PUSH-IT

Piloting Underground Seasonal Heat Storage In geothermal reservoirs

D5.3 Data Management Plan

Funded by
the European Union

Grant agreement number	101096566		
Project title	Piloting Underground Storage of Heat In geoThermal reservoirs		
Project acronym	PUSH-IT		
Start date	01-01-2023	DURATION	48 months
End date	31-12-2026		
Work Package No. and Title	WP5 Project coordination		
Deliverable number	D5.3		
Deliverable title	Data Management Plan		
Actual date of delivery	27-06-2023		
Nature	DMP	DISSEMINATION LEVEL	Sensitive
Deliverable Leader	Technische Universiteit Delft		
Author	Danitsja van Heusden-van Winden		
Co-authors and contributors	PUSH-IT Consortium		

Disclaimer

Funded by the European Union

Funded by the European Union under the Horizon Europe programme (grant no. 1011096566). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor CINEA can be held responsible for them.

Document history

Version	Description (Section, page number)	Author	Role (Organisation short name)
V0.1	First draft	Danitsja van Heusden	TUD
V0.1a	Review	Consortium partners	Consortium partners
V0.2	Second draft Processed of feedback received from reviewers / partners	Danitsja van Heusden	TUD
V0.2a	Review	Consortium partners	Consortium partners
V1.0	Final version	Danitsja van Heusden	TUD

Quality control

Author	Organisation short name	Role	Date
Danitsja van Heusden	TUD	Work Package Leader & Deliverable Leader	02-05-2023
Consortium partners	Consortium partners	Reviewers	16-05-2023 + 19-06-2023
Martin Bloemendal	TUD	Scientific Project Coordinator	27-06-2023
Danitsja van Heusden	TUD	Project Coordinator	27-06-2023

Table of Contents

List of tables

List of figures

List of Abbreviations

DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
HTPS	Hypertext Transfer Protocol Secure
LCSH	Library of Congress subject Headings
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting

1. Introduction

1.1. Purpose of this document

This Data Management Plan (DMP) provides a reference point for the consortium in terms of how data is dealt with and stored both during and after the end of the project. In particular, it describes how the DMP is in line with Guidelines on Findable, Accessible, Interoperable and Reusable (FAIR) Data Management in Horizon Europe[1] to make data findable, accessible, interoperable, and reusable, during and after the project. As such, this DMP is a dynamic document and will be updated as required throughout the project. If at any point consortium partners notice potentially conflicting data management procedures across different outputs of the PUSH IT project, they should inform the Coordinator (TU Delft) and Ethics lead (ARU).

This plan describes the Open Access strategy to disseminate scientific results and what data the project will identify and generate, how to handle it, and how the FAIR principles are implemented in the PUSH-IT project. In Table 1 the different forms/types of data relevant for the PUSH-IT project are described.

1.2. Starting points for PUSH-IT Data management

Partners will provide open access to newly generated data related to the public deliverables, non-protected results or results patented not to jeopardise commercial exploitation.

- Commercially sensitive data will be treated confidentially. Other datasets underlying research papers will be made openly available under a CC-BY or CC BY-NC(-SA): licence, depending on the source of the data, through the 4TU.RD no later than when papers are published. Datasets will be accompanied by DataCite metadata and be assigned a Digital Object Identifier (DOI). Whenever possible, open/common file formats will be used to make the data interoperable and reusable. Information needed to validate, reproduce, or reuse the research outputs will be shared in a PUSH-IT file accompanying each dataset. The PUSH-IT community in Zenodo and 4TU.ResearchData will preserve PUSH-IT data for 10 years[2].
- Software and codes presented in academic papers will be shared on GitHub, published via 4TU.RD, and publicly available for re-use under an open license and citable and available with a DOI.
- Project deliverables will be publicly available through the PUSH-IT website and the EU portal, unless classified as confidential. The education materials will be openly available through an educational platform such as CC license or edusources.nl. TUD will maintain the project website during 5 years after project lifetime.
- Personal data will be treated confidentially and handled according to the General Data Protection Regulation (GDPR) Directive.
- Research data collected with human participants will only be published on the open data repository when explicit consent has been obtained, and the data has been appropriately anonymised (which in some cases may mean removing significant portions).

[1] EC, H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. 2016.

[2] Delft, T. TU Delft & Faculty policies. 2021

The partners are entitled to use published or declassified knowledge or results from the project for research and teaching purposes. The PUSH-IT website will also contain an overview/archive of all published data and/or links hereto.

Table 1: Summary of the PUSH-IT Data Management Plan

Activity	Data collected or generated	Re-used data	Data sharing platform	Size of new data	Access to new data
Knowledge production in demo/follower-sites		Experimental data, code	4TU.RD, GitHub	Order of 10 Tb	Open access
Public engagement, societal benefits, and risks	Transcripts from interviews with PUSH-IT team members and partner organisations Transcripts from interviews and/or focus groups with community members Surveys with PUSH-IT team members Surveys with members of the public Transcripts from interviews with regulatory stakeholders	Not applicable	4TU.RD, (de-identified data after data anonymisation or aggregation / no personal data shared)	Order of 10 Gb	Restricted access because of GDPR.
Simulations	Network data, heat production, heat consumption	Subsurface data, local outcrop data, experiments on subsurface samples	4TU.RD GitHub.	Order of 10 Tb	Open access

Activity	Data collected or generated	Re-used data	Data sharing platform	Size of new data	Access to new data
Water quality	Geochemistry and microbiology	Existing public thermodynamic databases and free codes Existing on site measurements (if any)	4TU.RD GitHub.	Order of 10 Gb	Open access
Spatial multi criteria analysis	Geographical data	Geological data, network data, heat production, heat consumption	4TU.RD GitHub.	Order of 10 Tb	Open access
Smart DH network controller development and demo	Heat consumption and indoor temperature data in high resolution (15 minute to 1 hour)	Not applicable	VITO FlexHarvester platform	Order of 10 TB	Confidential
PUSH-IT project related files	**Contracts, deliverables, reports, minutes, presentations**	Not applicable	MS TEAMS	Up to 25 TB per site or group	Restricted access because of GDPR.

2. Data Summary

2.1. Will you re-use any existing data and what will you re-use it for?

Existing data will be re-used/re-generated in some cases. Table 2 provides an overview of the data that will be re-used.

Table 2: PUSH-IT re-used/re-generated data

Data collected or generated	Re-used data	Reason
Geological and monitoring data	Geological and monitoring data From experiments/measurements Geological data, network data, heat production, heat consumption	Knowledge production in demo/follower-sites. To extend our knowledge on subsurface characteristics, allowing to make better system design and control of heat flow in the underground. As well as improved insights on heat transport and distribution in the underground for the simulations.
Transcripts from interviews with PUSH-IT team members and partner organisations **Transcripts from interviews and/or focus groups with community members** **Surveys with PUSH-IT team members** **Surveys with members of the public** **Transcripts with regulatory stakeholders**	Not applicable	For this activity, PUSH IT does not currently plan to process any previously collected data. All activities will involve primary data (and therefore informed consent). In the unlikely event that the PUSH IT team does seek access to pre-existing data within the boundaries of this specific activity, consent must still be obtained from the original participants in the same manner as with new data if it is to be included in PUSH IT project-specific outputs and records. This should happen before any data is transferred to the project.
Geological data, network data, heat production, heat consumption	Subsurface data, local outcrop data, experiments on subsurface samples	Simulations

Data collected or generated	Re-used data	Reason
Geochemistry and microbiology	Measurements on-sites Existing public thermodynamic databases and codes	Knowledge production in demo/follower-sites. To extend our knowledge on subsurface characteristics (and notably on water properties), allowing to better anticipate and manage water issues in the reservoir, in wells and in shallower aquifers
Geographical data	Geological data, network data, heat production, heat consumption	Spatial multi criteria analysis
Heat consumption and indoor temperature data in high resolution	Historical time series of heat consumption related data, indoor temperature data, heat production data	Training of black- and gray-box models used by the smart DH network controller development

2.2. What types and formats of data will the project generate or re-use?

This project will generate and reuse different types of datasets.

Existing data will be re-used/re-generated. Table 3 provides an overview of the types of data that will be re-used.

Table 3: PUSH-IT types of re-used/re-generated data

Data collected or generated	Type	Format
Geological and monitoring data	Pictures and descriptions files on logging/cuttings/experiments, Raw and processed data from logging/monitoring	word, xls, pdf, picture-formats, ddl, txt, xls
Transcripts from interviews with PUSH-IT team members and partner organisations **Transcripts from interviews and/or focus groups with community members**	Verbatim transcripts of interviews Survey results	word, xls

Data collected or generated	Type	Format
Surveys with PUSH-IT team members **Surveys with members of the public** **Transcripts with regulatory stakeholders**		
Geochemistry and microbiology	Observational data (measurements on sites, samplings, and laboratory measurements) will be acquired to characterize water and solid properties: – Major and minor elements will be characterized for water samplings (sites' owners, GFZ, BRGM) – Microbiological analyses will also be carried out for ATES reservoir and for shallow aquifers (when relevant) (BRGM) – Scalings will be characterized by XRF and XRD for Berlin and Bochum (GFZ) Geochemical modelling will re-use the results of TH modelling provided by other partners. Geochemical calculations will be done with public thermodynamic databases (Phreeqc.dat and Thermoddem.dat) and open codes (Phreeqc). During the project, potential updates of public thermodynamic database will be made available. BRGM and GFZ will release input/output of geochemical simulations which will be suitable for reuse.	Xls, txt, csv, word, picture-format, other specific formats (e.g. dat).
Heat consumption and indoor temperature data in high resolution	Time series of: – Indoor temperatures in buildings – Supply and return temperatures, flow rates, heat consumption of heat meters of buildings – Supply and return temperatures, flow rates, heat consumption of heat meters of (branches of) DH networks – Heat and power production of production units in DH networks – Heat exchange with thermal storages	csv, xls, direct API, json

2.3. Purpose of the data and relation to the objectives of the project?

Table 4: PUSH-IT purpose of re-used/re-generated data

Data collected or generated	Purpose
Geological and monitoring data	Extend our knowledge on subsurface characteristics (e.g., well-logging), allowing to make better system design and control of heat flow in the underground. As well as improved insights on heat transport and distribution in the underground (e.g., DTS). This is needed to develop new technologies for drilling and completion, and optimise operation efficiency
Transcripts from interviews with – **PUSH-IT team members** – **partner organisations** – **focus groups with community members** **Surveys with:** – **PUSH-IT team members** – **members of the public** **Transcripts with regulatory stakeholders**	Team interviews and surveys, conducted at beginning and end of the project: To explore knowledge and perceptions of effective community engagement as it changes over time Interviews with local community members: to enable us to produce tailored engagement methods for each site and develop good practice in the field of community engagement with geothermal specifically and renewable energy technologies more widely Surveys with members of the public: To develop our understanding of public awareness, knowledge and attitudes toward geothermal energy storage and geothermal operators Interviews with regulatory stakeholders: to understand and evaluate the regulatory barriers and enablers around underground geothermal storage across countries
Geochemistry and microbiology	The overall aim of task 3.3 is to anticipate, monitor and manage water quality issues. The measurements, laboratory experiments, simulation data/code will be used to assess possible impacts in the UTES reservoir, in shallower aquifers and in wells and facilities.
Heat consumption and indoor temperature data in high resolution	The data is used to train the models of the smart DH controller algorithms and is used for the evaluation of the performance of this controller.

It should be noted that the size for each type of data will be updated once it's better known.

2.4. Expected size of the data

Table 5: PUSH-IT expected size of data

Data collected or generated	Expected size
Geological and monitoring data	Delft: <10 TB
Transcripts from interviews with PUSH-IT team members and partner organisations **Transcripts from interviews and/or focus groups with community members** **Surveys with PUSH-IT team members** **Surveys with members of the public** **Transcripts with regulatory stakeholders**	Less than 10Gb
Geochemistry and microbiology	BRGM: < 10Gb
Heat consumption and indoor temperature data in high resolution	> 10 GB

2.5. Origin/provenance of the data

Table 6: PUSH-IT origin of data

Data collected or generated	Origin
Geological and monitoring data	Re-used data originates from local data sources of subsurface composition databases: borehole logs, subsurface models Generated data consist of own borehole and logging data, temperature measurements, water/geo-chemistry. Interpreted dataà inter/extrapolations of temperature distribution, from (numerical)models
Transcripts from interviews with PUSH-IT team members and partner organisations **Transcripts from interviews and/or focus groups with community members**	Interviews with team members, with regulatory stakeholders across the countries where the sites are located, and with community members near the location of each site. Surveys with team members and regional surveys with members of the public in each of the countries where the sites are located

Data collected or generated	Origin
Surveys with PUSH-IT team members **Surveys with members of the public** **Transcripts with regulatory stakeholders**	
Geochemistry and microbiology	Thermodynamic databases and open codes Phreeqc come from previous scientific works. Site-specific data originates from measurements carried out within the sites deployment (either from earlier steps or within the PUSH-IT project). Generated data will consist of dedicated additional measurements (e.g. microbiological diversity characterization), of laboratory experimental results, and of input/output modelling results.
Geographical data	
Heat consumption and indoor temperature data in high resolution	The data is recorded by heat meters in the buildings, the DH network, heat production units, storages.

2.6. To whom might your data be useful ('data utility'), outside your project?

Table 7: PUSH-IT potential target audience for data

Data collected or generated	Potential target audience for data
Geological and monitoring data	The obtained data is useful for 2 types of users: 1. Local parties that need detailed information from the site for any activity they want to do in the subsurface requiring detailed insight in subsurface composition and characteristics 2. Parties that want to also apply a heat storage system in a geothermal reservoir under (partly) similar conditions.
Transcripts from interviews with PUSH-IT team members and partner organisations **Transcripts from interviews and/or focus**	Different parts of the data will be relevant for different types of users: - Data on the team's understanding of community engagement will be useful for reflection for the team itself but also for other interdisciplinary teams on their journey to work collaboratively and on

Data collected or generated	Potential target audience for data
groups with community members **Surveys with PUSH-IT team members** **Surveys with members of the public** **Transcripts with regulatory stakeholders**	understanding knowledge awareness and gaps on the topic. - Data on regulations will be of interest to industry parties and to regulators themselves as it will give insights in barriers and enablers for the future of geothermal storage. - Data on community perceptions will be of interest to researchers, industry, and regulators to understand awareness and perceptions of geothermal across the relevant locations and be able to respond to them accordingly. - Localised data from interviews with community members will be of interest to community members themselves, and to researchers and industry parties who are concerned with community engagement.
Geochemistry and microbiology	The obtained data is useful for: - Local parties that need detailed information from the site for any activity they want to do in the subsurface requiring detailed insight in subsurface composition and characteristics - Parties that want to also apply a heat storage system in a geothermal reservoir under (partly) similar conditions. - Scientific community interested in geochemical/microbiological results.
Heat consumption and indoor temperature data in high resolution	The individual personal consumption data will not be shared outside the project. Anonymized and/or accumulated data might be used in publications/conference presentations on the results of the controller.

3. FAIR data

3.1. Making data findable, including provisions for metadata

3.1.1. Making data findable, including provisions for metadata: Will data be identified by a persistent identifier?

Data/code that can be publicly released will be published via the PUSH-IT community in Zenodo[3] and preferably in 4TU.ResearchData. Both are trusted data repositories who ensure the published datasets will be findable and accessible in the long-term (>10 years from the publication date). These repositories assign a persistent identifier (Digital Object Identifier DOI), which makes the datasets citable and persistently available. The 4TU.ResearchData repository in particular also uses OPeNDAP for NetCDF data which facilitates the findability and inspection of NetCDF records.

Data/code that will not be publicly released (i.e. commercially sensitive data) will be stored within the respective owner's institutional archive, accessible to the employees of the respective institution.

3.1.2. Making data findable, including provisions for metadata: Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

During the project, metadata will be maintained and updated in the NetCDF files. The simulation code also provides output following this convention.

For the publishing of the data/code, both publishing repositories - PUSH-IT community in Zenodo and 4TU.ResearchData - use standard citation metadata.

3.1.3. Making data findable, including provisions for metadata Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Relevant keywords following the Library of Congress subject Headings (LCSH) will be added in the citation metadata of the Zenodo and 4TU.ResearchData repositories.

3.1.4. Making data findable, including provisions for metadata Will metadata be offered in such a way that it can be harvested and indexed?

Datasets that can be publicly released will be published via the PUSH-IT community in Zenodo and 4TU.ResearchData. Both are non-discipline specific repositories and use a general citation metadata standard embedded in schema.org. This means the datasets will be indexed in Google Dataset Search.

All metadata in the PUSH-IT community in Zenodo and institutional repositories can be harvested via the OAI-PMH protocol and are findable through portals like BASE, Google Dataset Search, and DataCite Search.

[3] https://zenodo.org/communities/push-it/?page=1&size=20

Regarding the Knowledge and Data Hub, the harvesting and indexing of the derived metadata will be discussed in the course of the research.

All data (underlying the published papers) will be made openly available through 4TU.ResearchData which uses schema.org metadata, meaning that all datasets are indexed in Google Dataset Search.

3.2. Making data accessible

3.2.1. Making data accessible- Repository: Will the data be deposited in a trusted repository?

Datasets that can be publicly released will be published via the PUSH-IT community in Zenodo and in 4TU.ResearchData. Both are trusted repositories which ensure the archived datasets are findable and accessible in the long-term (>10 years from publication date). In addition to that, both allow for private and public repositories. For the purposes of this project, the datasets that can be publicly released will be archived in public repositories (without any embargo periods).

Making data accessible – Repository: Have you explored appropriate arrangements with the identified repository where your data will be deposited?

As mentioned before, two repositories will be used: PUSH-IT community in Zenodo and 4TU.ResearchData.

Zenodo will be used for the publication of smaller datasets ~50 GB max (free-of-charge) and datasets where no TU Delft researcher is involved in its generation/collection.

The **4TU.ResearchData** will be used for large datasets and/or publication of processed NetCDF files, since TU Delft researchers can upload up to 1 TB of data per year free-of-charge and the 4TU.ResearchData repository has extra features for the discoverability and reuse of NetCDF data (i.e. OPeNDAP).

Regarding special arrangements with the repositories themselves, we do not foresee any issues. This Data Management Plan has been drafted with the support of the Data Steward of the Faculty of civil engineering and geosciences of the TU Delft.

The Data Steward works closely with the 4TU.ResearchData repository and has informed us about the policies and procedures required for data publication. If needed, the Faculty Data Steward will refer us to professionals at 4TU.ResearchData to discuss possible additional arrangements.

3.2.2. Making data accessible – Repository: Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Both the PUSH-IT community in Zenodo and 4TU.ResearchData ensure the archived datasets are assigned an identifier (DOI).

3.2.3. Making data accessible – Data: Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific

beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Not all data will be made publicly available. Data/code that will remain close is the commercially sensitive data or personal data. Thus, such data will be shared among relevant collaborators for the purposes of this project only. This data includes the raw observational data, simulation code and part of the simulation data.

Research data involving human participants, such as raw interview and survey data containing personal information, will be mostly gathered within WP2. These kinds of data will only be stored and accessible in the restricted WP2 internal project repository in order to guarantee the security and privacy of such data. Moreover, it will only be published on the open data repository when explicit consent has been obtained, and the data has been appropriately anonymised (which in some cases may mean removing significant portions).

Data that will be publicly released via the PUSH-IT community in Zenodo and 4TU.ResearchData are the processed observational database and part of the simulation data including some of the optimized trajectories which can be used to exemplify the results of the project. All the non-commercially sensitive datasets and codes (underlying the published papers) will be made openly available through the Zenodo and 4TU.ResearchData repositories at the time of the publication of corresponding research papers resulting from this study."

Only data produced by the project partners will be addressed in this Data Management Plan.

3.2.4. Making data accessible – Data: If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

We do not foresee any embargo regarding the data/code collected/generated in this project.

3.2.5. Making data accessible – Data: Will the data be accessible through a free and standardized access protocol?

4TU.ResearchData uses the Hypertext Transfer Protocol Secure (HTTPS) protocol which is based on TCP/IP, which is an open protocol for most of the internet.

Zenodo uses HTTP and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) mechanism for repository interoperability.

3.2.6. Making data accessible – Data: If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

There will not be any restrictions on use for the datasets that will be publicly released via PUSH-IT community in Zenodo and 4TU.ResearchData.

Regarding the datasets that will remain at the respective owner institution, contact points will be provided in this Data Management Plan in case access is required (e.g. regulatory purposes).

3.2.7. Making data accessible – Data: How will the identity of the person accessing the data be ascertained?

Identification of the person accessing the data is not necessary for the public repositories of the PUSH-IT community in Zenodo nor for 4TU.ResearchData.

Regarding the datasets that are commercially sensitive, the respective owner institution has the right to decide under which circumstances access will be given to others if requested.

3.2.8. Making data accessible – Data: Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

There is no need for a data access committee, but we highlight that the anonymisation process may mean removing portions of text in order to protect the identity of the individuals. For task T2.3, participants will be recruited due to their professional roles, and they may be asked if they are happy to be named (as an individual/organisation) in the general acknowledgments section of PUSH-IT outputs. The inclusion of names (individual/organisation) in project outputs is dependent on informed consent being obtained, and a clear procedure will be outlined in any amendment to the ethical approval already granted to the project.

3.2.9. Making data accessible –Metadata: Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

For the datasets that will be publicly released, the metadata is openly available under the CC-BY license and can be accessed directly from the dataset repository's landing page.

For the datasets that will not be publicly released, information about the data will be presented in the related articles (which will be published following the HE Open Science policy).

3.2.10. Making data accessible –Metadata: How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

The published datasets and the respective (citation) metadata will be preserved for > 10 years from the end of the publication data.

Datasets that will not be published will also be archived for the long-term following the owner institution archiving policies.

3.2.11. Making data accessible –Metadata: Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

The details on the requirements of required special software are provided in the meta data of any model/data published. It is the user of the data own responsibility to obtain required software to run models.

3.3. Making data interoperable

3.3.1. Making data interoperable: What data and metadata vocabularies, standards, formats, or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

One of the research outputs of this project is the observational database which will collect and make interoperable measurement data coming from different instruments. Depending on the data format, the database will be preferably in NetCDF format which has embedded metadata following CF conventions for the naming of variables.

3.3.2. Making data interoperable: In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?

We will use common terminology used in the field to describe the data. No new ontologies nor vocabularies will be created.

3.3.3. Making data interoperable: Will your data include qualified references[4] to other data (e.g. other data from your project, or datasets from previous research)?

A qualified reference is a cross-reference that explains its intent. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://data.4tu.nl/Delft_University_of_Technology).

For each published dataset, any other related datasets (e.g. code repository) will be tagged (via the DOIs) in the References citation metadata item of the repository. This also includes the DOI of the related article(s). Likewise, the DOI of the dataset will be added to the related article(s) in a Data Availability Statement or section alike as well as cited in the reference list to allow tracking of citations.

3.4. Increase data re-use

3.4.1. Increase data re-use: How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

All documentation needed to validate data analysis and facilitate data re-use will accompany the data via a README file.

3.4.2. Increase data re-use: Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

The data/code published via PUSH-IT community in Zenodo and 4TU.ResearchData repositories will be released via permissive open licenses such as CC-BY for data and the software licenses, such as MIT and GPL.

[4] A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta) data resources to enrich the contextual knowledge about the data. (Source : https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)

3.4.3. Increase data re-use: Will the data produced in the project be useable by third parties, in particular after the end of the project?

Most of the data (and code) produced in this project will be useable by third parties, being publicly released via a trusted repository (such as PUSH-IT community in Zenodo and preferable in 4TU.ResearchData repositories).

As mentioned before, part of the data/code will be commercially sensitive and thus will remain restricted access for long-term, available to be reused privately at the respective owner institution.

3.4.4. Increase data re-use: Will the provenance of the data be thoroughly documented using the appropriate standards?

Each collected dataset will be properly documented with both embedded (e.g. file headers) and supporting documentation (e.g. README files) as explained in a previous answer. This will be done regardless of whether the datasets are published or not, as commercially sensitive data will also be reusable within the respective owner institution.

3.4.5. Increase data re-use: Describe all relevant data quality assurance processes.

The following checklist will be used as a reference for data quality:

- Datasets should be deposited in a proper repository if published, or in a secure institutional storage if not published.
- Datasets should be in proper directory (tree) structure with consistent file/directory naming convention. Basic rules:
    - do not use only numbers to name files
    - avoid use of white spaces in file/directory names
    - separate data files from code files (e.g. in different sub directories)
    - distinguish datasets per processing level (raw data, processed data, visualizations, finalized data)
    - utilizing version control systems, such as git, for tracking the histories of data and code and facilitating team work.
- Datasets must have at least a README file with basic structure: title, authors (and ORCIDs), description of data (units, instrumentation, software, etc.), description of the directory (tree structure and naming convention), caveats, references, copyright, and license.

In addition to that, for published datasets:

- citation metadata items (such as authors, description, title, etc.) should describe the dataset accordingly.
- Citation metadata keywords should follow a standard vocabulary (e.g. LCSH).
- Uploaded files should not be corrupted.
- All files mentioned in the README should be present in the repository.
- There should be a license file.
- The DOI of the related article(s) and dataset(s) should be added in the citation metadata. Likewise, the DOI of the dataset should be added in the related article(s) and dataset(s).
- Project partners will be consulted regarding confidentiality before publication of potentially infringing datasets.

3.4.6. Increase data re-use: Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

See the answers to the questions in chapter 4.

4. Other research outputs

All digital outputs other than the ones described above will be documented and stored according to the data elements described in this DMP. This includes data and code (incl. models). The management of these outputs is described in the answers of this DMP. There will not be physical research outputs.

4.1. Other research outputs

All digital outputs will be documented and stored accordingly. This includes data and code (incl. models). The management of these outputs is described in the answers of this DMP.

There will not be physical research outputs.

4.2. Fair police on other research outputs

Not applicable.

5. Allocation of resources

5.1. Costs for data storage

The relevant/required partner institutions provide the necessary infrastructure for the data/code management related to the PUSH-IT project. This includes:

- software and processing power required to carry out the simulations;
- version control software;
- data storage during the project (institutionally managed servers);
- data archiving in public archives (TU Delft researchers can upload up to 1 TB of data to the 4TU.ResearchData per year free of charge, and Zenodo allows uploads of up to 50 GB each free-of-charge);
- data archiving for the data that will remain under closed-access due to commercial reasons (institutionally managed servers).

5.2. Costs for making data or other research outputs FAIR

No costs have been considered regarding research data/code management services/infrastructure.

5.3. Coverage of open data costs

Currently, non-applicable. Unless the deposited data exceeds the project repository's free of charge data limit budget allocation should be considered for larger data storage.

5.4. Who will be responsible for data management in your project?

Regarding responsibilities, researchers will manage the data related to their respective Work Packages.

Each partner institution will be responsible for the proper management of the data and code they generate/collect. Table 8 lists the team members who will be overseeing the data/code management at their respective institutions and ensure timely transfer of those data to central PUSH-IT repositories.

Table 8: Data management per institution

Beneficiary name	Data management team member
Technische Universiteit Delft	Xinyan Fan, Data Steward at Delft university of Technology, Faculty of Civil Engineering and Geosciences Martin Bloemendal, Assistant Professor at Delft university of Technology, Faculty of Civil Engineering and Geosciences
Equans Nederland	Matthijs Scholten
KWR WATER	Gerard van den Berg
Technische Universität Darmstadt	Ingo Sass

Beneficiary name	Data management team member
Fraunhofer Gesellschaft zur Förderung der Angewandten Forschung	David Bruhn
Helmholtz-Zentrum Potsdam - Deutsches GeoForschungszentrum	Katrin Kieling
Vlaamse Instelling voor Technologisch Onderzoek	Dirk Vanhoudt
Česká geologická služba	Vít Peřestý
Huisman Geo	Cees Willems
Bureau de Recherches Géologiques et Minières	Annick Loschetter
Universiteit Utrecht	Wen Liu
Grondboorbedrijf Haitjema	Ronald Folders
Univerzita Karlova	Josef Vlček
Université de Genève	Andreas Schmalz
Anglia Ruskin University Higher Education Corporation	Melanie Rohse
United Kingdom Research and Innovation	Andres Gonzalez Quiros
Geothermal Engineering	Amy Peach-Gibson
University of Exeter	Iain Soutar

All researchers will be encouraged to request advice to their respective support staff in case necessary.

5.5. Long term preservation of data

In PUSH-IT, long-term preservation of data is expected to be at least 10 years from the moment the datasets are finalized. This is ensured for the datasets that will be published via PUSH-IT community in Zenodo and preferable in 4TU.ResearchData. Regarding the data/code that will be shared via the Knowledge and Data Hub, the specifics on the archiving process (including e.g., open-access policy, data availability, responsibilities, internal/external access, GDPR compliance for user credentials, etc.) will be designed and developed during the project.

6. Data security

6.1. Provisions for data security

During the project, researchers will be actively working on the data/code on work laptops/stations provided by their respective institution and remotely maintained by the respective ICT Department. Researchers will be encouraged to make master copies of sensitive data on institutional drives, accessed remotely via secure protocols (e.g., SFTP). This will ensure sensitive data is backed up on a daily basis following the respective institution's ICT security protocols. For the development of code, researchers will be encouraged to use version control and have remote repositories in institutional instances for sensitive code (e.g. TU Delft Gitlab) or Github for non-sensitive developments.

Once datasets (including code) are ready to be added/deployed in the Knowledge and Data Hub, UST-HLRS will provide guidance and means for data transfer from the institutional networks to that of the developed platform (e.g., GridFTP and UFTP). The data/code managed within the Hub will be maintained following UST-HLRS security and privacy policies.

Datasets underlying the results of the related (journal) articles and that will not be shared via the Hub, will be published via PUSH-IT community in Zenodo and 4TU.ResearchData following their security archiving policies.

6.2. Safely stored in trusted repositories

The data storage facilities used are well known and reputable repositories. PUSH-IT community in Zenodo and preferable in 4TU.ResearchData will securely store PUSH-IT data.

7. Ethics

7.1. Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Some PUSH-IT research activities will be taking place in non-EU countries and participants may come from non-EU countries. Data from PUSH-IT activities may also be shared with partners and transcribers/translators in non-EU countries. Additionally, as many project deliverables will be publicly available on the project website, the results of the project will be disseminated in non-EU countries as well. Relevant information sheets must always state clearly who data may be shared with.

All those with access to PUSH-IT data, regardless of location, must confirm they will adhere to the relevant project data management requirements, in particular EU data protection law.

7.2. Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

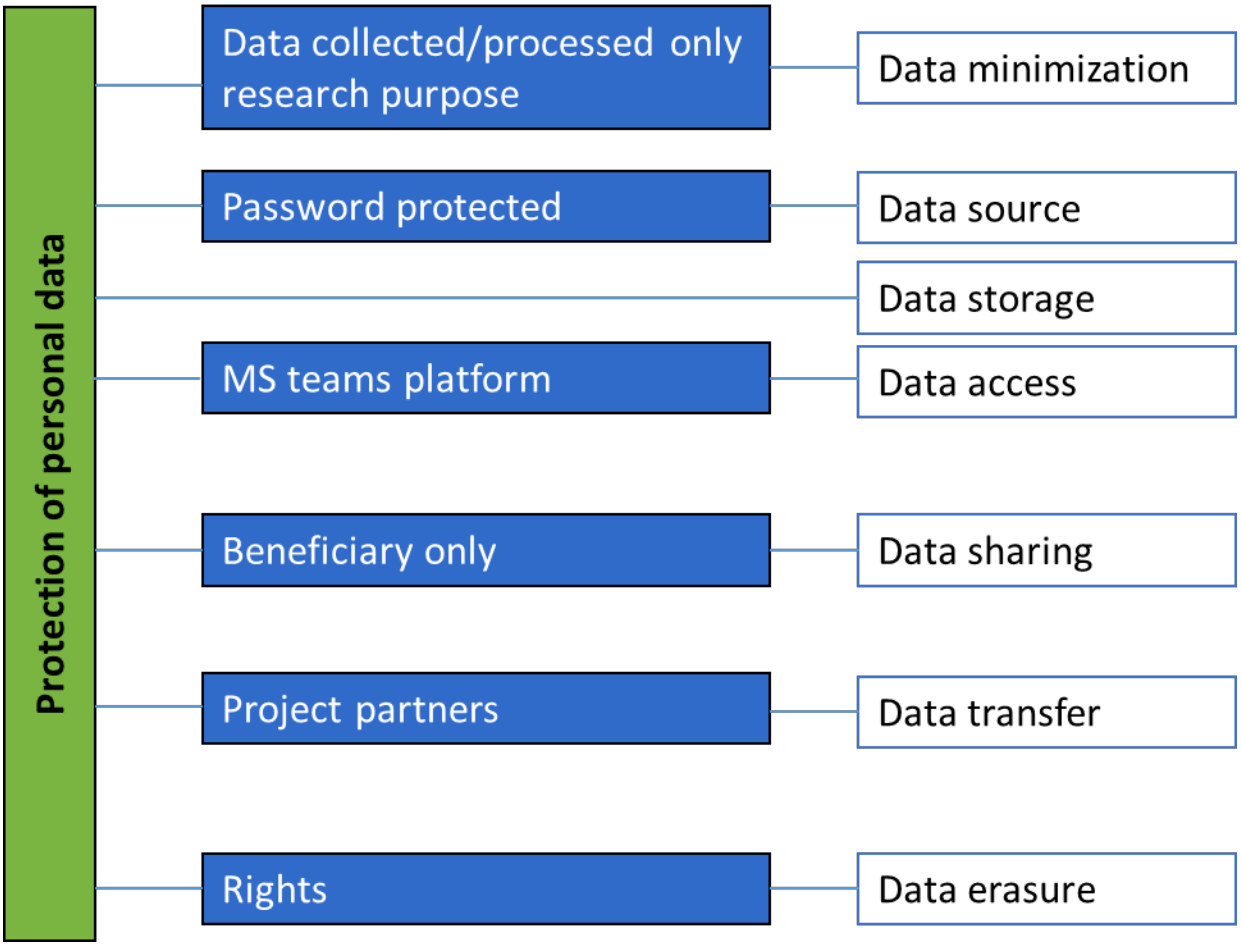
For each activity, ethical approval will be sought and in recruiting participants, all consortium partners will follow the principles of informed consent, described in deliverable D5.2 Quality and risk management report (and informed by the principles laid out in the Social Science and Humanities-led Horizon Europe project SSH CENTRE).

7.3. Protection of Personal Data

'Personal data processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction (art. 4.1 GDPR[5] of the EU regulation). D7.3 Data Management Plan offers a general framework for data collection, storage, protection, retention, and destruction according to EU legislation, directives, and opinions. Here we provide information about the way in which the PUSH-IT consortium has agreed to process personal data.

[5] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Figure 1: Processing of personal data in PUSH-IT



The GDPR requires adherence to several data processing principles (see art. 5.1 GDPR). For each of these principles we will explain how PUSH-IT has implemented or will implement them:

a. Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

The legal basis for processing of personal data for PUSH-IT research is in art. 6.1 GDPR: "the data subject has given consent to the processing of his or her personal data for one or more specific purposes". The informed consent procedure and template are detailed in D5.2 Quality and risk management report.

b. Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');

The purpose of the processing of personal data in the PUSH-IT projects is the performing scientific research in open data. Prior to the personal data processing, the objective of the research will be explained in writing to the participant (see D5.2 Information project sheet) and again explained in the activity that requires the processing of personal data.

c. Personal data shall be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

In section 4.1 we specify how PUSH-IT is adhering to the data minimisation principle.

d. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased, or rectified without delay ('accuracy');Data subjects will be asked to agree with the personal data collected and further processed. In this way the accuracy of the personal data is ensured. Relevant developments occurring during the course of the project that may impact the accuracy of the provided personal data will be shared and discussed with data subject, who will have the opportunity to update the personal data.
e. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation'); The personal data will be stored in a secure way for 10 years. All data will be anonymized, links between personal data of the participant and the collected data will be destroyed. Personnel in contact with identifiable data are restricted to solely the individual early stage researcher and their supervisor(s). Research data will be available for at least 20 years from the time of data deposition, as this is the current retention period of Zenodo.
f. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

PUSH-IT secure data storage

TU Delft (as coordinator) will keep copies of all final documents, reports, and deliverables of the project. These will be stored in a secure restricted access project drive, at Microsoft Teams backed-up daily and maintained by TU Delft IT. The drive will be accessible only to relevant team members via a secure protocol.

7.3.1. Data minimization principle

PUSH-IT adopts the data minimization principle[6], according to which data that is collected and processed is used only for the purpose of the research. To ensure the rights and freedoms of data subjects and research participants, the guiding principle is that at all times:

- appropriate disclosures to research participants via notices are made;
- provision of information is on a voluntary basis;
- explicit consent is obtained from research participants to make sure that they agree with usage of their information for the purpose of the research activity they are involved in;
- information will be treated confidentially and stored on secured file systems
- information will not be enforced by involved PUSH-IT researchers or research leaders.

Moreover, any participant may request the removal of their data, and such requests should be followed up adequately.

[6] The data minimisation principle is expressed in Article 5(1)(c) of the GDPR which provides that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed"

In those specific cases where potentially sensitive data would be collected, it will be anonymized before it is processed.

For those cases where personal data will be collected, the following data protection requirements must be fulfilled:

1. The data that is intended to be processed must be relevant and limited to the purposes of the research project (in accordance with the data minimization principle);
2. A description of the technical and organizational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be provided;
3. A description of the security measures that will be implemented to prevent unauthorized access to personal data or to the equipment used for processing it must be provided;
4. A description of the anonymization/pseudonymisation techniques that will be implemented must be provided;
5. Detailed information on the informed consent procedures concerning data processing must be provided;
6. Detailed information on access rights of the data subject must be provided;
7. Detailed information on the rectification and erasure procedures must be provided.

7.3.2. Data sources

Secondary use of personal data: Contact lists

With respect to the personal information obtained from the public domain (e.g., information from social media which is gathered to make a contact list), the consortium will inform the related individuals about this processing activity (collection) and the purposes of processing their data as far as possible. In addition, the consortium partner involved in the personal data processing is responsible to inform the research participants about their right to remove data and to give the participants the opportunity to withdraw their data. An example of the secondary use of data is the contacting list of the project partners and stakeholders in the project, which is made through web searching, contacting people, communications, and other stakeholders' activities. The project partners and stakeholders of PUSH-IT whose personal data is collected are informed about this data processing (collection of data) and are given the chance to choose not to be a part of the communication list of the project.

Recordings: Audio / video

In the cases that audio and video recording devices are used within PUSH-IT, the individual members of the research team are responsible for how audio and video recordings are used within PUSH-IT studies, and no one outside of the research team has direct access to any of these data. After the data transcription, the media will be deleted from digital recorders and stored digitally within a password-protected folder on the network drives of the participating institutions.

Permission for taking photos and videos will be requested before these are made. Consent should be provided for any exceptional cases, including at least any use of personal media and other content in the external dissemination of the research project. The project data information sheet and the informed consent form are appended into Annex 2 of this deliverable respectively.

Workshops, project meetings, training events and other events

For the organization of project events such as workshops, dissemination events and consortium meetings identifying data of the participants will be stored, including, e.g., name, organization, work address, email address, and possibly work telephone number. Solely information needed for the organization of the event will be gathered and stored privately and confidentially, while fully bound by applicable EU and national regulations. Non-identifying synthetic high-level data on,

e.g., the number of participants, ratios of industry versus academia, ratio of positive participant feedback, may be published online on the project website and in press releases. The complete data will be stored only as long as necessary for the project purposes.

If any pictures of audio/video recordings of audience are planned to be taken during a PUSH-IT event, participants will be informed in advance, and participants who do not wish to have their own image included will not be included and/or provided the possibility to participate from a separate area. Also, in case of online meeting, participants will be informed if screenshots will be taken in order to memorize the event in the form of a photo. Participants will have the opportunity to switch their cameras off in case they do not wish to be included in the photo. In case of online and teleconference meetings, tools that meet adequate levels of data exchange safety and security are selected, such as Microsoft Teams.

Interviews, surveys & questionnaires

The participants in the project activities, like interviews, surveys, and questionnaires, will be asked to give their informed consent (either in writing or verbally, which will be recorded, and in line with the ethics approval granted to the project) that describes the goals of the activity and how the collected information will be used. Participants will always have the right to withdraw from an activity, and this is detailed in the information provided to the participants ahead of their taking part in the project.

7.3.3. Data storage and data access

All collected data in PUSH-IT will be stored digitally within internal protected project repository in Microsoft Teams titled PUSH-IT. This Microsoft Teams group is hosted by Delft University of Technology (TU DELFT) and all members of the beneficiaries have been invited to join the Microsoft Teams Channel PUSH-IT. The channel is protected and can be accessed only by authorised individuals.

There is also a possibility to store the data obtained in PUSH-IT on a project storage drive protected by security facility provided by TU Delft, such as SURFDrive[7]. SURFDrive is a personal cloud service for the Dutch education and research. This storage drives of TU Delft are designed to store, backup, organise, annotate, and share research data with project partners all over the world.

On all data storage tools, the collected data will be stored securely on the server of the researcher institute in a way only accessible to the authorized personnel (controlled access to the data). When data are no longer required, they will either be deleted or kept in the anonymous mode when they serve historical, statistical, or scientific uses.

Sensitive data will be securely stored for the duration of each project in a trusted and access-controlled digital repository maintained by the Project Coordinator. The anonymised data will be stored on secure servers with a backup frequency of every month during the project and only accessible to those people who are authorized to get this information.

The anonymization of data is done by the responsible researcher and the anonymised data will be stored on the secure server. During the anonymization process, all direct identifiers, as well as the link between data and the individual, will be removed. Additionally, data will be stored pseudonymously (i.e., one attribute in a record will be replaced by another in a way that individuals are only identifiable under predefined conditions).

[7] https://www.surf.nl/en

7.3.4. Data sharing and data transfer

Data obtained in PUSH-IT will only be shared among the consortium partners except in some specific cases (data related to an article, confidential health data, projects involving third parties with specific agreements, etc.). Data will only be accessed by authorized personnel. Stakeholders (including the primary PUSH-IT target groups: 1. Scientific community; 2. Academia; 3. District heat network operators, geothermal storage site operators including specialised engineering and software companies; 4. Investors; 5. Policy makers, Licensing authorities, municipalities (local, regional, national, EU-level); 6. General & localized) publics / citizen) may express their interest in reusing the data during or after the project to support further research projects or products and services.

In any case, the aim is to make relevant data fully accessible after publication of the corresponding article. All data and publications will be submitted under Open Access conditions in Zenodo, which allows assigning all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citable and persistently available

Personal data transfer includes the communication, sending or otherwise making personal data available for third parties outside the national borders. The researchers in PUSH-IT do foresee the necessary transfer of personal data inside and/or outside the EU. In this case, research participants will be informed about: (1) data transfer, (2) the purpose of data transfer, and (3) what is happening with their data. Their consent will be recorded as evidence of a data transfer agreement.

7.3.5. Data erasure

For any type of data processing as well as data transfer, participants will be informed about their rights with a project information sheet about the purpose of the project and in the informed consent form about their right of erasing personal data relating to them and about the abstention from further dissemination of such data.

Consent to participate in the PUSH-IT research is based on information provision on the one hand and obtaining informed consent (which may be given electronically, or audio/video recorded) from a participant on the other. Regarding research personnel, their consent for participating in the research is implemented by the information provision during the recruitment process and/or by accepting the work that they will carry out at the relevant PUSH-IT beneficiary. The project information sheet and consent form can be found in D5.2 Quality and risk management report.

Applications for erasure of personal data will be carried out without delay and without any explanation. In case the personal data has been made public, the consortium will take all reasonable steps, including technical measures, to erase the data.

7.3.6. Conclusion

As PUSH-IT Consortium we seek to comply with the ethics requirements with regard to the protection of personal data. This means that:

- We have described procedures for data protection compliance in the grant agreement.
- All beneficiaries in PUSH-IT have confirmed that they have appointed a Data Protection Officer (DPO) and the contact details of the DPO will be made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project will be kept on file.
- Detailed information on the informed consent procedures regarding data processing will be kept on file.
- In case of further processing of previously collected personal data, an explicit confirmation that the beneficiary has lawful basis for the data processing and that the appropriate

technical and organisational measures are in place to safeguard the rights of the data subjects will be kept on file.

8. Other issues

PUSH-IT is committed to comply with the funder's research data management requirements of FAIR data, following the aim to publish *as* open as possible and as closed as necessary. The project will be also conducted in line with national Code of Conducts of each partner / country, which contains a framework for good research practice. Using these as a framework, the following policies regarding data/code management will also apply at the respective partner institutions listed in Table 9.

Table 9: Policies data/code management applied by institutions

Beneficiary name	Policies regarding data /code management
Technische Universiteit Delft	TU Delft Research Data Framework policy TU Delft Research Software policy Netherlands Code of Conduct for Research Integrity
Equans Nederland	
KWR WATER	
Technische Universität Darmstadt	
Fraunhofer Gesellschaft zur Förderung der Angewandten Forschung	
Helmholtz-Zentrum Potsdam - Deutsches GeoForschungszentrum	
Vlaamse Instelling voor Technologisch Onderzoek	VITO data privacy policy
Česká geologická služba	Data management policies are described in several internal documents (in Czech only), brief overview is here, metadata storage is available here. https://micka.geology.cz/en/ Generally the data are as open as possible.
Huisman Geo	
Bureau de Recherches Géologiques et Minières	Scientific work will be licensed in CC BY 4.0, when possible. Results produced by BRGM will be as open as possible, the only restrictions will come from the data
Universiteit Utrecht	
Grondboorbedrijf Haitjema	
Univerzita Karlova	Charles University Policy
Université de Genève	

Beneficiary name	Policies regarding data /code management
Anglia Ruskin University Higher Education Corporation	Privacy Information for Research Participants – ARU
United Kingdom Research and Innovation	UKRI Concordat on Open Research Data UKRI Data Protection Policy UKRI Information Management Policy NERC Data Policy – Guidance Notes BGS Research Data Management Policy BGS Research Ethics and Research Integrity Policy BGS Digital Preservation Policy
Geothermal Engineering	
University of Exeter	Research participants | About our site | University of Exeter

The EU aims to have a net-zero greenhouse gas (GHG) economy by 2050, with 55% reduction on 1990 levels by 2030. At present, heating and cooling represent around 50% of the final energy demand in Europe and are mainly supplied by fossil fuel derived energy. It is therefore essential for heating and cooling to decarbonise to achieve EU ambitions.

A challenge for decarbonizing heat systems is the size of the seasonal mismatch between demand for heat and heat generation from sustainable sources – this mismatch is much larger than the equivalent intermittency in electricity supply and demand. The two main solutions to address this mismatch are: (i) to install a large capacity, so that peak demands can be met even at low production levels; or (ii) to store energy for later use if it is not needed at time of conversion. Many sustainable heat supply systems are characterised by high capital expenditure and low operational costs. Therefore, an installed capacity tailored at peak demand is not cost effective, while extending the annual operation period is advantageous for meeting energy needs, reducing levelised cost of energy (LCOE) and decarbonisation. Optimal utilisation of sustainable heat requires storing large amounts of heat to account for seasonal supply and demand fluctuations. Various technologies have been proposed for large-scale heat storage in geothermal reservoirs and low temperature storage is routinely applied. PUSH-IT focuses on extending storage temperature ranges to high temperatures. We will tackle remaining barriers, demonstrate applicability, increase public engagement, and optimise and de-risk operations. We will showcase three technology options that are fit for a wide variety of geological conditions covering most locations in Europe.

PUSH IT

Funded by
the European Union