

FAIR data and FAIR principles

National End-Users Training Event, 6th June
2022

Chrysovalantis Constantinou,
The Cyprus Institute



- ❑ Data, the Modern Landscape
- ❑ Which data are FAIR?
- ❑ What are the FAIR principles?
- ❑ Making your data FAIR
- ❑ Further studying material

The modern landscape of data

- ❑ The Internet is a source of a vast amount of data produced by both researchers and the public (2.5×2^{60} bytes of data are produced daily)
- ❑ Large scientific experiments (CERN, Fermilab, etc.) produce vast amounts of data
- ❑ Big tech also produces big data
- ❑ These data can be utilized in various research or industrial ventures
- ❑ Secure storage and sharing of these data might lead to
 - ❑ New discoveries
 - ❑ Enable new collaborations
 - ❑ Enable the verification of previously obtained results
 - ❑ Reduce research duplication

Which data are FAIR?

- ❑ Research/Industrial/Other data are only useful for research or other usage when they are well documented, described, structured etc

- ❑ FAIR data are data which are:
 - ❑ Findable
 - ❑ Accessible
 - ❑ Interoperable
 - ❑ Reusable

What are the FAIR principles?


[Open Access](#) | [Published: 15 March 2016](#)

The FAIR Guiding Principles for scientific data management and stewardship

[Mark D. Wilkinson](#), [Michel Dumontier](#), ... [Barend Mons](#)  [+ Show authors](#)

[Scientific Data](#) **3**, Article number: 160018 (2016) | [Cite this article](#)

455k Accesses | **4020** Citations | **1956** Altmetric | [Metrics](#)

 An [Addendum](#) to this article was published on 19 March 2019

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them,

Findable

- ❑ Metadata and data should be easy to find by both humans and computers
- ❑ Machine readable metadata are essential for automatic discovery of datasets and services
- ❑ (Meta)data are assigned globally unique and persistent identifier(s)
- ❑ Data are described with rich metadata
- ❑ Metadata clearly and explicitly include the identifier of the data they describe
- ❑ (Meta)data are registered or indexed in a searchable resource

Accessible

- ❑ Once a user finds the required data, they need to know how they can be accessed
- ❑ (Meta)data are retrievable by their identifier using a standardised communications protocol
- ❑ The protocol is open, free, and universally implementable
- ❑ The protocol allows for an authentication and authorisation procedure, where necessary
- ❑ Metadata are accessible, even when the data are no longer available

Interoperable

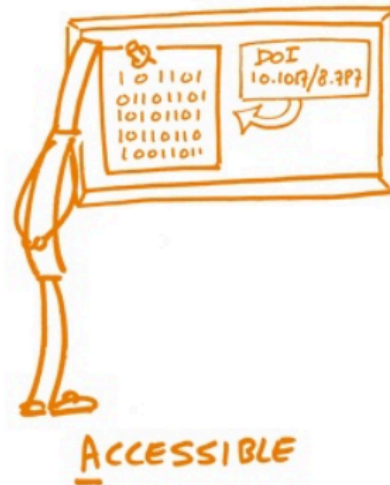
- ❑ The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows
- ❑ (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- ❑ (Meta)data use vocabularies that follow FAIR principles
- ❑ (Meta)data include qualified references to other (meta)data

Reusable

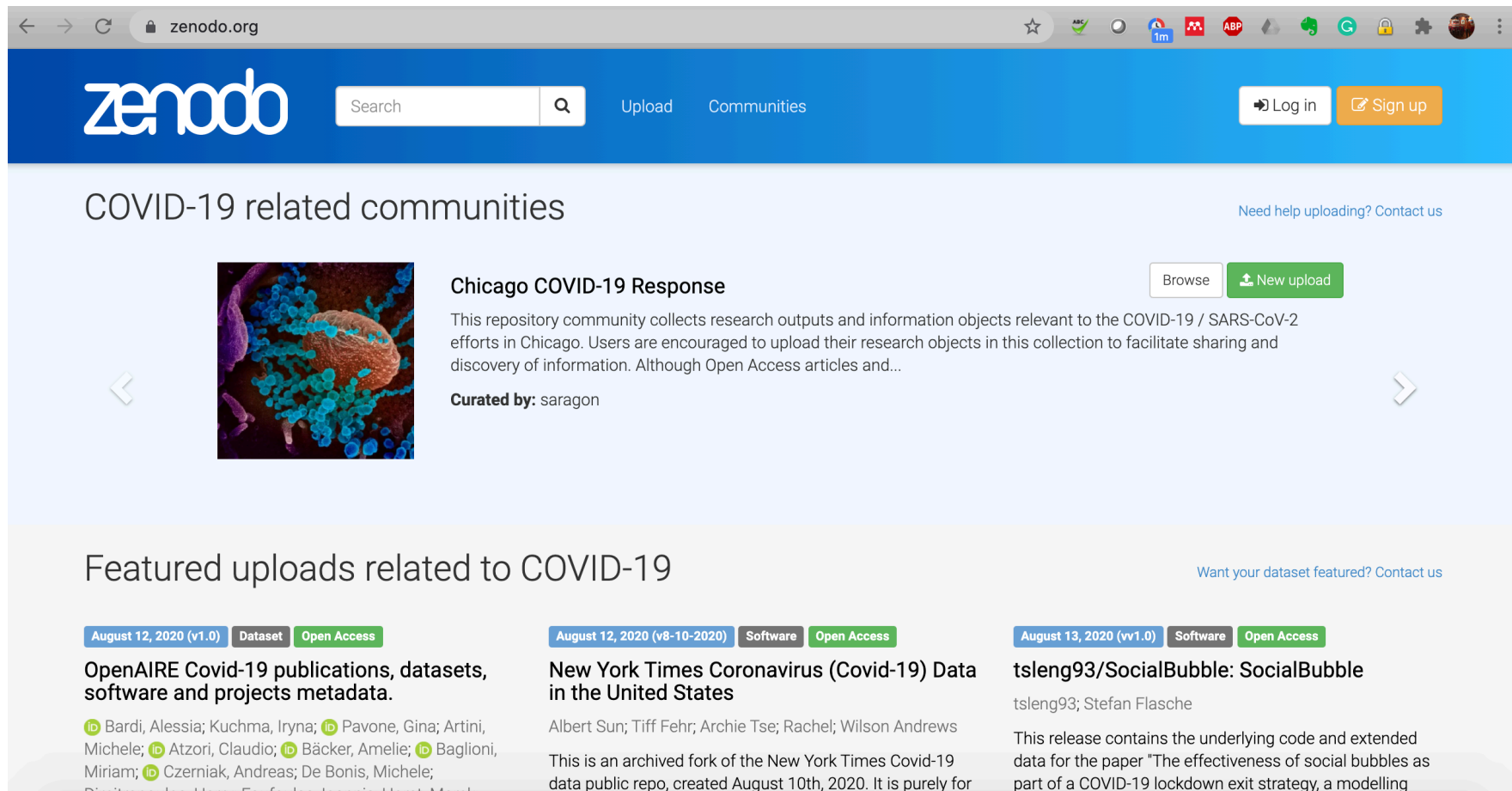
- ❑ The ultimate goal of FAIR is to optimise the reuse of data
- ❑ To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings
- ❑ Meta(data) are richly described with a plurality of accurate and relevant attributes
- ❑ (Meta)data are released with a clear and accessible data usage license
- ❑ (Meta)data are associated with detailed provenance
- ❑ (Meta)data meet domain-relevant community standards

FAIR Principles (the cartoon)

FAIR DATA PRINCIPLES

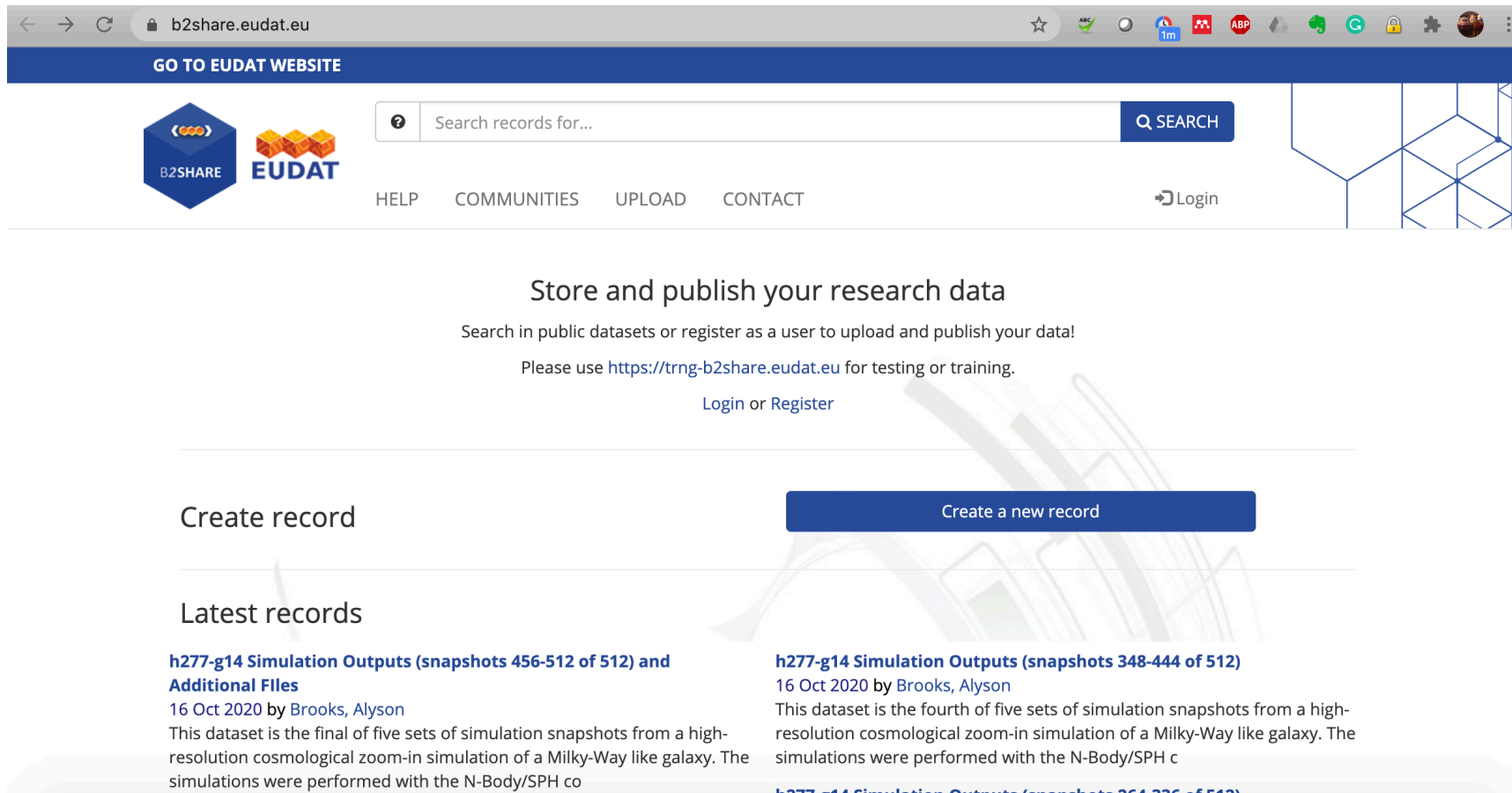


FAIR principles in practice (making your data FAIR)



The screenshot shows the Zenodo website interface. At the top, there is a navigation bar with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. On the right side of the navigation bar, there are 'Log in' and 'Sign up' buttons. Below the navigation bar, the main content area is titled 'COVID-19 related communities'. A featured community is 'Chicago COVID-19 Response', which includes a description, a 'Browse' button, and a 'New upload' button. Below this, there is a section for 'Featured uploads related to COVID-19'. Three uploads are listed: 'OpenAIRE Covid-19 publications, datasets, software and projects metadata', 'New York Times Coronavirus (Covid-19) Data in the United States', and 'tsleng93/SocialBubble: SocialBubble'. Each upload includes a date, version, and tags like 'Dataset', 'Software', and 'Open Access'.

FAIR principles in practice (making your data FAIR)



The screenshot shows the b2share.eudat.eu website. At the top, there is a navigation bar with the text "GO TO EUDAT WEBSITE". Below this, the B2SHARE and EUDAT logos are displayed. A search bar with the placeholder "Search records for..." and a "SEARCH" button is present. Navigation links for "HELP", "COMMUNITIES", "UPLOAD", and "CONTACT" are visible, along with a "Login" button. The main content area features the heading "Store and publish your research data" and a sub-heading "Search in public datasets or register as a user to upload and publish your data!". A note advises using <https://trng-b2share.eudat.eu> for testing or training, with a "Login or Register" link. Below this, there are sections for "Create record" with a "Create a new record" button, and "Latest records". Two record entries are shown, both titled "h277-g14 Simulation Outputs (snapshots 456-512 of 512) and Additional Files" by Brooks, Alyson, dated 16 Oct 2020. The first entry's description is partially visible, mentioning a high-resolution cosmological zoom-in simulation. The second entry's description is also partially visible, mentioning a high-resolution cosmological zoom-in simulation of a Milky-Way like galaxy.

- ❑ For optimal long-term archiving, files should not be compressed and should avoid proprietary formats. Only unencrypted files should be published and archived.

- ❑ Examples of open file formats are:
 - ❑ Text: TXT, ODT, PDF/A, XML
 - ❑ Tabular data: CSV, TSV
 - ❑ Image: TIFF, PNG, JPG 2000, SVG, WebP
 - ❑ Audio: WAV, FLAC, OPUS
 - ❑ Video: MPEG2, Theora, VP8, VP9, AV1, Motion JPG 2000 (MJ2)
 - ❑ Binary hierarchical data: HDF5

Why making your data FAIR?

- ❑ Making research data FAIR is beneficial for researchers, research communities, research infrastructure facilities, and research organizations.
- ❑ FAIR data:
 - ❑ Help to gain maximum potential from data, and overall maximum impact of research, increasing visibility and citations
 - ❑ Improve the reproducibility and reliability of research
 - ❑ Help in staying aligned with international standards and approaches
 - ❑ Engage in new partnerships with researchers, business, policy and broader communities
 - ❑ Enable new research questions to be answered
 - ❑ Use new innovative research approaches and tools

- ❑ The European Open Science Cloud (EOSC) is implementing open science in Europe
- ❑ The European Open Science Cloud strives to ensure that European scientists/citizens will be able to take advantage of a data-driven science
- ❑ It also strives to cultivate an open science culture and disseminate the FAIR principles
- ❑ It will provide a platform for European research, including a web of FAIR research data and services.
- ❑ NI4OS-Europe will provide a bridge for EOSC in 15 partner countries of the Balkan and Mediterranean region

- ❑ Handbook for open science: <https://book.fosteropenscience.eu/en/>
- ❑ NI4OS training platform: <https://training.ni4os.eu/my/>
- ❑ NI4OS webpage: <https://ni4os.eu/>
- ❑ B2Share: <https://b2share.eudat.eu/>
- ❑ re3data: <https://www.re3data.org/>

- ❑ Personal email: ch.constantinou@cyi.ac.cy

Thank you!

