

# Data 101 - Webinar – What is metadata and what does it do?

*April 24th, 1:30PM-3:00PM*

This webinar, organised in the scope of the FedOSC project, included two parts: first a high-level 30-minute introduction about the concept of metadata for research data by the FedOSC data stewards; then, a series of presentations by metadata standard specialists to address the question of added value of metadata and discipline-specific metadata standards for research visibility, documentation, interoperability, and more generally, FAIRness.

Questions were received through the registration form and live during the event. All questions were addressed live.

## *Live Q&A summary*

- *As a researcher, am I expected to be aware of the metadata standards applicable to my work as early as the data collection phase?*

It is always recommended to start consulting data stewards as early as possible in any research project or activity. Additionally, it is likely that any given researcher has already encountered metadata standards sometimes without even realizing it, either by browsing major databases in their research field, or by looking up in data repositories. Aligning with wide, community-endorsed standards right from the start is best practice to support research interoperability.

- *Can metadata pertain to personal data and how should it be managed?*

Metadata “data on data” and is therefore subject to the same management strategy. Metadata can be made visible, kept confidential, or even embargoed, fully or partially, following the “as open as possible, as closed as necessary” principle. As a rule of thumb, personal metadata is usually less sensitive than personal data, but all datasets should be subject to a confidentiality analysis.

- *How exhaustive should a metadata record be?*

When documenting for storage and preservation, the more metadata, the better! In most cases, machine-readable metadata does not need to be generated manually or separately from the dataset: structural and descriptive metadata is likely to be embedded in most community-based file formats, which contain specific descriptors. In other cases, metadata records can be as simple as structured codebooks (in text or text-based formats such as XML, JSON, ...) associated with an Excel file to explicit units and variables or with an image to explicit its additional properties.

When creating sharing or publication metadata, the minimum set of metadata needed to find, cite and understand the dataset is a title, a creator, a date, a description (abstract) and a Persistent Identifier. Those are usually generated when uploading the dataset on a dissemination platform such as a data repository.

In general, the more complete the metadata, the higher the FAIRness of the data.

- *Where can I register health metadata in a way that is compatible with HealthDCAT-AP?*

Most metadata catalogs are compliant with HealthDCAT-AP and in any case, at the time of recording the webinar, it will soon become mandatory for all health data holders. At Sciensano, dedicated staff, services and infrastructure support HealthDCAT-AP compliance as a service for members of its institutions (see [healthdata.be](https://healthdata.be)). Other (health) data sharing infrastructure is likely compliant or will be mandated compliance, external resources such as [data.gov.be](https://data.gov.be) or [data.europa.eu](https://data.europa.eu)

- **How does one go about starting to develop a new metadata standard?**

Speakers insist that a metadata standard is supposed to serve the needs of a community, and that creating a new standard should not be done unless there is a critical gap in existing standards. Even then, creating a metadata standard is a lengthy process that usually involves ratification processes, standardisation bodies, and presents a risk of fragmentation of metadata models; even if a new community need emerges, it is always preferred to address it by using a combination or an extension of pre-existing, already recognised tools and standards. Energy is therefore better spent on guiding and educating research communities on existing metadata standards and their use. Most metadata standard maintenance bodies rely on community interaction in the first place.

### *Further Browsing and Reading*

During the Q&A, speakers and participants have shared the following resources for further exploring the topic.

- Derycke, Pascal, Beatriz J. Barros, Nienke M. Schutte, Charles-Andrew Vande Catsyne, and Martina Bargeman Fonseca. 2025. “Designing DCAT-AP Extensions for Common European Data Spaces: The EHDS HealthDCAT-AP Case Study.” In *Posters, Demos, Workshops, and Tutorials at the 21st International Conference on Semantic Systems (SEMANTiCS 2025)*. CEUR Workshop Proceedings, Vol. 4064. <https://ceur-ws.org/Vol-4064/NXDG25-paper5.pdf>.
- More info about Biodiversity Information Standards (TDWG): <https://www.tdwg.org/>
- More info about GBIF: <https://www.gbif.org/>