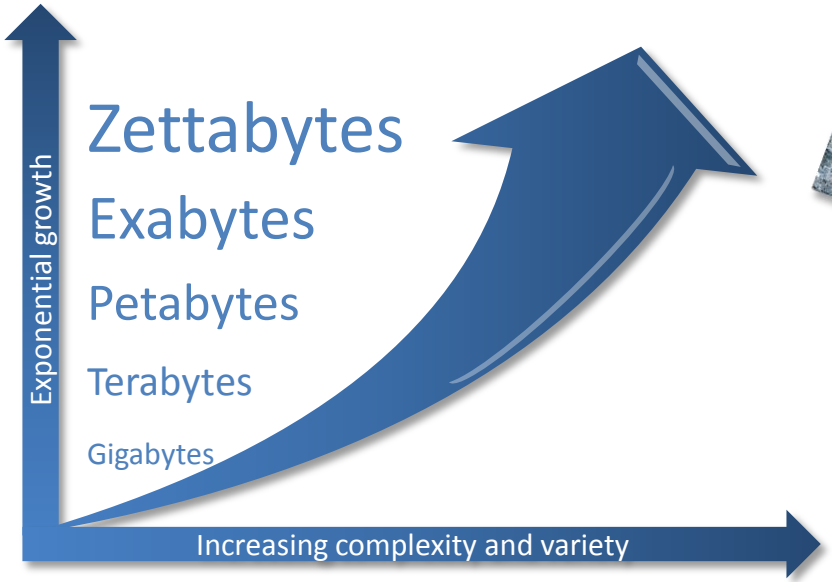
EUDAT

Standardized storage services for European research communities

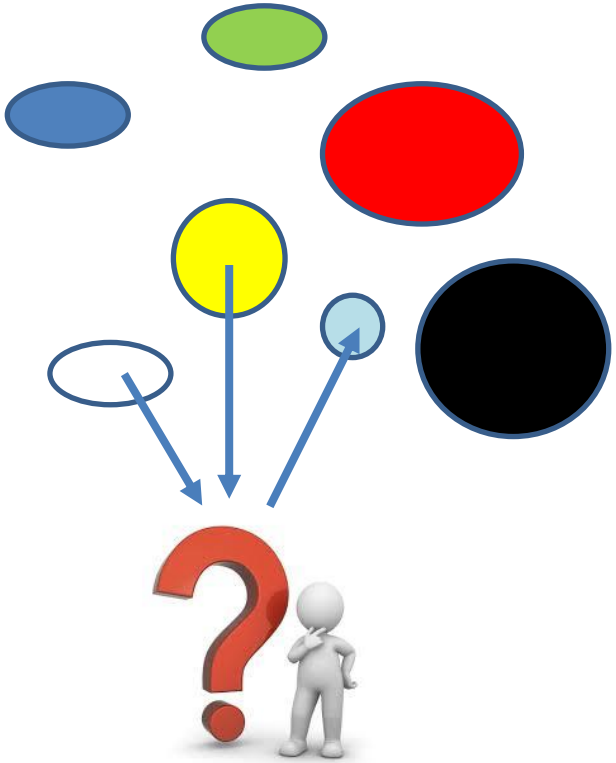
Dr. Ari Lukkarinen
CSC – IT Center for Science Ltd

Data ”Deluge”



EUDAT

Isolated Data Repositories



EUDAT

Data Reuse ?

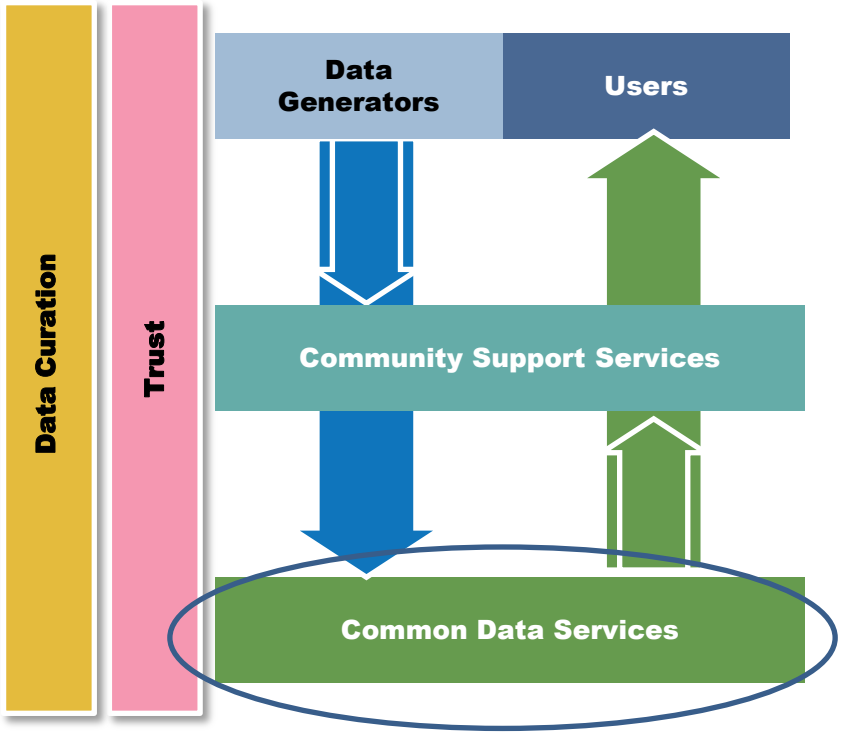
- EXIST – Bits will have to be somewhere
- SEARCHABLE – How to find the data ?
- PERMISSIONS – Data is mostly open, but
- FORMAT – Can I use xyz format ?

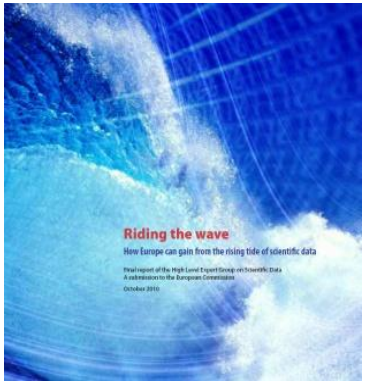
- POINTER - Data citation, PIDs
- APIs - ftp, http, REST, etc.

- OLD REPOSITORIES !

EUDAT

Collaborative Data Infrastructure





EUDAT

Consortium

26 European partners

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

KIT Karlsruhe Institute of Technology

INGV

CERN

epcc

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

DKRZ DEUTSCHES KLIMARECHENZENTRUM

INES Centre Informatique National de l'Enseignement Supérieur

CINECA

CSC

RedIRIS

Max-Planck-Institut für Meteorologie

JÜLICH FORSCHUNGSZENTRUM

rzg RECHEN-ZENTRUM GARCHING

Science & Technology Facilities Council

maatG

CERFACS

umweltbundesamt ENVIRONMENT AGENCY AUSTRIA

POZNAŃ

SNIC

Trust-IT Services Ltd Communicating ICT to markets

UNINETT sigma

SURF SARA

UCL

EUDAT

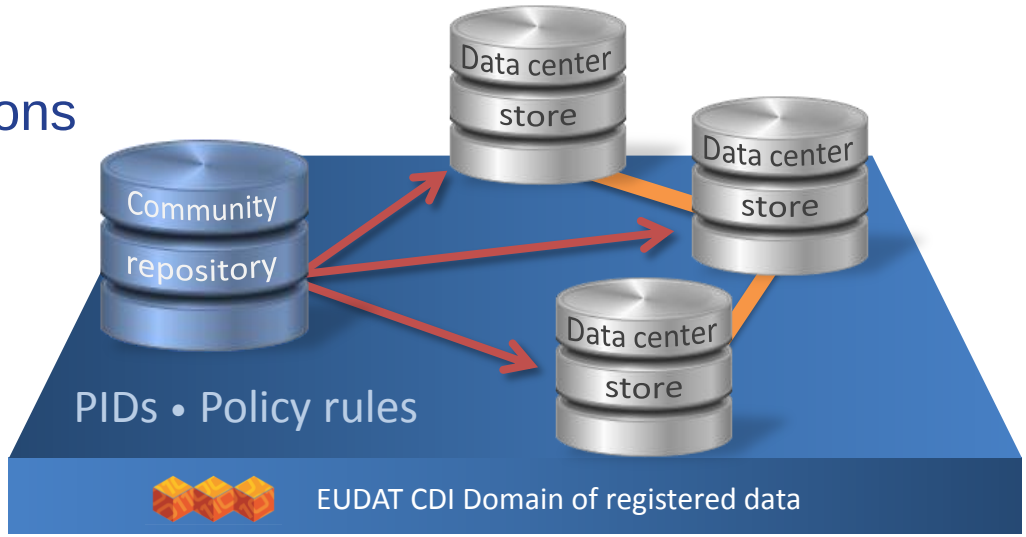
User Driven

- **Five research communities initially on board**
  - Coming from wide range of disciplines
  - Requirement gathering
  - Involved in developing, testing and operating the first services
- **Several others joined and/or are involved in discussion and service pilots**
  - Requirements gathering
  - Customisation of services
  - Pilots

EUDAT

Safe Replication Service (B2SAFE)

- Robust, safe and highly available data replication service for small- and medium- sized repositories
  - To guard against data loss in long-term archiving and preservation
  - To optimize access for user from different regions
  - To bring data closer to powerful computers for compute-intensive analysis

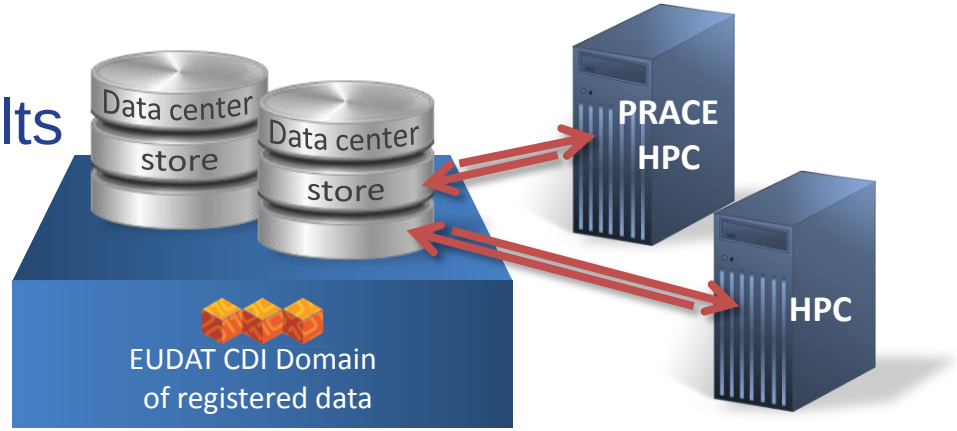


http://eudat.eu/safe-replication | eudat-safereplication@postit.csc.fi

EUDAT

Data Staging Tools (B2STAGE)

- Support researchers in transferring large data collections from EUDAT storage to HPC facilities
- Reliable, efficient, and easy-to-use tools to manage data transfers
- Provide the means to re-ingest computational results back into the EUDAT infrastructure

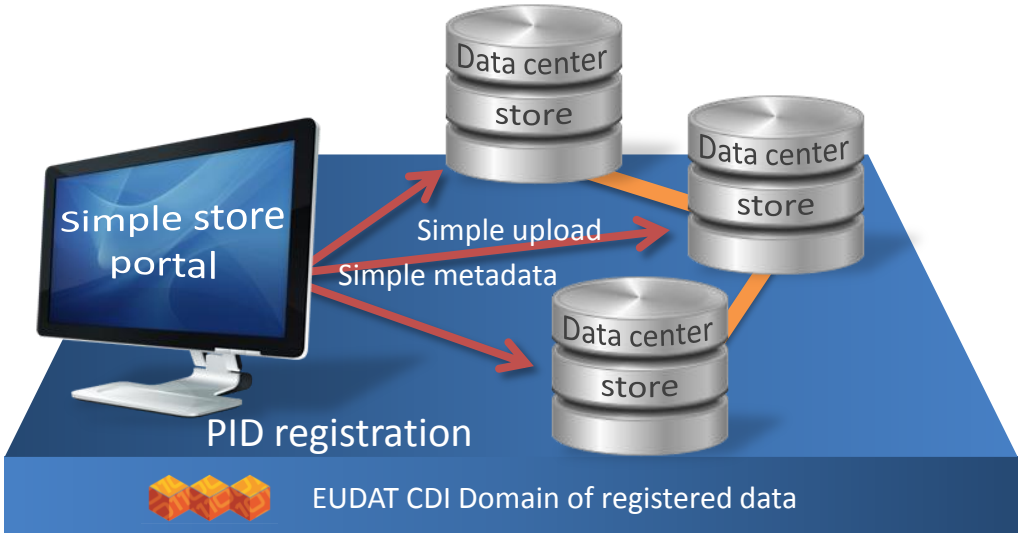


http://eudat.eu/datastaging | eudat-datastaging@postit.csc.fi

EUDAT

Simple Store Service (B2SHARE)

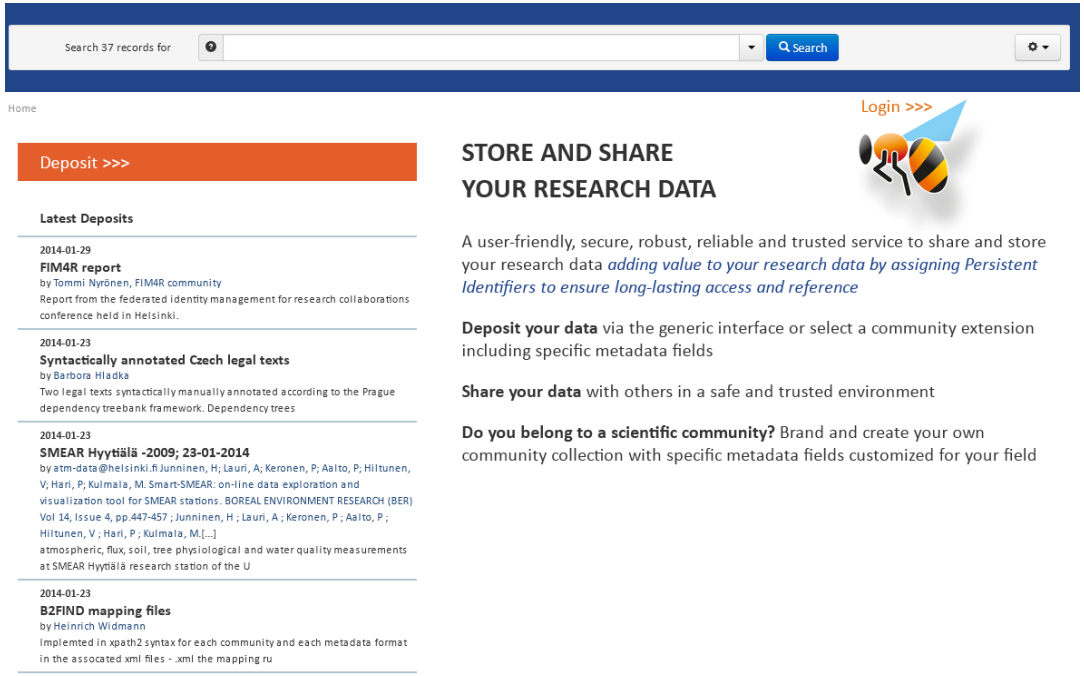
- Allow registered users to upload ”long tail” data into the EUDAT store
- Enable sharing objects and collections with other researchers
- Utilise other EUDAT services to provide reliability and data retention



http://eudat.eu/simplestore | eudat-simplestore@postit.csc.fi

EUDAT

B2SHARE: an Open Research Data Repository



- Easy deposit service for *storing* and *sharing* research data
- Targets the “long tail” of “small” data
- Offered to all European researchers and citizen scientists

https://b2share.eudat.eu/

EUDAT

A four-click service

Step 01

Drag and drop files here

Select files

Start upload

Filename	Size	Status	Remove
3 CSC_TUTA_REPORT_10-2011.pdf	390 KB		

Step 02

Select a domain or project

Generic | Ontology | Hydro-Meteorology | Linguistics | Biomedical Research

EUDAT | euon | DRIHM | CLARIN

Community specific metadata fields

Metadata fields common to all

Add basic details

Generic

Add more details?

Title *

Description *

Creator

Open Access ON

Licence

Publisher

Publication Date

Tags

Linguistics

Language Code *

Country/Region

Resource Type *

Project Name

Quality

* indicates required field

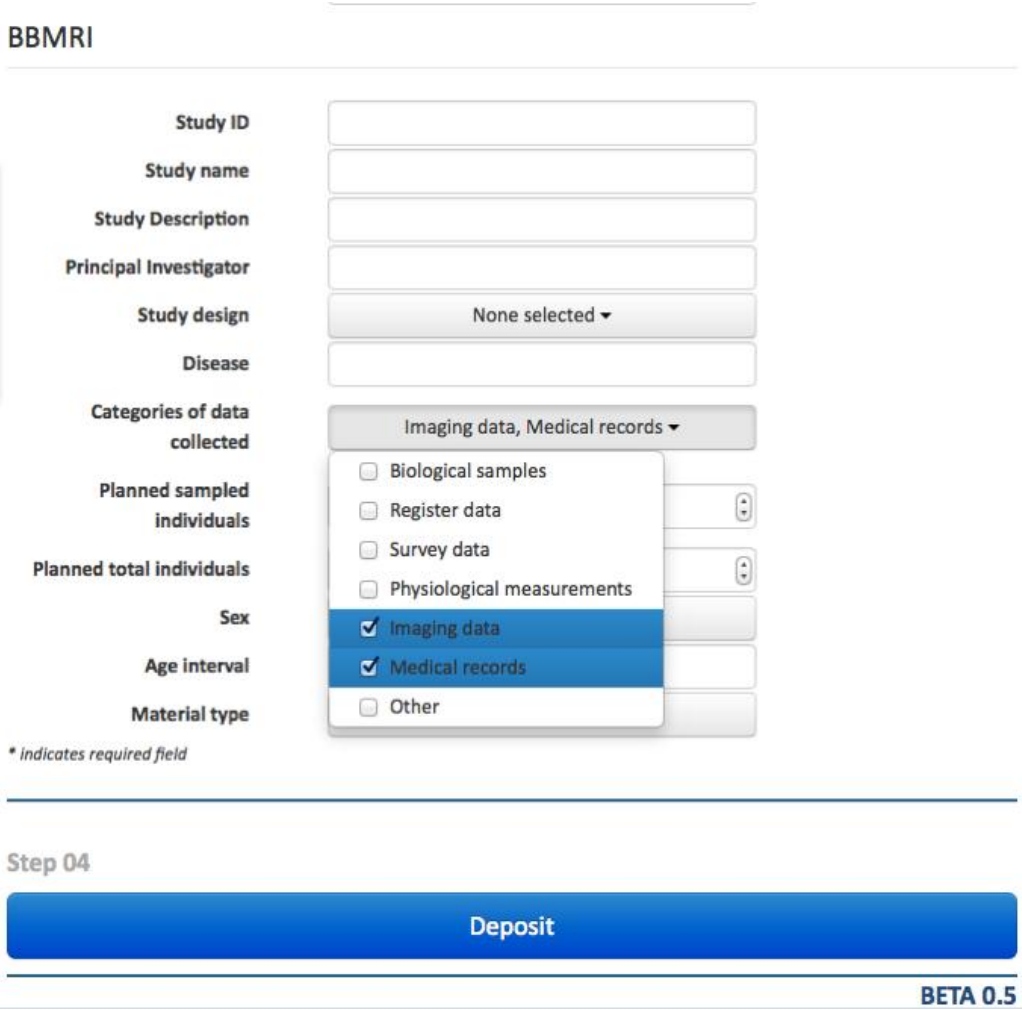
Step 04

Deposit

EUDAT

Example of Community Extension

The type of data collected in the study, and if biological samples are part of the study. Can be one or several of the following values: Biological samples, Register data, Survey data, Physiological measurements, Imaging data, Medical records, Other



EUDAT

Metadata Service (B2FIND)

- Easily find collections of scientific data – generated either by various communities or via EUDAT services
- Access those data collections through the given references in the metadata to the relevant data stores
- Europeana of scientific data

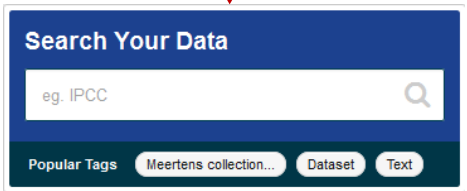
http://eudat.eu/metadata | eudat-metadata@postit.csc.fi

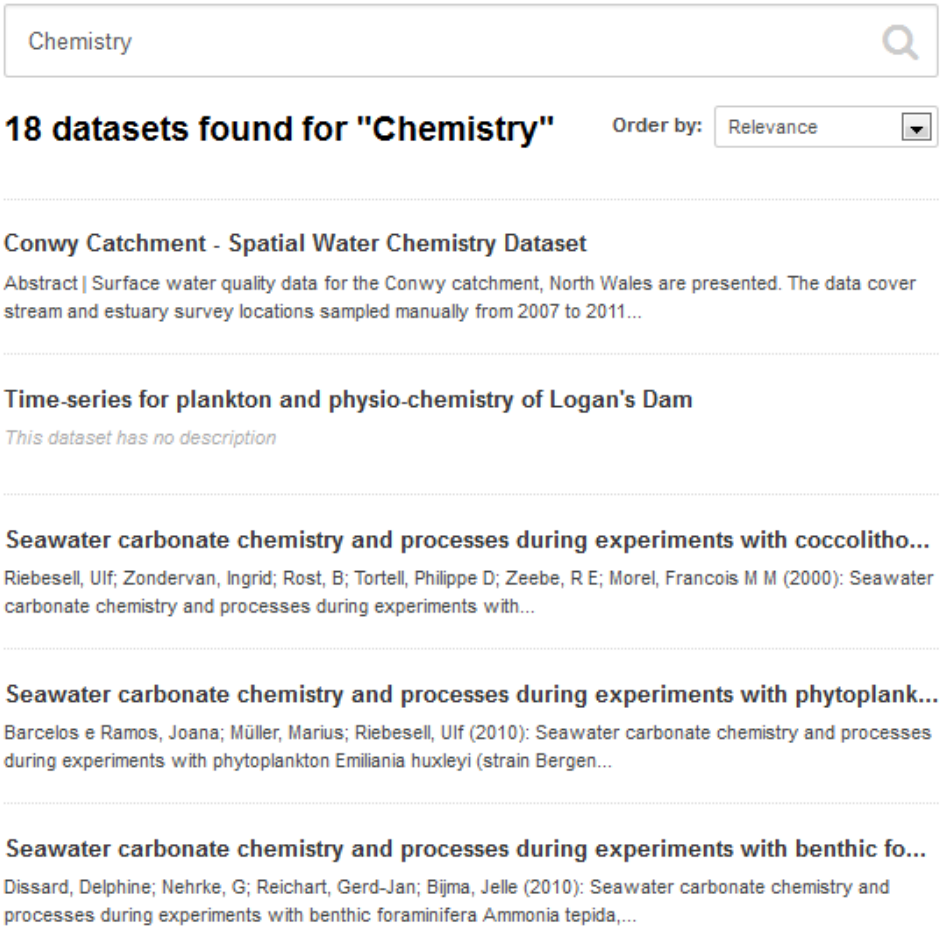
EUDAT

Search and browse datasets

Search and browse all data sets via Keyword searches

Results displayed in easy to read format and listed in order of relevance to your search





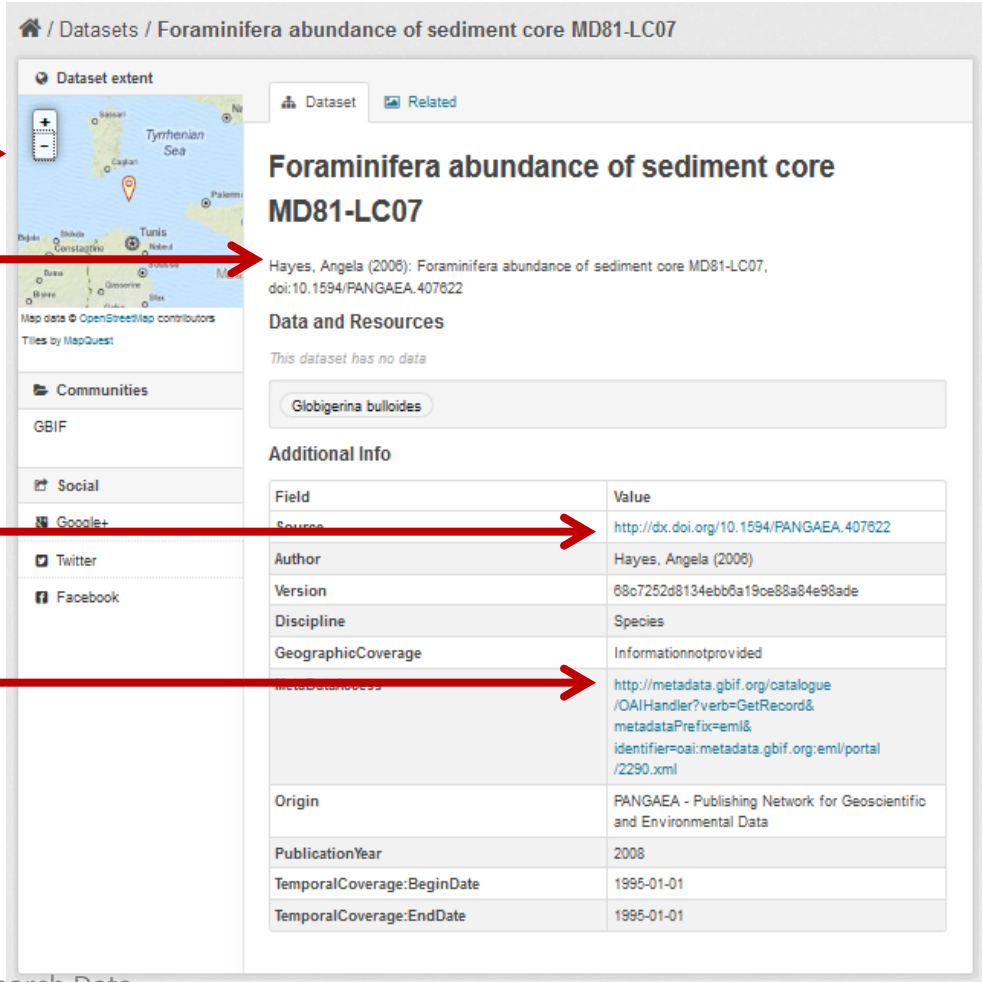
Detailed information on all data sets

Spatial coverage

Abstract

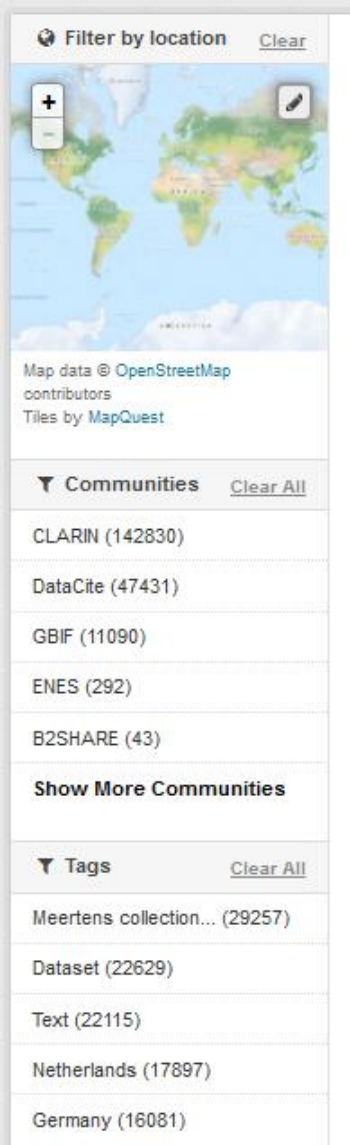
Detailed metadata

- Author
- Version
- Discipline
- Source
- Format
- Geographical description
- Metadata access
- Origin
- Publication Time stamp
- Year of publication
- Rights



Search features

- **Keyword search**
- **Filter by**
  - Communities
  - Location
  - Tags
  - Author
  - Discipline
  - Language
  - Origin
  - Year of publication
  - Format



Download datasets

Simply click on data set "Metadata access" to access data.

English House Condition Survey, 1986: Physical Survey

Data and Resources

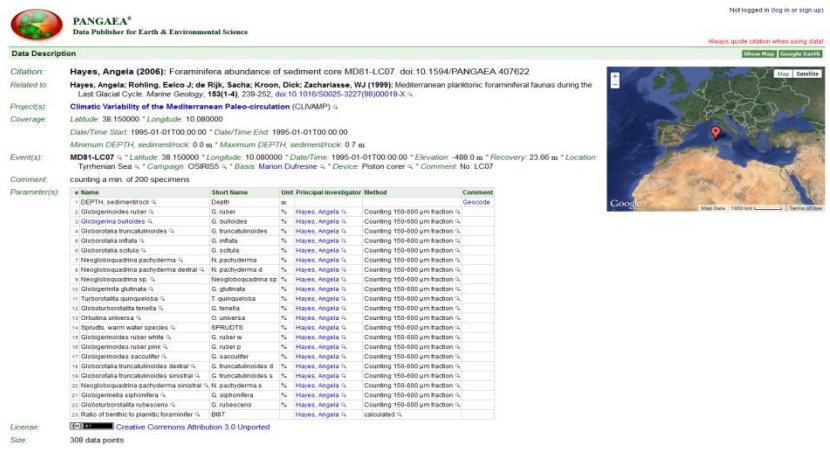
This dataset has no data

Dataset

Additional Info

Field	Value
Source	http://dx.doi.org/10.5255/UKDA-SN-2995-1
Author	Department of the Environment. Building Stock Research Division
Version	8283dd3c063cee5e9161278c3a1945e1
MetaDataAccess	http://oai.datacite.org/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:oai.datacite.org:58034
Origin	Colchester, Essex: UK Data Archive
PublicationTimestamp	1995-07-01T11:59:59Z
PublicationYear	1995

Resolved link to data object



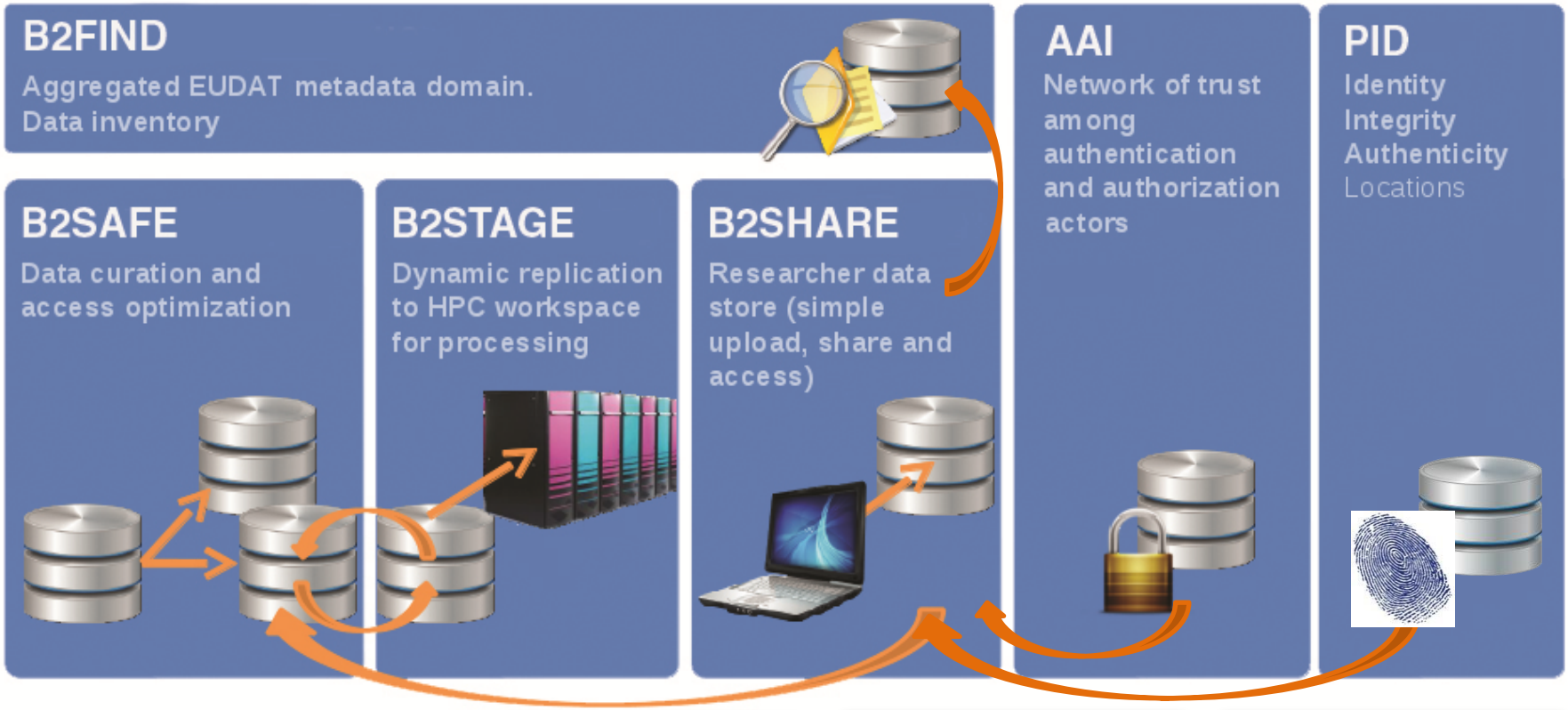
Resolved link to harvested metadata


```
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <responseDate>2014-01-20T09:43:23Z</responseDate>
    <request identifier="oai:metadata.gbif.org:eml/portal/2290.xml" metadataPrefix="eml" verb="GetRecord">http://metadata.gbif.
  - <GetRecord>
    - <record>
      - <header>
          <identifier>oai:metadata.gbif.org:eml/portal/2290.xml</identifier>
          <datestamp>2011-08-17T15:58:07Z</datestamp>
        </header>
      - <metadata>
        - <!--
            These documents can be harvested into a metacat instance using the harvest list at
          /datasets/resource/harvestList.xml
          -->
        - <eml:eml packageId="eml.gbif.package.2290" system="gbif" xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 h
          - <dataset>
              <alternateIdentifier>2290</alternateIdentifier>
              <title>Foraminifera abundance of sediment core MD81-LC07</title>
            - <creator>
              - <organizationName>
                  PANGAEA - Publishing Network for Geoscientific and Environmental Data
```

NEXT STEPS

- Review current design
- New services
- Form pilots to production
  - Certifications (Data Seal of Approval)
  - Uniform service production practices (ITIL, FED-SM)
  - Policies, legal issues
  - Sustainability (cost models, etc.)

Integrated Solution



New services to come

EUDAT

Thank you

EUDAT