

Open science and FAIR principles

1st Capacity Building Event, 29th October 2020

Dr. Chrysovalantis Constantinou,

The Cyprus Institute,

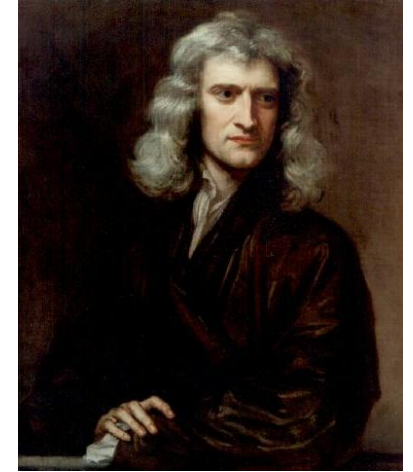
NI4OS-Europe Work Package 6 Leader



- ❑ Brief history of science and publishing
- ❑ What is Open Science?
- ❑ What are FAIR principles?
 - ❑ Making your data FAIR
- ❑ Examples
 - ❑ Further studying, material, etc.

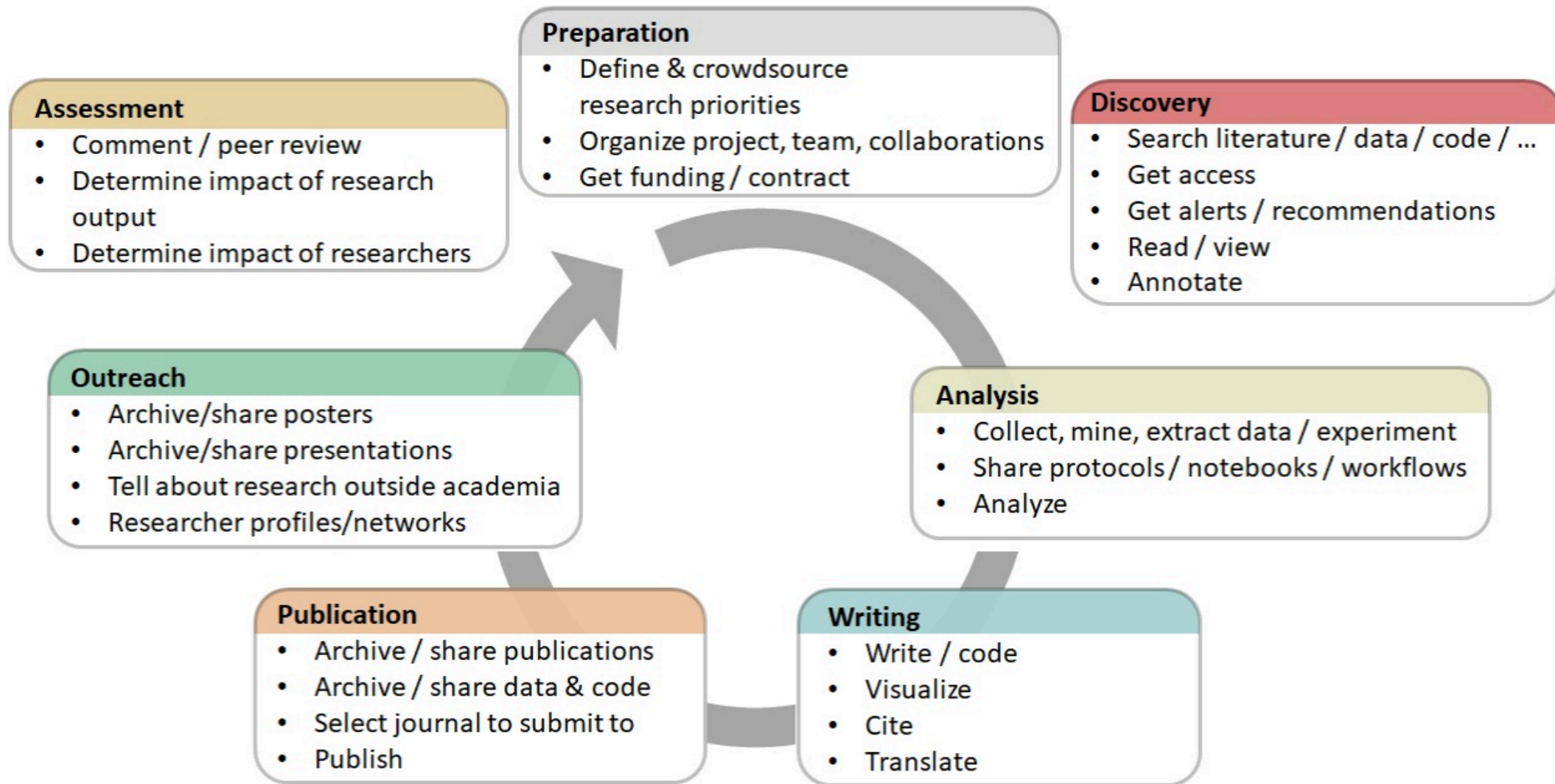
A brief history of science and publishing

- ❑ The scientific revolution starts with Copernicus (1543) and ends with the publication of Newton's Principia (1687)
- ❑ **Before the advent of scientific journals**, scientists had little to gain and much to lose by publicizing scientific discoveries
- ❑ Scientists received funding from patrons who had no interest in publishing because they wanted to profit from potential discoveries
- ❑ Publications were coded so there was no way to trace the origin of a discovery
- ❑ Leibniz and Newton's calculus is the most famous example



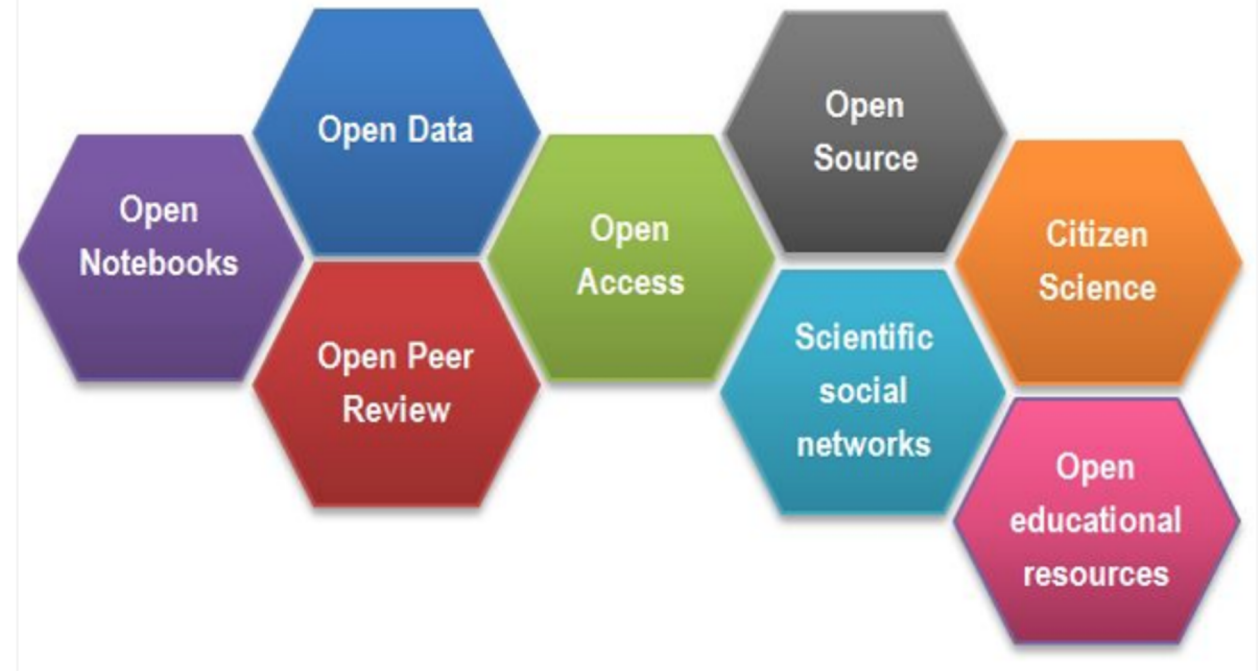
- ❑ The system of funding by patrons becomes unsustainable
- ❑ The Royal Society of England and the French Academy of Sciences are founded in 1660 and 1667 respectively
- ❑ Research results start to get published
- ❑ Scientists start collaborating
- ❑ The advent of the internet challenges how research outputs can be published
- ❑ Publications, software, data, and repositories are now widely available
- ❑ Open science is coined in 1998. Exploration towards that direction starts

What is open Science? (the research cycle)



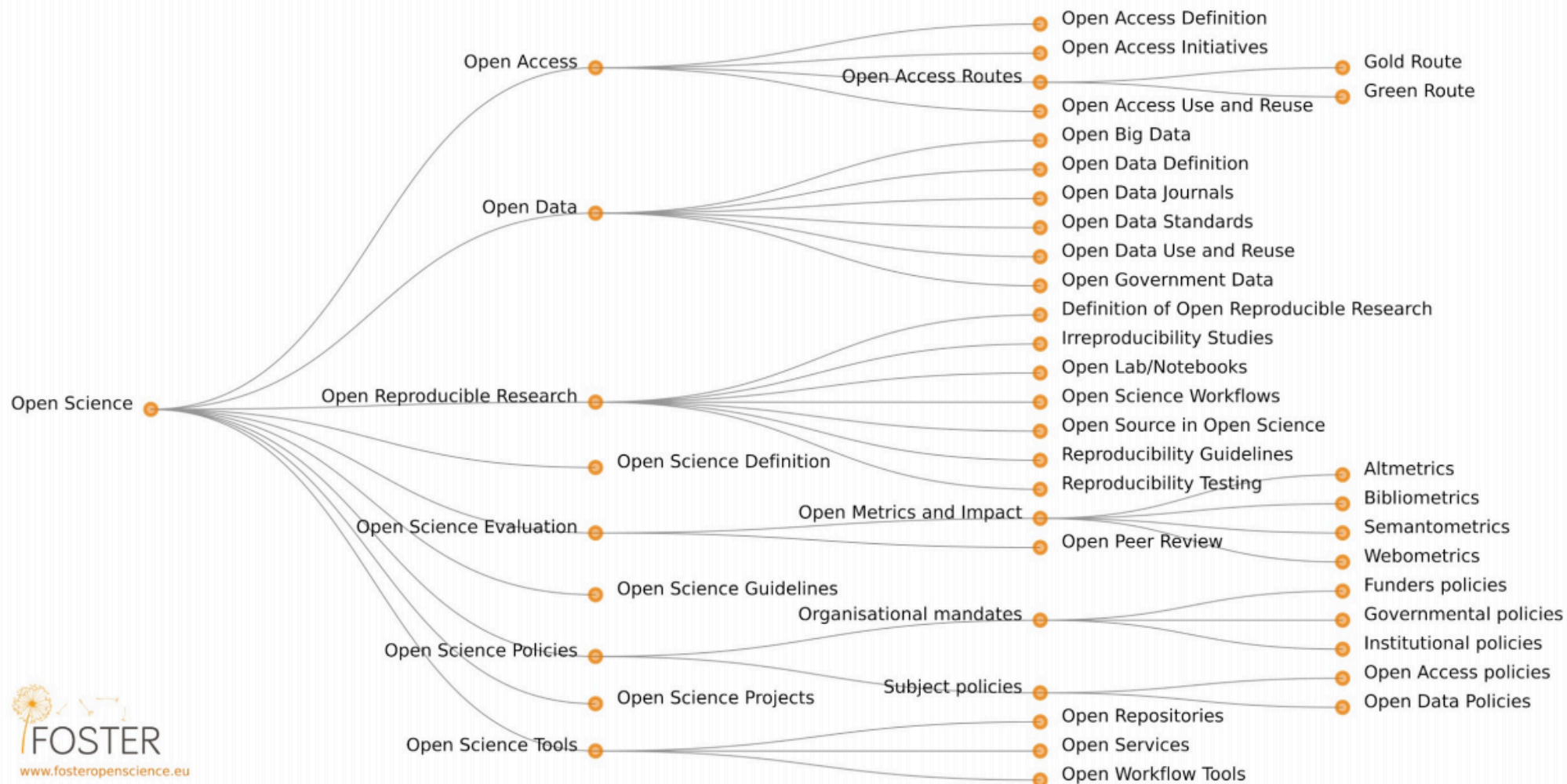
What is Open Science? (the general definition)

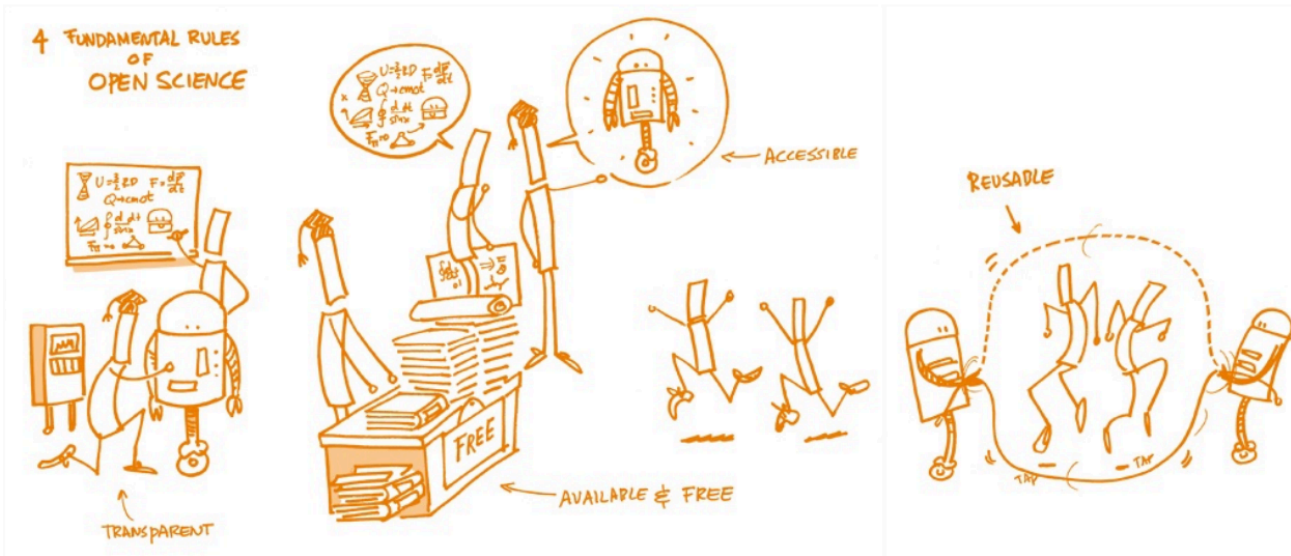
- ❑ Open Science strives to make scientific research accessible to an inquiring society
- ❑ It's core principles are increased transparency, re-use, reproducibility, etc
- ❑ It's core practises are open access to: research publications, data, source code, software, repositories, etc
- ❑ Open Science **is NOT** restricted to the natural sciences
- ❑ Using the term **Open Research** allows for the inclusion of Social Sciences, Life Sciences, Engineering, the Humanities, etc



What is open Science? (the diagram)

Open Science Taxonomy



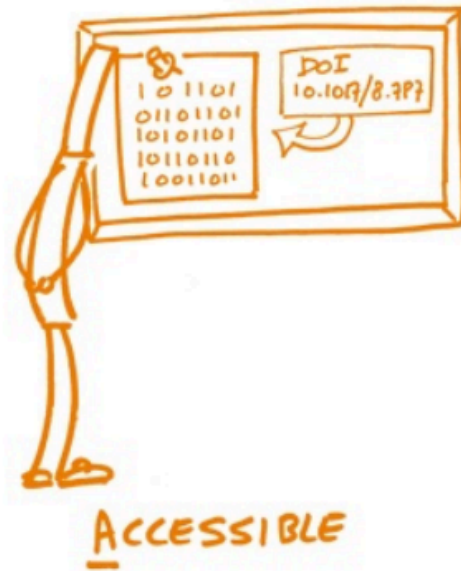


- ❑ To implement open science initiatives we need a set of principles
- ❑ These principles are to be followed by research outputs aiming to become part of open science initiatives
- ❑ These rules are known as FAIR principles

- ❑ The increased production and availability of online resources requires data to be created with longevity in mind
- ❑ FAIR is a set of principles, not a standard, according to which data must be Findable, Accessible, Interoperable, and Reusable
- ❑ FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data
- ❑ Humans increasingly depend on computational support to deal with the increased volume and complexity of data (every day, 2.5 Exabytes of data are produced!)

FAIR Principles (the cartoon)

FAIR DATA PRINCIPLES



FAIR Principles (the NI4OS-Europe logo)



- ❑ **Findable:** Data and metadata should be easily findable by both humans and computers.
- ❑ **Accessible:** Users need to know how data can be accessed, possibly including authentication and authorization.
- ❑ **Interoperable:** Data usually need to be integrated with other data. Also, data need to interoperate with applications or workflows for analysis, storage, and processing.
- ❑ **Reusable:** The ultimate goal of FAIR is to optimize 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.

❑ Findable

- ❑ Metadata are assigned a globally unique persistent identifier (PID)
- ❑ Data are described with rich metadata
- ❑ Metadata clearly and explicitly include the identifier of the data they describe
- ❑ Metadata are registered or indexed in a searchable resource

❑ Accessible

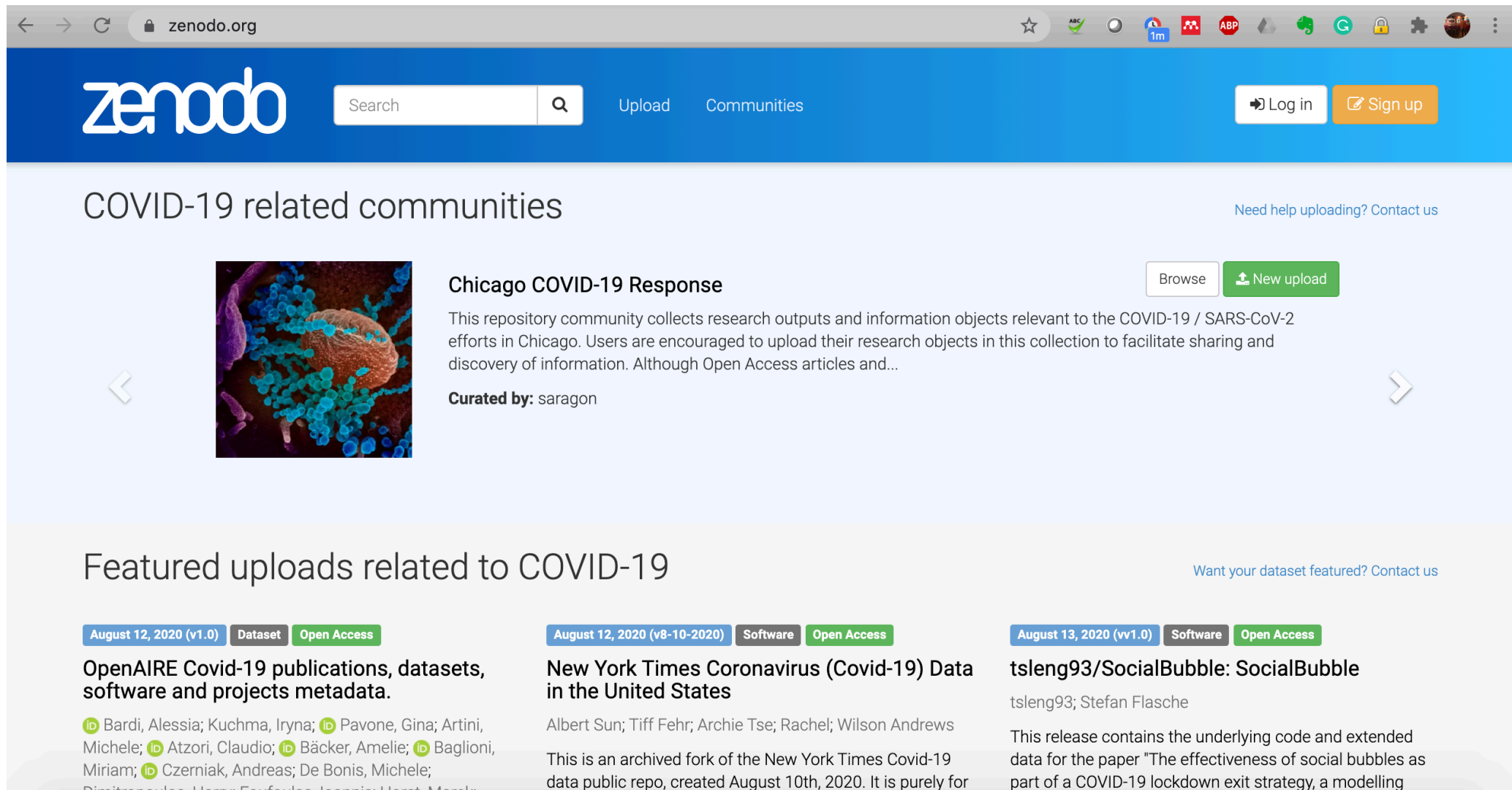
- ❑ Metadata are retrievable by their identifier using a standardized communications protocol
- ❑ The protocol is open, free, and universally implementable
- ❑ The protocol allows for an authentication and authorization procedure (where necessary)
- ❑ Metadata are accessible, even when the data are no longer available

❑ Interoperable

- ❑ Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation
- ❑ Metadata use vocabularies that follow FAIR principles
- ❑ Metadata include qualified references to other metadata

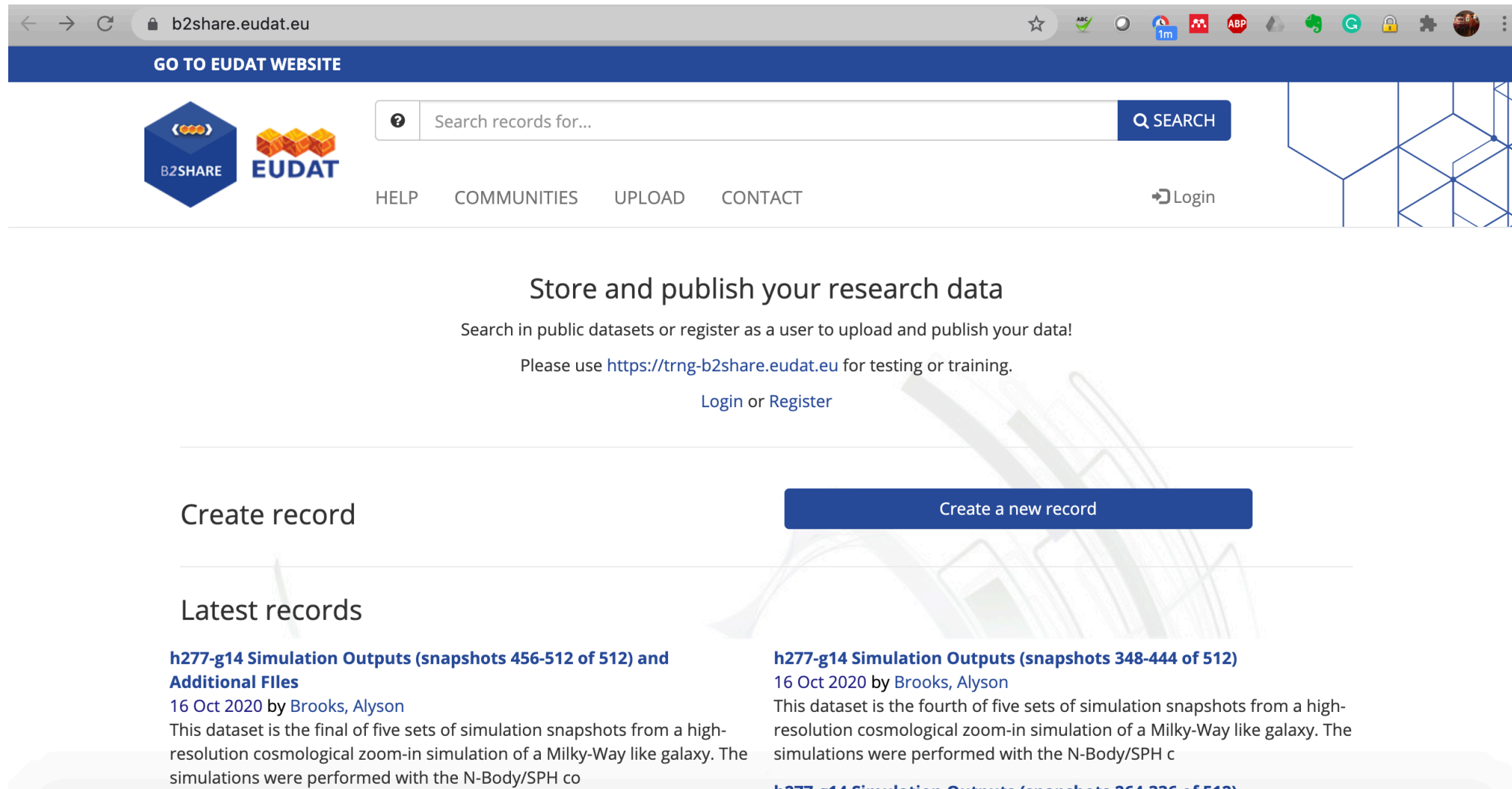
❑ Reusable

- ❑ Metadata are richly described with a plurality of accurate and relevant attributes
- ❑ Metadata are released with a clear and accessible data usage license
- ❑ Metadata are associated with detailed provenance
- ❑ Metadata meet domain-relevant community standards



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' with three items: '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 item includes a date, version, and tags like 'Dataset', 'Open Access', 'Software', etc.

FAIR principles in practice (making your data accessible)



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, there are logos for B2SHARE and EUDAT. A search bar is present with the placeholder text "Search records for..." and a "SEARCH" button. Navigation links include "HELP", "COMMUNITIES", "UPLOAD", "CONTACT", and "Login". 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 link is provided: "Please use <https://trng-b2share.eudat.eu> for testing or training." Below this is a "Login or Register" link. A prominent blue button labeled "Create a new record" is displayed. The "Latest records" section shows two entries for "h277-g14 Simulation Outputs (snapshots 456-512 of 512) and Additional Files" by Brooks, Alyson, dated 16 Oct 2020. The description for the first entry is partially visible: "This dataset is the final of five sets of simulation snapshots from a high-resolution cosmological zoom-in simulation of a Milky-Way like galaxy. The simulations were performed with the N-Body/SPH co".

- ❑ 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

- ❑ 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

What does the European Open Science Cloud have to do with all these? How is NI4OS-Europe related?

- ❑ 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/>
- ❑ FAIR checklist: [link](#)
- ❑ Argos: <https://argos.openaire.eu/home>
- ❑ Metadata: <https://rd-alliance.github.io/metadata-directory/>
- ❑ Zenodo: <https://zenodo.org/>
- ❑ Personal email: ch.constantinou@cyi.ac.cy

Thank you!

