

Data Management Plan

Author: V. Turewicz, GeoSphere Austria / G. Goetzl, GeoSphere Austria

Deliverable: D1.5 Data Management – first revision / **Version:** 1 (initial) / **Status:** Draft

Submission Date: 06 April 2023

Supporting Work Package: WP1

Supporting Team: Veronika Turewicz, Anna Bruestle, Stefan Hoyer, Gregor Goetzl

Verified by: Coordinator / **Approved by:** General Assembly

Confidentiality level: Public

Contact: V. Turewicz veronika.turewicz@geosphere.at



www.saphea.eu



**Funded by
the European Union**

This article/publication is based upon work from the project SAPHEA, funded by the European Union's HORIZON EUROPE research and innovation programme under the Grant Agreement number 101075510



GeoSphere
Austria



EGEC
GEOTHERMAL



VIA University
College



UNIVERSITÀ
DI TORINO



AGH
UNIVERSITY OF SCIENCE
AND TECHNOLOGY



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria



Geothermal
Engineering Ltd

„SAPHEA will tackle this challenge to promote more geothermal energy supply heating and cooling networks to become a key element of the green and sustainable transformation of the European energy sector.“

Gregor Götzl – main proposer

List of abbreviations

Abbreviation	Full name
CHP	Combined heat and power plants
DH	District heating and cooling systems
geoHC	Geothermal Heating and Cooling networks supplied by geothermal energy as a source, sink or storage for heat
GHG	Greenhouse Gas emissions
H&C	Heating and cooling
PPP	Public – private partnerships
RES	Renewable energy sources
UTES	Underground thermal energy storage
kWth and MWth	Kilowatt thermal and Megawatt thermal of the geoHC capacity installed.

Change Log

VERSION	PUBLICATION DATE	CHANGE
1	06.04.2023	

Disclaimer:

This article/publication is based upon work from the project SAPHEA, funded by the European Union work programme HORIZON EUROPE under the Grant Agreement number 101075510. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Contents

1.	Introduction.....	7
1.1.	Objectives.....	7
2.	Data Management Plan.....	8
2.1.	Data summary	8
2.1.1.	Types of data managed in SAPHEA.....	8
2.1.2.	Data from external sources.....	9
2.2.	FAIR Data.....	10
2.2.1.	Best practices and data management guideline	10
2.2.2.	Data quality process.....	11
2.2.2.1.	Data inventory.....	11
2.2.2.2.	Data processing.....	12
2.2.2.3.	Data definition and comparability.....	13
2.2.3.	Data vocabulary and metadata.....	13
2.2.4.	Name conventions.....	13
2.2.5.	Data formats.....	14
2.2.6.	Data preservation.....	14
2.2.7.	Project feedback	15
2.2.8.	Data enhancement	15
2.2.9.	Data license.....	16
2.2.10.	Data provenance and quality.....	16
2.2.11.	Data versioning.....	16
2.2.12.	Data identification.....	17
2.2.13.	Sensitive data.....	17
2.2.14.	Data access.....	18
2.2.15.	General data protection regulation (GDPR).....	19
2.2.16.	Conclusion of best practices and guidelines.....	19
2.3.	Allocation of Resources.....	19
2.3.1.	Assessment of project costs for data management.....	19
2.3.2.	Project tasks and responsibilities related to data management.....	20

3. Conclusion.....	21
4. References.....	21
5. Annex.....	23

1. Introduction

SAPHEA is a HORIZON Coordination and Support Action (CSA) that addresses the uptake of multivalent heating and cooling networks supplied by geothermal energy and underground storage ('geoHC networks') by creating a durable digital market hub. The SAPHEA Market Hub contains toolboxes and guidelines to support stakeholders in early-stage investment decisions and strategic planning activities and addresses market actors in districts or municipalities across Europe. Therefore, the project combines, adapts and expands existing open-source tools considering a set of market-ready and emerging technological concepts linked to geoHC networks.

The main objectives of the project are:

- Establishing a digital access point for decision-making support and consultancy on geothermal energy use in heating and cooling networks
- Adapting and upgrading existing datasets and tools for early-stage investment decisions and strategic planning of energy suppliers, communities and municipalities
- Fostering supportive market frameworks to facilitate future investments into 'geoHC networks'
- Reducing social gaps and barriers to bring 'geoHC networks' closer to regional stakeholders across Europe
- Empowering future investors and operators of 'geoHC networks'

Future users of the SAPHEA Market Hub (local authorities, community services and energy suppliers) will be empowered by targeted training to draft development scenarios and roadmaps considering the specific geological and socio-economic boundary conditions in their respective regions. Dedicated communication activities aim to lead to a lively network around the SAPHEA Market Hub of public and private market actors as well as researchers beyond the lifetime of SAPHEA.

1.1. Objectives

All deliverables and outputs related to publishing spatial datasets and knowledge related to document or file repositories will be governed and documented by a Data Management Plan (DMP). This document is the **first revision** of the Data Management Plan. It provides comprehensive guidelines and templates for documenting (see Annex) and permanently storing output datasets and explicit knowledge distributed via the SAPHEA Market Hub. It follows the "Guidelines on FAIR Data Management in

Horizon 2020" [1]. It, therefore, includes the organisation, storage, preservation, security, quality assurance, allocation of persistent identifiers (PIDs) and rules, and procedures for data sharing, including licensing. This first revision will be adapted, if required, during the implementation of SAPHEA and finally provided at the end of the project, including a metadata catalogue linked to the final version of the document repository as part of the SAPHEA Market Hub (D 7.2).

SAPHEA aims to comply with the principles of open research in line with the FAIR (Findable, accessible, interoperable and reusable) principles in full. The project implements and promotes open-science practices by:

- a) Adopting FAIR data principles in all project activities and outputs (e.g. FAIR data management),
- b) Developing an open-access database and open-source toolbox based on previous open science projects,
- c) Disseminating the project results via open access data platforms (e.g., EGDI infrastructure of EuroGeoSurveys [2]) and journals, including peer-reviewed scientific publications;
- d) Supporting existing initiatives on open science (e.g. OpenAire [3], European Open Science Cloud [4]);
- e) Participating to open science events (e.g. Open Science Fair);
- f) Fostering the cultural change for open science practice by training target users (public administrations, energy utilities, researchers and academia) to use an open science tool;
- g) Cooperating with European research platforms (EERA Data [5]).

2. Data Management Plan

2.1. Data summary

2.1.1. Types of data managed in SAPHEA

The data used in SAPHEA are geographical datasets at various spatial scales in the form of raster or vector data (NUTS 3 to NUTS 1 level). By default, pan-European up to 100 x 100 meters raster data will be provided, and new datasets developed in the project will cover the entirety of Europe, if possible, with lower spatial resolution. However, the decision support tool will offer import functions to use local data in a

higher spatial resolution by users. The following stakeholders were identified as target groups that will benefit from the SAPHEA Market Hub:

- Public administrations and public service providers (planners, authorities)
- Public and private energy suppliers and related service providers
- Policymakers and sectoral agencies dealing with energy and climate strategies
- Communities and councils, including residents, involved or affected by geoHC networks
- Financial service providers and energy contractors
- Energy communities in the HC sector
- Research and academia linked to the HC sector
- International civil society organisations, including interest groups and research networks linked to the HC sector
- EU institutions and policymakers linked to the HC sector
- Financial service providers and energy contractors
- International projects addressing research topics linked to geoHC networks
- Interested lay public (indirect)

Most target groups may benefit from the decision support tool by performing spatial techno-economic analyses, which support early-stage investment decisions and strategic energy planning at a local level (focus on NUTS 3) while allowing for upscaling to regional and national levels. Furthermore, the SAPHEA Market Hub will offer networking and information tools for stakeholders. SAPHEA initiates basic infrastructure and aims to transfer the SAPHEA Market Hub into a persistent consultancy infrastructure.

2.1.2. Data from external sources

All datasets created in the context of SAPHEA are based on already available datasets. One of the project's main outcomes is to adapt and upgrade existing datasets and tools towards better integration of geoHC networks. Therefore, the team will collect and assess publicly accessible pan-European datasets (e.g., GEODH project [6], European Geological Data Infrastructure - EGDI [2]) and local datasets in involved study areas to provide an inventory of available geoscientific datasets relevant for investment decision support and strategic planning. All relevant, publicly accessible geoscientific, technological pan-European datasets will be processed

based on created workflows. The datasets will be complemented by publicly accessible data related to relevant socio-economic criteria.

Furthermore, the existing pan-European, open-source tools , Hotmaps [7] ,Enermaps (EnerMaps Data Management Tool ') [8] and, TUW-Geophires [9] will be adapted, expanded and combined with the decision support tool. An existing web portal, currently developed in the framework of the COST Action CA18219 [10], will be expanded towards the SAPHEA Market Hub by integrating the additional toolbox.

2.2. FAIR Data

2.2.1. Best practices and data management guideline

The DMP should include information on how research data are made findable, accessible, interoperable and reusable (FAIR). For this purpose, the consortium should develop a repository to deposit open data and make it accessible for users, apply the open data to validate the results in scientific research and deal with other data reasonably and as intended.

The following web-based tools and repositories will be used to manage the generated and external data:

1. Temporary data storage and exchange cloud: The coordinator of SAPHEA (GeoSphere Austria) provides a data storage and exchange cloud. All project team members have reading and writing access to the cloud. The platform is used to exchange datasets between project team members during the implementation of SAPHEA. The shared cloud must not be used to store documents related to the implementation of SAPHEA, including report-based deliverables.
2. Document exchange platform (Microsoft Teams [11]): A document exchange platform based on MS Teams is provided by the Project Team Member e-think. The platform is meant for private and project internal use only to easily share documents and contacts between the partners during the project (e.g. deliverable drafts, project management documents, presentation slides, project-related literature, etc.).
3. ZOTERO Group library [12]: All Project Team Members use the joint literature database tool ZOTERO to exchange and manage external literature (e.g. relevant articles and

reports). The literature database will later be implemented in the knowledge repository provided through the SAPHEA Market Hub.

4. GitLab repository [13]: A GitLab repository will be used to store output datasets permanently, and each dataset will have a dedicated Git repository. Zenodo [14] will be used to archive the repository and assign digital object identifiers (DOI) to make items easily citeable.
5. SAPHEA Market Hub: The SAPHEA Market Hub (deliverable D7.2) will be a web portal adapted from the COST Action CA18219 [10] and represent the core output of WP7. The SAPHEA Market Hub will offer the digital environment for all tools and products developed in SAPHEA. It will include geographical tools (GIS-based interactive map of characterised good practices, decision support and strategic planning toolbox), consultancy applications (electronic guidelines, fact sheets of geoHC networks) and a knowledge and document repository. The decision support and strategic planning toolbox will be based on open-source codes and will be openly accessible. It will constitute a crucial product in the SAPHEA Market Hub.

SAPHEA connects all published datasets and explicit knowledge (e.g., literature and information aligned with the FAIR principles). This includes external repositories, which are supposed to host all spatial datasets, publicly accessible literature databases, and websites. In addition, the SAPHEA Market Hub will only contain tools and explicit knowledge created by the project in case they are not embedded in external repositories.

As the DMP follows the "Guidelines on FAIR Data Management in Horizon 2020" [1], a standard schema for metadata, name conventions, dataset description, and data sharing and preservation will be addressed in the next sections.

2.2.2. Data quality process

The data quality process contains the following topics: Data inventory, data processing and data definition and comparability.

2.2.2.1. Data inventory

A GitLab repository serves as the data inventory location. For this purpose, the output datasets, which will be generated in SAPHEA, will be published in the GitLab public repository.

Datasets will be described and documented by metadata and created with guidance from the W3C Proposed Recommendations on the Data on the Web Best Practices [15]. It is recommended to provide standardised structured information for data documentation.

All open datasets gathered and generated by SAPHEA will comply with the INSPIRE Directive 2007/2/EC protocols and standards [16]. Therefore, the consortium will use the protocols defined by the Open Geospatial Consortium to publish open datasets.

2.2.2.2. Data processing

Most data processing will be conducted in WP2 (Positioning, data assessment and fact-finding) and WP3 (Spatial data-based decision support toolbox).

In one of the first tasks (Task 2.3), the project team will provide an inventory of available geoscientific datasets relevant to investment decision support and strategic planning. This inventory will provide interfaces to all thematic work packages (WP3 to WP5), with special attention to the harmonisation of geoscientific input parameters and easy-to-apply GIS-based data workflows for the decision support toolbox (WP3). As a next step, all collected relevant, publicly accessible geoscientific, technological, pan-European datasets will be processed based on the created workflows. This will be done in collaboration with the '*EuroGeoSurveys*' organisation, which offers a comprehensive data repository for Europe and will be involved as a stakeholder in the project. The datasets will be complemented by publicly accessible data related to relevant socio-economic criteria previously identified. Furthermore, the datasets will be complemented with guidelines and data preparation protocols for individual data inputs.

In WP3, relevant open-source applications and tools will be characterised and evaluated. This will include tools such as the EnerMaps Data Management Tool [8], the open-source platform of the EU H2020 project "Hotmaps" [7] and the economic geothermal evaluation tool "TUV-Geophires" [9], previously developed by the project partner TUV (Technical University of Vienna). Interfaces between individual calculation modules (e.g., the link between EMDT and TUV-Geophires) will be programmed towards a chain-based toolbox. This will also include interfaces for importing and processing geographical input datasets at various spatial scales. A set of key datasets will be provided in a pre-beta development phase for internal runtime tests. The pan-European datasets created in WP2 will then be integrated into the

toolbox. In a final sprint, the toolbox will be amended by a core team of developers based on the feedback received during testing in the study areas.

2.2.2.3. Data definition and comparability

Datasets will be collected from different providers that use standardised data formats and units. However, this does not mean that the data will be fully comparable. The project aims to harmonise the gathered geoscientific input parameters by adjusting and applying differences and inconsistencies to the created workflows. Comparability can also be accomplished by collecting datasets with common time references or at least with reference to the most recent year.

2.2.3. Data vocabulary and metadata

One important aspect of adapting the FAIR principles is data interoperability. Therefore, agreeing on a shared vocabulary is important to make data understandable for all users. It is recommended to reuse already existing standard terms and often-used vocabulary. The keywords used by the project to describe the datasets will be defined in the Wiki platform (Task 7.3) of SAPHEA.

Moreover, publicly accessible data must be provided with metadata that give essential information about the origin, creator, purpose, spatial coverage, terms of use, etc. For this purpose, the “datapackage.json” [17] will be used as a data package format to provide metadata that can be read both by human and machine.

2.2.4. Name conventions

Reasonable name conventions are essential for making data findable. SAPHEA will have different naming conventions for project documents and data.

Documents: The documents are managed in a hierarchical file tree that follows the general work plan of the project. First, there is the title of the work package, which will contain a dedicated directory for each task. The prefix of the directory title must include the absolute number of the respective task or deliverable (e.g., D21-Short title). The title of the file must indicate the presented content.

Data: Project data should not be associated with the work packages and tasks of SAPHEA. The first segment will be defined by the spatial resolution of the dataset (e.g. NUTS0) for each data repository. Then keywords, which describe the data and are defined in the Wiki platform (Task 7.3) of SAPHEA, will be used to name the data.

2.2.5. Data formats

SAPHEA collects data that are geographical datasets at various spatial scales in the form of vector (NUTS 3 to NUTS 1 level) or raster data (by default pan-European up to 100 x 100 meters raster data is provided, but users will later be able to import and use their own data). The datasets will be assessed, and relevant datasets will be processed based on created workflows. As a result, at least 10 publicly accessible geographical datasets will be created.

Any output dataset will be deposited in its own GitLab repository, which will consist of the dataset and the related metadata and documentation. It will be useful for manual and automatic use (machine-readable). Hence, SAPHEA follows the standard Data Package format defined by Frictionless data [18] and adapted during Hotmaps and EnerMaps projects.

Therefore, each repository should follow the provided structure below. This is also very important for data integration into the toolbox. Data integration scripts skip the repository if the structure is different.

Basic information will be provided by a **README file**, which contains the title, description, process for data creation and references for the input data source. Furthermore, a `datapackage.json` [16] file will be used as a data package format for machine-readable purposes. The JSON file will contain fields for the metadata information. **Raw data** will be available preferably in a text format (e.g. CSV format) and not in the form of shapefiles or other file types. Hence, no dedicated software is required to access the data. Finally, **script** files will be included to load and process data if necessary.

2.2.6. Data preservation

Data preservation is important to ensure that data remain reusable beyond the lifetime of SAPHEA. Hence, maintenance and data backups are essential. Moreover, datasets will safely be stored in a repository for long-term preservation.

The produced datasets will be published in the GitLab public repository and can be used freely and openly. The durable SAPHEA Market Hub will host all products developed in the project, including the data repository. Additionally, the datasets created will be provided to EuroGeoSurveys for publishing at the European Geological Data Infrastructure (EGDI) [2], a comprehensive European data repository.

It is expected to back up the database weekly, which will change depending on the volume of data. Since SAPHEA collects and generates geographical datasets at various spatial scales, the data size can vary. There are different possibilities to avoid exceeding the size limit of the repository. Vector data usually have a small volume and can be stored as uncompressed text documents in a CSV format. If data have a higher resolution, another file format can be considered. Raster data can potentially reach very large sizes. Consequently, compressed formats may be used, like a GeoTiff compressed binary format, which will be managed by the Git Large File Storage tool [19]. This tool is able to divide very large files into smaller binary pieces.

Since the data are publicly accessible, downloading data on external personal computers will not be monitored.

2.2.7. Project feedback

The different technological concepts of geoHC networks, which will be identified and characterised, will be reviewed by potential future users (stakeholders at regional European levels) through a workshop and an online feedback survey.

SAPHEA partners and stakeholders will test the toolbox by performing real conditions tests in the study areas. These preliminary training will be organised during a cross-thematic hybrid workshop (approx. month 18) and additional online webinars. Feedback will be collected via a digital survey and lead to a list of bugs and shortcomings for final amendments. In a final sprint, the toolbox will be amended by a core team of developers based on the feedback received during the testing in the study areas.

Furthermore, SAPHEA will capitalise on the network of the current COST Action CA18219 (covering more than 200 participants from more than 40 countries), which will be further expanded by involving at least 500 local to international market actors who register to the SAPHEA Market Hub. The SAPHEA Market Hub will offer networking tools with a user forum where feedback can be provided.

2.2.8. Data enhancement

Data will mainly be enhanced by adapting and upgrading existing datasets. This includes the assessment, harmonisation, and processing according to created workflows. The produced datasets will be integrated into the open-access Geological

Data Infrastructure (EGDI) [2] and further expanded or adapted in future activities of EuroGeoSurveys.

Feedback surveys (see chapter 2.2.7.) are also intended to provide data enhancement. Furthermore, there is a pull-request system on GitLab where users have the possibility to discuss, change and improve datasets.

2.2.9. Data license

Datasets must be accompanied by a license detailing the permissions for sharing and reusing data.

The licenses have to be checked for external datasets to ensure that data may be shared on the SAPHEA Market Hub. If a license does not accompany datasets, the data provider will be contacted to get information on how data may be used.

One of the project's main outcomes is to adapt and upgrade existing datasets, which means that all output datasets are based on already available data. Hence, checking the data license used to generate new datasets is essential. A proper license will be chosen to publish the created datasets. The dedicated repository for each dataset will specify the specific license of the data in the README and the JSON file, and the license can be chosen directly on Zenodo [14].

2.2.10. Data provenance and quality

The historical record of data is important information for data provenance and allows consumers to trace the data back to its source of origin. The metadata file (datapackage.json) will contain the author's contact information for the data in SAPHEA, allowing the original information sources to be obtained.

Furthermore, data quality documentation is essential for users to know if data will be reused for their purposes. Hence, metadata will include detailed information on data quality concerning any quality-related issues and include feedback from the author.

2.2.11. Data versioning

Data versioning is essential to inform users about the exact version they are working with. It allows for identifying the specific changes in the data, discovering newer versions and enabling comparisons.

Since all data will be published on the GitLab repository, the data will also be under the control of the Version Control System called Git [20]. Thus, the history and status

changes of the data will be stored, and a unique hash string will identify each status. The Git Large File Storage extension [19] (mentioned in chapter 2.2.6.) will be used for large files, which can be problematic when using Git. Moreover, the Git Journal tool [21] will be used, which automatically generates changelogs for each data release.

2.2.12. Data identification

Data identification is another important aspect of adapting the FAIR principles for making data easily citeable. This helps consumers to identify and reuse the data appropriately. SAPHEA uses Zenodo [14] to manually assign a unique digital object identifier (DOI) to the GitLab repository for each stable dataset release.

2.2.13. Sensitive data

Personal and confidential data are characterised as sensitive data and must be handled in accordance with the applicable EU, international or national regulations.

First, it is important to identify all sensitive data to ensure proper management. Internal project documents and personal data from project partners, like contact details, will only be shared on MS Teams, which is meant only for private and project internal use.

Data from the study areas that will be used to test the toolbox by performing real conditions tests will be shared between project partners and stored in a dedicated private GitLab repository, as well as other sensitive data created during the project.

Sensitive data like personal data (e.g people's names, gender, sector of work, country and contact details) or response data from experts, stakeholders and other interest groups generated during the project should be treated responsibly and securely. It is essential that sensitive data remain private and anonymous and are not disclosed to any third party (including participants involved in the project implementation) or published in any way without prior explicit approval. Such data are collected during several interactions and activities, which are described in the following sections.

Stakeholder and expert interaction: Stakeholder and expert interviews will be conducted in WP2, WP4 and WP5. For instance, the assessment and characterization of key supportive and hindering market drivers for geoHC networks will be complemented by experts and stakeholder interviews. Stakeholder interaction includes also a short survey for each stakeholder group to collect information on the current knowledge and awareness of geoHC networks, their potential impacts and

barriers for development. The territorial coverage of the assessment will be expanded by involving members of the COST Action CA18219.

Training and capacity building activities: Training and capacity building activities are intended to transfer knowledge on the geoHC Tool (WP3), as well as the project fact-finding (T2.5), reviews (T4.1, T4.2), blueprints (T4.3, T4.4), recommendations (T5.4), Market Uptake Hub (T7.3) and future geoHC uptakes in Europe. Training will be tailored to different target groups, focusing on regional stakeholders. Workshops (online, in-person or hybrid) will address both adopters and multipliers, with a train-the-trainer approach to allow and support multipliers with training themselves additional users/regions. Training events will support the science-to-policy dialogue by involving local communities of the study areas. Training will also provide a working space, to share experiences in a peer-to-peer learning mode.

Awareness raising and dissemination activities: Awareness raising and dissemination activities cover all measures to raise awareness, inform on the use of geoHC networks and disseminate the outcomes of the project. This includes activities linked to different communication channels like events and webinars, the website, and electronic newsletters. The website organized by EGEC will become the landing page of the Market Uptake Hub, referencing blogs on the activities and outcomes of SAPHEA. Electronic newsletters linked to the major milestones of the CSA will inform registered followers.

2.2.14. Data access

SAPHEA aims to develop an open-access database and open-source toolbox based on previous open science projects. Hence, the toolbox will be based on open-source codes and will be openly accessible. Users can access all modelling and output parameters used and produced in the toolbox.

All datasets released as open data are stored in public GitLab repositories and will be freely accessible without authentication. Classified sensitive and confidential data will be stored in private GitLab repositories and require user authentication. There is the possibility of an HTTPS authentication where a user name and a password are used or a private/public key via an SSH connection for strong authentication.

The temporary data storage and exchange cloud can only be accessed via a hyperlink and password provided by the project coordinator. All project team members have reading and writing access and must not share the access hyperlink and password

with any third parties. There will be a read-only external access area to documents and files for external stakeholders, including EAB members.

2.2.15. General data protection regulation (GDPR)

The Horizon 2020 Project SAPHEA is fully compliant with the General Data Protection Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC ('GDPR') [22]. The Communication Manager of SAPHEA is responsible for maintaining any contact list or register related to followers in line with GDPR requirements. A GDPR clause will be included in every registration form and an opt-out link will be included in the newsletter. Moreover, a data processing register will record what kind of data will be collected, where the data will be stored, who has access to it and how long the data will be stored before deleting it.

A signed consortium agreement, which was established at the beginning of the project, regulates all intellectual property rights (IPR) issues related to background knowledge relevant for the implementation of the planned work. In addition, a final project agreement (deliverable D 1.4) regulates all issues linked to IPR created during the project implementation and in the context of the results achieved.

2.2.16. Conclusion of best practices and guidelines

Regarding data management, the SAPHEA consortium will follow the recommendations for best practices and guidelines from the European Commission [1] and the proposed recommendations on data on the web best practices by the W3C communities [15]. All datasets collected, expanded and upgraded in the context of SAPHEA will be stored on a public or private git repository hosted on the GitLab platform, while the internal documents and contacts will be shared through MS Teams.

2.3. Allocation of Resources

2.3.1. Assessment of project costs for data management

For the implementation of SPAHEA, several free software and services will be used. This includes GitLab [13] as the data repository where data can be publicly stored. Zenodo [14] to archive the repository and assign digital object identifiers (DOI) for

making items easily citeable. Furthermore, the Frictionless Data Package format [18], which is also free to use, will be used in the data format process.

It is important to use free software and services to ensure data preservation, meaning that data remain reusable beyond the lifetime of SAPHEA. Hence, datasets will safely be stored in a repository for long-term preservation.

Nevertheless, there will be other expected costs, such as the working hours of researchers and data managers spent on the data. This includes, for instance, the collection, assessment, processing, storing, and documentation of data, as well as performing compliance checks, providing consultancy and support by the data manager. Furthermore, there will be costs for maintaining legal compliance in reusing datasets and advice on privacy law compliance in output datasets publication.

2.3.2. Project tasks and responsibilities related to data management

The coordinator of SAPHEA (GeoSphere Austria) is responsible for WP 1, which includes the DMP, and provides a data manager who supports the project team in complying with FAIR data principles and open science practices. The data manager is responsible for providing consultancy and support to the partners in preparing datasets governed by the DMP, and performing compliance checks on datasets delivered by the project team members.

TUM coordinates WP 2, which aims at performing data assessments and fact-finding missions. This includes the creation of publicly accessible pan-European geographical datasets. In terms of data and key aspects related to geoHC networks implementation, WP2 will feed into 3 thematic WPs, which cover the technology domain (WP3) and society domain (WP4 and WP5) of the SAPHEA Market Hub. Therefore, the whole team will be involved in activities related to WP2. AGH UST, for instance, will contribute to the development of concepts and data preparation related to geothermal energy use in geoHC networks. UNITO will play a vital role in defining the geoHC utilisation scenarios and data assessment with a special focus on underground thermal energy storage.

WP3 will focus on adapting existing tools and instruments for techno-socio-economic spatial data analyses for early-stage decision support and planning. TUW was involved in the development of tools (e.g., Hotmaps, EnerMaps Geophires) on which SAPHEA is built. This partner will therefore focus on coordinating the development of all decision support and planning tools related to spatial datasets (WP3). TUW will also assist the

project team in designing the products of the SAPHEA Market Hub, which is a core output of WP7, coordinated by GeoSphere Austria. Moreover, e-think will contribute to adapting the spatial geographical planning tools, and TUM will provide major contributions to developing the decision support and planning tools (WP3).

3. Conclusion

This document is the first revision of the SAPHEA Data Management Plan and will be delivered at the end of the sixth project month. It represents the identified data structure and offers templates for completing the metadata catalogue during the project. The DMP is a living document and will be adapted during the implementation of SAPHEA and finally provided at the end of the project, including a metadata catalogue linked to the final version of the document repository as part of the SAPHEA Market Hub (D 7.2).

4. References

- [1] European Commission, 'Guidelines on FAIR Data Management in Horizon 2020', Jul. 26, 2016. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [2] EuroGeoSurveys (EGS), 'European Geological Data Infrastructure (EGDI)', 2022. <https://www.europe-geology.eu/>
- [3] OpenAIRE EXPLORE, 'OpenAIRE', 2020. <https://www.openaire.eu/>
- [4] European Commission, 'EOSC Portal - A gateway to information and resources in EOSC', *European Open Science Cloud*, 2023. <https://eosc-portal.eu/>
- [5] European Union's Horizon 2020, 'EERA Data', *EERA Data*, 2023. <https://eeradata-project.eu/>
- [6] GeoDH Project, 'Geothermal District Heating', 2014. <http://geodh.eu/geodh-project/>
- [7] Hotmaps project, 'Hotmaps', *Hotmaps*, 2023. <https://www.hotmaps.eu/map>
- [8] The EnerMaps project, 'EnerMaps', 2022. <https://enermaps.eu/>
- [9] K. F. Beckers and K. McCabe, 'GEOPHIRES v2.0: updated geothermal techno-economic simulation tool', *Geotherm. Energy*, vol. 7, no. 1, p. 5, Dec. 2019, doi: 10.1186/s40517-019-0119-6.
- [10] COST network Geothermal-DHC, 'Geothermal DHC', *Geothermal DHC*, 2023. <https://www.geothermal-dhc.eu/>
- [11] Microsoft, 'Microsoft Teams'. 2017. [Online]. Available: <https://www.microsoft.com/de-at/microsoft-teams/group-chat-software/>
- [12] Roy Rosenzweig Center for History and New Media, 'Zotero'. 2016. [Online]. Available: www.zotero.org/download
- [13] GitLab Inc., 'GitLab'. 2012. [Online]. Available: <https://gitlab.com/>
- [14] CERN, 'Zenodo'. CERN, 2013. [Online]. Available: <https://zenodo.org/>
- [15] W3C, 'W3C Data on the Web Best Practices', 2017. <https://www.w3.org/TR/dwbp/>
- [16] European Commission, 'INSPIRE. Infrastructure for spatial information in Europe.', 2016. <https://inspire.ec.europa.eu/>
- [17] JSON-LD Community Group, 'JSON for Linking Data', 2016. <https://json-ld.org/>
- [18] Frictionless Data, 'Specifications for Packaging and Transporting Data', 2017. <https://specs.frictionlessdata.io/>
- [19] GitHub, 'Git Large File Storage'. 2022. [Online]. Available: <https://git-lfs.com/>

- [20] Git, 'Git - fast version control'. 2023. [Online]. Available: <https://git-scm.com/>
- [21] GitHub Inc., 'Git Journal'. 2023. [Online]. Available: <https://github.com/saschagrunert/git-journal>
- [22] European Parliament, 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC'. 2016. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>

5. Annex

The following table shows a template for documenting the output datasets.

Table 1: A template for documenting the output datasets created in SAPHEA.

Dataset title	Format	Size [MB]	Content description	Methodology / process for data generation	References to input data sources	Authors	DOI	Keywords	Licenses	Public value to whom?	IP/GDP R Sensitive