Co-ordination & Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing

The CHAIN-REDS Programme on Data Infrastructures

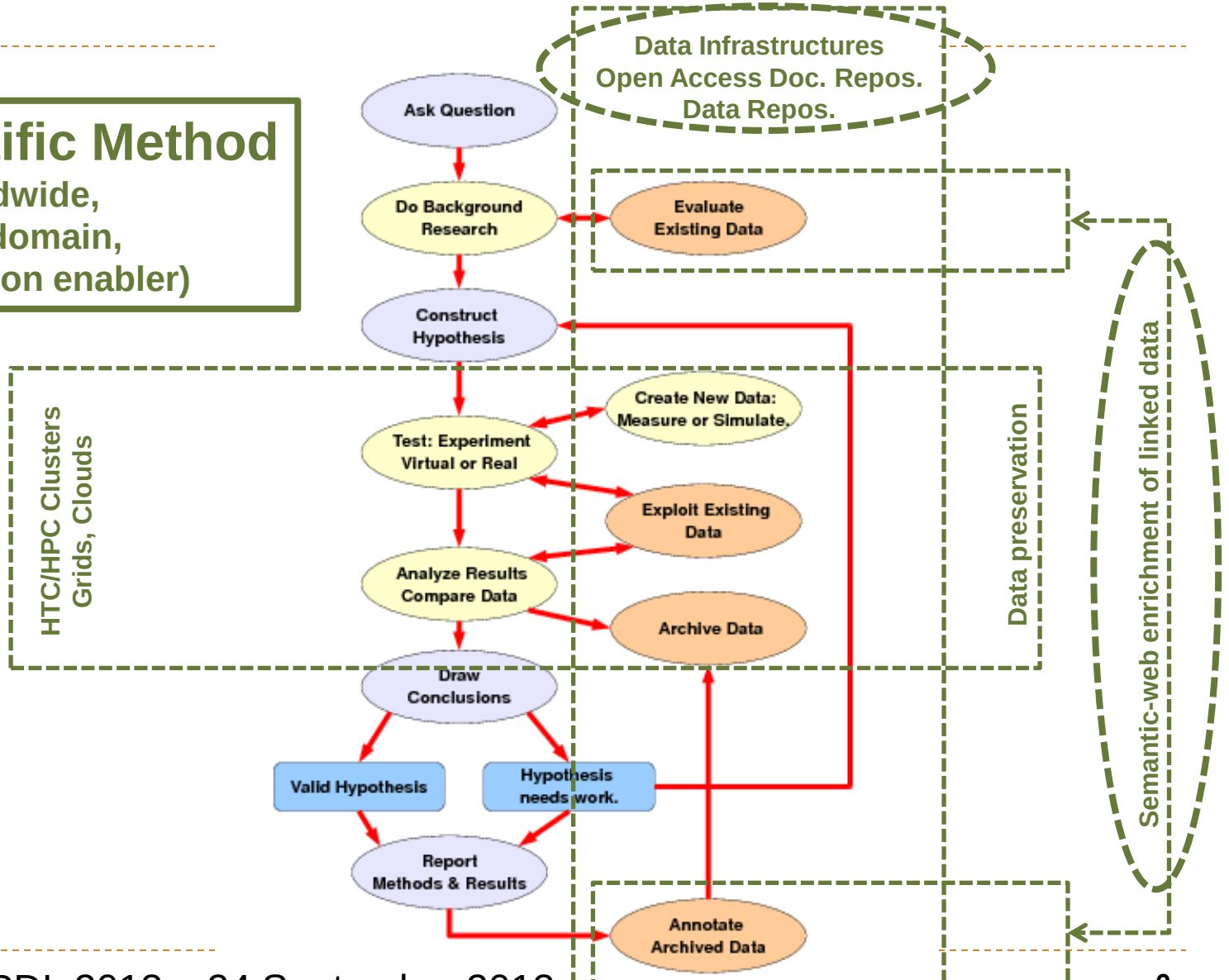
Panel @ TPDL 2013 – Malta, 24 September 2013

Roberto Barbera – University of Catania and INFN
(roberto.barbera@ct.infn.it)

Research Infrastructures – Proposal n. 306819

What do researchers do and how can e-Infrastructures help them to improve it ?

The Scientific Method
(worldwide,
cross-domain,
collaboration enabler)



The CHAIN-REDS Project
(www.chain-project.eu)

Started: 1 Dec 2012
Duration: 30 months
Targeted regions: Africa, Middle-East and Gulf Region, Latin America, China, India, and Far-East Asia

Objectives

- Extend and consolidate the international cooperation of Europe with other regions of the world in the domain of e-Infrastructures for R&E, thus reinforcing the scientific collaboration and broadening the impact of the European Research Area
- Promote, coordinate and support the effort of a critical mass of non-European e-Infrastructures for R&E to collaborate with Europe by addressing interoperability and interoperation of Grids and other DCIs such as potential upcoming Cloud federations and HPC centres
- Study the opportunities of data sharing across different e-Infrastructures and continents thus widening the scope of the existing CHAIN Knowledge Base to Data Infrastructures and Cloud implementations
- Promote trust-building towards open Scientific Data infrastructures across the world regions, including organisational, operational and technical aspects
- Demonstrate the relevance of intercontinental cooperation in several scientific data fields addressing existing and emerging Virtual Research Communities (e.g. Earth Science, Climate Change, Genomics, etc.) and propose pragmatic approaches that could impact the everyday work of the single researcher, even if not structured in the framework of a Virtual Research Community
- Provide guidance and recommendations for roadmaps for long-term global collaboration in e-Infrastructures and harmonisation of existing policies. These are envisaged to act as input to policy and decision-making mechanism, harmonised with the European Digital Agenda and Horizon 2020

CHAIN-REDS program for Data Infrastructures

- Identify standards to easily gather and access both Open Access Document Repositories (OADRs) and Data Repositories (DRs)
- Build a demonstrator to easy visualise and access OADRs and DRs (both geo-views and tab-views)
- Correlate OADRs and DRs to create linked data and discover new knowledge through semantic enrichment of metadata
- Promote Data Infrastructure standards and identify new OADRs and DRs from regions addressed by the project (Africa, Middle-East and Gulf Region, Latin America, China, India, Far-East Asia)
- Populate the demonstrator with these new repositories, add them to the semantic enrichment tool, and set-up at least two

Open Access Document Repositories (OADRs)

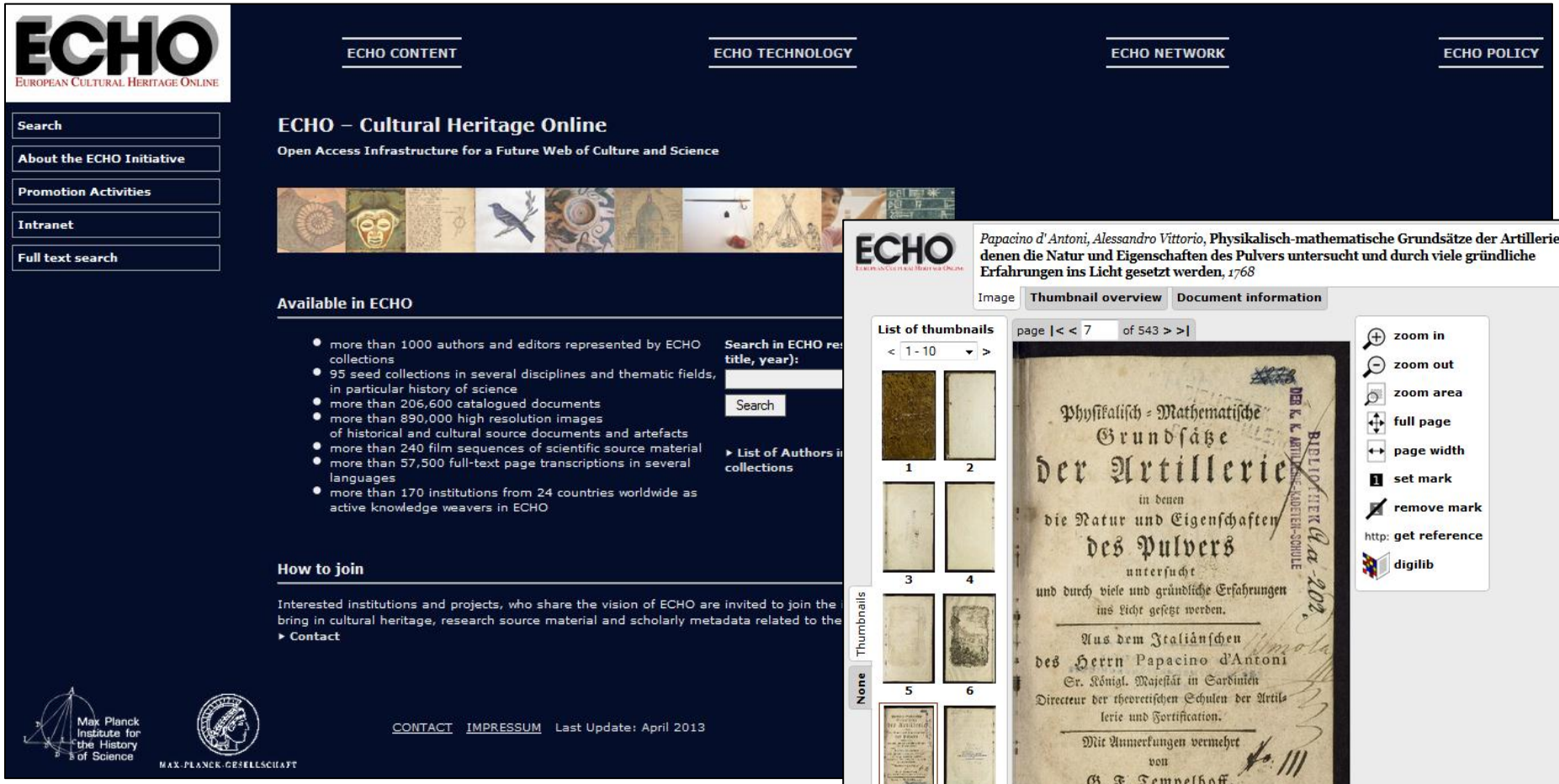
(www.chain-project.eu/oadr-map)

- ~2,500 repos
- >33 M docs

Registry of Open Access Repositories (ROAR)

driver OpenAIRE OpenDOAR

Open Access Document Repositories (OADRs)



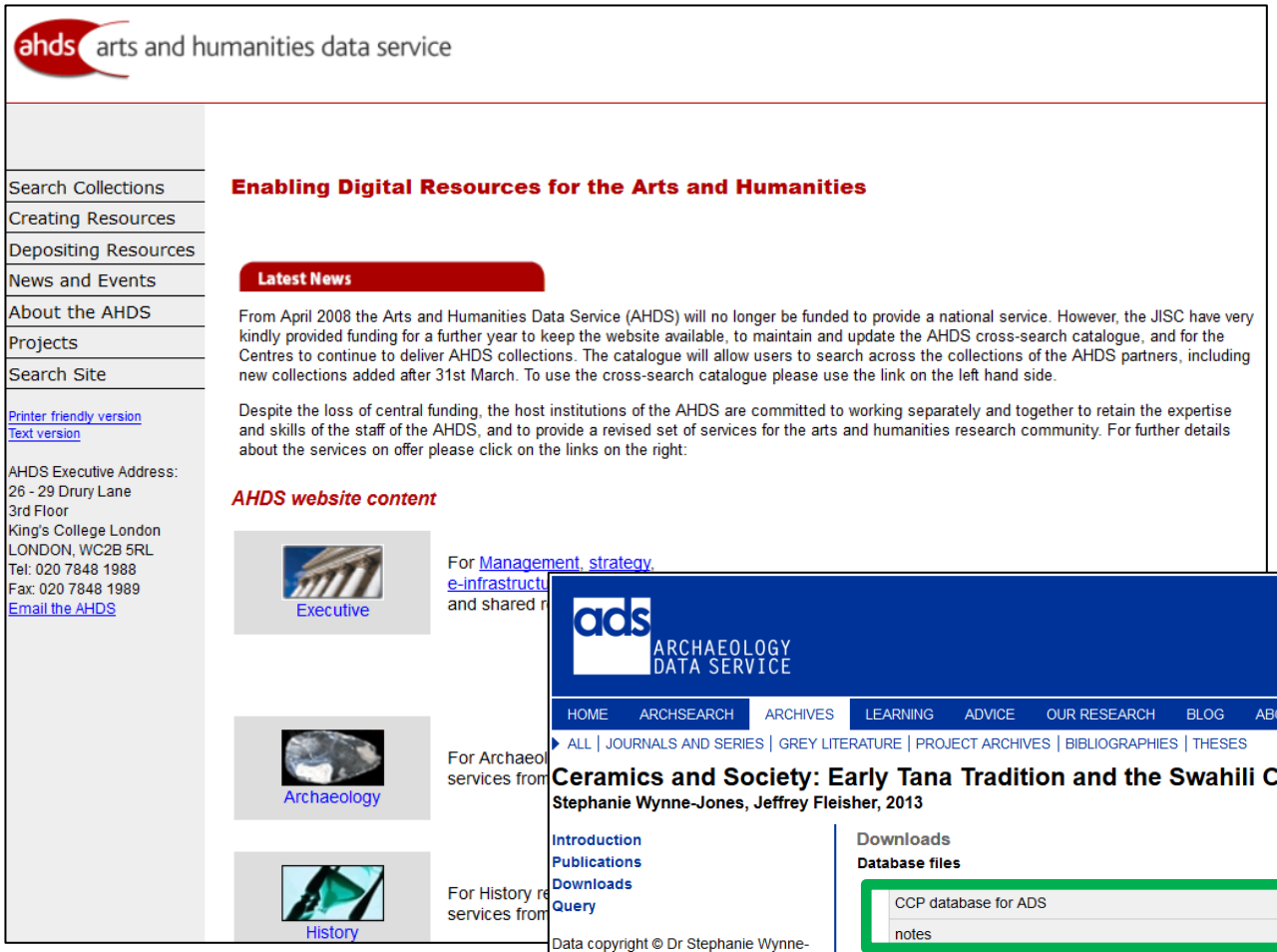
Panel @ TPDL 2013 – 24 September 2013

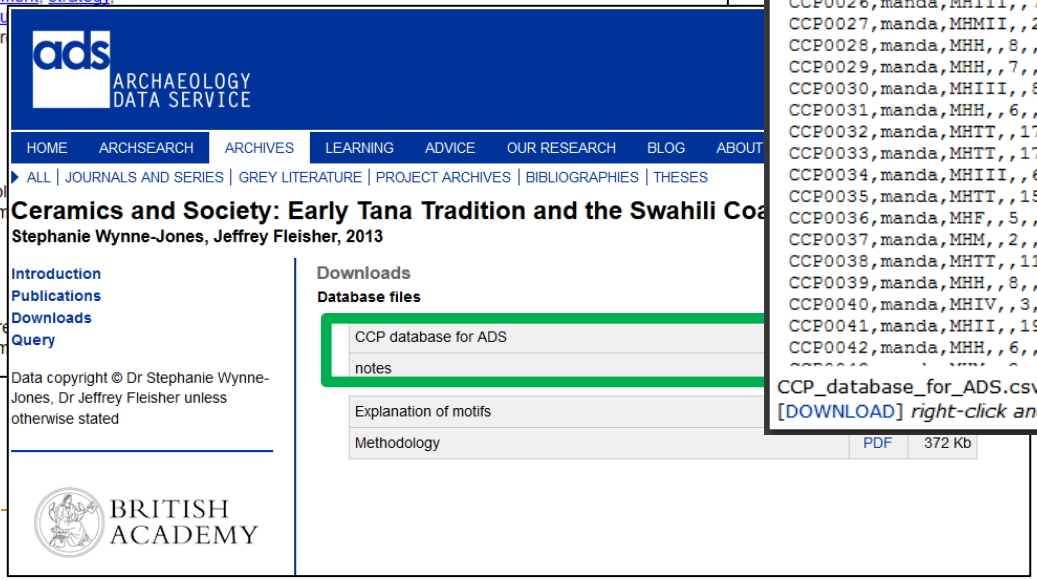
Data Repositories (DRs)

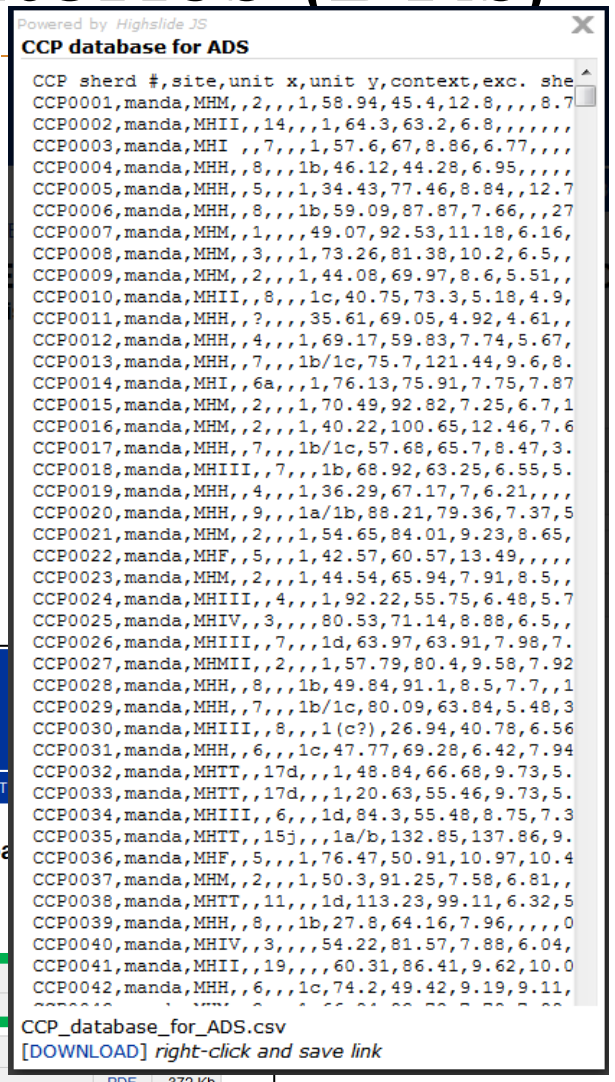
(www.chain-project.eu/oadr-map)

- ~ 600 repos
- Lots of data !

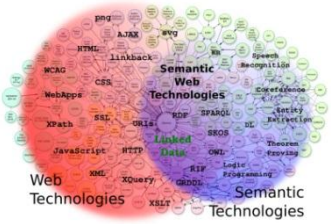
Data Repositories (DRs)







Linked Data



Linked-data search engine

Semantic-web enrichment

Harvester (running on grid/cloud)

Harvester (running on grid/cloud)

OADRs

OAI-PMH

Dublin Core

Registry of Open Access Repositories (ROAR)

End-points

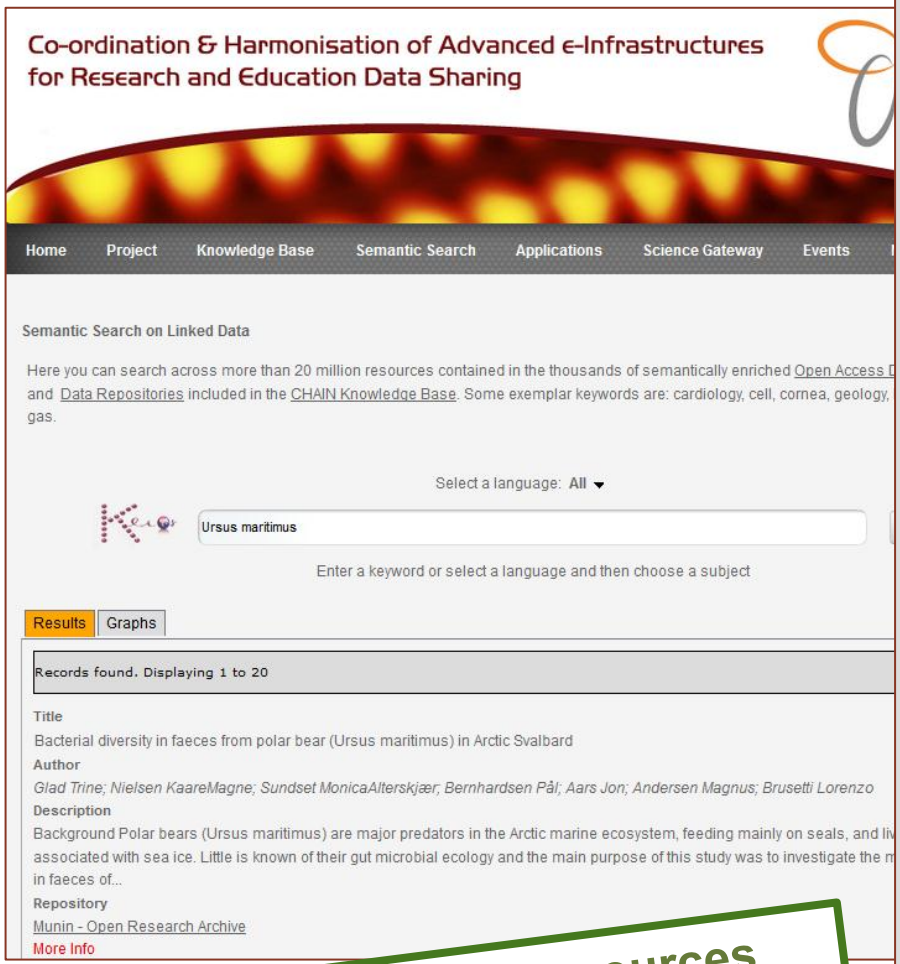
OAI-PMH

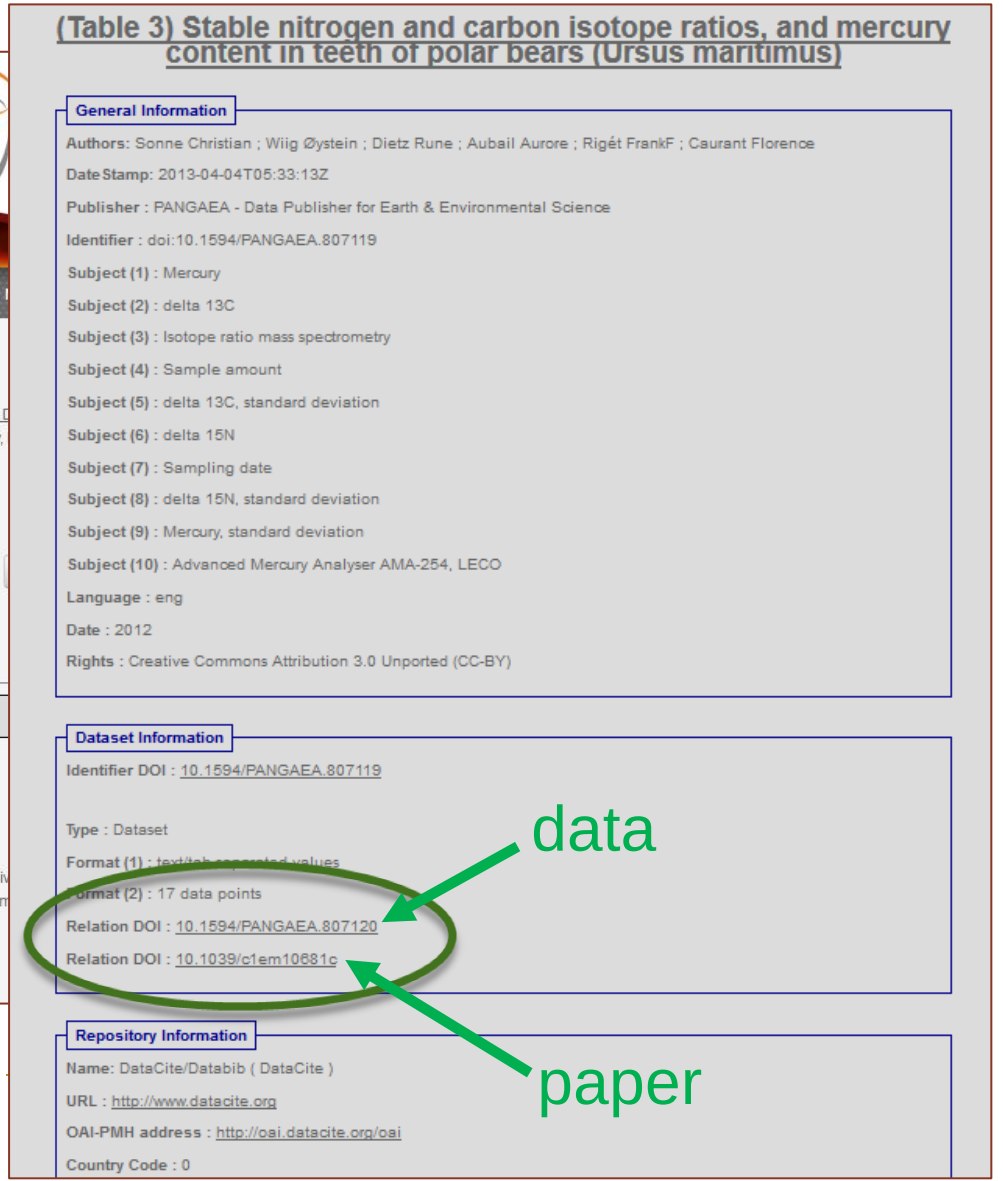
Dublin Core

Data Repos.

Linked data semantic search

(www.chain-project.eu/linked-data)

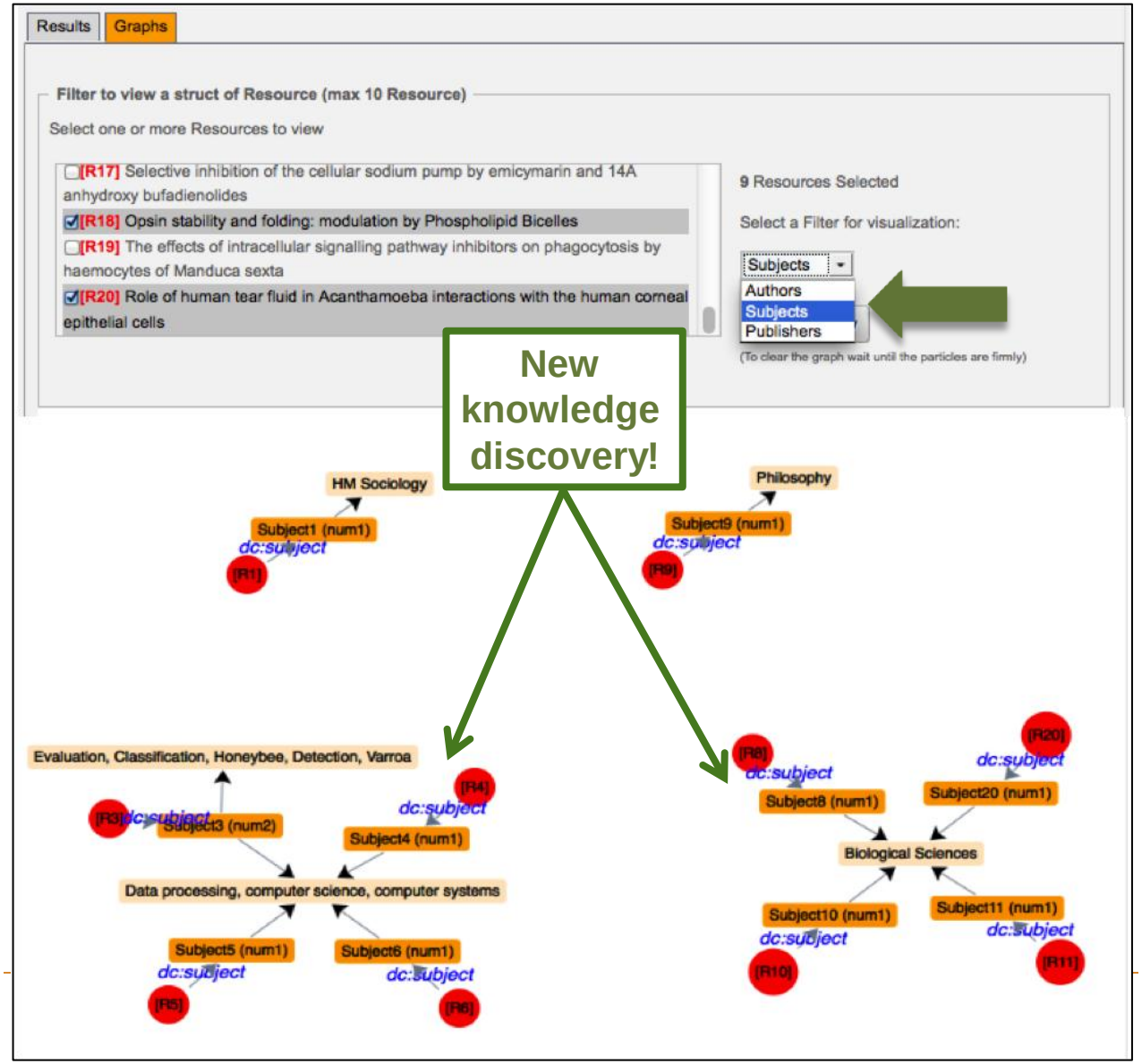




>22 M resources
~600 M triples

Linked data semantic search

(www.chain-project.eu/linked-data)



Summary and conclusions

- Data Infrastructures are becoming an essential component of e-Infrastructures
- Next years' biggest challenges will be
  - To build reliable, sustainable and up to date data discovery services
  - To uniquely correlate scientific papers with data used to write them with applications used to analyse them so to be able to go across the knowledge path both ways
- Semantic-web and linked-data technologies can play a major role in this context and CHAIN-REDS aims to promote these standards in the targeted regions
- OADRs' and DRs' managers/owners are welcome to share their data within the CHAIN Knowledge Base (both in Africa and Latin America this is already happening)

Co-ordination & Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing

Questions ?

Research Infrastructures – Proposal n. 306819