Contribution ID: **110** Type: **TALK**

Bringing samples to the digital data curation world - the FAIR WISH Project

Tuesday 5 November 2024 10:05 (20 minutes)

Persistent identifiers (PID) are an essential component of digital research data infrastructure. They are used to unambiguously identify, locate, and cite digital representations of a growing range of entities like publications, data, and others. Physical samples represent the basis for many research results and data and are at the same time deeply in the "long tail"of research data. The HMC project "FAIR Workflows to establish IGSN for Samples in the Helmholtz Association"(FAIR WISH) established standardised workflows for sample description and registration of International Generic Sample Numbers (IGSN) in the Earth Science community within the Helmholtz Association. The IGSN is a globally unique, citable PID for physical samples with discovery functionality in the internet.

FAIR WISH has developed (1) standardised and discipline-specific IGSN metadata schemes for different sample types within the Earth and Environment research, (2) workflows to generate machine-actionable IGSN metadata from different states of digitisation, (3) workflows to automatically register IGSNs, and (4) registered more than 35,000 curated metadata sets with IGSNs for the use cases of the three project partners GFZ, AWI and Hereon. We further fully documented the IGSN metadata schema of GFZ representing the current status quo (Brauser et al., 2024) and made the project results available at the website of GFZ Data Services (https://dataservices.gfz-potsdam.de) and the dedicated Zenodo FAIR WISH Community (https://zenodo.org/communities/fair_wish/).

The FAIR SAMPLES Template (Wieczorek et al, 2023) - the main project output - includes the new, extended (IGSN) metadata schema and the controlled vocabularies identified during FAIR WISH. The Excel-built FAIR SAMPLES enables metadata collection and batch upload of sample descriptions at various sample hierarchies (parent, children at different hierarchy levels) at once. The ability to fill the FAIR SAMPLES Template by individual researchers for a wide range of sample types makes the template flexible and widely applicable. The structured metadata, captured with the FAIR SAMPLES Template and converted into XML files, already represents an important step for the standardisation of rich sample descriptions and their provision in machine-actionable form. The new Software Tool SAMIRA: FAIR SAMPLES Template Processing (Frenzel, 2023), enables semi-automated workflows for IGSN registration.

This presentation will introduce the project and its outcomes and describe lessons learned. It will also look forward to possible solutions for fully automated and quality assured metadata generation and collection.

References:

Brauser, A.; Frenzel, S.; Mohammed, A.; Elger, K. (2024): GFZ Metadata Schema for International Generic Sample Numbers (IGSN) and documentation. V. 1.3. GFZ Data Services. https://doi.org/10.5880/GFZ.LIS.2024.001

Frenzel, S. (2024). FAIR WISH Software Tool: SAMIRA: FAIR SAMPLES Template Processing [Computer software]. GFZ Data Services. https://doi.org/10.5880/GFZ.LIS.2023.001

Wieczorek, M., Brauser, A., Frenzel, S., Heim, B., Baldewein, L., Kleeberg, U., & Elger, K. (2023b). FAIR WISH "FAIR SAMPLES Template. Zenodo, https://doi.org/10.5281/zenodo.10436276

Please specify "other"

In addition, please add 3 to 5 keywords.

FAIR samples, IGSN, metadata collection

Please specify ”other”

Researchers and Scientits and technicians who maintain and operate research infrastructure for data generation

For whom will your contribution be of most interest?

other (please specify below)

Please assign yourself (presenting author) to one of the following groups.

Data professionals and stewards

Primary author: ELGER, Kirsten

Co-authors: BRAUSER, Alexander (Deutsches GeoForschungsZentrum (GFZ) Potsdam); HEIM, Birgit (Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung); BALDEWEIN, Linda (Helmholtz-Zentrum Hereon); WIECZOREK, Mareike (Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung); FRENZEL, Simone Christina (GFZ); KLEEBERG, Ulrike

Presenter: ELGER, Kirsten

Session Classification: Session C1

Track Classification: Assessing and monitoring FAIR data: 1. Human actors in the FAIR data landscape