DMP title

Project Name D-2024-3193

Grant Title G0GE123N

Principal Investigator / Researcher Dirk Springael

Description

Managed aquifer recharge (MAR) refers to the purposeful recharge of water to aquifers for subsequent recovery, to increase groundwater availability and improve water quality while preventing resource deterioration. During MAR, water is infiltrated through soil and aquifer sediments, improving water quality. Although MAR is an established technology, the underlying microbial, geo-chemical, and hydrological processes during subsurface passage are not entirely understood, which limits predictions of expected water quality improvements. The ERA-NET EU project MARSURE aims at implementing and understanding next-generation hybrid MAR systems for groundwater restoration taking advantage of a range of controlled reactions and sequential multi-barrier pre- and *in situ* treatment concepts to achieve the needed reduction of pollutants while maintaining a low carbon footprint with no waste production including nutrients, trace metals, organic pollutants, pathogens, and antibiotic resistance genes (ARG). The project involves 7 partners from 6 EU countries. In the project, the KU Leuven group is responsible for tracking the microbial community dynamics in the MAR systems including composition and functionality, as a response to the implemented treatment conditions.

Institution KU Leuven

1. General Information

Name applicant

Dirk Springael

FWO Project Number & Title

G0GE123N: Hybrid 'Managed Aquifer Recharge' as an adaptation tool to ensure a resilient and healthy groundwater source

Affiliation

KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Sequencing data files (fastq format), around 200 (microbial community, metagenome sequences), approx. 6 TB, generated by short read/long read sequencing + 16S amplicon sequencing data files (fastqformat) from around 50 soil samples generated by Nanopore

sequencing, OTU table derived from the 16S amplicon sequences and containing the relative abundances of each Operational Taxonomic Unit (OTU) for each sample (in.csv format).

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

No

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

Yes. This is a EU project with different partners. A consortium agreement has been signed between the partners. There are no special restrictions in place.

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

Overall, (e-)labbooks will contain information on experimental design, protocols, sampling location, abbreviations used, structure of the data (including link with physical storage of data), and steps involved in data analysis and relevant analysis scripts (R scripts, MOTHUR/QIIME scripts). A clear coding for all data files related to the project will be used. In the concluding stage of the project, a master index file containing the combined information for all experiment will be compiled which will be archived and also stored on the personal harddrives/PC of the PI. Altogether, this should allow any secondary analyst to use the data accurately and effectively. More specifically, the following information will be given on the items described in section 2:

Metadata files on origin and characteristics of samples, will be provided with a clear description of the methods that were used to generate the data. The metadata will include information on the samples (where collected (GPS coordinates) and where the related sequence data can be found. Sequencing data files deposited in sequence data bases like EMBL will include the information/documentation required by the data base. OTU tables derived from 16S amplicon sequences and containing the relative abundances of each Operational Taxonomic Unit (OTU) for each sample will be deposited in an official nucleic acid database like EMBL and contain the documentation requested by the depository.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

No

Metadata files on origin and characteristics of samples, will be provided with a clear description of the methods that were used to collect the data. The metadata will include information on the samples (where collected (GPS coordinates) and information on where the sequence data can be found. Sequencing data files deposited in sequence data bases like EMBL will include the information/documentation required by the data base.

5. Data storage and backup during the FWO project

Where will the data be stored?

All data will be stored in the university's secure environment. Nucleic acid sequence data (plasmid, metagenome, 16S amplicon) will be submitted and stored in official nucleic acid data bases like EMBL upon publication.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

Yes. The amount of data foreseen to not encompass foreseen capacities

What are the expected costs for data storage and back up during the project? How will these costs be covered?

If any, not expected to be high. Costs will be covered by the FWO project itself.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be stored in the university's secure environment.

6. Data preservation after the FWO project Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

In the concluding stage of the project, a master index file containing the combined information for all experiments will be compiled which will be archived and also stored on the personal harddrives/PC of the PI. Altogether, this should allow any secondary analyst to use the data accurately and effectively. All data will be preserved for at least 5 years after completion of the project.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university's central servers (with automatic back-up procedures). for at least 10 years, conform the KU Leuven RDM policy. Nucleic acid sequence data will be stored at official public data bases like EMBL.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

No direct idea but not expected to be high. If costs, they will be covered by related research projects.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

No

Which data will be made available after the end of the project?

The full dataset (except the nucleic acid sequence data deposited at official repositories for sequencing results) will be deposited in a cvs format in KU Leuven RDR under a CC-BY license. The nucleic acid sequence data will be avalailble through official nucleic acid databases like EMBL.

Where/how will the data be made available for reuse?

In an Open Access repository through KU Leuven RDR.

When will the data be made available?

Upon publication of the research results, the full dataset will be uploaded and made available in a cvs format in RDR immediately afer the end of the project in case published. Others will be added upon publication. If not published within 1,5 years of project complation, all datasets will be made available.

Who will be able to access the data and under what conditions?

Open access data in RDR.

What are the expected costs for data sharing? How will the costs be covered?

No high costs expected. In case, costs will be covered by FWO project itself.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Rakeshkumar Yadav who acts as the Post-doc working on the project.

Who will be responsible for data storage & back up during the project?

Rakeshkumar Yadav who will act as the Post-doc working on the project.

Who will be responsible for ensuring data preservation and reuse ?

PI Dirk Springael

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsibility of updating & implementing this DMP.