

# Lecture 21 – Repositories

*Open Data Management & the Cloud*

(Data Science & Scientific Computing / UniTS – DMG)

# Real argument of the lesson



- Storage
- Repository
- Registry/catalogue
- Archive

What are?

How we use them?

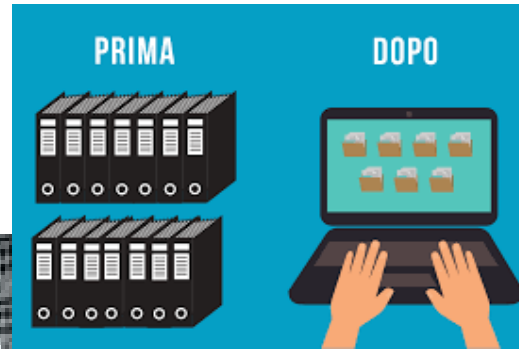
Examples

# Analogy





# Analogy



# Repository – What is it?



- Data storage / Data base entity/ies into which data has been specifically partitioned for an analytical or reporting purpose
  - Data warehouse
    - a large data repository that aggregates data
    - usually from multiple sources or segments of a business
    - without the data being necessarily related
  - Data lake (needs disambiguation)
    - a large data repository that stores unstructured data
    - classified and tagged with metadata
  - Data mart
    - subsets of the data repository
    - more targeted to what the data user needs and easier to use
  - Data library
  - Data archive
- Particular kind of setup within an overall IT structure, such as a group of databases, where an enterprise or organization has chosen to keep various kinds of data
- Metadata repositories store data about data and data bases
  - Where the data source is, how it was captured, and what it represents

# Repository – Some disambiguation



- Repository has additional functionalities compared with registry
  - Registries stores metadata
  - Repositories add relationships with related metadata types
  - Distinction is really loose / hardly enforced
- Disambigua to collections
  - Grouping of digital objects, within a scope
  - Collections are themselves resources in repositories
- What is the difference between archive and repository?

<https://wikidiff.com/archive/repository>

<https://www.usgs.gov/data-management/archive-vs-repository-there-difference>

<https://www.familytreemagazine.com/libraries-archives/library-repository-archives/>

<https://dictionary.archivists.org/entry/repository.html>

# Repository – Goals



- FAIR principles: Findable
  - “F4. (meta) data are registered or indexed in a searchable resource.”
- Keep a certain population of data isolated so that it can be mined
  - isolated: living in its one fragmented/distributed source
  - For greater insight
  - For business intelligence
  - To be used for a specific (reporting) need
- Isolation allows for easier and faster data reporting or analysis because the data is clustered together
  - Not necessarily by location
- Data is preserved and archived



- Data repositories
  - Index data
  - Provide access to
    - Data collections
    - Datasets
  - (usually) keep meta information dedicated to filtering
- Metadata repositories
  - (usually) don't provide storage space
  - Use linking solutions to point or access data
  - Provide rich metadata documents
    - (usually) standardized
    - More general in scope
      - discovery/filtering
      - relationships



# Repository – Usage




- There are more and more data open and available through data repositories: it becomes ever more challenging for researchers to find relevant data.
- Example
  - The Open Access Directory (OAD) is a compendium of simple factual lists about open access (OA) to science and scholarship, maintained by the OA community at large. By bringing many OA-related lists together in one place, OAD makes it easier for everyone to discover them, use them for reference, and update them.

# Repository – Usage



## ● There are more and more data open and available through data

### Data repositories

 This list is part of the [Open Access Directory](#).

- This is a list of repositories and databases for open data.
- Please annotate the entries to indicate the hosting organization, scope, licensing, and usage restrictions (if any). If a repository is open in some respects but not others, please include it with an annotation rather than exclude it.
- If you're not sure whether a given dataset or data collection is open, post your query to [Is It Open Data?](#)
- Related lists in OAD: [Disciplinary repositories](#) (primarily for texts, not data).
- For news about data repositories, including some newly launched repositories not yet listed here, follow the [oa.repositories.data](#) tag of the [Open Access Tracking Project](#).
- See also: [re3data.org](#). The re3data.org project intends to create a global registry of research data repositories.

### Archaeology

- *Also see* Social sciences.
- [Archaeology Data Service](#).
- [Fasti Online](#). Subdivided in Excavation, Restoration and Survey.
- [Open Context](#). From the [Alexandria Archive Institute](#).
- the [Digital Archaeological Record](#). From [Digital Antiquity](#).

### Astronomy

- *Also see* Physics.
- [Astronomical Data Archives Center](#). From the [National Astronomical Observatory of Japan](#).
- [Astrophysics Data System](#). From the [Smithsonian Astrophysical Observatory](#) (SAO) and [National Aeronautics and Space Administration](#) (NASA).
- [The Canadian Astronomy Data Centre](#). From the [National Research Council Canada](#).
- [National Space Science Data Center](#). From the US [National Aeronautics and Space Administration](#) (NASA).

### Biology

- *Also see* BCO-DMO, Marine Biology data, listed with Marine Sciences repositories.
- *Also see* DataONE, Entrez databases, KNB, and PANGAEA, listed under Multidisciplinary repositories.
- [The Arabidopsis Information Resource](#) - The Arabidopsis Information Resource (TAIR) maintains a [database](#) of genetic and [molecular biology data](#) for the model higher plant [Arabidopsis thaliana](#).
- [BOND](#) (Biomolecular Object Network Databank). From [Unleashed Informatics](#).
- [The Cell: An Image Library](#) Images of all cell types from all organisms, including intracellular structures and movies or animations demonstrating functions. This project relies upon the cell biology community to populate the library. The Cell: An Image Library™ is a freely accessible, easy-to-search, public repository of reviewed and annotated images, videos, and animations of cells from a variety of organisms, showcasing cell architecture, intracellular functionalities, and both normal and abnormal processes. The purpose of this

#### Contents [hide]

- 1 Archaeology
- 2 Astronomy
- 3 Biology
- 4 Chemistry
- 5 Computer Science
- 6 Energy
- 7 Environmental sciences
- 8 Geology
- 9 Geosciences and geospatial data
- 10 Linguistics
- 11 Marine sciences
- 12 Medicine
- 13 Multidisciplinary repositories
- 14 Physics
- 15 Social sciences

# Repository – Usage



- There are more and more data open and available through data repositories: it becomes ever more challenging for researchers to find relevant data.
- Example
  - The Open Access Directory (OAD) is a compendium of simple factual lists about open access (OA) to science and scholarship, maintained by the OA community at large. By bringing many OA-related lists together in one place, OAD makes it easier for everyone to discover them, use them for reference, and update them.
- Where to start from?
  - Use Google?
  - Use a specific “global” research index?
  - Find your domain starting point?



- **Data Discovery Paradigms** (Interest Group)
  - User Requirements and Recommendations for Data Repositories
    - <https://www.rd-alliance.org/group/data-discovery-paradigms-ig/outcomes/data-discovery-paradigms-user-requirements-and>
- Purpose
  - Help data repositories improve the findability of data in their repository
- Approach
  - Collected use cases describing users' needs, and the contexts of these needs, when searching for data
- Outcome
  - Identified requirements for data discovery in repositories
  - Proposed a set of recommendations



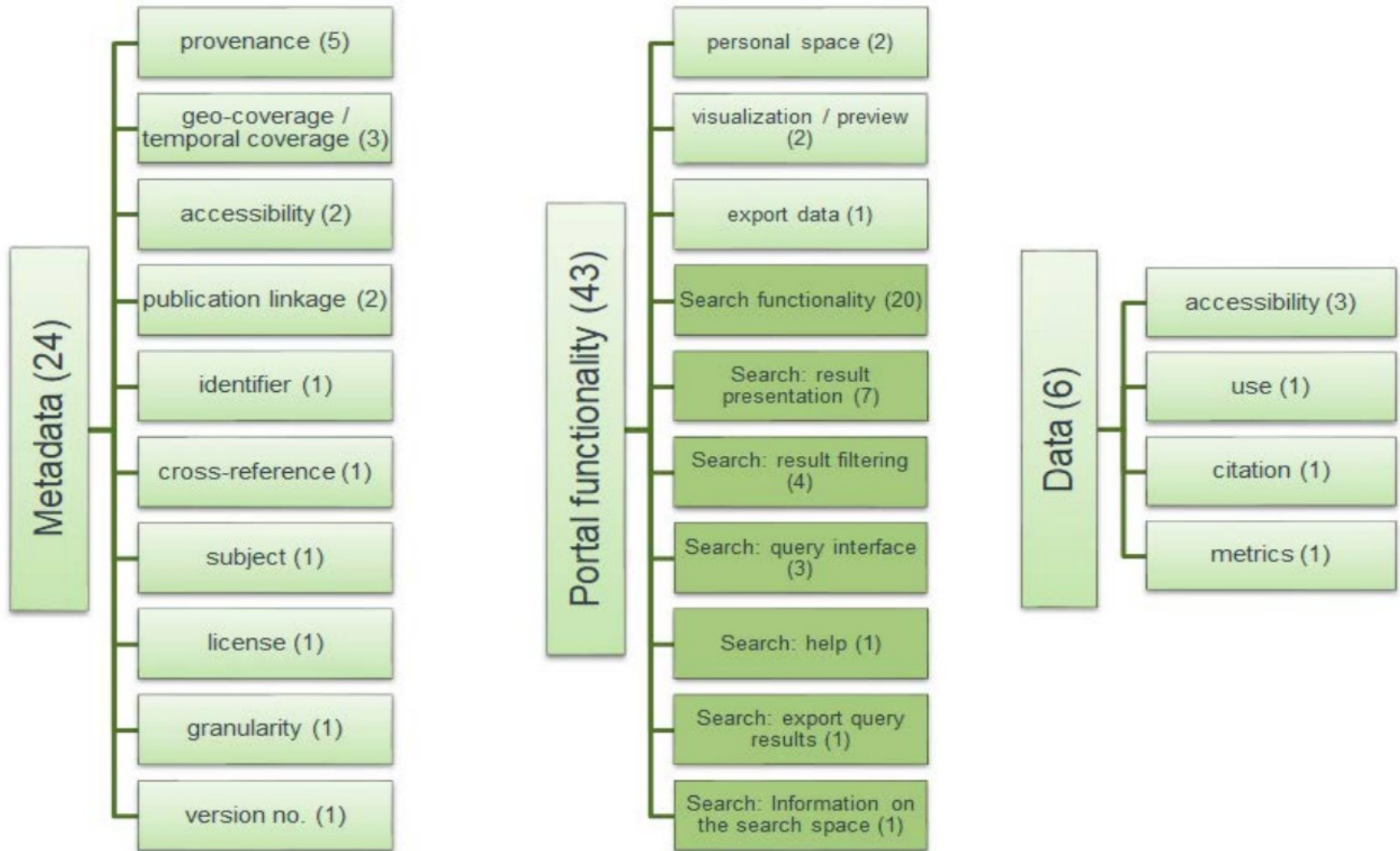
- Data repositories can adapt methodologies and learnt experiences from the design of web-based information systems, and digital library in particular
- Some data repositories have been following the path of user-centred system design principle gathering use cases and requirements
- Some design and evaluation criteria may apply only to a specific repository
- Some trans-repository criteria can be generalised and may serve as guidelines for other data repositories
- DDP-IG study attempts to identify which criteria are of common relevance



# RDA – DDP-IG (3) – Use cases



## classification scheme summary



# RDA – DDP-IG (4) – Requirements



- 1) Indication of data availability
  - Search usually dropped if no clear indication of data availability
- 2) Connection of data with person / institution / paper / citations / grants
  - allows for ranking of datasets, comparative studies, manuscript direct connection
- 3) Fully annotated data (including granularity, origin, licensing, provenance, ...)
  - validate the use of a dataset in a particular study
- 4) Filtering of data based on specific criteria on multiple fields at the same time
  - Support targeted studies
- 5) Cross-referencing of data (same or different repositories)
  - avoid duplication, maximise efficiency and access
- 6) Visual analytics / inspection of data / thumbnail preview
  - quick visual filtering from a results set to validate dataset use
- 7) Sharing data (whole dataset/particular records/bibliographic information) in a collaborative environment
  - common space of keeping both data and their versions across time
  - quick check on latest changes
  - share bibliographic information
- 8) Accompanying educational / training material
- 9) Portal functionality similar to other established academic portals
  - subject/visual search, free text search, build query functionality, subscription, ...

# RDA – DDP-IG (5) – Recommendations



- 1) Provide a range of query interfaces to accommodate various data search behaviours
- 2) Provide multiple access points to find data
- 3) Make it easier to judge relevance, accessibility and reusability of a data collection
- 4) Make Individual metadata records readable and analysable
- 5) Be able to share and output bibliographic references
- 6) Provide feedback about data usage statistics
- 7) Be consistent with other repositories
- 8) Identify and aggregate metadata records that describe the same data object
- 9) Make metadata records easily indexed and searchable by major web search engines
- 10) Follow API search standards and community adopted vocabularies for interoperability

# RDA – DDP-IG (6) – Recommendations



	REQ1: Data availability	REQ2: Connection of data	REQ3: Annotations	REQ4: Filtering	REQ5: Cross-referencing	REQ6: Inspection of data	REQ7: Collaborative environment	REQ8: Similarity across portals	REQ9: Training material
REC 1: Query interfaces			✓		✓		✓		
REC 2: Multiple access points		✓		✓		✓		✓	
REC 3: Summarize search results	✓		✓			✓			
REC 4: Metadata records readable		✓	✓						
REC 5: Bibliographic references							✓		
REC 6: Usage statistics			✓						
REC 7: Consistency								✓	
REC 8: Identify duplicates		✓			✓				
REC 9: Findability from web SEs	Support data searches from web search engines								
REC 10: Interoperability	The Fair Data Principles								

Ten simple rules for finding data

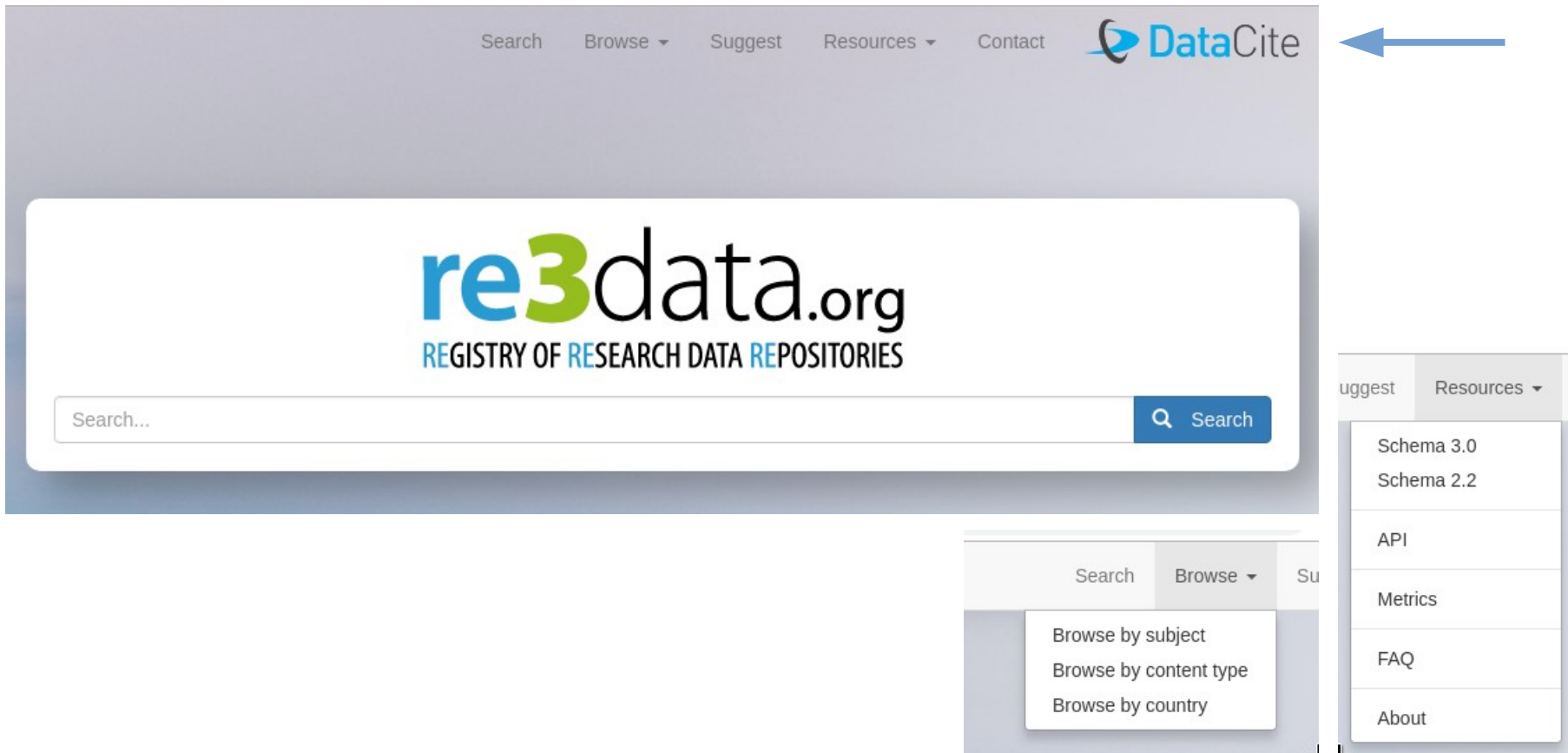
# Repositories - examples



- Repositories can be
  - Domain specific
  - General purpose
- Following slides will show examples of
  - Community driven efforts
  - Project based solutions
  - Super-national driven entities
  - Document based repositories
  - Linked data solutions



- REgistry of REsearch REpositories (re3data.org)
- “By offering detailed information on more than 2,000 research data repositories, re3data has become the most comprehensive source of reference for research data infrastructures globally”



Search Browse Suggest Resources Contact DataCite

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

Search... Search

Suggest Resources

- Schema 3.0
- Schema 2.2
- API
- Metrics
- FAQ
- About

Search Browse Su

- Browse by subject
- Browse by content type
- Browse by country

## Filter

Reset all

### Subjects

- Natural Sciences (1)
  - Physics (1)
    - Astrophysics and Astronomy (1)

### Content Types

### Countries

### API

### OAI-PMH (1)

other (1)

### Data access

### Database access

### Database licenses

### Data licenses

### Data upload

### Data upload restrictions

### Enhanced publication

### Institution responsibility type

### Institution type

### Keywords

### PID systems

### DOI (1)

other (1)

### Provider types

### Quality management

### Repository languages

### Syndications

### Repository types

Search...

Search

Toggle short help

Sort by ▾

← Previous 1 Next →

Found 1 result(s)

## GAVO Data Center

German Astrophysical Virtual Observatory Data Center

Subject(s)

Physics Astrophysics and Astronomy Natural Sciences

Content type(s)

Standard office documents Images Scientific and statistical data formats Structured graphics Databases Raw data Software applications

Country

Germany

The GAVO data center at Zentrum für Astronomie der Universität Heidelberg

Repository details

## GAVO Data Center

General Institutions Terms Standards

Persistent identifier system(s) other  
DOI

Data citation guideline <http://dc.g-vo.org/hsol/q/q/howtocite>

Enhanced Publication yes

Quality management unknown

### Application programming interfaces (3)

API type other

URL [http://dc.zah.uni-heidelberg.de/\\_system\\_/adql/query/form](http://dc.zah.uni-heidelberg.de/_system_/adql/query/form)

API type other

URL <http://dc.zah.uni-heidelberg.de/tap>

API type OAI-PMH

URL <http://dc.zah.uni-heidelberg.de/oai.xml>

### Alerting services (1)

Type of alerting service RSS

Alerting service <http://dc.g-vo.org/regrss>

The screenshot shows the OpenAIRE website interface. At the top, there is a navigation bar with a home icon and the following menu items: EXPLORE, PROVIDE, CONNECT, MONITOR, and DEVELOP. Below this, a secondary navigation bar includes SERVICES, SUPPORT, OPEN SCIENCE IN EUROPE, and ABOUT, along with a search icon. The main content area features the OpenAIRE logo and the word 'EXPLORE' in orange. To the right of the logo are links for SEARCH, SHARE, LINK, and CONTENT PROVIDERS.

Publications	Funder	Project	Publication Date
Research Data	European Commissi... (294,518)	Programs on Critical ... (1,813)	2015 (1,792,868)
Software	National Institutes of... (203,439)	ASIA (1,401)	2014 (1,763,781)
Other Research Products	National Science Fou... (178,206)	COLLMOT (1,128)	2016 (1,756,358)
Projects	Wellcome Trust (65,458)	XSEDE: eXtreme Scie... (1,055)	2017 (1,581,232)
Content Providers	Research Council UK (52,793)	HIPEAC (1,047)	2013 (1,580,518)
Organizations	<a href="#">View more</a>	<a href="#">View more</a>	<a href="#">View more</a>
	Access Mode	Type	Language
	Open Access (23,583,069)	Article (12,395,853)	English (12,226,453)
	Restricted (279,004)	Other literature type (3,712,093)	Undetermined (1,954,593)
	Closed Access (207,158)	Preprint (1,960,594)	Russian (1,564,268)
	not available (24,387)	Doctoral thesis (1,421,576)	Japanese (1,442,111)
	Embargo (7,518)	Research (1,357,914)	Portuguese (1,155,990)
		<a href="#">View more</a>	<a href="#">View more</a>
	Community	Content Provider	
	EGI Federation (18,047)	Europe PubMed Cen... (4,971,143)	
	FET FP7 (9,098)	JAIRO (1,896,451)	
	FET H2020 (2,104)	arXiv.org e-Print Arc... (1,438,640)	
	Research Data Allian... (28)	LAReferencia - Red F... (1,268,782)	
		CyberLeninka - Russi... (1,254,605)	
		<a href="#">View more</a>	

- (openaire.eu)
- EOSC-related initiative
- Multi-faceted
- Includes a Repository
  - And APIs



- (eudat.eu)
  - EU funded initiative
    - Connected to EOSC
- Data discovery



## B2FIND

Service Area: Data discovery

B2FIND is a **discovery service based on metadata steadily harvested from research data collections from EUDAT data centres and other community repositories**. The service offers faceted browsing and it allows in particular to discover data that is stored through the B2SAFE and

B2SHARE services.

👤 *User, Repository Manager, Community Manager*

**Filter by location** Clear

Map data © OpenStreetMap contributors  
Tiles by Stamen Design (CC BY 3.0)

**Filter by time** Clear

Start:

End:

**Publication Year** Clear

to

**Communities** ▼

**Tags** ▼

**Creator** ▼

**Discipline** ▼

**Language** ▼

**Publisher** ▼

**693,600 datasets found** Order by:

---

**Alpine3D simulations of future climate scenarios CH2014**

Overview The CH2014-Impacts initiative is a concerted national effort to describe impacts of climate change in Switzerland quantitatively, drawing on the scientific resources...

---

**Fatal avalanche accidents in Switzerland since 1995-1996**

This data collection contains information concerning all accidents by snow avalanches causing at least one fatality in Switzerland. The data set commences on 01/10/1995. After...

---

**Number of avalanche fatalities per hydrological year in Switzerland since 193...**

This dataset contains the statistics on the number of avalanche fatalities per hydrological year in Switzerland. The data set commences with the beginning of the hydrological...

---

**Precipitation Scaling Data Set (Vögeli et al., Frontiers)**

Dataset (Model input, snow distribution and validation) for the precipitation scaling paper, which should be cited along with the data set citation. This data is useful for...

---

**DISCHMEX - Impact of extreme land-surface heterogeneity on micrometeorology o...**

This dataset contains eddy-covariance measurements in the ablation period of 2014-2016. Measurements were taken from two turbulence towers over a long-lasting snow patch, which...

---

**DISCHMEX - High-resolution daily snow ablation rates in an Alpine environment...**

We recorded snow ablation maps with a terrestrial laser scanner (TLS, Riegl-VZ6000) at the Gletschboden area. The TLS position is located approximately 30 vertical meters above...

---

**Forest Access Roads 2013 Walderschliessungsstrassen LFI3**

In 2013–2014, a survey was conducted in Switzerland to update the Forest Access Roads geo-dataset within the framework of the Swiss National Forest Inventory (NFI). The...



A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

(fairsharing.org)

View as Table | View as Grid

Sort by: Best Match

**Recommended Records**

Recommended

**Associated Publication?**

No Publication Has Publication

**Claimed?**

No Maintainer Has Maintainer

**Record Status**

Uncertain Deprecated In develop Ready

**Standard Type**

- Model/Format 21
- Reporting Guideline 19
- Terminology Artifact 3
- Identifier Schema 1

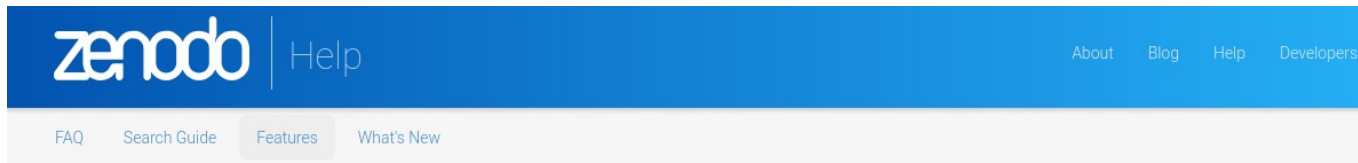
**Domains**

- Web Service 10
- Data Model 7
- Report 4
- Resource Metadata 4

44 records in view

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Collection/Recommendation	Status
	IVOA Identifier	IVOA Identifier	Standard	Astrophysics And Astronomy	Centrally Registered Identifier	Not applicable	None	IVOA Registry Interfaces StandardsRegExt VOResource VODataService RegTAP Plus 2 more...	None	International Virtual Observatory Alliance (IVOA)	R
	IVOA Registry Interfaces	IVOA Registry Interfaces	Standard	Astrophysics And Astronomy	None	Not applicable	None	StandardsRegExt VOResource IVOA RM SimpleDALRegExt VODataService Plus 2 more...	None	International Virtual Observatory Alliance (IVOA)	R
	IVOA Standard for Unified Content Descriptors	IVOA UCD	Standard	Astrophysics and astronomy	None	Not applicable	None	UCD1+ Vocabulary Vocabularies in the Virtual Observatory VOUnits	None	International Virtual Observatory Alliance (IVOA)	R
	IVOA Credential Delegation Protocol	IVOA Credential Delegation Protocol	Standard	Astrophysics And Astronomy	Web Service	Not applicable	None	IVOA Web Services Basic Profile VOSI PDL VOSpace UWS Plus 1 more...	None	International Virtual Observatory Alliance (IVOA)	R
	IVQA-Web Services-Basic Profile	IVQA-Web Services-Basic Profile	Standard	Astrophysics And Astronomy	Web Service	Not applicable	None	PDL SSO - Authentication VOSpace IVOA Credential Delegation Protocol UWS Plus 1 more...	None	International Virtual Observatory Alliance (IVOA)	D
	IVOA Support Interfaces	VOSI	Standard	Astrophysics And Astronomy	Web Service	Not applicable	None	PDL VOSpace UWS IVOA Web Services Basic Profile SimpleDALRegExt Plus 2 more...	None	International Virtual Observatory Alliance (IVOA)	R

- (zenodo.org and about.zenodo.org/)
  - [https://zenodo.org/record/802100#.YbshJ9so\\_RY](https://zenodo.org/record/802100#.YbshJ9so_RY)
  - CERN based
  - OpenAIRE connected
  - General content repository, mainly papers/proceedings/presentations...



## Introducing Zenodo!

(All) Research.  
Shared.

– your one stop research shop!

all research outputs from across all fields of research are welcome! Zenodo accepts any file format as well as both positive and negative results. We choose to promote peer-reviewed openly accessible research, and we curate the uploads posted on the front-page.

Citeable.  
Discoverable.

– be found!

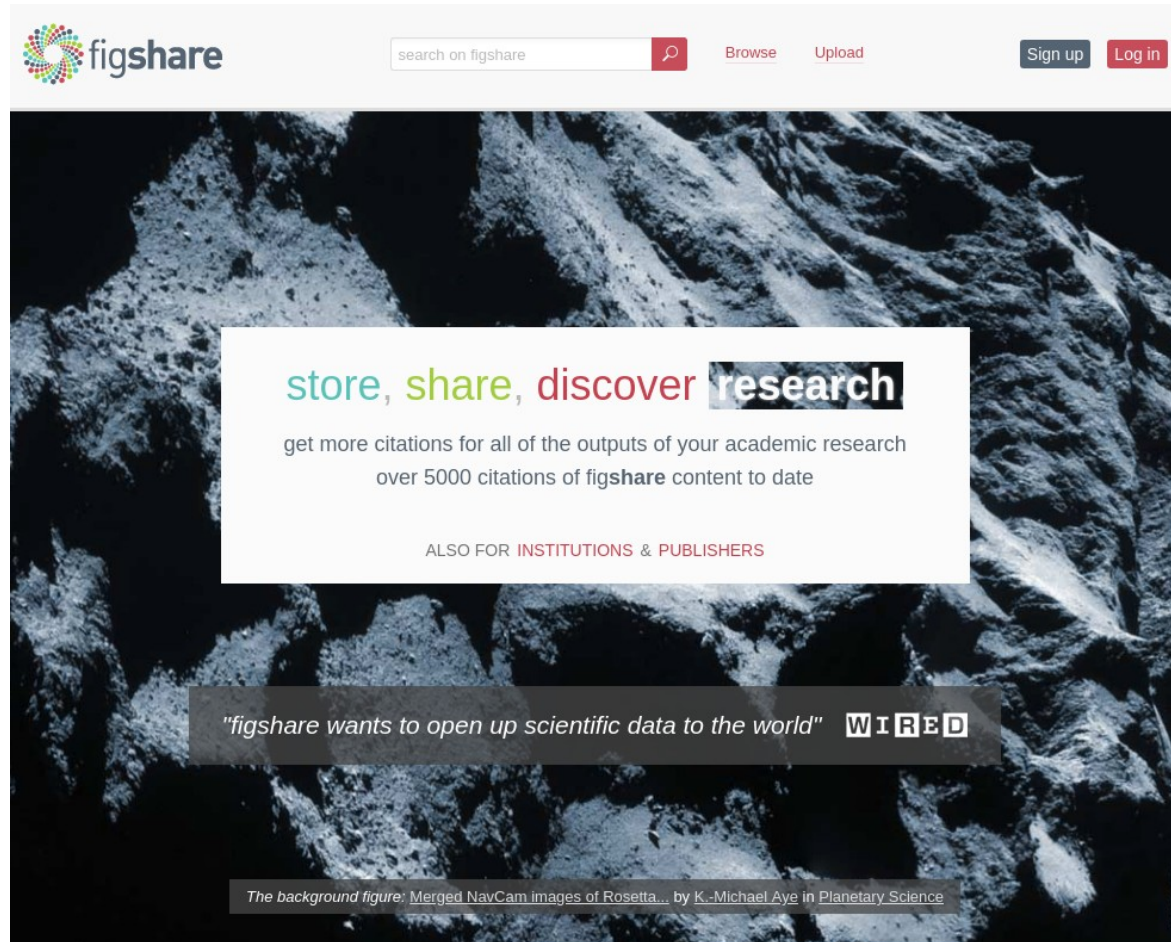
Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. Zenodo further supports harvesting of all content via the OAI-PMH protocol.

Communities

– create your own repository

Zenodo allows you to create your own collection and accept or reject uploads submitted to it. Creating a space for your next workshop or project has never been easier. Plus, everything is citeable and discoverable!

- (figshare.com)
- Publication driven
- Support service
- [upload limits]



store, share, discover **research**

get more citations for all of the outputs of your academic research  
over 5000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

"figshare wants to open up scientific data to the world" **WIRED**

The background figure: Merged NavCam images of Rosetta... by K.-Michael Aye in Planetary Science

simplifying your research workflow

Upload > Manage > Share > Publish

- (datadryad.org)
  - The Dryad Digital Repository is a curated resource that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad provides a general-purpose home for a wide diversity of datatypes.
  - Nonprofit

The screenshot shows the Dryad website homepage. At the top, there is a navigation bar with the Dryad logo and links for 'About', 'For researchers', 'For organizations', 'Contact us', 'Log in', and 'Sign up'. Social media icons for Twitter, Facebook, and RSS are also present. The main content area features a large green button labeled 'Submit data now' with a link to 'How and why?'. Below this is a search bar with the text 'Enter keyword, author, title, DOI, etc.' and a 'Go' button, along with a link to 'Advanced search'. The 'Browse for data' section has two tabs: 'Recently published' and 'Popular'. Under 'Recently published data', two entries are listed: one by Clark EG, Kanauchi D, Kano T, Aonuma H, Briggs DEG, and Ishiguro A (2018) from the *Journal of Experimental Biology*, and another by Cravens ZM and Boyles JG (2018) from *Oecologia*. On the right side, there is a section titled 'Latest from @datadryad' showing a tweet from Dryad (@datadryad) about molecular insights into chronic kidney disease.

- Directory of Open Access Journals
  - (doaj.org)
  - Journals as primary records
    - Articles out of them



The screenshot shows the DOAJ website homepage. At the top left is the DOAJ logo and the text "DIRECTORY OF OPEN ACCESS JOURNALS". To the right is an orange button labeled "SUPPORT DOAJ". Below this is a navigation bar with links: Home, Search, Browse Subjects, Apply, News, About, For Publishers, API, and Login. The main content area features a search box with the text "Search DOAJ" and a magnifying glass icon. Below the search box are checkboxes for "journals" and "articles", and a link for "[Advanced Search]". To the right of the search box is a statistics box showing: 12,372 Journals, 9,415 searchable at Article level, 129 Countries, and 3,566,892 Articles. Below the search box is a box with the heading "DOAJ (Directory of Open Access Journals)" and a paragraph of text: "DOAJ is a community-curated online directory that indexes and provides access to high quality, open access, peer-reviewed journals. DOAJ is independent. All funding is via donations, 40% of which comes from sponsors and 60% from members and publisher members. All DOAJ services are free of charge including being indexed in DOAJ. All data is freely available." Below this is another paragraph: "DOAJ operates an education and outreach program across the globe, focussing on improving the quality of applications submitted." To the right of the main content area is a sidebar with a language selection dropdown labeled "Seleziona lingua" and a list of links: FAQs, OAI-PMH, XML, Widgets (indicated by a blue arrow), Open Access Resources, Transparency & Best Practice, Download metadata, Journals Added/Removed, and New Journals Feed.



- Open access to 1,475,672 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance, Statistics, Electrical Engineering and Systems Science, and Economics
  - (arxiv.org)
- User driven repository
- Used mainly for pre-prints and ongoing work
  - Author-paper relationship “limited”

The screenshot shows the top section of the arXiv.org website. On the left is the Cornell University Library logo. On the right, it says 'We gratefully acknowledge support from the Simons Foundation and member institutions'. Below this is a red banner with the 'arXiv.org' logo on the left and a search bar on the right. The search bar contains the text 'Search or Article ID' and 'All fields' with a dropdown arrow and a search icon. Below the search bar are links for '(Help | Advanced search)'. In the top right corner of the banner is a 'Login' link.

Open access to 1,475,672 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance, Statistics, Electrical Engineering and Systems Science, and Economics

Subject search and browse:

6 Nov 2018: [December 2018, January 2019 holiday schedule announced](#)

5 Sept 2018: [arXiv looks to the future with move to Cornell CIS](#)

See cumulative "[What's New](#)" pages. Read [robots beware](#) before attempting any automated download

- The SAO/NASA Astrophysics Data System
- (adsabs.harvard.edu)



## The SAO/NASA Astrophysics Data System

Search

[Search](#) [Browse](#) [Help](#)

## Welcome to the Digital Library for Physics and Astronomy

This site is hosted by the **High Energy Astrophysics Division** at the **Harvard-Smithsonian Center for Astrophysics**

Authors: (Last, First M, one per line)  SIMBAD  NED  ADS **Objects**

Exact name matching  Object name/position search  
 Require author for selection  Require object for selection  
 (  OR  AND  simple logic ) (Combine with:  OR  AND )

Publication Date between   and    
 (MM) (YYYY) (MM) (YYYY)

Enter **Title Words**  Require title for selection  
 (Combine with:  OR  AND  simple logic  boolean logic )

Enter **Abstract Words/Keywords**  Require text for selection  
 (Combine with:  OR  AND  simple logic  boolean logic )

Return  items starting with number

ads Feedback ORCID About Account

QUICK FIELD: Author First Author Abstract Year Fulltext All Search Terms

author:"huchra, john" Search

Your search returned 127 results

1 2012yCat..21990026H 2012/06  
 VizieR Online Data Catalog: The 2MASS Redshift Survey (2MRS) (Huchra+, 2012)  
 Huchra, J. P.; Macri, L. M.; Masters, K. L. [and 17 more](#)

2 2012ApJS..199...26H 2012/04 cited: 284  
 The 2MASS Redshift Survey—Description and Data Release  
 Huchra, John P.; Macri, Lucas M.; Masters, Karen L. [and 17 more](#)

3 2011ASSP...24...11H 2011  
 Astronomical Publishing: Yesterday, Today and Tomorrow  
 Huchra, John

4 2009AAS...21322401H 2009/01  
 The International Year of Astronomy 2009 Overview  
 Huchra, John; Isbell, D.; Deustua, S. E. [and 1 more](#)

5 2007ASPC...377...3H 2007/10 cited: 1  
 Keynote Address: Science Libraries in the Information Age  
 Huchra, J. P.

6 2006apri.meet..215H 2006/01  
 The State of the Universe Report  
 Huchra, J.

7 2005ASPC...329..135H 2005/06 cited: 46  
 The 2MASS Redshift Survey and Low Galactic Latitude Large-Scale Structure  
 Huchra, J.; Jarrett, T.; Skrutskie, M. [and 7 more](#)

8 2005IAUS..216..170H 2005/01 cited: 20  
 2MASS and the Nearby Universe  
 Huchra, J.; Martinbeau, N.; Jarrett, T. [and 7 more](#)

0 selected  
 Add papers to library

Years Citations Reads

■ refereed ■ non refereed

Limit results to papers from 1972 to 2012 Apply

# RofR (IVOA)



- Registry of Registries
  - (rofr.ivoa.net)
- Distributed
- Domain driven
- API based
  - GUI exist
    - Limited

International Virtual Observatory Alliance  
IVOA Registry of Registries



[Register/Validate a Registry](#) | [Guide for Registry Providers](#) | [Registry Specifications](#) | [IVOA Registry Working Group](#)

## Welcome to the Registry of Registries

The Registry of Registries (RofR, pronounced *rover*) is a web portal provided on behalf of the International Virtual Observatory Alliance (IVOA) and overseen by the IVOA Registry Working Group. It is targeted to VO registry providers and VO application developers that wish to interact with registries.

The key service provide by the RofR is an IVOA publishing registry that lists all publishing registries known to the IVOA. When a resource metadata harvester harvests from these publishing registries, they can discover all published VO resources around the world. The design and recommend uses of the RofR is documented in the IVOA Note, The Registry of Registries.

If you maintain a publishing registry and you are ready to let it be known to the outside world, you can register it here. Before you are allowed to register, you must demonstrate that it conforms to the IVOA Registry Interfaces standard. Note, that you can use the registry validator to check your registry without actually registering it.

### Looking for Registries?

Click on [+] below to see the corresponding list.

#### **[+] Full Searchable Registries**

These registries claim to harvest from publishers regularly and therefore should have records for all resources known in the VO. This list is generated from a cached list that is updated every four hours by a query to a full searchable registry.

WFAU Publishing Registry

**IVOA Identifier:** <ivo://wfau.roe.ac.uk/org.astrogrid.registry.RegistryService>

**Search service endpoint:** [http://publishing-registry.roe.ac.uk:80/astrogrid-registry\\_v1\\_0/services/RegistryQueryv1\\_0](http://publishing-registry.roe.ac.uk:80/astrogrid-registry_v1_0/services/RegistryQueryv1_0)

STScI Searchable Registry

**IVOA Identifier:** <ivo://archive.stsci.edu/nvregistry>

**Search service endpoint:** <http://vao.stsci.edu/directory/ristandardservice.asmx?>

EURO-VO Full Harvestable Registry

**IVOA Identifier:** <ivo://esavo/registry>

**Search service endpoint:** <http://registry.euro-vo.org/services/RegistrySearch>

**RegTAP service endpoint:** <http://registry.euro-vo.org/regtap/tap>

#### **[+] Currently Registered Publishing Registries**

These publishing registries have successfully register with the RofR after a full validation. These are the registries that the full searchable registries are pulling records from. This list is generated on-the-fly via a query to the RofR's harvesting interface.

*For support, please contact [ivoa-rofr@cfa.harvard.edu](mailto:ivoa-rofr@cfa.harvard.edu)*

# Do still exist single “archives”?



- A lot
  - Domain specific
  - Custom based
  - Some trying to reach open interoperation
    - Depending on the domain (usually)
- They're usually listed as items inside more general repositories
- They can provide quite powerful analysis/discovery interfaces
  - Because of the specific scenario they expose
- Some astrophysics VO examples
  - ESA (<http://sky.esa.int/>)
  - ESO (<http://archive.eso.org/scienceportal/home>)
  - MAST (<https://archive.stsci.edu/>)
  - CADC (<http://www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/en/>)

# Repository Certification



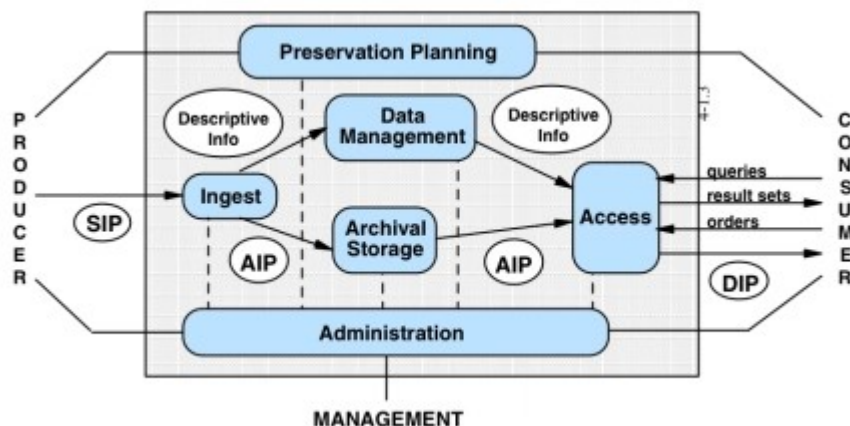
- National and international funders will (likely) mandate
  - Open data
  - Data management policies for the long-term storage and accessibility of data
  - Access to funded data products and proper data management plans (DMP)
- Need to store shared data in a trustworthy data repository
  - managed, curated, and archived to preserve the initial investment in collecting them
- Sustainability of repositories raises a number of challenging issues in different areas
  - Organizational, technical, financial, legal, etc.
- Certification can be an important contribution to ensuring the reliability and durability of data repositories
- By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.



# Core Trust Seal



- (www.coretrustseal.org)
  - Core certification involves a minimally intensive process whereby data repositories supply evidence that they are sustainable and trustworthy
  - A repository first conducts an internal self-assessment, which is then reviewed by community peers
  - Such assessments help data communities to improve the quality and transparency of their processes, and to increase awareness of and compliance with established standards
  - This community approach guarantees an inclusive atmosphere in which the candidate repository and the reviewers closely interact
- [https://www.coretrustseal.org/wp-content/uploads/2017/01/Core\\_Trustworthy\\_Data\\_Repositories\\_Requirements\\_01\\_00.pdf](https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf)



Open  
Archive  
Information  
System