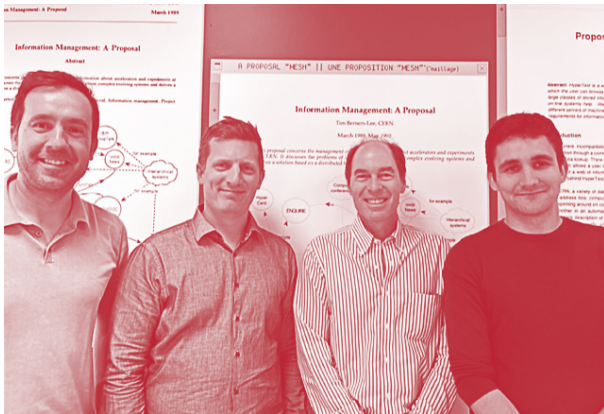


NEWS FROM zenodo



ZENODO DATA REPOSITORY: PROVIDING PRACTICAL SOLUTIONS FOR DATA STORAGE AND DATA SHARING

Authors: **Lars Holm Nielsen, Jose Benito Gonzalez Lopez, Tim Smith, and Alex Ioannidis**

Affiliations: European Organization for Nuclear Research (CERN)

doi: [10.54920/SCTO.2022.RAWatch.7.25](https://doi.org/10.54920/SCTO.2022.RAWatch.7.25)

Properly managing, preserving, and sharing data can be a daunting task, especially for busy researchers who are constantly confronted with new tasks and requirements from funders and their institutions. Zenodo is a general-purpose data repository that enables researchers, scientists, project managers, and institutions to share, preserve, and showcase multidisciplinary research results (data, software, publications, and other research objects) that are outside the scope of existing institutional or subject-based repositories. Based in the trustworthy CERN data centre, Zenodo is a service provided by researchers to researchers contributing to open science by capturing research objects and making them FAIR (findable, accessible, interoperable, and reusable). This article addresses some of the challenges of data storage and data sharing, such as finding the right place to store data, citing data properly, and using hybrid data sharing solutions. It also demonstrates how using a data repository like Zenodo can help researchers address these challenges.

ZENODO IN A NUTSHELL

The idea of the [Zenodo](#) data repository was conceived when the European Commission (EC) decided that, in order to support its nascent open data policy, it needed a catch-all repository to ensure that every EC funded research output could have a home. In the vanguard of the open access and open data movements in Europe, the EC commissioned the [OpenAIRE](#) project to build

this repository. As an OpenAIRE partner and pioneer in open source, open access, and open data, the [European Organization for Nuclear Research](#) (CERN) had the capabilities to create the repository, and Zenodo was launched in May 2013. Zenodo is currently being used by more than 200,000 researchers and 7,000 communities from around the world.

FINDING THE RIGHT PLACE FOR DATA

Researchers often ask where they should deposit/archive data and why their own hard drive or server is not suitable. Unfortunately, places under the control of an individual researcher are probably the worst choices for archiving data because the task of ensuring they stay operational and accessible often rapidly falls off priority lists as research is completed. Archiving and preserving data are tasks for professionals that require considerable knowledge and both the appropriate technical and organisational infrastructure. This is important not only to guarantee the safekeeping of research data but also to ensure that research data that was previously not citable and discoverable becomes so.

The most suitable place for depositing/archiving data is a repository that can best serve the data and its user

community. Often, the best solution ends up being a domain-specific repository that has the necessary domain expertise to make the data as useful as possible for its user community and that also has appropriate funding and organisational structures. Data, however, exist in many shapes and forms, and many intermediary or non-standard research outputs do not neatly fit in a domain-specific repository. That is why Zenodo exists. As a generic repository, Zenodo can step in when there is no appropriate domain or institutional data repository. And because it accepts research data in any shape and form, it ensures there is always a safe place for the long tail of science. In addition, as a generic repository, Zenodo can often better transcend domains by making data findable and accessible outside the normal boundaries of a researcher's own domain.

CITING DATA PROPERLY

Once an appropriate data repository has been identified, a follow-up question that often arises is: How should data be cited? There is no straightforward answer to this. It often depends on the data itself as well as the community and publishing standards of a specific domain. The most important – and quite often the most overlooked – aspect of citing data, though, is to ensure that a persistent identifier is included when citing data (e.g. a digital object

identifier (DOI)). A persistent identifier not only ensures that the data used is uniquely identified and provides access to the data itself, but discovery systems also require a persistent identifier to be able to properly attribute citations. Currently, DOIs are the persistent identifiers that can be most easily integrated into existing scholarly communication infrastructures and that are understood inside and outside a specific domain.

KEEPING DATA AS OPEN AS POSSIBLE AND AS RESTRICTED AS NEEDED

Sharing clinical trial data has strict regulatory requirements. Even when consent for data sharing and further use has been obtained and data have been anonymised as required by law, data can be difficult to share due to the risk of future cross-correlation. This is why Zenodo supports restricted and controlled access records. In addition, sometimes researchers hoard data locally, hoping to exploit their data set for future projects. Unfortunately, when the data are eventually deposited into a repository, descriptions may have been forgotten, processing steps

overlooked, and most likely people with key knowledge have moved on to other positions. That is why Zenodo allows for the depositing of closed access records, which makes it possible for researchers to deposit and describe their data when the information is still fresh in their minds and later flip the switch to open access. Zenodo also provides features that allow data to be selectively shared as needed, for instance by requiring a justification and the researcher's approval (scientific collaboration, licenses, intellectual property protection, etc.).

SHARING RESEARCH DATA: GIVE IT A GO!

Overall, sharing research data can be a complex and daunting task. Finding the right place to store data, citing data correctly, and making data openly available can be especially difficult for clinical trial data. Therefore, Zenodo's best advice is to always start thinking early on about FAIR data before it is too late. And try exploring [Zenodo's](#) features, since it is quite likely that solutions for some of your needs have already been found and implemented by others!