Data Publishing and Data Linking

Introducing SCHOLIX

Howard Ratner, Executive Director, CHORUS

on behalf of STM

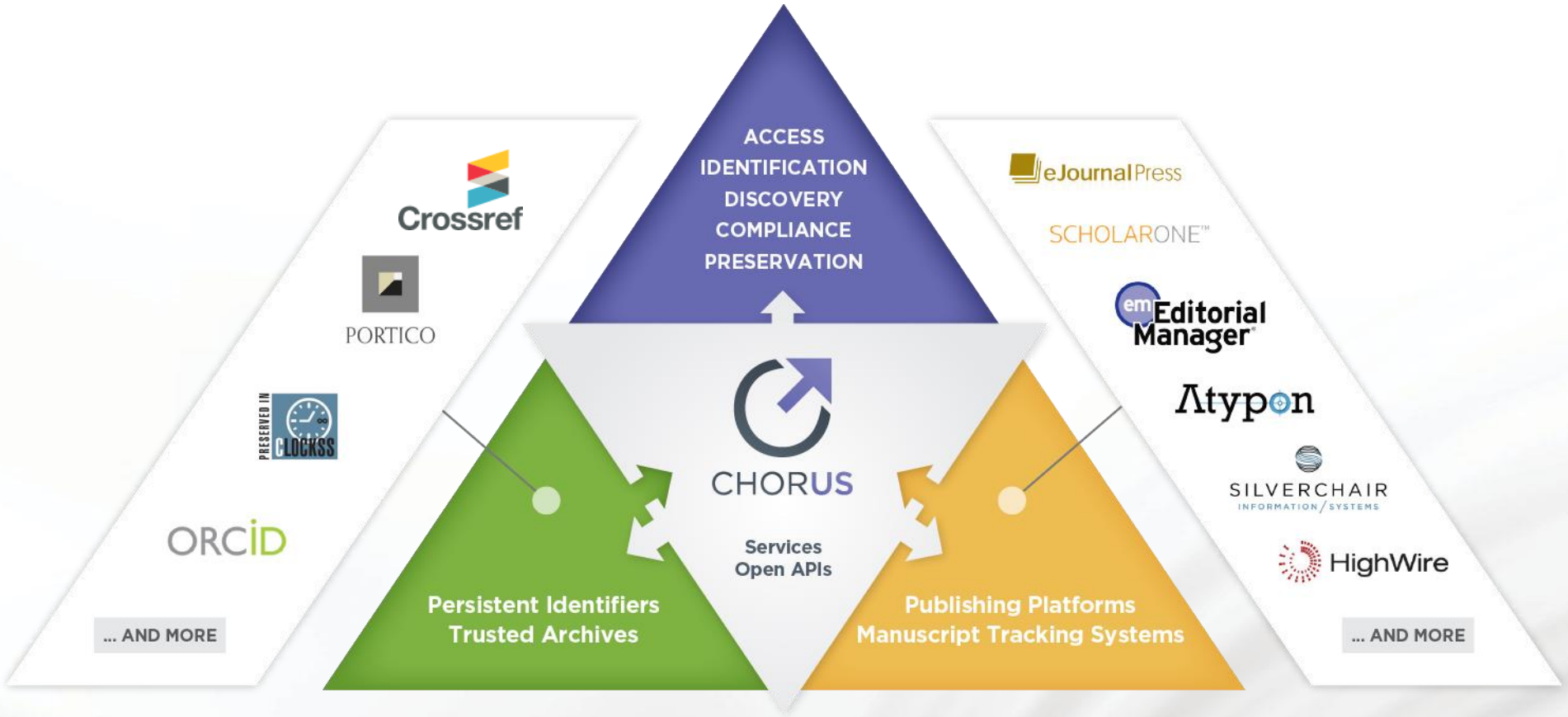
Workshop on Data Citation

12 July 2016, National Academy of Sciences, Washington DC

www.stm-assoc.org

Cost-effective Public Access Solution

Builds on Existing Infrastructure

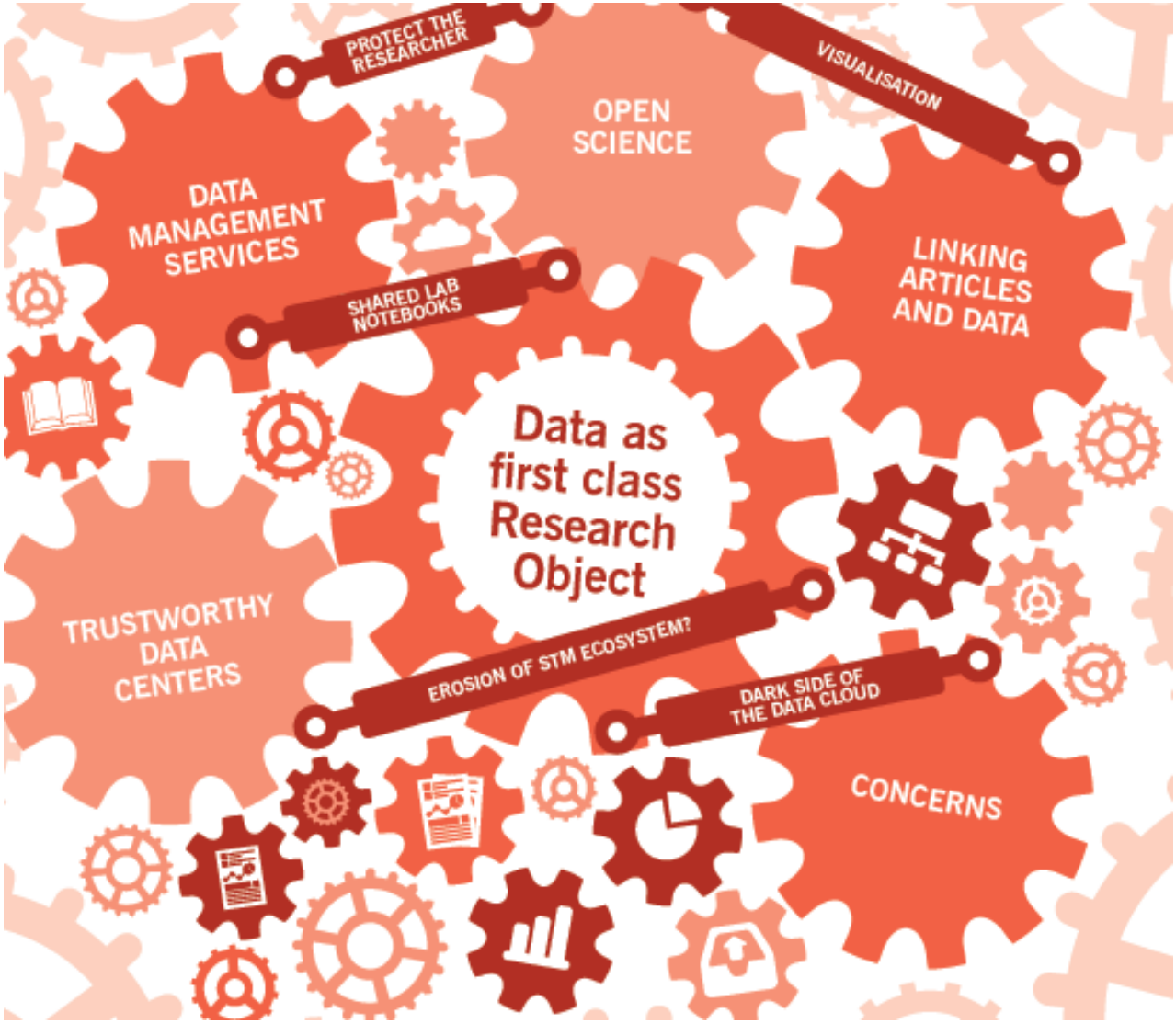


CHORUS is part of CHOR, Inc., a non-profit 501(c)(3) membership organization

What is STM?

- Worldwide representative of scholarly communication community
- Diverse membership
  - not-for-profit, societies, university presses and commercial
  - Approximately 125 publishing houses are members
- Members responsible for nearly **two-thirds** of all annual article output and **over half** of active peer-reviewed titles

Publishing Data as first class objects



Source:
2015 STM
Tech Trends

Following the FAIR principles

- STM community endorses the FAIR principles for Data Publishing:
  - Findable
  - Accessible
  - Interoperable
  - Re-usable
- Key to making data citation work is linking data and literature
- STM has worked hard within RDA to achieve goal via Interest Group on Data Publishing
- Now ready for launch: SCHOLIX

The SCHOLIX Principles:
A framework for Scholarly Link Exchange

Linking research data with literature is of great value, but existing solutions are disconnected, which is inefficient and limits value for researchers

Why link data and literature?

1. Increase visibility & discoverability of data (and articles)
2. Place data in the right context to enable proper re-use
3. Support credit attribution mechanisms

The Scholix Principles:
A framework for Scholarly Link Exchange

The SCHOLIX Principles will enable us to build the web of literature – research data links

- The SCHOLIX Principles is a set of best practice guidelines to link research data and literature
- SCHOLIX represents the consensus achieved by the various stakeholder groups in the research data landscape – including data centers, publishers, Crossref, DataCite, OpenAIRE, libraries, industry, and many others
- SCHOLIX came out of the ICSU-WDS / RDA Data Publication Services working group

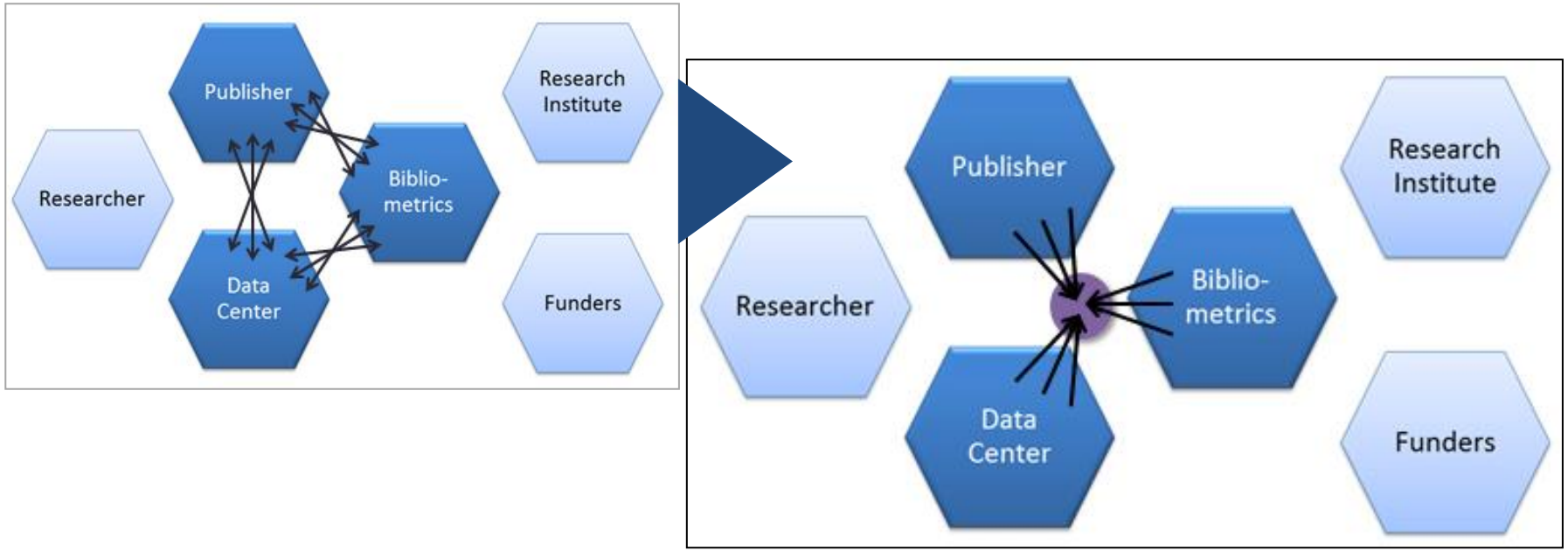
SCHOLIX scholix.org

ICSU WORLD DATA SYSTEM

STM

SCHOLIX: The current problem

How to move from (mostly) bilateral arrangements to a **one-for-all service model infrastructure** for research data linking with publications: operational prototype now available.



SCHOLIX Best Practice Guidelines

Shared Conceptual Model

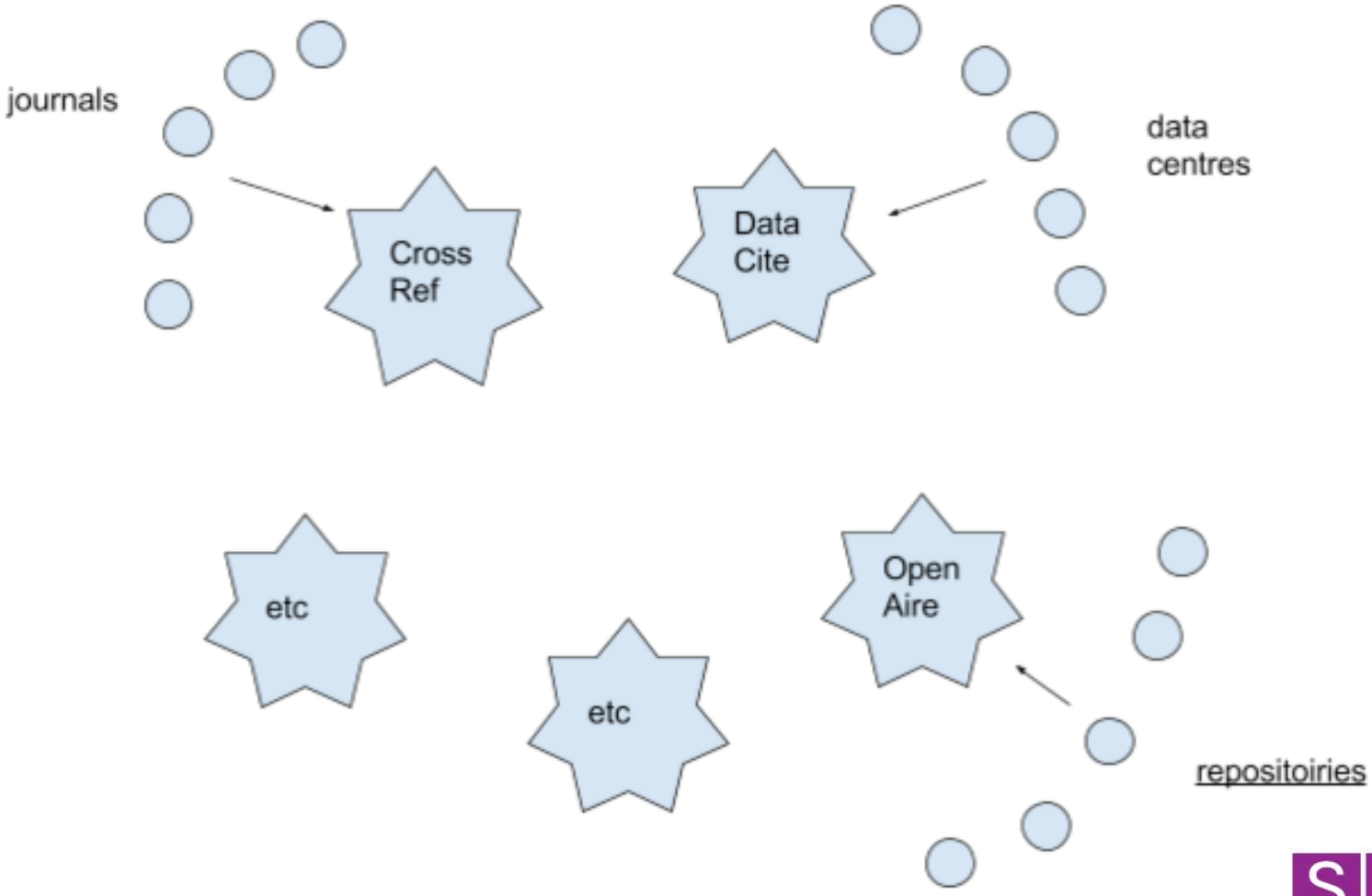
Information Model

Encoding Options

Exchange Options

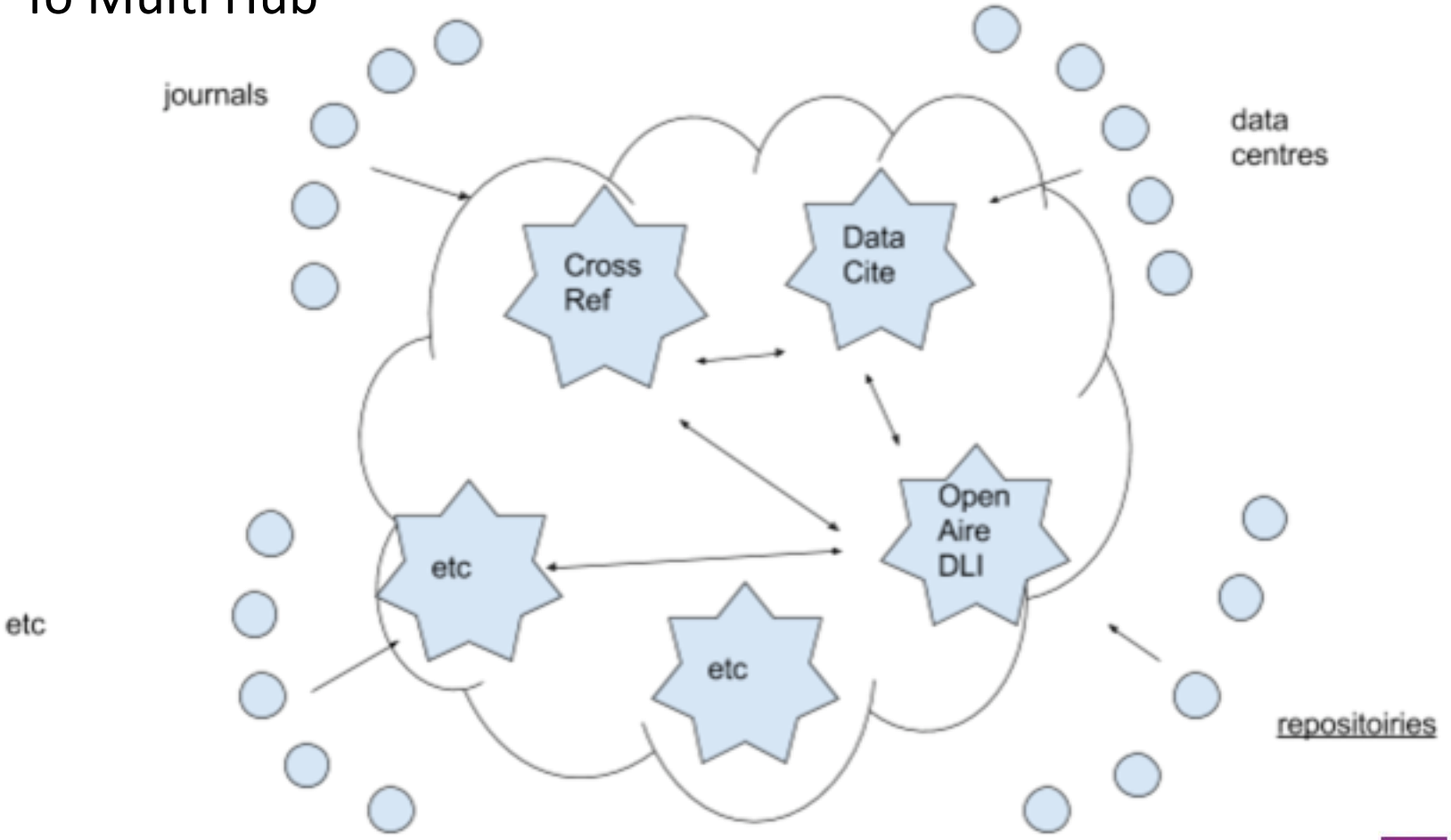
SCHOLIX Shared Conceptual Model

From Silos

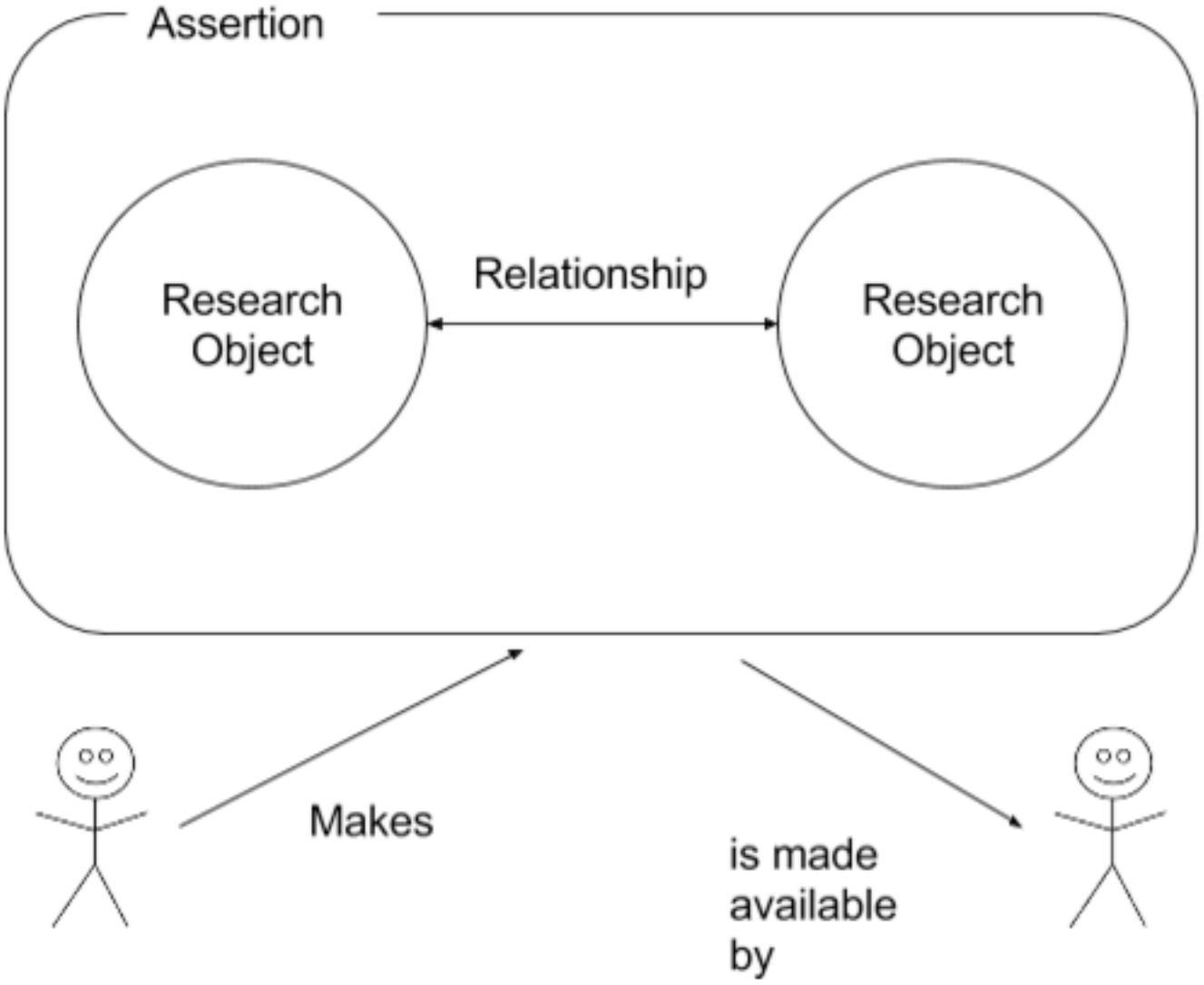


SCHOLIX Shared Conceptual Model

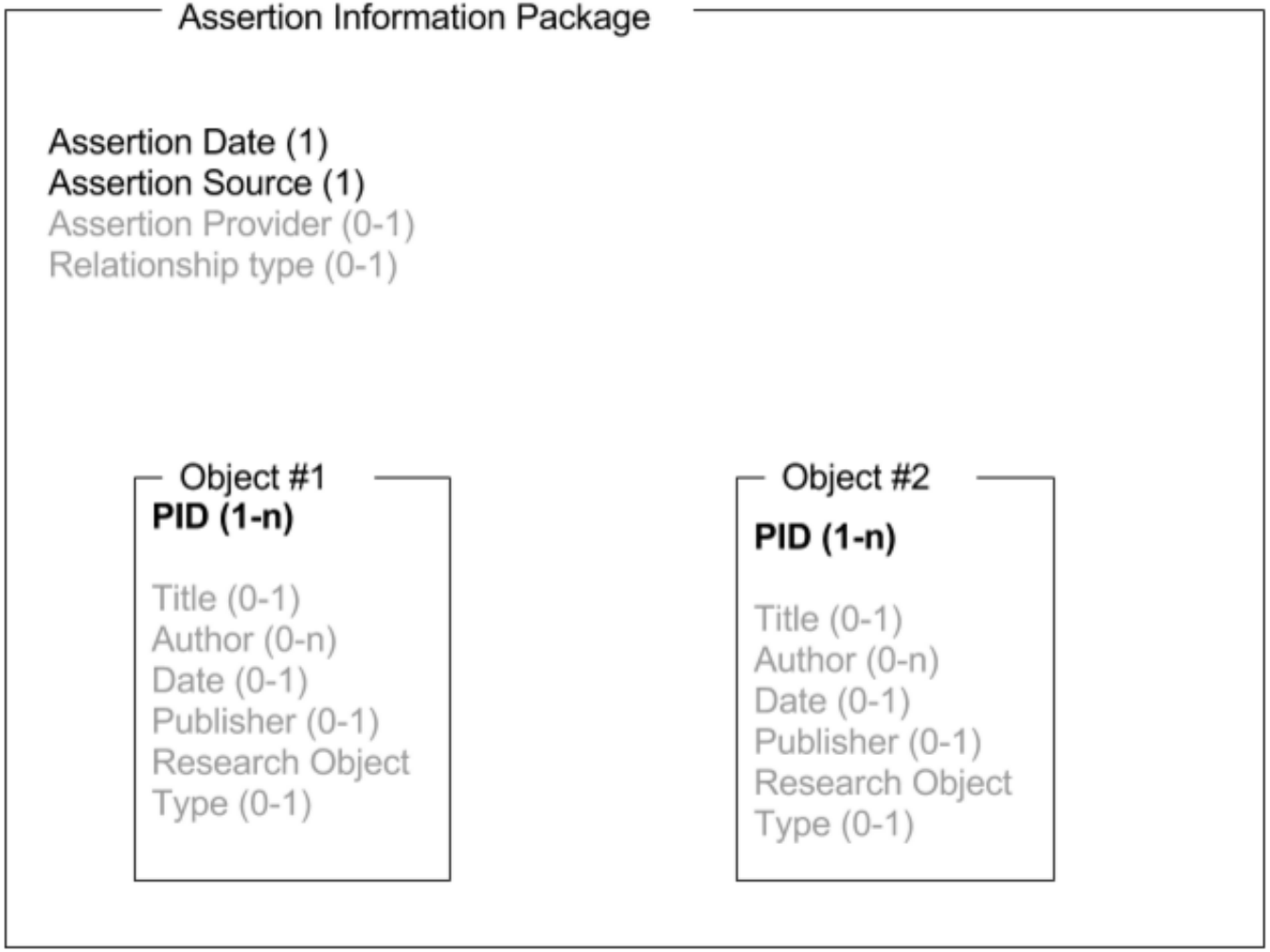
To Multi Hub



SCHOLIX Shared Conceptual Model



SCHOLIX Information Model



SCHOLIX Information Model: Standards

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

SCHOLIX Encoding Options

- Current framework does not specify any particular schemas or exchange formats
- Investigating DiSCOs (Distributed Scholarly Compound Object) from RMap project to encode the information in an RDF assertion information package

SCHOLIX Exchange Options

- Current framework does not specify particular exchange protocols
  - All hubs support JSON through open documented APIs registered on SCHOLIX website
- Future – framework will support:
  - well documented open RESTful APIs
  - handful of community supported protocols (e.g., OAI-PMH, etc.)
  - machine readable hub registry with notifications
  - plan to investigate LinkBack and webmention methods for notificactions

SCHOLIX Protoype now available

DLI Service (Data Literature Interlinking)

Articles and data links from:

- 3TU.Datacentrum
- Australian National Data Service (ANDS)
- Cambridge Crystallographic Data Center (CCDC)
- CrossRef
- DataCite
- Elsevier
- Interdisciplenary Earth Data Alliance (IEDA)
- Interuniversity Consortium for Political and Social Research (ICPSR)
- Institute of Electrical and Electronics Engineers (IEEE)
- OpenAire
- PANGAEA
- RCSB Protein Data Bank
- Springer Nature
- Thomson Reuters

ABOUT API DATASOURCES STATISTICS CONTACT US

Query: penguins Search Total Items: 57 Total pages: 7

< 1 2 3 4 5 >

Typology

dataset	30
publication	27

Content Provider

Datasets in DataCite	43
PANGAEA	27
CrossRef	15
EuropePMC	12
Thomson Reuters	6
Springer Nature	2

Penguins of Antarctica

registry-migration.gbif.org

11 Related Datasets 0 Related Publications 0 Other Relations Original Object [Datasets in DataCite , DataCite Resolver]

Tab. 3: Counting of seals and penguins

Schneppenheim, Reinhard - 1980-01-04T15:00:00/1980-02-18T14:30:00

1 Related Datasets 0 Related Publications 1 Other Relations Original Object [Datasets in DataCite , PANGAEA]

Data from: New fossil penguins (Aves, Sphenisciformes) from the Oligocene of New Zealand reveal the skeletal plan of stem penguins

Ksepka, Daniel T. , Fordyce, R. Ewan , Ando, Tatsuro , Jones, Craig M.

2 Related Datasets 1 Related Publications 0 Other Relations Original Object [Datasets in DataCite]

["New fossil penguins (Aves, Sphenisciformes) from the Oligocene of New Zealand reveal the skeletal plan of stem penguins"]

Ksepka Daniel T. , Fordyce R. Ewan , Ando Tatsuro , Jones Craig M. - 2012-02-28T19:11:49Z

3 Related Datasets 0 Related Publications 0 Other Relations Original Object [Datasets in DataCite , CrossRef]

Next Steps

- Promoting adoption and implementation of the SCHOLIX Framework
- Presentations to stakeholders to implement the SCHOLIX Framework
- Crucial participants:
  datacenters, publishers, funders, researchers
- Your help and input is very welcome!

Questions?

Adrian Burton, Co-Chair, ICSU-WDS/RDA Data-Publication Services WG Director, Australian National Data Service (ANDS), adrian.burton@ands.org.au

Hylke Koers, Co-Chair, ICSU-WDS/RDA Data-Publication Services WG Head of Content Innovation, Elsevier, h.koers@elsevier.com

Howard Ratner, Executive Director, CHORUS
hratner@chorusaccess.org

Eefke Smit, Director of Standards and Technology, STM
smit@stm-assoc.org

Try the prototype - now with more than 7 M links available:
http://dliservice.research-infrastructures.eu/#/