



FÓRUM  
GESTÃO DE DADOS  
DE INVESTIGAÇÃO

# How can institutions better manage their research data

Digital Commons  
Data

Data Monitor

Juan Garcia  
Elsevier



# As a research institution, you want to:

- **Meet funder guidelines** with F.A.I.R. data
- **Store, archive** and **preserve** research data
- Encourage **RDM best practices** for greater impact
- **Enable collaboration** on data by project teams
- Showcase data outputs for **impact and reporting**

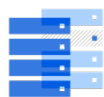
And prepare for an Open Data future



**Digital Commons**  
Data

# Digital Commons Data: Benefits to the Institutions and their researchers

Digital Commons Data gives your institution the tools you need to successfully drive forward your Research Data Management program, with powerful features for researchers, administrators and data curators to store, manage, curate, share and preserve data.



Store data in secure flexible storage, or connect to your institutional storage, solving your storage needs effortlessly



Showcase data alongside other outputs, to maximise reach and impact



Report on research data with interactive dashboards, for visibility on inventory, usage and engagement



Preserve data for long-term availability and reuse

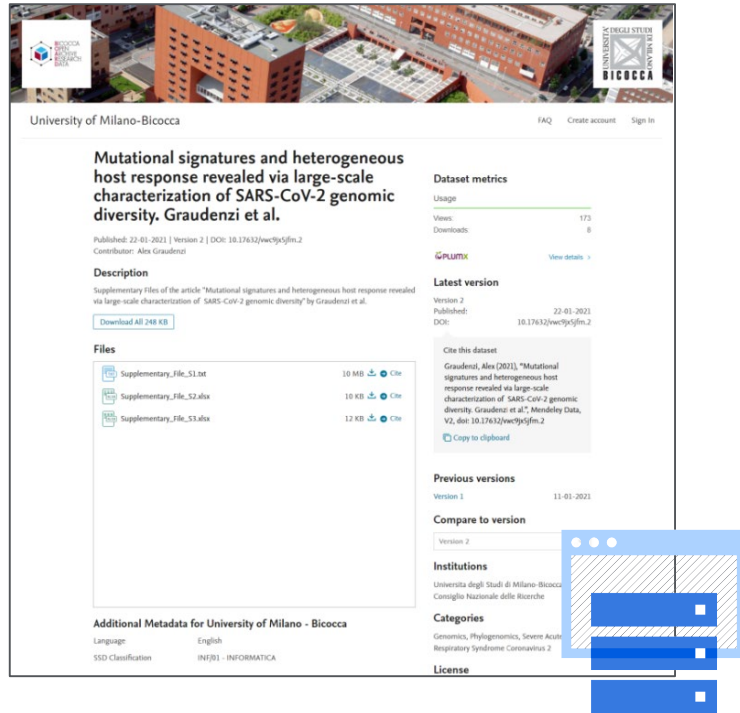


Enable collaboration within and beyond the institution, to facilitate the research process



Put your library team in full control, with powerful tools to manage, administer and curate data

## Digital Commons Data



The screenshot shows a Digital Commons dataset page for the University of Milano-Bicocca. The dataset is titled "Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. Graudenzi et al." and is published on 22-01-2021. The page includes a description, a list of files (Supplementary\_File\_S1.txt, Supplementary\_File\_S2.xlsx, Supplementary\_File\_S3.xlsx), and various metadata sections such as Dataset metrics, Latest version, Previous versions, Compare to version, Institutions, Categories, and License. A blue box highlights the 'Institutions' section, which lists 'Università degli Studi di Milano-Bicocca' and 'Consiglio Nazionale delle Ricerche'.

University of Milano-Bicocca

### Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. Graudenzi et al.

Published: 22-01-2021 | Version 2 | DOI: 10.17612/wc9tjfm.2  
Contributor: Alex Graudenzi

**Description**  
Supplementary files of the article "Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity" by Graudenzi et al.

[Download All 248 KB](#)

**Files**

File Name	Size	Download	View
Supplementary_File_S1.txt	10 MB	Download	View
Supplementary_File_S2.xlsx	10 KB	Download	View
Supplementary_File_S3.xlsx	12 KB	Download	View

**Dataset metrics**  
Usage:   
Views: 173  
Downloads: 8

**Latest version**  
Version 2  
Published: 22-01-2021  
DOI: 10.17612/wc9tjfm.2

**Previous versions**  
Version 1  
11-01-2021

**Compare to version**  
Version 2

**Institutions**  
Università degli Studi di Milano-Bicocca  
Consiglio Nazionale delle Ricerche

**Categories**  
Genomics, Phylogenomics, Severe Acute Respiratory Syndrome Coronavirus 2

**License**

- A trusted data repository to store, share and archive datasets to comply with funder requirements and meet the FAIR principles.
- Links your institution's data to the wider ecosystem, facilitating discovery and reproducibility via DOIs, DataCite, open APIs, deep indexing etc.
- RDM best practices through advanced data curation tools such as versioning, customized metadata, and moderation workflows.
- A cloud-hosted platform that meets industry standards for secure storage, hosting, and preservation.

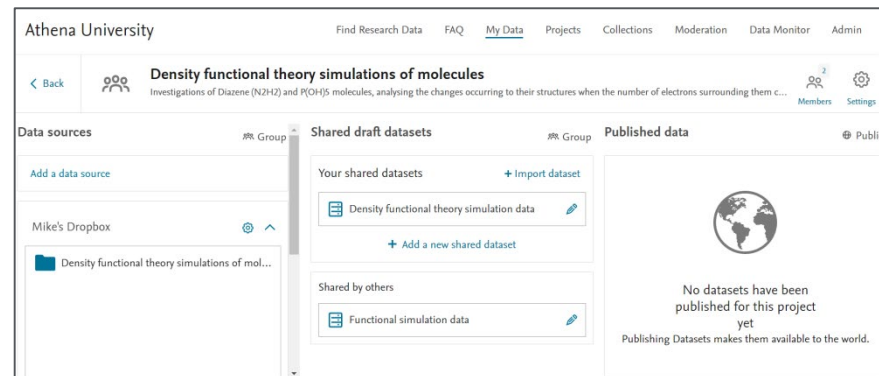
# DC Data provides a secure, compliant environment for sharing project data

## Projects in DC Data allow researchers to:

- Invite collaborators into a project, regardless of their physical location and institutional affiliation
- Import data from existing cloud or university storage, and share files seamlessly
- Prepare to share research data, in accordance with the FAIR data principles

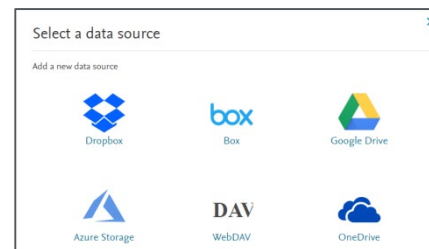
## Benefits for the institution:

- Facilitate project teams collecting and collaborating on data securely
- Track progress of projects



The screenshot shows the project page for 'Density functional theory simulations of molecules' on the Athena University platform. The page is divided into three main sections: 'Data sources', 'Shared draft datasets', and 'Published data'. Under 'Data sources', there is a link to 'Add a data source' and a folder named 'Mike's Dropbox' containing a file 'Density functional theory simulations of mol...'. The 'Shared draft datasets' section shows 'Your shared datasets' with one dataset 'Density functional theory simulation data' and an option to 'Add a new shared dataset'. Below this, 'Shared by others' shows a dataset 'Functional simulation data'. The 'Published data' section is currently empty, displaying a globe icon and the message: 'No datasets have been published for this project yet. Publishing Datasets makes them available to the world.'

*The project environment allows teams to securely share data files, and collaboratively prepare datasets to publish.*



The 'Select a data source' dialog box offers options to 'Add a new data source'. It features six icons for different cloud storage services: Dropbox, Box, Google Drive, Azure Storage, DAV WebDAV, and OneDrive.

*External cloud storage can be connected to projects.*



# Tracking institutional research data

## Data Monitor

Institutions need solutions that help them easily find and keep track of vast amounts of research data published by their researchers to:

- Maintain a clear **overview** of **research impact**
- Oversee the management of **institutional policies**
- Safeguard **compliance** with **funder mandates**



Source: <https://blog.research-plus.library.manchester.ac.uk/2019/02/15/from-couch-to-almost-5k-raising-research-data-visibility-at-the-university-of-manchester/>

## Simplifying research data visibility

Actively monitoring thousands of data repositories researchers use often presents a challenges for institutions.

### Our solution



Provides accurate **visibility into the entire research data output** of institutions

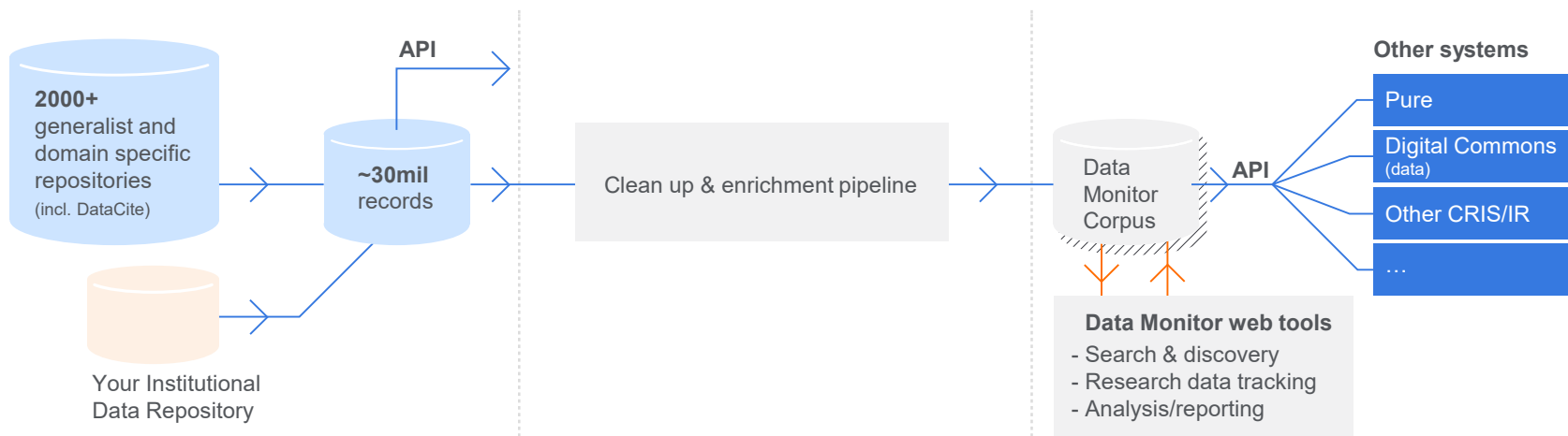


Integrates **~30 million research data records** from 2000+ generalist and domain specific repositories

Research data repositories	Number of datasets
Zenodo	1817618
USGS Mineral Resources	995208
Cambridge Structural Database	909175
GEOROC	478309
PANGAEA	403887
RCSB-PDB	138270

Above: Example of repositories from Data Monitor corpus as of April 2021.

# How Data Monitor works



## Step 1

- **Harvest** metadata from 2000+ public repositories
- **Normalize** the metadata, following the OpenAIRE metadata schema
- **Accessible** via a free API

## Steps 2

- **Remove** non-reliable/non-data sources, duplicates, datasets with dead links, non-data/SPAM records
- **Enrich** metadata with publications/author/institutional linking and more by triangulating other sources (DataCite, Scholix, Scopus) + text mining and NLP

## Step 3

Data Monitor Corpus is **accessible** via the Data Monitor API (for integration) and Web Tools



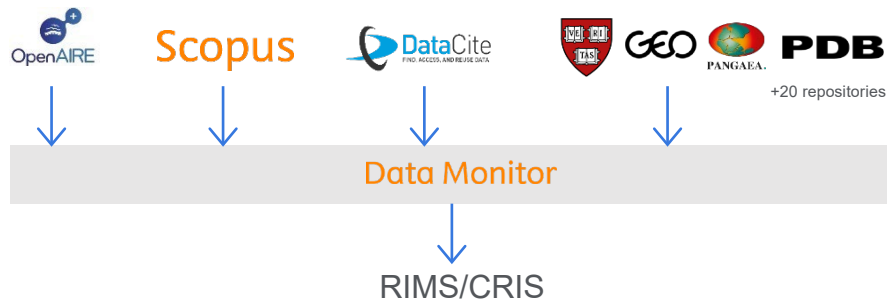


# Integrating with industry-leading research tools

- Connects institutions to ~30 million research data records from 2000+ generalist and domain-specific repositories
- Integrates with services used by researchers and institutions worldwide

 ~30 mil  
data records

 2000+  
domain repositories





E.WEDEL-LARSEN@ELSEVIER.COM

# Thank you

Digital Commons  
Data

Data Monitor

