Plan Overview

A Data Management Plan created using DeiC DMP

Title: Soil Microbial Communities: Antimicrobial Resistance, Activity and Functional Characterization

Creator: Simon Meski

Principal Investigator: Carla Paredes-Cruz

Affiliation: Roskilde Universitet / Roskilde University

Template: DCC Template

ORCID iD: 0000-0001-9604-5630

Project abstract:

This PhD project investigates soil microbial communities, with particular emphasis on antimicrobial resistance, microbial activity, and functional characterization. The research combines experimental microbiology with molecular and analytical approaches to investigate microbial biomass, metabolic activity, antimicrobial resistance determinants, and protein expression in different soil environments.

Methods include laboratory-based microbial assays, molecular analyses, and mass spectrometry-based metaproteomics. Metagenomic datasets and bioinformatics tools are used to support protein identification and functional annotation.

The project generates various types of research data, including laboratory observations, experimental measurements, instrument-generated raw data, nucleotide and protein sequences, and computational analysis results. These datasets are documented and managed to support reproducibility, scientific interpretation, and future reuse.

ID: 8441

Start date: 01-10-2026

End date: 31-01-2029

Last modified: 03-11-2026

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Soil Microbial Communities: Antimicrobial Resistance, Activity and Functional Characterization

Data Collection

What data will you collect or create?

Experimental microbiology: Soil characterization (e.g., pH), microbial biomass measurements, leucine incorporation assays, cultivation experiments, antimicrobial susceptibility testing and PCR analyses.
Metagenomics: Raw high-throughput sequencing data, quality-controlled sequencing reads, metagenomic assemblies, predicted genes and proteins, taxonomic profiles, antimicrobial resistance gene annotations and functional characterization of microbial communities.
Metaproteomics: Raw LC-MS/MS data, peptide and protein identifications, quantitative protein measurements and functional annotations.
Bioinformatics: Computational analysis outputs, reference databases, sequence alignments, taxonomic and functional classifications, and integrated metagenomic and metaproteomic analyses.

How will the data be collected or created?

Data will be generated through soil sampling, laboratory experiments, metagenomic sequencing, LC-MS/MS proteomics and bioinformatics analyses, following established protocols.
Data will be organized in experiment-specific folders using consistent sample identifiers and file names. Raw and processed data will be kept separate, with GitHub used for version control of analysis scripts.
Quality will be assessed using experimental controls, replicates, instrument calibration and method-specific quality checks. Experimental deviations will be recorded.

Documentation and Metadata

What documentation and metadata will accompany the data?

Data will be accompanied by SOPs, Experimental Records (methods, observations, measurements and deviations) and Experimental Reports (analyses, results and interpretations), maintained in physical and digital form.
Metadata will include sample identifiers, sampling dates and locations, experimental conditions, instrument settings, software and analysis parameters. Dataset descriptions will identify contributors, creation dates, file formats and access conditions.
Established metadata standards, such as MIxS for environmental metagenomics, will be followed where applicable to support interpretation and reuse.

Ethics and Legal Compliance

How will you manage any ethical issues?

The project primarily involves environmental soil samples and does not involve human participants or personal data. Therefore, informed consent and anonymization are not required.
Potential ethical concerns related to antimicrobial resistance research will be addressed through appropriate biosafety procedures, responsible data handling and compliance with RUC guidelines and the Danish Code of Conduct for Research Integrity.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Research data will be managed in accordance with RUC's policies on data ownership and intellectual property rights.
Publicly available third-party datasets will be reused according to their respective licences and citation requirements. Research data and code will be shared under appropriate open licences where possible, with publication delayed if necessary to protect intellectual property or unpublished results.

Storage and Backup

How will the data be stored and backed up during the research?

Experimental documentation and smaller datasets are stored in RUC's institutional OneDrive, with automatic cloud synchronization and backup. Physical copies of SOPs, experimental records and reports are also maintained.
Large metagenomic and proteomic datasets will be stored on institutional research storage with appropriate backup arrangements. The researcher is responsible for ensuring that data are transferred and backed up regularly. Later in the project, genomic and proteomic datasets are also to be stored via NIH's GenBank repository and EMBL-EBI PRIDE repository.

How will you manage access and security?

During the research, data will be protected through institutional account authentication and restricted access to authorized researchers. Collaborators will receive access through RUC-approved sharing systems.
Upon publication, relevant metagenomic and proteomic datasets will be made publicly accessible through repositories such as NCBI GenBank, ENA and EMBL-EBI PRIDE. Until public release, access will be controlled according to repository policies.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Raw metagenomic and proteomic datasets, processed results, experimental records, SOPs and analysis scripts will be retained for their scientific value and reproducibility.
These data may support validation of published findings, future comparative studies and reanalysis using improved bioinformatics methods.
Research data will be retained in accordance with RUC policies and Danish legal requirements. Publicly deposited datasets will be preserved long-term through established repositories. Other data will be retained for the required period, with redundant files excluded where appropriate.

What is the long-term preservation plan for the dataset?

Metagenomic datasets will be preserved in NCBI SRA/GenBank or EMBL-EBI ENA, and metaproteomic datasets in PRIDE. Analysis scripts will be archived through GitHub.
Experimental documentation and other research data will be retained according to RUC's retention policies and applicable Danish archiving requirements.
Data will be prepared for preservation through standardized formats and accompanying metadata. Public repositories are generally free to use, while any additional institutional storage costs will be assessed as needed.

Data Sharing

How will you share the data?

Research data will be made publicly available upon publication through established repositories: metagenomic data via NCBI SRA/GenBank or EMBL-EBI ENA, and metaproteomic data via PRIDE. Analysis scripts will be shared through GitHub.
Datasets will be linked to scientific publications through persistent identifiers (DOIs or accession numbers), allowing researchers to find, access and cite the data. Additional processed data may be shared through supplementary materials or repositories such as Zenodo, under appropriate open licences.

Are any restrictions on data sharing required?

No major restrictions are anticipated, as the project primarily involves environmental samples and non-personal data.

Responsibilities and Resources

Who will be responsible for data management?

The PhD researcher will be primarily responsible for data collection, organization, documentation, storage and sharing, under the supervision of the PhD supervisors.

What resources will you require to deliver your plan?

1) a computer. 2) OneDrive.