Plan Overview

A Data Management Plan created using DMPonline

Title: Geothermal Project on TU Delft Campus

Creator: Lora Armstrong

Data Manager: Lora Armstrong

Affiliation: Delft University of Technology

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

Project abstract:

The Geothermal Project on TU Delft Campus (Geothermie Delft, former DAPWELL) is a unique project in which a sustainable geothermal source is exploited on the TU Delft Campus. TU Delft, Aardyn (former Hydreco Geomec), Energie Beheer Nederland (EBN) and Shell are all partners in the project. The goal is to do more scientific research on geothermal energy as well as make use of the earth’s energy to heat buildings. The geothermal project on TU Delft campus will become part of a large-scale urban laboratory, to study optimal production scenarios and monitoring techniques in order to achieve the highest possible energy efficiency. Research and monitoring infrastructure will be used to investigate the fundamental scientific challenges that are presently limiting the development of geothermal energy. This could not be realized by using an existing geothermal plant that is built for commercial energy production because the monitoring equipment cannot be installed once operation has started and drill cores are not taken in commercial wells. The geothermal source on the TU Delft campus will be built with extensive monitoring and testing instrumentation, both downhole and at the surface. The scientific infrastructure for this project is funded by the EPOS-NL project, and associated research projects (e.g. WarmingUP, EASYGO) will utilize this facility.

ID: 73662

Start date: 15-11-2021

End date: 15-11-2029

Last modified: 10-09-2024

Copyright information:

Geothermal Project on TU Delft Campus

General Information

Name applicant and project number

This DMP is for a sub-project within a larger NWO-funded research infrastructure. The parent project, EPOS-NL, already has a separate DMP that was submitted to NWO in 2018. The main applicant for the EPOS-NL project is Prof. M. R. Drury (Utrecht University) and the project number is 184.034.005.

The sub-project addressed in this DMP, the geothermal project on TU Delft Campus, is lead by Dr. Phil Vardon who is the head of the scientific programme. Dr. Susanne Laumann is the project manager. This DMP was initially created by the former data manager Dr. Lora Armstrong, and it is updated and implemented by Dr. Liliana Vargas Meleza, the current data engineer. This DMP is for internal purposes and is not required to be submitted to NWO.

We have opted to use the latest version of the NWO DMP template (Template version 3, publish 29 Jan 2021) because it is more thorough than the older version that was in use when the project was funded (Template version 1). Some additional sections have been added below in order to give the necessary background for the geothermal project on TU Delft campus.

Project organization

Partners

TU Delft, Aardyn, Energie Beheer Nederland (EBN) and Shell are partnering together to instal a geothermal source on the TU Delft campus. In order to operate it, the partners have jointly formed a company called Geothermie Delft (GTD for short; see https://geothermiedelft.nl/).

Scientific programme

The scientific programme is overseen by scientific staff from Geoscience and Engineering Department at TU Delft.

Communication between partners about the scientific programme

An Integrated Operations & Research Council (IORC) is being set up for the coordination and communication activities related to the scientific programme. The aim of the IORC is the close coordination and cooperation between operational procedures and R&D activities and the active contribution of the GTD shareholders and the operator to the scientific programme.

The IORC consists of a representative of every shareholder of GTD (Aardyn, EBN, Shell, TU Delft), a representative of the operator and the R&D programme manager of the TU Delft.

Project phases

With respect to data collection, there are 3 project phases:

- Pre-production phase: Before the well is drilled. Data collected during this phase can be used to compare with data collected after the well is installed and operational.
- Implementation phase: This phase comprises the actual drilling, coring, installation and testing of the well.
- Production phase: This phase begins once the well is operational and is producing heat.

Data contribution to the European EPOS project

The EPOS-NL project (https://epos-nl.nl/) is a national research infrastructure in the Netherlands funded by NWO. EPOS-NL intends to contribute data to the European Plate Observing System (EPOS; https://www.epos-eu.org/), a larger, EU-funded project that integrates solid earth science data and facilities at the European level. Different communities (each called a 'Thematic Core Service', or TCS) exist within the wider EPOS project. Each community deals with a specific theme and manages the collation and sharing of data related to that theme. EPOS is also in the process of developing a central data portal that will combine data from all of the different communities in one place. The contributions (facility access and data access) that EPOS-NL is expected to make to the EPOS project can be viewed in a diagram on the EPOS-NL website under the 'About > Contribution to EPOS' section.

At the time the EPOS-NL project was funded, it was intended that the geothermal source would contribute data to 'Geo-Energy Test Beds for Low Carbon Energy' (GETB; https://www.epos-eu.org/tcs/geo-energy-test-beds-low-carbon-energy), an emerging community within EPOS. In of April 2021, GETB requested to merge with a more established EPOS community, Multi-Scale Laboratories (MSL; https://www.epos-eu.org/tcs/multi-scale-laboratories). It is envisioned that GETB would be re-invented within MSL as a subdomain relating to large scale and living laboratories. Pending approval by the MSL steering committee and the EPOS leadership, GETB data services are now expected to be developed within the MSL community using existing infrastructure.

Data sharing

The data sharing model within MSL is compatible with what was already planned for the geothermal project on TU Delft campus, i.e., sharing data via data publication in an institutional data repository (4TU.ResearchData in the case of DAPWELL). MSL maintains a data catalog (https://epos-msl.uu.nl/) that harvests metadata from different institutional repositories, making it easy to explore datasets from different MSL subdomains in one central location. MSL also maintains discipline-specific metadata schemas and vocabularies for each of its subdomains. It is expected that such a metadata schema would be developed in the future for GETB.

Finally, it should be noted that all data produced by the geothermal project will be openly shared via publication in data repositories and data centers, as described in this data management plan, regardless of the development or existence of relevant EPOS data infrastructure and services.

Name of data management support staff consulted during the preparation of this plan and date of consultation.

- Lolke Boonstra, ICT Innovation. Consulted various dates 2020-2021 (consultation ongoing)
- Marta Teperek, Head of Research Data Services (TU Delft Library) & Director of 4TU.ResearchData (consulted September, 2020)
- Kees den Heijer, former CEG Faculty data steward, consulted April 2021
- Lora Armstrong, CEG Faculty data steward. Consultation ongoing as per October 2022

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- No

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

The table below contains details on the data that is expected to be produced. For much of the production data, the collection schedule is flexible and it is not planned to collect the data continuously over the lifetime of the project. Rather it will be collected only when needed to meet science objectives, which will be refined over the course of the project. Once the geothermal well enters the production phase and data is being generated, this section will be updated in consultation with the scientific team.

Project phase(s)	Data Category	Data source/Description	Format	Size or amount
Pre-production & Production	Digital- monitoring data	Geophones- seismic waveform data	mSEED	6.7 GB/day; 0.6 TB/year (collected for 90 days)
Pre-production & Production	Digital- monitoring data	EM surface network- 10 receivers + source	To be determined	500 GB/month; 2 TB/year (collected for 4 months)
Implementation	Digital- monitoring data	Logging data	LAS	MBs-GBs
Implementation	Digital- monitoring data	Well testing data	Tabular (.xls, .csv, .txt or similar)	MBs
Implementation	Digital- monitoring data	Real time operations data 'RTOC'	Tabular (.xls, .csv, .txt or similar)	GBs
Production	Digital- monitoring data	Distributed Acoustic Sensing (DAS)	SEG-Y	400-500 GB/day for main well, 90 GB/day for monitoring well; 40 TB/year total (collected for 90 days in main well, 21 days in monitoring well)
Production	Digital- monitoring data	Distributed Temperature Sensing (DTS)	LAS	MBs-GBs
Production	Digital- monitoring data	Distributed Pressure Sensing (DPS)	LAS	MBs-GBs
Production	Digital- monitoring data	Distributed Strain Sensing (DSS)	To be determined	MBs-GBs
Production	Digital- monitoring data	Hydrogeochemistry bypass	Tabular (.xls, .csv, .txt or similar)	MBs-GBs
Production	Digital- monitoring data	Well production data	Tabular (.xls, .csv, .txt or similar)	MBs-GBs
Pre-production	Physical sample(s)	Groundwater samples	Physical Samples	To be determined
Implementation	Physical sample(s)	Cuttings	Physical Samples	To be determined
Implementation	Physical sample(s)	Rock cores	Physical Samples	300 m of cores maximum
Implementation	Physical sample(s)	Down hole fluid sampling	Physical Sample(s)	1 to several samples total
Implementation	Physical sample(s)	Sand	Physical Samples	To be determined
Production	Physical sample(s)	Water samples from sample taps	Physical Samples	Samples taken at most once per day; 100s of samples per year (collection schedule not yet known)
Pre-production	Digital- derived from sample	Groundwater analyses	Tabular (.xls, .csv, .txt or similar)	KB/sample
Implementation	Digital- derived from sample	Mud logs	Tabular (.xls, .csv, .txt or similar)	MBs-GBs
Implementation	Digital- derived from sample	Core logs	Text (PDF)	MBs-GBs
Implementation	Digital- derived from sample	Core photos	Image (JPEG)	500 GBs
Implementation	Digital- derived from sample	Core scans	.IMA, .TIFF	2 TBs
Implementation	Digital- derived from sample	X-ray diffraction analyses of samples	Tabular (.xls, .csv, .txt or similar)	MBs
Implementation	Digital- derived from sample	X-ray fluorescence analyses of samples	Tabular (.xls, .csv, .txt or similar)	MBs
Production	Digital- derived from sample	Cation analyses of water samples	Tabular (.xls, .csv, .txt or similar)	KB/batch of analyses; KBs-MBs per year
Production	Digital- derived from sample	Anion analyses of water samples	Tabular (.xls, .csv, .txt or similar)	KB/batch of analyses; KBs-MBs per year
Production	Digital- derived from sample	Gas analyses of water samples	Tabular (.xls, .csv, .txt or similar)	KB/batch of analyses; KBs-MBs per year
Production	Digital- derived from sample	Manual measurements on water samples	Tabular (.xls, .csv, .txt or similar)	KB/sample; KBs-MBs per year

1.3. How much data storage will your project require in total?

- >1000 GB

The storage required is estimated to be approximately 43-45 TB per year.
Physical samples (cores, sand, cuttings and fluids) will be stored at TU Delft for at least 5 years. Half of each core will be transferred to the national core repository of the Dutch Geological Survey (TNO) in Zeist for archiving as required by the Dutch Mining Act.

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

All datasets published at 4TU.ResearchData or the SURF Data Repository will be accompanied by README files providing documentation necessary for data re-use data. Guidance provided by 4TU.ResearchData will be followed when preparing the README files. The files will include definitions of all variables used in the datasets, as well as units of measurement. Additional metadata may also be included in the documentation, and is described in section 2.2.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

4TU.ResearchData
All scientific data except for seismic waveform data and potentially DAS data will be made openly available through 4TU.ResearchData. 4TU.ResearchData is a trusted and certified research data repository (it has a Data Seal of Approval certification). All datasets will be accompanied by rich metadata (adhering to the DataCite metadata standard) to ensure that they are findable. In addition, to further aid their discoverability, keywords describing the datasets will be added. 4TU.ResearchData also uses schema.org metadata, meaning that all datasets are indexed in Google Dataset Search. Every dataset will be also assigned a Digital Object Identifier (DOI), to make them citable and persistently available.
Discipline-specific metadata developed within the EPOS project may also be included with the data, preferably in custom metadata fields defined on the 4TU.ResearchData repository. Use of this metadata is dependent on future community efforts within the EPOS project to define metadata schemas for 'Geo-Energy Test Beds', the EPOS community related to DAPWELL and other living laboratories. If the required metadata schemas are developed within the EPOS project, they will make it possible for EPOS-maintained data catalogs and portals to harvest the metadata for DAPWELL datasets, leading to improved discoverability of the data within the geologic community.
If possible, International Geo Sample Numbers (IGSNs) will be acquired for all samples collected as part of the project. IGSNs are persistent identifiers for physical samples. They will be included in the documentation of all datasets relating to physical samples that are published on 4TU.ResearchData. The System for Earth Sample Registration (SESAR) has been contacted about acting as an allocating agent for DAPWELL IGSNs, but this is not yet confirmed because SESAR is developing a new business model that is not yet ready to be shared. If IGSNs are allocated by SESAR, metadata relating to each registered sample will also be made publicly available in the SESAR sample catalog. Each sample receives its own page in the online catalog where the associated metadata is displayed (see an example here). These pages can also be linked in publications, making sample metadata easily findable via relevant journal articles. The metadata schema used for sample registration is available on GitHub. For more information about the benefits of using IGSNs, see Lehnert et al. (2020). In the case that it is not possible to acquire IGSNs, unique identifiers will still be assigned to samples within the project and included in the documentation and datasets.
Metadata developed internally within the geothermal well project will be included in the documentation accompanying datasets where appropriate. This includes metadata related to lithology, instrumentation, and measurement types/locations/depths/dates/times.
ORFEUS data center (seismic waveform data)
Seismic waveform data shared via the ORFEUS data center will be accompanied by Station XML metadata (http://www.fdsn.org/xml/station/), following the schema developed by the International Federation of Digital Seismograph Networks.
SURF Data Repository (DAS data, if required)
If it is not possible to share DAS data via the 4TU.ResearchData repository, it will be published in theSURF Data Repository accompanied by DataCite metadata and a DOI.

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

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

Monitoring data will initially be transferred to a 'control room' located near the well. Interrogators to receive and process fibre optic data will also be located there, along with local data storage. Monitoring data will be transferred via a dedicated connection from the control room to a data center maintained by TU Delft ICT services, where it will be stored and automatically backed up during

research. Other types of data will also be stored in the data center, but will not need to be transferred via the control room.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Not applicable (no sensitive data)

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- No

4.2 How will ownership of the data and intellectual property rights to the data be managed?

After the partners form the Geothermie Delft company to operate the well, an R&D agreement will be signed that includes agreements on data ownership and IP rights. It is expected that TU Delft will have ownership of all scientific data collected in the geothermal project, whereas the well operator will own any operational data collected. During the active phase of research, the data manager/engineer, in consultation with the scientific head and programme manager, will oversee the access to scientific data (and other outputs), as well as any requests for access from external parties. Related projects that require use of the scientific data, for example EASYGO, will be given full access to the relevant data. Data will be released publicly no later than at the time of publication of corresponding research papers.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

The project has the potential to generate large volumes of data, requiring careful consideration of what data are selected for long-term preservation. Some data types will be fully preserved (in particular, data collected during the implementation phase, and data derived from samples). The decision about what monitoring data to select for long-term preservation will be undertaken by the scientific team after initial data collection begins. Data underlying scientific publications will always be preserved.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- Yes

Operational data from the project is expected to be owned by the well operator (this will be formally defined in the R&D agreement, see question 4.2) and cannot be made public without their consent. This includes real time operations data and well production data listed in Table 1.
Due to the large volumes of scientific data that could potentially be generated in the project, as well as the potentially long lifetime of the project, it is likely that not all scientific data will be selected for long-term preservation (see previous question). Only scientific data selected for long-term preservation will be made publicly available.

5.3 What data will be made available for re-use?

- Other (please specify)

Scientific data selected for long-term preservation (including all data underlying publications) will be made available for re-use.

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available as soon as article is published

All research data underpinning research papers will be published no later than at the time of the publication of corresponding research articles. Data may be embargoed until the primary scientific publication or (publication series) describing the results is published in a peer-reviewed international journal, in which case metadata will still be made available at the time of data publication. Any embargo period will be consistent with the overall EPOS data policy (which allows an embargo period of 2 years from the time of data publication).

Data deposited at 4TU.ResearchData is preserved for at least 15 years.

Data deposited in the SURF Data Repository will be preserved for a minimum of 10 years.

5.5 In which repository will the data be archived and made available for re-use, and under which license?

All data except for seismic waveform data and potentially DAS data will be published at 4TU.ResearchData, which is a trusted and certified research data repository (Data Seal of Approval certification). All datasets will be licensed under a CC-BY license which requires attribution/credit for the original creation, while at the same time ensures broadest possible re-use. All datasets will be accompanied by rich and descriptive metadata, compliant with the DataCite metadata schema, to ensure that all datasets are findable and accessible online. See https://data.4tu.nl/info/en/ for more information.

Seismic waveform data from geophones will potentially be shared via KNMI's ORFEUS data center (https://www.orfeus-eu.org/data/odc/), which provides access to data through the web interfaces and web services of the European Integrated Data Archive (EIDA; http://orfeus-eu.org/data/eida/). The waveform data will be accompanied by Station XML metadata (http://www.fdsn.org/xml/station/), following the schema developed by the International Federation of Digital Seismograph Networks. As described on the KNMI Seismic & Acoustic Data Tools page (http://rdsa.knmi.nl/), a CC-BY license is applied to data unless otherwise specified.

If it is not possible to share DAS data via the 4TU.ResearchData repository, it will be published in theSURF Data Repository, which specializes in storing large volumes of data in the TB-PB size range. Any data stored in this repository will be accompanied by DataCite metadata and a DOI, and published under a CC-BY license.

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

There are currently no plans to generate software as part of this project.

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

4TU.ResearchData is able to archive 1TB of data per researcher per year free of charge for all TU Delft researchers. This free storage is sufficient for publishing most data types in the project in a FAIR manner. 4TU.ResearchData also provided a letter of support for the EPOS-NL proposal indicating their willingness to facilitate the publication of data from the project. Additional funds have been included in the EPOS-NL project budget for data-related costs, which can be used for costs related to the publication of DAS data.

In terms of personnel, the dedicated data manager/engineer hired in the project will be responsible for data management in the project.