Contribution ID: **132**   Type: **Poster**

Life Cycle of Metagenomic Research Data Management

Monday 13 October 2025 19:10 (20 minutes)

Effective research data management (RDM) is essential for ensuring that data adheres to the FAIR principles of Findability, Accessibility, Interoperability, and reusability. In this session we will examine how these principles drive the life cycle of metagenomic data at the Uniklinikum University in Aachen, from data generation to long-term storage and reuse.

The process begins when biological samples are sent for sequencing, generating raw data that is then processed using a reproducible Snakemake workflow developed by the Clavel Lab. This workflow is documented and publicly available in GitHub: GitHub - ClavelLab/genome-assembly: A Snakemake workflow assembling bacterial genomes according to the standard operating procedure in the Clavel Lab.
The output of the workflow includes the assembled genomes (in gzipped FASTA format), plasmid sequences (results/plasmids/isolate.plasmids.fa.gz), and a structured metadata table in CSV format. This metadata captures key information about the genome assemblies and conforms to the minimal metadata standards recommended by NFDI4Microbiota. The associated metadata profile is created using the Metadata Profile Generator, ensuring that each field is linked to appropriate ontology terms for improved interoperability. The profile is then made available for use in Coscine, the platform used at the UKA for storing and archiving research data along with its linked metadata profile.

In the next stage, the metagenomic data and its metadata are prepared for storage and reuse. This process is facilitated by the data steward using Python to automate the extraction of metadata from the CSV file. The extracted metadata is added to the metadata form and then both the file and metadata form are uploaded to specified resources in Coscine.

Each resource in Coscine has a unique persistent identifier, ensuring the findability of the data. Data stored in Coscine remains accessible for at least ten years following the conclusion of the research project, in accordance with good scientific practice. Researchers and collaborators can access the data using institutional credentials or ORCiD. Project-based permissions enable secure sharing of data and collaboration.

By walking through each stage of this workflow—from data generation to archiving—this presentation demonstrates how the FAIR principles are applied in practice to support transparent, sustainable, and reusable metagenomic research at UKA.

Primary authors: Dr PAUVERT, Charlie (RWTH/UKA/NFDI4Microbiota); MAGEL, Maja (University Hospital RWTH Aachen); Dr CLAVEL, Thomas (RWTH/UKA/NFDI4Microbiota)

Co-author: GONZALEZ, Catherine (RWTH/UKA)

Presenter: GONZALEZ, Catherine (RWTH/UKA)

Session Classification: Poster Session

Track Classification: SciDataCon Persistent Themes: Data Stewardship