



Linking data to publications: lessons learned from a pilot implementation

Donatella Castelli
CNR-ISTI

Outline

1. The context: ICORDI & RDA
2. Linking research resources: prototype & interoperability issues

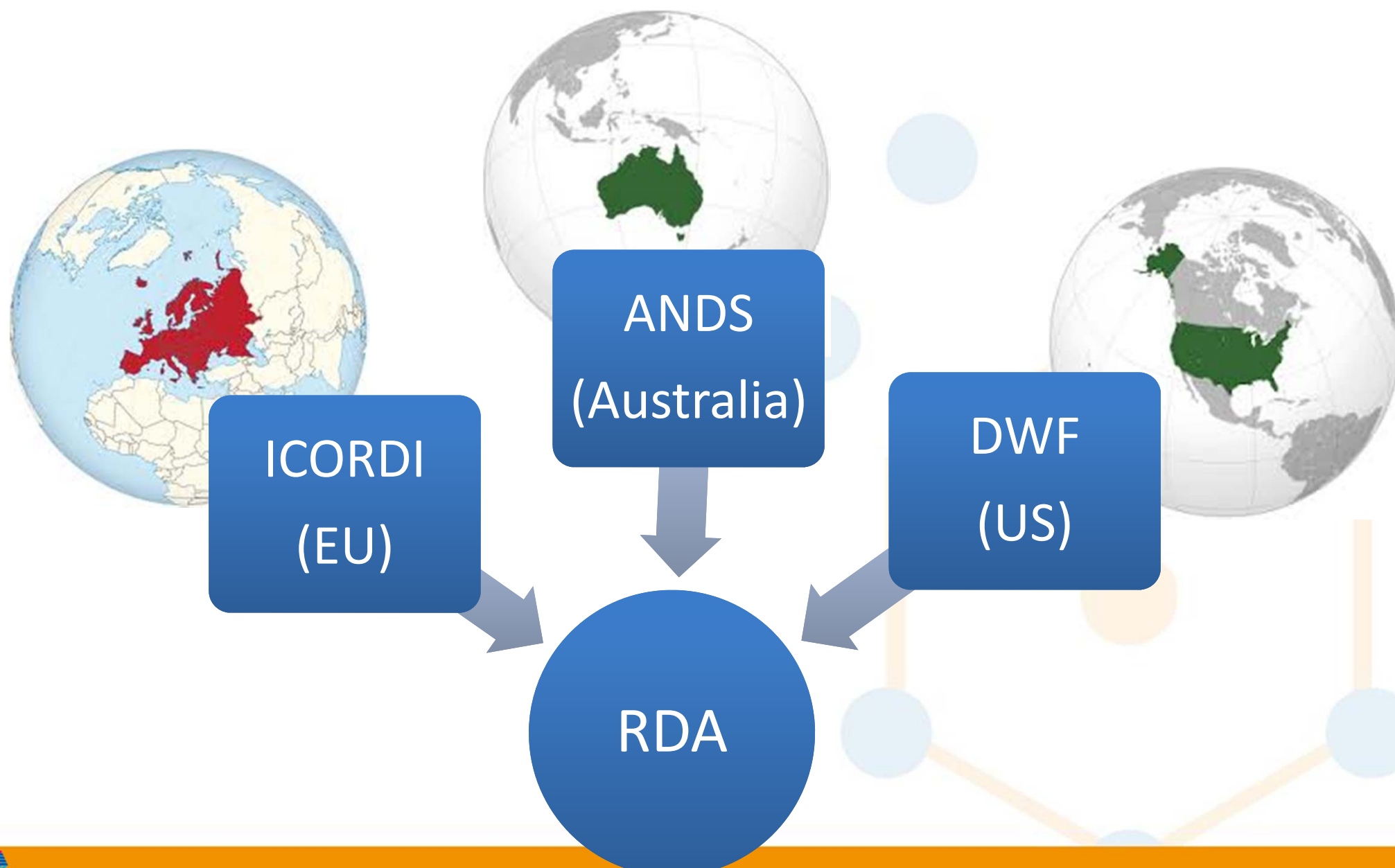
ICORDI

Objective: “To establish a coordination platform between Europe and USA to discuss and improve the interoperability of today’s and tomorrow’s scientific infrastructures of both continents and to extend this to the global levels”

 Started in September 2012

 <https://www.icordi.eu/>

Merging of multiple Initiatives



Research Data Alliance



“To facilitate specific, short-term efforts that accelerate the sharing and exchange of research data”

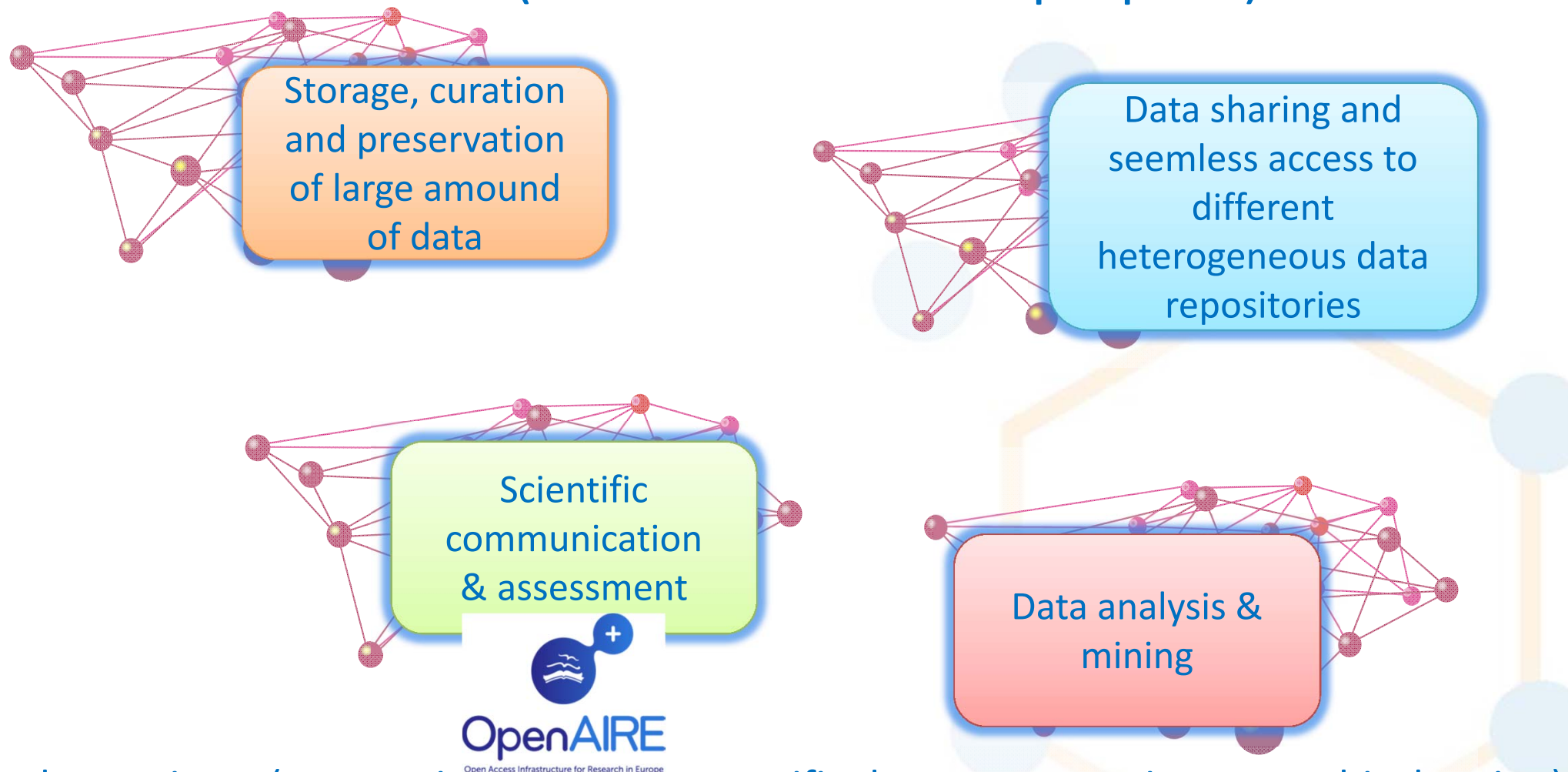
-  **Working groups** serving as accelerants to data sharing practice and infrastructure interoperability
 -  Adopted standards
 -  Deployed infrastructure
 -  Adopted policy
 -  Implemented best practice, etc.

 <http://rd-alliance.org>

The context:

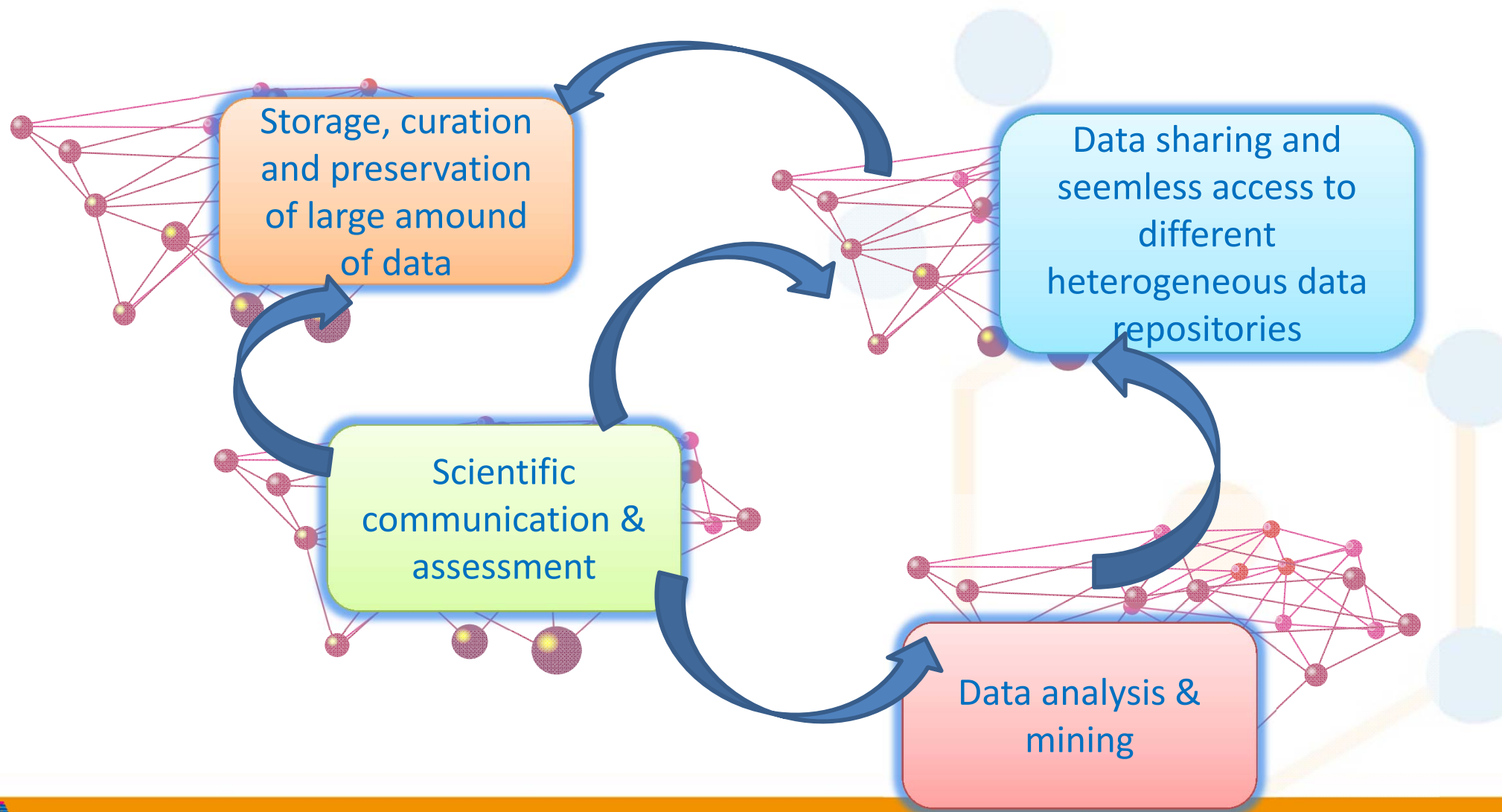
many e-Infrastructures

(infrastructure for a purpose)



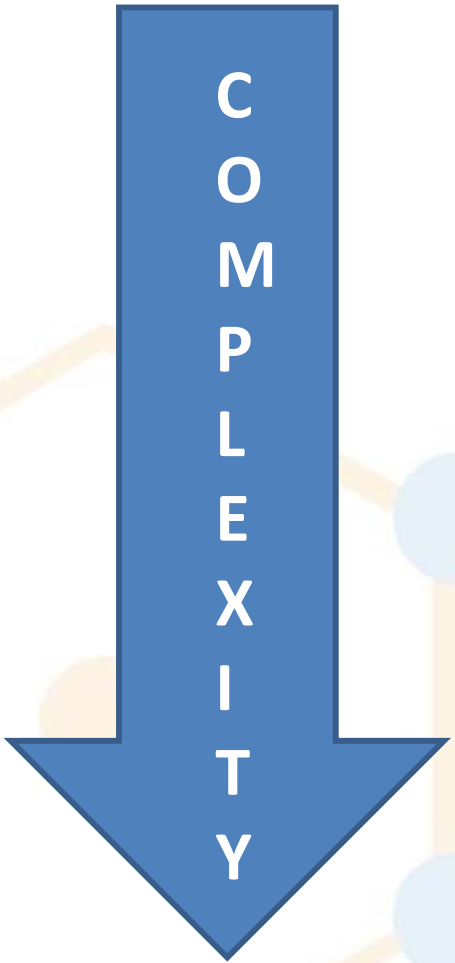
+ other variants (e.g. particular domain, specific data type, certain geographical region)

The addressed challenge: e-Infrastructure ecosystem



Interoperability is the key issue

-  Discover resources
-  Access and understand resources
-  Use (e.g. analyse, process, deploy, etc.) resources

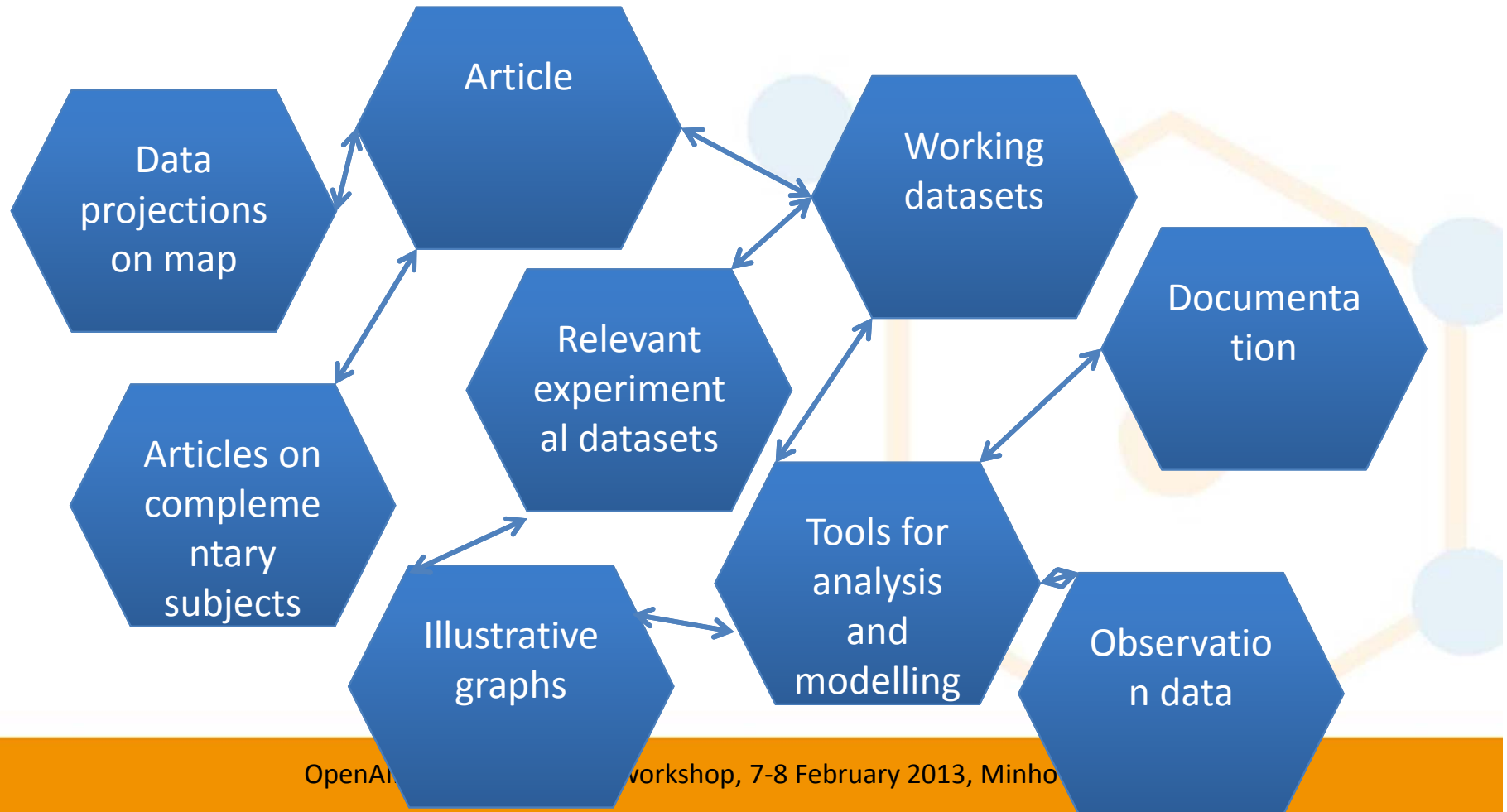


C
O
M
P
L
E
X
I
T
Y

OpenAIRE mandate:

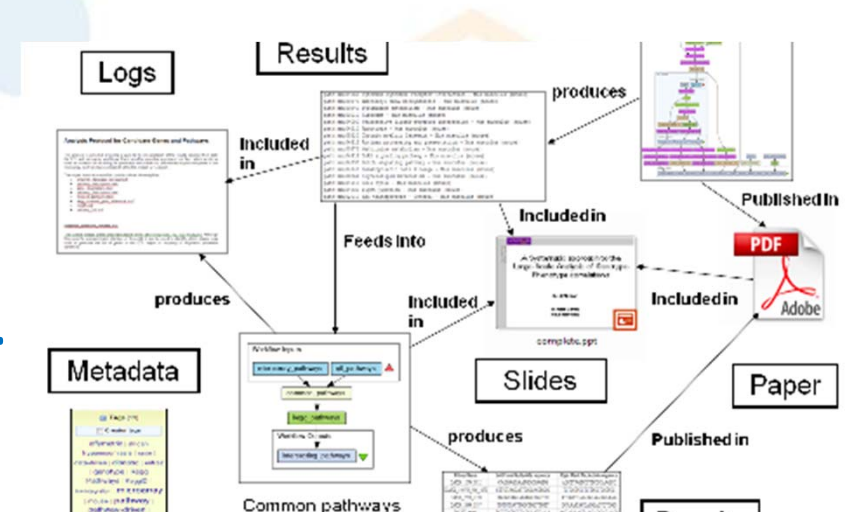
e-Infra for scientific information

Research processes require to retrieve, access and reuse a large variety of resources often made available from scientists working in different domains



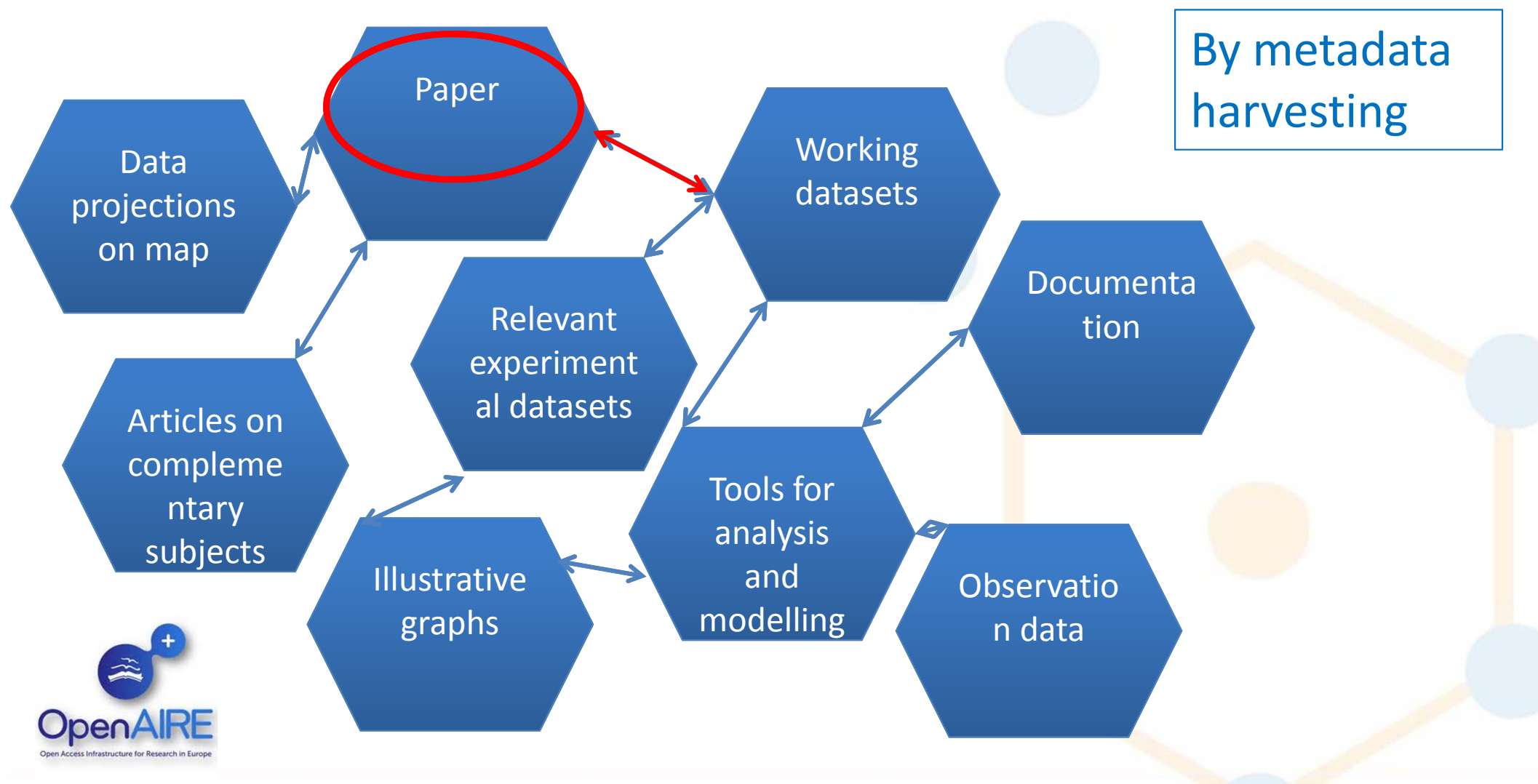
"New publications": relationships explicitly specified by the author

- ❖ *Structured Publications* (e.g. Elsevier Article of the Future)
- ❖ *Compound Objects* (e.g. enhanced publications, modular articles)
- ❖ *Live articles* (e.g. D4Science dynamic reports)
- ❖ *Experiment-oriented publication* (e.g. Scientific Publication Package, My-Experiment research objects)
- ❖



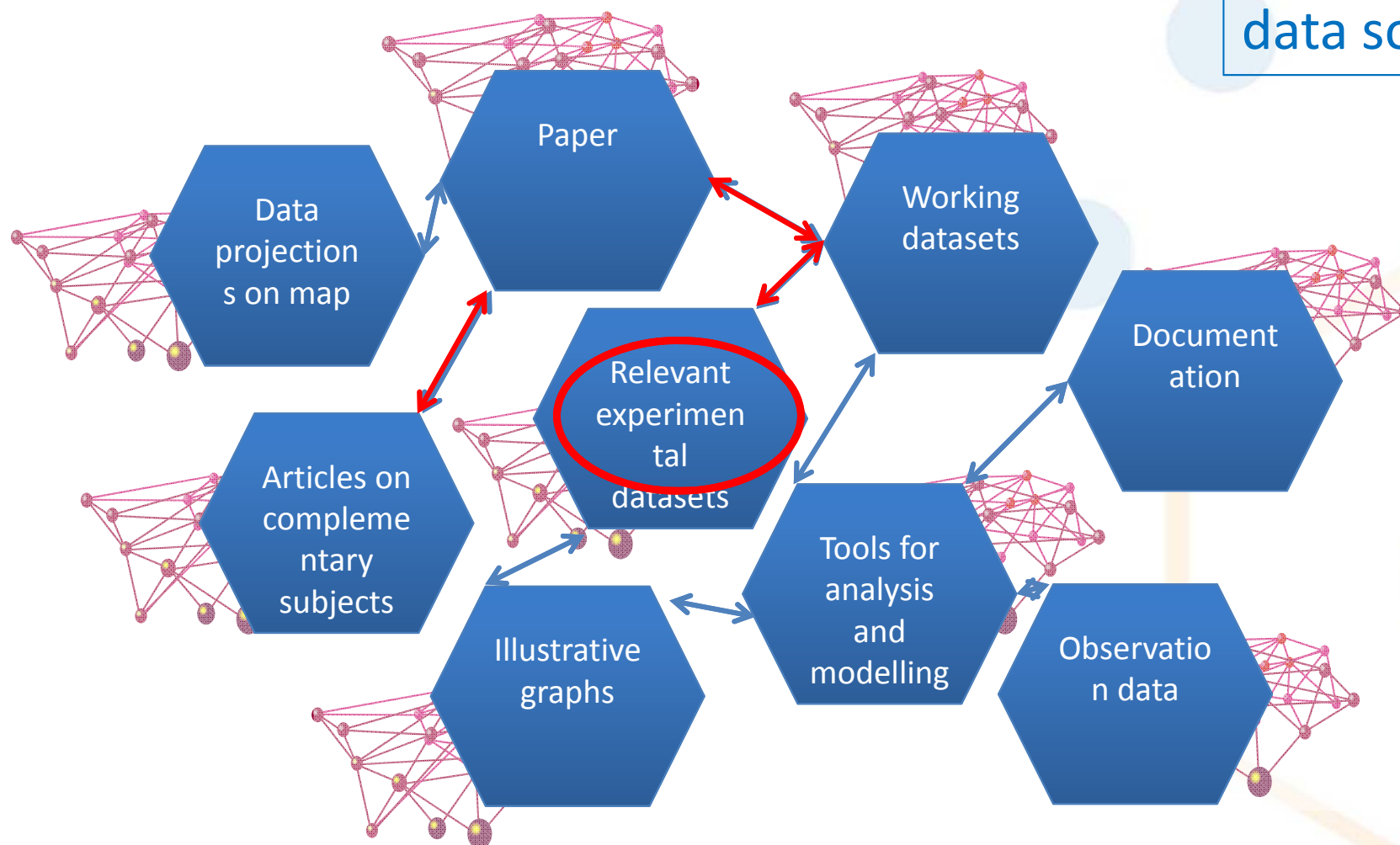
OpenAIRE now:

“Automatic” identification of relationships



The ICORDI Prototype

By querying different
data sources/infra



Search in a data source /aggregator

enAIRE - Data Searchery



Homepage - Starting datasource:

Collections:

- ☐ 3TU Datacentrum
- ☐ Aalborg University Library
- ☐ Alfred-Wegener-Institut
- ☐ altrelettere
- ☐ AMA Service GmbH
- ☐ Applied Rheology
- ☐ Archaeology Data Service
- ☐ Australian Antarctic Division
- ☐ Australian National Data Service
- ☐ Beijing Genomics Institute
- ☐ Bibliothek des Wissenschaftsparks Albert Einstein
- ☐ Biodiversity and Sustainable Management of a Megadiverse Mountain Ecosystem in South Ecuador
- ☐ Biodiversity Institute of Ontario
- ☐ Bundesanstalt für Gewässerkunde
- ☐ CDL
- ☐ Cellular Therapy and Transplantation
- ☐ CERN
- ☐ CERN Document Server
- ☐ CORSSA - Community Online Resource for Statistical Seismicity Analysis
- ☐ CSIRO
- ☐ CSSE Uni Melbourne
- ☐ Curtin University
- ☐ Dagstuhl
- ☐ DataCite
- ☐ datalino
- ☐ DESY - Deutsches Elektronen-Synchrotron
- ☐ Deutsche Gartenbauwissenschaftliche Gesellschaft



Results

OpenAIRE - Data Searchery

Home

eukaryotic cell

Hits on datacite (total: 59)

Prev

Next

"3-D Ultrastructure of : Electron Cryotomography of an Entire Eukaryotic Cell-7"



Gregory P. Henderson, Lu Gan, Grant J. Jensen

Subjects:

Description:

Publisher: Figshare

DOI: 10.6084/M9.FIGSHARE.60374

Published in: 2011

"Proteins injected by the bacterial pathogen "Bartonella" subvert eukaryotic cell signaling"



Guye-Vuillème, Patrick

Subjects:

Description: Diss. Phil.-Nat. Univ. Basel, 2006. - Ref.: Christoph Dehio, Michael Hall, Guy Cornelis

Publisher: Basel, [s.n.]

DOI: 10.5451/UNIBAS-004044162

Published in: 2006

"Picophytoplankton cell counts at station TT049_1-1"



Olson, Robert

Subjects: Synechococcus, Prochlorococcus, Phytoplankton, ultra eukaryotic, Bottle number, TT049, Thomas G. Thompson, Joint Global Ocean Flux Study (JGOFS)

Description:

Publisher: PANGAEA - Data Publisher for Earth & Environmental Science

DOI: 10.1594/PANGAEA.117335

Published in: 2003

"Picophytoplankton cell counts at station TT049_10-2"



Olson, Robert

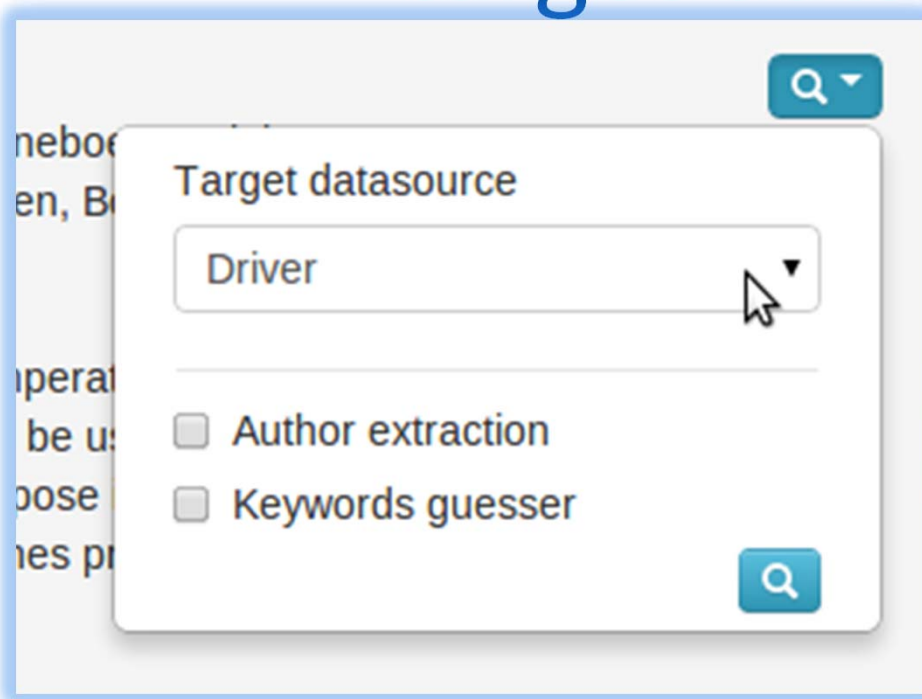
Subjects: Synechococcus, Prochlorococcus, Phytoplankton, ultra eukaryotic, Bottle number, TT049, Thomas G. Thompson, Joint Global Ocean Flux Study (JGOFS)

Description:

Publisher: PANGAEA - Data Publisher for Earth & Environmental Science

DOI: 10.1594/PANGAEA.117336

Selecting the target data source



Visualization of the result

OpenAIRE - Data Searchery

Home

eukaryotic cell

Hits on datacite (total: 59)

Prev

Next

"3-D Ultrastructure of : Electron Cryotomography of an Entire Eukaryotic Cell-7"

Gregory P. Henderson, Lu Gan, Grant J. Jensen

Subjects:

Description:

Publisher: Figshare

DOI: 10.6084/M9.FIGSHARE.60374

Published in: 2011

Target datasource

Driver

- ☒ Author extraction
- ☒ Keywords guesser

"Proteins injected by the bacterial pathogen "Bartonella" subvert eukaryotic cell signaling"

Guye-Vuillème, Patrick

Subjects:

Description: Diss. Phil.-Nat. Univ. Basel, 2006. - Ref.: Christoph Dehio, Michael Hall, Guy Cornelis

Publisher: Basel, [s.n.]

DOI: 10.5451/UNIBAS-004044162

Published in: 2006

"Picophytoplankton cell counts at station TT049_1-1"

Olson, Robert

Subjects: Synechococcus, Prochlorococcus, Phytoplankton, ultra eukaryotic, Bottle number, TT049, Thomas G. Thompson, Joint Global Ocean Flux Study (JGOFS)

Description:

Publisher: PANGAEA - Data Publisher for Earth & Environmental Science

DOI: 10.1594/PANGAEA.117335

Published in: 2003

"Picophytoplankton cell counts at station TT049_10-2"

Olson, Robert

Subjects: Synechococcus, Prochlorococcus, Phytoplankton, ultra eukaryotic, Bottle number, TT049, Thomas G. Thompson, Joint Global Ocean Flux Study (JGOFS)

Description:

Publisher: PANGAEA - Data Publisher for Earth & Environmental Science

DOI: 10.1594/PANGAEA.117336

Hits on driver (total: 1) Query

Prev

Next

"3-D Ultrastructure of O. tauri: Electron Cryotomography of an Entire Eukaryotic Cell"

Henderson, Gregory P. - Gan, Lu - Jensen, Grant J.

Subjects: Research Article

Description: The hallmark of eukaryotic cells is their segregation of key biological functions into discrete, membrane-bound organelles. Creating accurate models of their ultrastructural complexity has been difficult in par

Publisher:

DOI: 1-96340a52-e709-4692-b2dc-

66424fdc10c1_UmVwb3NpdG9yeVNlcnZpY2VSZXRvdXJjZXR5cGU=:oai:ukpmc.ac.uk:1

Published in:

Prev

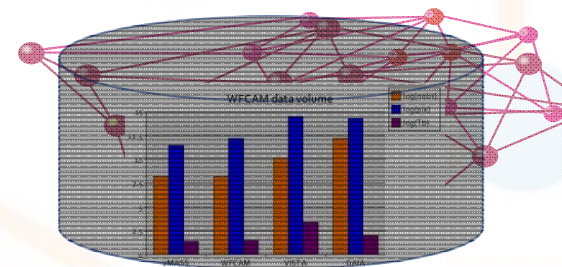
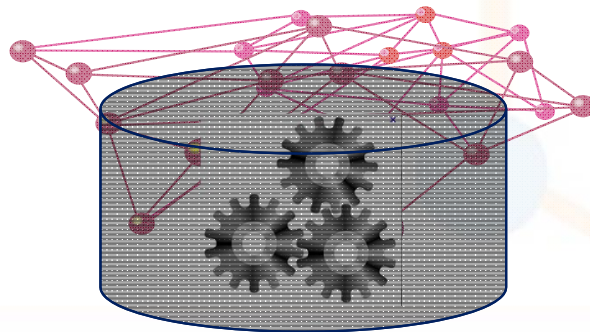
Next

Working in progress & plans

- ⬢ Configurable criteria
- ⬢ More selective algorithms
- ⬢ Authority Registries in identifying relationships
 - ⬢ of general applicability (e.g. ORCID)
 - ⬢ applicable to specific domains/collections (e.g. Catalogue of Life)

Main Challenge: Interoperability

- ✧ Discovery in an open set of data sources/data infrastructures
- ✧ Dichotomy between digital libraries/repositories, data centers & sw repositories



THANK YOU

donatella.castelli@isti.cnr.it