



# Scholix Metadata Schema for Exchange of Scholarly Communication Links

[www.Scholix.org](http://www.Scholix.org)

**Version 3.0**

21 November 2017

Members of the Metadata Working Group:

Amir Aryani

Geoffrey Bilder

Catherine Brady

Ian Bruno

Sandro La Bruzzo

Adrian Burton

Helena Cousijn

Tom Demeranville

Michael Diepenbroek

Martin Fenner

Florian Gräf

Wouter Haak

Paolo Manghi

Gerry Ryder

Uwe Schindler

Phil Shaw

Joe Wass

# Table of Contents

<b>Introduction</b>	<b>3</b>
Scholix Community Participation	3
<b>The Scholix Information Model</b>	<b>3</b>
<b>The Scholix Metadata Schema</b>	<b>6</b>
The Scholix Metadata Schema: Tables of Properties	7
Table 1 - Link Information Package Properties	7
Table 2 - Source and Target Object properties	11
<b>Appendix</b>	<b>15</b>
Controlled Vocabularies	15
3.1 Relationship Type Name	15
8.1 Object Type Name	15
Code examples	16
Schema updates	17

# Introduction

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.

The current vehicle for the Scholix initiative is the Scholarly Link Exchange Working Group, a joint initiative of the [Research Data Alliance](#) and the [World Data System](#).

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 Community Participation

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).

# The Scholix Information Model

The Scholix initiative facilitates exchange of information about links between data and literature. The fundamental information concerns the relationship between two objects as sketched in Figure 1. This information about the link is issued by a party or chain of providers.

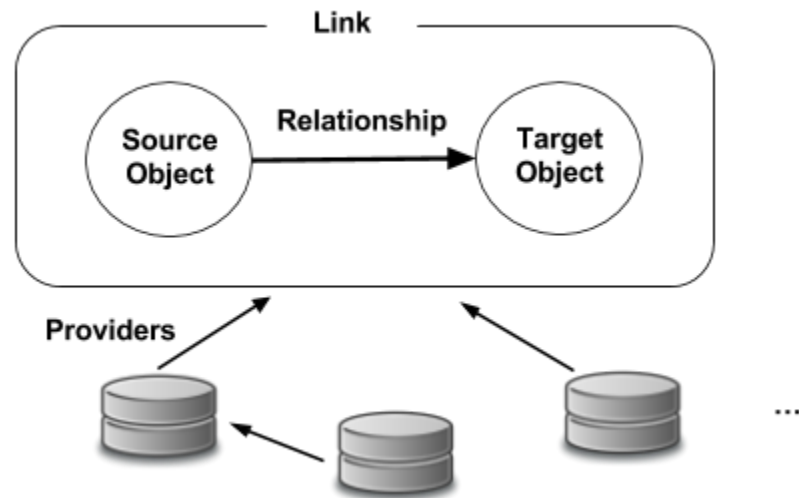


Figure 1. Scholix Link: main concepts

A Scholix Link Information Package contains:

- Information about the two objects
- Information about the nature of the link and the link package itself (date, issuer, rights)

## 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)

Figure 2 - Link Information Package: high-level information model

# The Scholix Metadata Schema

The Scholix schema is the list of all the properties in a Scholix link information package with their definitions, structure and occurrence rules.

Each property has a cardinality or occurrence (the number of times the element may or must occur or re-occur):

- 0..N = optional and repeatable
- 0..1 = optional and not repeatable
- 1..N = required and repeatable
- 1 = required and not repeatable

Many of the properties are optional; the schema is designed to allow bulk exchange of link information with a minimum of information relevant to the link being mandatory. For example if an identifier of a dataset can be resolved to provide its name, creator, publisher and publication date, then only the identifier is mandatory. The other properties are necessary but optional, to be provided if the identifier system doesn't support such resolution easily. A handful of properties form the mandatory core of a Scholix information package: the identifier and resource type of the two objects, their relationship, with the date and provider of this link information package.

This bias to simplicity and lightness is balanced with the need for end users of the link information (e.g. web sites that want to display information about a linked object) for a minimum of “usable” information (e.g. publisher, title, clickable links).

Some of the properties are “structured”, which means they do not characterize a value but rather a nested set of other properties with values. Structured properties are indicated by a *[structured property]* tag.

The full details of the properties of the Scholix schema with all their definitions and examples are below in Tables 1 and 2.

## The Scholix Metadata Schema: Tables of Properties

**Table 1** - Link Information Package Properties

ID	Property Label	Occ	Element name	Property definition	Allowed values	Examples	Guidance on usage
1	<b>Link Publication Date</b>	1	LinkPublicationDate	Date when this Link Information Package was first formally issued from this current Provider	[date]	"1997-10-23"	<p>This date might typically come from the current Link Provider's system date for its receipt or creation [and hence availability] of the information in this Link Information Package.</p> <p>This is not the publication date of the research objects (see property 11). As this is provenance information about the issuance of this Link Information Package, year-month-day is more relevant than simple year.</p> <p>Consider using an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].</p>
2	<b>Link Provider</b>	1..N	LinkProvider	The source(s) of this Link Information Package	[structured property]		<p>The party actually generating this Link Information Package is mandatory. Use multiple Link Provider elements only if you want to encode the current provider [mandatory] <i>as well as</i> any antecedent sources [optional]. This may be the case: if the link information has been gathered by an aggregator (eg DataCite) from an antecedent source (eg a data centre); or if a chain of scholix hubs were to transfer link information</p>

						packages to each other. The purpose of this property is to provide a simple provenance context.	
2.1	<b>Link Provider Name</b>	1	Name	The name of the Link Provider	[text]	"DataCite"	
2.2	Link Provider Identifier	0..N	Identifier	A unique string that identifies the Link Provider	[structured property]		
2.2.1	<b>Link Provider ID</b>	1	ID	The identifier string	[ID]	"475826.a"	
2.2.2	<b>Link Provider ID Scheme</b>	1	IDScheme	The scheme or namespace of the identifier string	[text]	"grid"	Consider using the "namespace" value from the identifier.org registry
2.2.3	Link Provider ID URL	0..1	IDURL	A URL form of the Link Provider ID.	[URL]	"https://grid.ac/institutes/grid.475826.a"	This element is intended to deliver end-users directly to a web page.  Optional only if the Link Provider ID [2.2.1] is already expressed in URL form.
3	<b>Relationship Type</b>	1	Relationship Type	The nature of the relationship between the source object and target object in this Link Information Package	[structured property]		
3.1	<b>Relationship</b>	1	Name	The relationship	<i>Controlled List</i>		See Appendix for definitions,



	<b>p Type Name</b>			type chosen from a Scholix controlled vocabulary	<i>Values:</i> IsReferencedBy References IsSupplementTo IsSupplementedBy IsRelatedTo		<p>examples and usage notes for the values in the controlled list.</p> <p>The Scholix Controlled List for Relationships is based on the main DataCite and CrossRef vocabulary terms, plus one generic, and are intended to provide a common set of standard values across the Scholix initiative reflecting common scenarios and supporting common user needs.</p> <p>Relationship Type Name is mandatory; if unsure, use the generic value "IsRelatedTo".</p>
3.2	Relationship Sub-type	0..1	SubType	The sub-type of "3.1 Relationship Type name"	[text]	"basedOnData"	<p>Optionally, can be used to convey different or more nuanced relationships than those allowed in the Scholix Controlled List for "3.1 Relationship Type Name".</p> <p>If a value for "3.3 Relationship Sub-type Schema" is provided, the value of 3.2 should be part of any controlled vocabulary or terms specified in that schema.</p>
3.3	Relationship Sub-type Schema	0..1	SubTypeSchema	The name of the schema or controlled list from which "3.2 Relationship Sub-type" is sourced	[URL]	"http://www.crossref.org/schema/4.4.1"	Use of a standard schema for additional relationships is recommended.

4	License URL	0..1	LicenseURL	The URL of the license for the Scholix Link Information Package	[URL]	<p>“http://creativecommons.org/publicdomain/zero/1.0”</p> <p>“http://creativecommons.org/licenses/by/4.0 “</p>	<p>CC0 is recommended, but other options are possible.</p> <p>This is not the licence of any of the research objects themselves, rather the rights statement under which the link metadata (this Link Information Package) is released.</p>
5	<b>Source Object</b>	<b>1</b>	Source	Root element relative to all properties describing the link’s source object. For the expansion of these properties see Table 2.	[structured property]		
6	<b>Target Object</b>	<b>1</b>	Target	Root element relative to all properties describing the link’s target object. For the expansion of these properties see Table 2.	[structured property]		

**Table 2 - Source and Target Object properties**

Expansion of properties “5. Source” and “6. Target” in Table 1.

ID	Property	Occ	Element name	Property definition	Allowed values	Examples	Guidance on usage
7	<b>Object Identifier</b>	1	Identifier	A unique string that identifies the object	[structured property]		
7.1	<b>Object ID</b>	1	ID	The identifier string	[ID]	“6EKT” “10.17632/rc6rwf7c8n.1”	
7.2	<b>Object ID Scheme</b>	1	IDScheme	The scheme or namespace of the identifier string	[text]	“pdb” “doi”	Consider using the namespace value from the identifier.org registry
7.3	Object ID URL	0..1	IDURL	An internet resolvable form of the identifier	[URL]	“https://www.rcsb.org/pdb/explore/explore.do?structureId=6EKT” (PDB) “https://doi.org/10.17632/rc6rwf7c8n.1” (optional)	This property is intended to deliver end-users directly to a web page.  Optional only if the Object ID [7.1] is already in URL form (eg CrossRef and DataCite DOIs).
8	<b>Object Type</b>	1	Type	Describes the nature of the object (its intended usage)	[structured property]		
8.1	<b>Object</b>	1	Name	The object type	<i>Controlled</i>		See Appendix for definitions, examples

	<b>Type Name</b>			chosen from a Scholix controlled vocabulary	<i>List Values:</i> literature dataset		and usage notes for the values in the controlled list.
8.2	Object Sub-type Name	0..1	SubType	A sub-type of “8.1 Object Type Name”	[text]	“journal article” “report” “book chapter” “conference abstract” (literature object type)	Used to convey specific types of literature or dataset objects.  If a value for “8.3 Object Sub-Type Schema” is provided, the value should be part of the relative vocabulary.
8.3	Object Sub-type Schema	0..1	SubTypeSchema	The name of the schema or controlled list from which “2.2 Object Sub-type” is sourced	[text]	“CASRAI”	
9	Object Title	0..1	Title	The name of the object	[text]	“On the Nature of Things”	
10	Object Creator	0..N	Creator	Party responsible for the creation of the object	[structured property]		
10.1	Creator Name	1	Name	The name of the Object Creator	[text]	“John H., Smith”  “National Aeronautics and Space Administration”	Recommended format for a person: {First Name, Last Name}
10.2	Creator Identifier	0..N	Identifier	A unique string that identifies the Object Creator	[structured property]		
10.2.1	Creator ID	1	ID	The identifier	[ID]	“0000-0002-1825-0097”	

				string		(ORCID) "0000 0004 4907 1619" (ISNI)	
10.2.1	Creator ID Scheme	1	IDScheme	The scheme or namespace of the identifier string	[text]	"ORCID"  "ISNI"	
10.2.3	Creator ID URL	0..1	IDURL	An internet resolvable form of the identifier	[URL]	https://orcid.org/0000-0002-1825-0097  http://www.isni.org/isni/000000449071619	This element is intended to deliver end-users directly to a web page.  Optional only if the Creator ID [10.2.1] is already in URL form.
11	Object Publication Date	0..1	PublicationDate	The date the object was formally issued, published or distributed	[date]	"1997-10-23"	If the publication date cannot be determined, use the date of registration. If there is no standard publication year value, use the date that would be preferred from a citation perspective.  Use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
12	Object Publisher	0..1	Publisher	The publisher of the source or target object	[structured property]		
12.1	Publisher Name	1	Name	The name of the publisher of the object	[text]	"Cambridge Crystallographic Data Centre"  "Data in Brief"	This can be (for example) a data publisher, or an article publisher, e.g. domain repository, data repository, a journal name, or an institutional repository
12.2	Publisher	0..N	Identifier	A unique string	[structured property]		

	Identifier			that identifies the publisher			
12.2.1	Publisher ID	1	ID	The identifier string	[ID]	“0000 0001 2180 7418” (ISNI) “2352-3409” (ISSN)	
12.2.2	Publisher ID Scheme	1	IDScheme	The scheme or namespace of the identifier string	[text]	“ISNI” “ISSN”	
12.2.3	Publisher ID URL	0..1	IDURL	An internet resolvable form of the identifier	[URL]	“http://isni.org/isni/0000-0001-2180-7418” “http://road.issn.org/issn/2352-3409-data-in-brief”	This element is intended to deliver end-users directly to a web page. Optional only if the Publisher ID [12.2.1] is already in URL form.

# Appendix

## Controlled Vocabularies

### 3.1 Relationship Type Name

Table 3: Valid values for Property 2.1 Relationship Type Name - the relationship between source object (A) to target object (B).

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.

Note: the controlled vocabulary list for this relationship is limited on purpose. There are potentially many more nuanced relationship types possible, but this goes beyond the main purpose of creating a large and comprehensive collection of Scholix relations.

## 8.1 Object Type Name

Table 4: Valid terms for Property 8.1 Object Type Name - the nature of the object [its intended usage].

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

## Code examples

Swagger API: <http://api.scholexplorer.openaire.eu/v1/ui/>

Github repository with serialisation examples (json, xsd): <https://github.com/scholix>



## Schema updates

The WDS-RDA Scholarly Link Exchange Working Group is responsible for reviewing and updating this schema on a regular basis. The development of the schema began in 2016. Versions 1 and 2 were issued as draft schemas. Version 3 is the first schema to be used operationally. Version 3.0 is intended to remain stable until the Berlin RDA conference in 2018. Future schemas aim to be backwardly compatible. Comments and suggestions for changes can be sent to [info@scholix.org](mailto:info@scholix.org).

Author	Date	Update description	Version
Paolo Manghi	29/11/2016	<ul style="list-style-type: none"> <li>• Changed property 1 in Table 1</li> <li>• Changed property 2 of Table 2</li> </ul>	0.0
Paolo Manghi	20/12/2016	<ul style="list-style-type: none"> <li>• One mandatory PID for source and target object</li> <li>• Relationships semantics from mandatory vocabulary</li> </ul>	1.0
Paolo Manghi	21/01/2017	<ul style="list-style-type: none"> <li>• Make Link Publisher optional and name changed to Link Provider</li> <li>• Make Link Provider mandatory</li> <li>• Add optional License URL field</li> <li>• Remove glossary</li> </ul>	2.0
Wouter Haak, Adrian Burton, Ian Bruno, Paolo Manghi, Martin Fenner	4/10/2017	<ul style="list-style-type: none"> <li>• Added ID URIs where an ID element is request</li> <li>• Relationship type: strict list of terms from DataCite vocabulary</li> </ul>	2.1
Wouter Haak, Paolo Manghi (during Pisa workshop)	5/11/2017	<ul style="list-style-type: none"> <li>• Edited Relationship types: removed InverseRelationship, add original relationship details, included mandatory Scholix relationship</li> <li>• Added examples where relevant</li> <li>• Added editable schema tables</li> </ul>	2.1
Adrian Burton, Catherine Brady, Wouter	14/11/2017	Complete review	3.0

Haak		Complete content in all cells of schema tables with examples and definitions. Re-naming "type" properties for uniformity between property 2 and property 8. Change "Id schema" to "ID Scheme" in all identifier structured properties	
------	--	---	--