in Estonia and Uzbekistan and 15 in Ukraine. The 2-hour interviews were designed to record information about the interviewee's personal background as well as their involvement in the changes occurring in the past 10 years, assessment of the current situation, and thoughts on the future directions of their countries.

doi: [10.3886/icpsr36802](https://doi.org/10.3886/icpsr36802)

Sources

- [Datasets in Datacite](#)

Publisher:

- [Inter-University Consortium for Political and Social Research](#)



[LINK TO DATASETS \(1\)](#) [LINK TO PUBLICATIONS \(9\)](#) [OTHER LINKS \(1\)](#)

dataset

Version 1

Anderson, Barbara A., Kennedy, Michael D., Hopf, Ted, Kamp, Marianne, Malanchuk, Oksana,

doi: [10.3886/icpsr36802.v1](https://doi.org/10.3886/icpsr36802.v1)

Sources: [Datasets in](#)

LINK TO DATASETS (1)

LINK TO PUBLICATIONS (9)

OTHER LINKS (1)

publication

Gender, work and economic restructuring in a transcarpathia (Ukraine) village

doi: 10.1080/00905990500193261

Sources: Datasets in
Datacite

publication

The gendered perception of social problems in post-Soviet Estonian society: A qualitative perspective

doi: 10.1080/00905990500193188

Sources: Datasets in
Datacite

publication

Post-Soviet identities in formation: Looking back to see where we are now

doi: 10.1080/00905990500193147

Sources: Datasets in
Datacite

publication

Changing social priorities and the increased salience of the economy in Estonia*

doi: 10.1080/00905990500193170

Sources: Datasets in
Datacite

publication

Symbolic boundaries and national borders: The construction of an Estonian Russian identity

doi: 10.1080/00905990500193196

Sources: Datasets in
Datacite

publication

Gender ideals and income realities: Discourses about labor and gender in Uzbekistan

doi: 10.1080/00905990500193287

Sources: Datasets in
Datacite

publication 2005

Gender, work and economic restructuring in a transcarpathia (Ukraine) village

Dickinson J. A.

2005-09-20T19:22:05Z

doi: [10.1080/00905990500193261](https://doi.org/10.1080/00905990500193261)

Sources

- [Crossref](#)
- [Datasets in Datacite](#)

Publisher:

- [Informa UK Limited](#)



LINK TO DATASETS (2) LINK TO PUBLICATIONS (0) OTHER LINKS (0)

dataset

Version 1

Anderson, Barbara A., Kennedy, Michael D., Hopf, Ted, Kamp, Marianne, Malanchuk, Oksana,

doi : [10.3886/icpsr36802.v1](https://doi.org/10.3886/icpsr36802.v1)

Sources: Datasets in Datacite

dataset

Identity Formation and Social Problems in Estonia, Ukraine, and Uzbekistan: Focus Group and Oral History Transcripts, 1996-1998

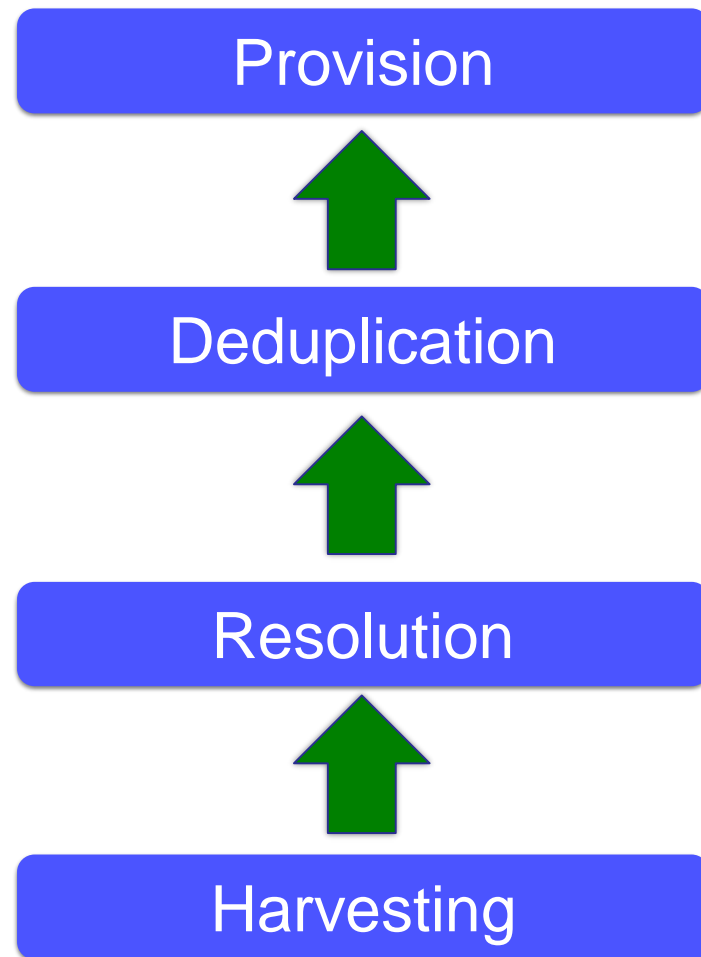
Anderson, Barbara A., Kennedy, Michael D., Hopf, Ted, Kamp, Marianne, Malanchuk, Oksana,

doi : [10.3886/icpsr36802](https://doi.org/10.3886/icpsr36802)

Sources: Datasets in Datacite

scholexplorer.openaire.eu

LINKS!



Scholexplorer

ABOUT SCHOLIX API STATISTICS

Scholexplorer
The Data Literature Interlinking Service

Type your search!

Search

Search by persistent Identifier

Search PID



Deployment co-funded
by RDA

Use-cases: API and User-Interface



API: link search/resolution

WebUI: link discovery/navigation

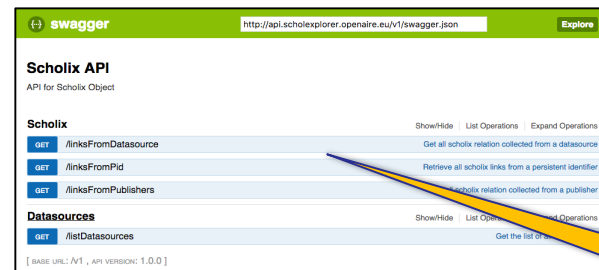
ScienceDirect

Scopus®

zenodo

EarthChem

QPSR



5 Billion
Accesses since
November 2017

Harvesting of links
(every hour)

SCHOLIX API

56+ Mi bi-
directional links
between 8+ Mi
datasets



Scholexplorer APIs

`http://api.scholexplorer.openaire.eu/`



`http://api.scholexplorer.openaire.eu/v1/swagger.json`

Explore

Scholix API

API for Scholix Object

Scholix

Show/Hide | List Operations | Expand Operations

GET

`/linksFromDatasource`

Get all scholix relation collected from a datasource

GET

`/linksFromPid`

Retrieve all scholix links from a persistent identifier

GET

`/linksFromPublishers`

Get all scholix relation collected from a publisher

Datasources

Show/Hide | List Operations | Expand Operations

GET

`/listDatasources`

Get the list of all Datasources

[BASE URL: /v1 , API VERSION: 1.0.0]



Use-cases: JSON dump

Datathons/
Hackatons

The screenshot shows a Zenodo record page for the dataset 'OpenAIRE ScholeXplorer Service: Scholix JSON Dump'. The page includes the Zenodo logo, a search bar, and the user profile 'paolo.manghi@isti.cnr.it'. The record is dated March 16, 2018, and is marked as 'Released' and 'Open Access'. It has 77 views and 16 downloads. The authors listed are Sandro La Bruzzo and Paolo Manghi. The dataset is associated with DataCite and CrossRef, and sponsored by World Data System - ICSU. A description states: 'This datasets contains the GZ-compressed dump of the information space of the OpenAIRE ScholeXplorer service. The dataset consists of 62+ Million literature-dataset and dataset-dataset links between 9+ Million objects, where links are encoded as records in Scholix format (schema Version 3). Links were collected from publishers (CrossRef), data centers (DataCite), and institutional/thematic repositories (OpenAIRE). Uncompressed, the links are organized in 30 files, each of ~2,5GB, for a total of ~75GB.' A table lists the files, showing one file named 'scholix_dump-2018-03-16.tar.gz' with a size of 7.8 GB. The publication date is March 16, 2018, and the DOI is 10.5281/zenodo.1200253. Keywords include 'scholix', 'scholixplorer', 'article-dataset links', and 'scholarly communication'. Grants from the European Commission (OpenAIRE2020 - Open Access Infrastructure for Research in Europe 2020) and RDA Europe (Research Data Alliance - Europe 3) are also listed.



ELSEVIER

La Bruzzo, S. & Manghi, P. (2018). **OpenAIRE ScholeXplorer Service: Scholix JSON Dump** - March 2018 [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1200253>

May 5, 2019

Dataset Open Access

OpenAIRE ScholeXplorer Service: Scholix JSON Dump

Sandro La Bruzzo; Paolo Manghi

Other(s)

DataCite; CrossRef

Sponsor(s)

World Data System - ICSU

This datasets contains the GZ-compressed dump of the information space of the [OpenAIRE ScholeXplorer service](#). The dataset consists of 126+ Million literature-dataset and dataset-dataset links between 12+ Million objects, where links are encoded as records in Scholix format (schema [Version 3](#)). Links were collected from publishers (CrossRef, EventData), data centers (DataCite), and institutional/thematic repositories (OpenAIRE). the links are organized in 29 compressed files, each of ~500MB, for a total of ~15GB,.

Files (15.5 GB)

Name	Size	
scholix_dump-2019-05-05.tar	15.5 GB	Download
md5:93bcf242cb68c620bfeadbdc79353b1a		

650

views

140

downloads

[See more details...](#)



[See more details](#)

Tweeted by 40
On 1 Facebook pages

Indexed in

OpenAIRE

Publication date:

May 5, 2019

DOI:

DOI [10.5281/zenodo.2674330](#)

Keyword(s):

[scholix](#) [scholexplorer](#) [article-dataset links](#)
[scholarly communication](#)

<https://zenodo.org/record/2674330#.XPWXc49S83F>

Thank you!

Iryna Kuchma

iryna.kuchma@eifl.net

