This draft document is currently under revision by the European Commission (EC) and has not yet been validated or approved by the EC. The content provided herein is subject to change, and the information presented may not represent the final position or official stance of the EC.

Horizon 2020
Programme

GO-VIKING

Research and Innovation Action (RIA)

This project has received funding from the Euratom research and innovation programme 2021-2025 under Grant Agreement No 101059603

Start date : 2022-06-01 Duration : 48 Months

Data Management Plan

Authors : Dr. Papukchiev ANGEL (GRS)

GO-VIKING - Contract Number: 101059603

Project officer: Panagiotis MANOLATOS

Document title	Data Management Plan
Author(s)	Dr. Papukchiev ANGEL
Number of pages	21
Document type	Deliverable
Work Package	WP7
Document number	D7.3
Issued by	GRS
Date of completion	2022-08-31 09:21:39
Dissemination level	Public

Summary

Preparation of a Data Management Plan for the GO-VIKING project.

Approval

Date	By
2022-08-31 10:13:00	Mrs. Marianne GROS (LGI)
2022-08-31 10:14:30	Dr. Papukchiev ANGEL (GRS)

Data Management Plan

Deliverable 7.3 of the GO-VIKING project (101060826)

Version N°1.0
August 2022

Angel Papukchiev – GRS

Funded by
the European Union

Disclaimer

The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Document information

Grant Agreement / Proposal ID	101060826
Project Title	Gathering expertise On Vibration ImpaKt In Nuclear power Generation
Project Acronym	GO-VIKING
Scientific Coordinator	Angel Papukchiev, Angel.Papukchiev@grs.de, GRS
Project starting date (duration)	1st June 2022 – 31st May 2026 (48 Months)
Related Work Package	WP7
Related Task(s)	7.4
Lead Organisation	GRS
Contributing Partner(s)	-
Due Date	31.08.2022
Submission Date	
Dissemination level	PU

History

Date	Version	Submitted by	Reviewed by	Comments
31.08.2022	V1.0	GRS	LGI	

Table of contents

1. Description of GO-VIKING research 7
1.1 Research objective of the project 7
1.2 Project structure 7
2. Background of the Data Management Plan and some definitions 8
2.1 Background 8
2.2 Definitions 8
2.3 Versions of the Data Management Plan 10
3. GO-VIKING data 10
3.1 Experimental research data 10
3.2 Numerical research data 11
3.3 Other datasets 11
3.4 General framework for data collection within GO-VIKING 11
4. FAIR Data Management in GO-VIKING 14
4.1 Making data findable 14
4.2 Making data openly accessible 15
4.3 Making data interoperable & increasing re-use 16
5. GO-VIKING and Open Science 18
6. Data Management Plan implementation 18
7. Ethical aspects 18
7.1 General Data Protection Rules 18

List of figures

Figure 1: Knowledge Management – Definitions and hierarchy 9

List of tables

Table 1: Table specifying the content of a dataset 15
Table 2: Example of licenses 16

Abbreviations and Acronyms

Acronym	description
BPG	Best Practice Guidelines
CFD	Computational Fluid Dynamics
CSM	Computational Structural Mechanics
EC	European Commission
DMP	Data Management Plan
EU	European Union
FA	Fuel assembly
FAIR	Findable, Accessible, Interoperable and available for Re-use
FIV	Flow-Induced Vibrations
FSI	Fluid-Structure Interaction
GDPR	General Data Protection Rules
GO-VIKING	Gathering expertise On Vibration ImpaKt In Nuclear power Generation
UQ	Uncertainty Quantification
NPP	Nuclear Power Plant
OA	Open Access
ROM	Reduced Order Model
WP	Work Package

Summary

This document describes the Data Management Plan (DMP) to be implemented within the European Horizon Project "Gathering expertise On Vibration ImpaKt In Nuclear power Generation (GO-VIKING)". Large amount of experimental and numerical data will be generated during the four years of project duration. The DMP outlines the conditions for data preservation, adherence to FAIR principles, publication, and makes a distinction between potentially sensitive or confidential information and open access data.

Keywords

Data Management Plan, FAIR, public, experimental, numerical, data

1. Description of GO-VIKING research

1.1 Research objective of the project

The overall objective of GO-VIKING is to increase the expertise and improve the tools and skills of the European nuclear stakeholders for the analysis of complex flow-induced vibrations (FIV) phenomena. This will maintain and further enhance the nuclear plant operation and safety.

The objective will be accomplished by:

- **Generation of high-resolution numerical and experimental data of FIV in single and multiphase flows.**
- **Development and validation of high- and medium-resolution, as well as fast-running practical tools for the FIV analysis**
- **Implementation of efficient methods for uncertainty propagation in the FIV analysis results**
- **Synthesis of best practice guidelines for FIV analysis in accordance with the needs of the stakeholders**
- **Targeted training of graduates and young experts as well as practitioners from stakeholders in FIV analysis and modelling techniques**

The GO-VIKING project will improve the safety of contemporary reactors and the design evaluation of new concepts by making available new experimental results and improved numerical approaches for the evaluation of FIV to the stakeholders.

1.2 Project structure

The GO-VIKING project consists of seven work packages (WP). In WP1 the relevant FIV phenomena in NPPs will be reviewed. The generated output will be provided and taken into account as far as possible in all new experimental activities in order to close as much as possible existing gaps. After reviewing the available FSI methods and taking into consideration the needs of the stakeholders, important findings and conclusions will be provided to WP2-WP5, where fluid-structure interaction (FSI) methods and models are developed and validated. Within these work packages, models, numerical, and experimental data for relevant geometrical configurations in single and multiphase flows will be generated. This data will also be used as an input for the development of fast-running methods in WP5. Uncertainty quantification (UQ) methods for FSI simulations, based on fast-running methods, will be developed. The gained experience in all other technical WPs will be synthesized in Best Practice Guidelines on the Use of FSI Methods for the Evaluation of Flow-Induced Vibrations in Nuclear Applications. Within WP6, the project results will be disseminated. Practitioners from stakeholders, as well as students and postgraduates, will be educated and trained in the application of the developed methods and tools. The GO-VIKING project management activities are enclosed in WP7.

2. Background of the Data Management Plan and some definitions

2.1 Background

This document presents the GO-VIKING Data Management Plan (DMP), it is a deliverable that has been required by the European Commission (EC).

According to the EC, Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

- peer-reviewed scientific research articles (published in scholarly journals);
- research data (data underlying publications, curated data and/or raw data).

The rationale is that data management is not a goal itself, but a way that leads to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. Therefore, projects must aim at improving and maximising access to and reuse of research data generated, while balancing openness and protection of scientific information, commercialisation and Intellectual Property Rights, privacy concerns, etc.

DMPs are a key element of good data management. As part of making research data findable, accessible, interoperable, and reusable (FAIR), a DMP should include information on the data life cycle:

- the handling of research data during and after the project,
- what data will be collected, processed or generated,
- what methodology and standards will be applied,
- whether data will be shared/made open and how,
- how data will be curated and preserved.

While open access to research data has become applicable by default in Horizon Europe, the Commission acknowledges that there could be good reasons to keep the research data confidential.

2.2 Definitions

Before explaining the consortium strategy in terms of Data Management, several terms must be defined:

- Data: Data refers to unstructured facts and figures, which are not organised in any way and which provide no further information regarding patterns, context, etc. For instance, data on production, demand, results from technical tests and so on, is unstructured data.

- Information: For data to become information, it must be contextualized, categorized, calculated and condensed. Information thus paints a bigger picture; it is data with relevance and purpose. It may convey a trend in the environment, or perhaps indicate a pattern of sales for a given period of time.

- Knowledge: Knowledge is closely linked to doing and implies know-how and understanding. The knowledge possessed by every individual is a product of his/her experience and encompasses the norms by which he/she evaluates new inputs from his/her surroundings. For instance, knowledge is related to the know-how acquired in R&D projects, commercial activities or the expertise that is inherent to each partner.

Figure 1 shows the definition and the hierarchy in the knowledge management.

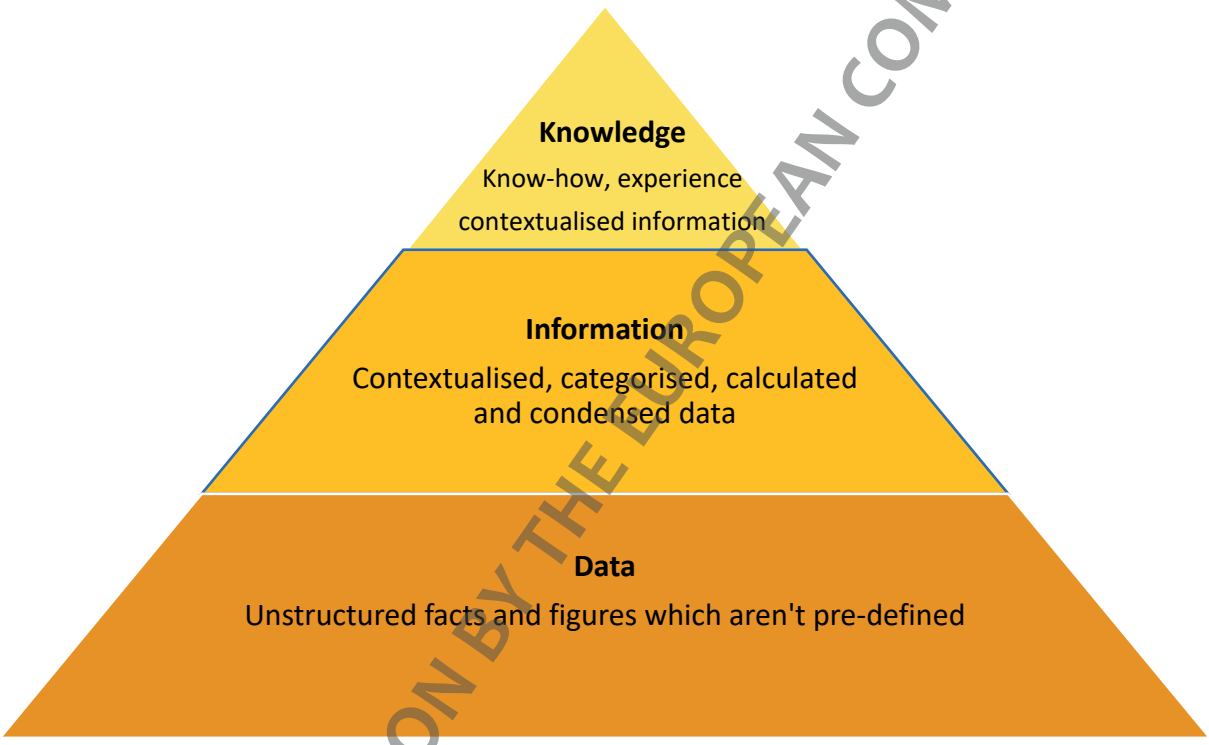

Figure 1: Knowledge Management – Definitions and hierarchy

This present DMP will mainly deal with how the data will be managed and will mention the links with knowledge. Below are some further definitions, concerning data:

- Data codebook: A codebook is an essential document that informs the data user about the study, data file(s), variables, categories, etc., that make up a complete dataset. The codebook may include a dataset's record layout, list of variable names and labels, concepts, categories, cases, missing value codes, frequency counts, notes, universe statements, and so on.

- Dataset: a dataset is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable. The data set lists values for each of the

variables, such as height and weight of an object, for each member of the dataset. The data set may comprise data for one or more members, corresponding to the number of rows.

2.3 Versions of the Data Management Plan

According to the EU's guidelines regarding the DMP (European Commission[1], 2022), the document may be updated, if necessary, during the project lifetime. The minimum requirement is that the DMP is updated at least for each periodic evaluation of the project.

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses. The DMP should, therefore, have a clear version number and include a timetable for updates.

3. GO-VIKING data

Within the GO-VIKING project, a significant amount of both numerical and experimental data will be produced. This is the case for WP2 to WP4, while in WP5 and WP6 only numerical data will be generated. In WP1, existing and new information will be used to generate knowledge.

3.1 Experimental research data

Experimental data from six different facilities will be used in the GO-VIKING project:

- **Cantilever Rod (performed/to be performed at University of Manchester)**
- **AMOVI (performed/to be performed at CEA)**
- **TITAN (performed/to be performed at CEA)**
- **ALAIN (performed at Framatome GmbH)**
- **GOKSTAD (to be performed at The von Karman Institute for Fluid Dynamics)**
- **TREFLE (to be performed at IRSN)**

Experimental data from the first four facilities already exists. Within the project, new experimental data will be generated at the AMOVI, TITAN, and Cantilever Rod facilities. New experimental data for code validation will be also generated at the GOKSTAD and TREFLE experimental rigs.

The new experimental research data will be collected using appropriate Data Acquisition Systems with a properly developed software. The data will cover the main subject areas of the research: structural vibrations induced by fluid flow in different geometries and flow conditions (single- and multi-phase, axial and cross-flow). Data will be gathered for structural displacements, flow velocities, void fractions, pressures, etc. The formats used for the datasets will be agreed among the partners performing experimental and numerical activities. The data will be mainly in .dat, .txt, .csv or .xlsx formats. It is hard to estimate the total volume of the data to be generated, since the length of the experiments and the data acquisition, as well as important details such as the number of vibrating tubes in some experiments with tube arrays still needs to be discussed and specified by the partners. Reuse of the existing Cantilever Rod, AMOVI, TITAN, and ALAIN data is planned. The data will be stored on local and network hard disk drives, bands, clouds, platforms (FLEXX), etc.

The datasets will be processed and analyzed with appropriate tools such as Excel, GNU Octave, Jupyter Notebook, gnuplot, as well as with Python scripts, and routines in Fortran, C++ or other programming languages. If necessary, the use of further open research data might be considered

in GO-VIKING. Potential reutilization will be enabled and quality of the data ensured by careful documentation of data collection methods as well as the contents of the datasets.

3.2 Numerical research data

Research data will be generated with computational fluid dynamics (CFD) and computational structural mechanics (CSM) programs (e.g. NEPTUNE_CFD, ANSYS CFD, STAR CCM+, OpenFOAM, ANSYS Mechanical, ABAQUS, etc.). The data will cover the main subject areas of the research: structural vibrations induced by fluid flow in different geometries and flow conditions (single- and multi-phase, axial- and cross-flow). Data will be gathered for structural displacements, flow velocities, void fractions, pressures, etc. The formats used for the datasets will be agreed among the partners performing experimental and numerical activities. The data will be, for example, in .dat, .txt, .csv, .xlsx or code dependent proprietary binary formats. It is hard to estimate the total volume of the data to be generated, since important details such as number of vibrating tubes in some experiments with tube arrays or simulation length still need to be discussed and specified by the partners. Moreover, some of the experimental facilities are still in the planning phase and the exact structure of the measurement equipment (number of sensors) is not yet defined. This brings further uncertainties in the estimation of the numerical data that has to be generated and stored.

The datasets will be processed with code specific post-processing programs (e.g. ParaView, ANSYS Post, etc) and analyzed with appropriate tools such as Excel, GNU Octave, Jupyter Notebook, gnuplot, as well as with Python scripts, and routines in Fortran, C++ or other programming languages.

3.3 Other datasets

Relevant inputs and documentation of the experimental setups and simulation cases will be stored in MS Word, PDF or .txt formats. The descriptions will be redacted to protect the confidential /commercial data.

3.4 General framework for data collection within GO-VIKING

In this section, the data types to be collected in each WP of the GO-VIKING project are presented

WP1 – State of the art and stakeholder outreach

The main objectives of WP1 are to document the most important FIV phenomena, relevant for the operational safety of nuclear reactors, to review currently available experimental data and numerical methods for the analysis of these phenomena, to identify industry needs and regulatory expectations in terms of tools and methods for the analysis of FIV phenomena, and to specify requirements for advanced numerical methods and tools for the analysis of FIV phenomena.

Data from the free literature as well as from discussions with the stakeholders will be collected. This data will be in a MS Word, Excel or PDF formats. The size of the data should not exceed several hundreds of megabytes.

WP2 – Flow-induced vibrations in fuel assemblies

The main objectives of WP2 are to develop advanced simulation methods and tools to facilitate the understanding of FIV in fuel assemblies (FA); to utilize existing experimental data on FIV under axial

flow for validation of numerical simulation tools; to assess simulation approaches with respect to physical models, numerical schemes and coupling algorithms for the simulation of FIV in FA. Further objectives are the synthesis of guidance for practical use of FSI tools as well as to increase their efficiency through development and implementation of simplified models; and to implement efficient coupling strategies.

The experimental data from ALAIN (existing) and Cantilever Rod (existing and new) facilities will be collected in text, CSV, binary, or other proprietary formats. The generation of additional data at the Cantilever Rod facility might be connected with modifications in the instrumentation and the test section. Therefore, an estimate of the size of the experimental data in this WP is currently not possible.

The numerical data from the simulations will be collected, among others, in text, CSV, binary, MS Word, MS Excel, or PDF formats. An estimation of the overall data size is not easy, mainly because of the unknown total number of numerical simulations that will be performed, and the size of the generated binary files. The numerical data size also strongly depends on the numerical setup and the number of output variables and their generation frequency in the transient numerical runs.

WP3 – Flow-induced vibrations in steam generators

The main objectives of WP3 are to develop methods to provide reliable assessment of structural vibrations, occurring in heat exchangers and steam generators under cross-flow conditions; to provide high-resolution computations and utilize existing and generation of new, high-resolution experimental data on FIV under single phase cross flow for validation of the numerical tools; to develop and implement simplified models of the investigated structures (tube bundles) to increase the overall efficiency of the FSI calculations; to validate and assess the advantages and drawbacks of the applied simulation approaches with respect to turbulence models, numerical schemes, coupling algorithms etc. for the simulation of FIV in cross flow conditions; and to provide input to WP5 for the development of ROMs.

The experimental data from AMOVI (existing and new) and GOKSTAD (new) facilities will be collected in text, CSV, binary, or other proprietary formats. The generation of additional data at the AMOVI facility might be connected with modifications in the instrumentation and the test section. Therefore, an estimate of the size of the experimental data in this WP is currently not possible.

The numerical data from the simulations will be collected, among others, in text, CSV, binary, MS Word, MS Excel, or PDF formats. An estimation of the overall data size is not easy, mainly because of the unknown total number of numerical simulations that will be performed, and the size of the generated binary files. The numerical data size also strongly depends on the numerical setup and the number of output variables and their generation frequency in the transient numerical runs.

WP4 – Flow-induced vibrations in multiphase flows

The main objectives of WP4 are the development of advanced simulation tools to provide deeper understanding of two-phase flow-induced vibrations of nuclear reactor components and their validation against available experimental results; the utilization of existing and new experimental data on two-phase FIV under axial and cross flow for validation of numerical simulation tools; the increasing the efficiency of coupled FSI simulations through development and implementation of simplified structural models to represent complex structures of nuclear power plant components and implementation of efficient coupling strategies; the assessment of simulation approaches with

respect to applicability of corresponding underlying two-phase models, numerical schemes, coupling algorithms etc. for the FIV simulation, and deriving guidance for practical use.

The experimental data from the Cantilever Rod (new), TITAN (existing and new) and TREFLE (new) facilities will be collected in text, CSV, binary, or other proprietary formats. The generation of additional data at the TITAN facility might be connected with modifications in the instrumentation and the test section. Some construction and instrumentation details are still open for discussion, therefore, a precise estimate of the size of the experimental data in this WP is currently not possible. Due to the two-phase flow specifics of the experiments, larger datasets are expected in this WP.

The numerical data from the simulations will be collected, among others, in text, CSV, binary, MS Word, MS Excel, or PDF formats. An estimation of the overall data size is not easy, mainly because of the unknown total number of numerical simulations that will be performed, and the size of the generated binary files. The numerical data size also strongly depends on the numerical setup and the number of output variables and their generation frequency in the transient numerical runs.

WP5 – Fast-running methods, uncertainty quantification and best practice guidelines

The objectives of WP5 are the development and implementation of fast-running FSI methods for efficient analysis of FIV phenomena; the validation of fast-running methods and identify their limitations; the development, implementation, and application of approaches to quantify the uncertainty in the FIV evaluation in reactor components; and the elaboration of guidance to vendors, operators, and regulators on robust and efficient FIV assessments.

The experimental data, utilized and generated in WP2- WP4, will be used to produce new numerical data. This will be collected, among others, in text, CSV, binary formats. An estimation of the data size is currently not possible, mainly because of the unknown total number of numerical simulations that will be performed and the size of the generated binary files. Their size also strongly depends on the number of output variables and their generation frequency in the transient numerical runs.

WP6 – Education, training and dissemination

The main objectives of WP6 are the training of young scientists, master and PhD students on FSI related to FIV in nuclear applications; the development of educational material for FIV; the presentation and distribution of the BPGs developed in WP5; the dedicated code training with code benchmarks; the communication and dissemination; the elaboration of the detailed plan and of the communication material; the organisation of an international event with key stakeholders.

The experimental and numerical data utilized and generated in WP2- WP4 will be used to produce new numerical data. This will be collected, among others, in text, CSV, binary formats. Training Instructions will be generated in MS Word, MS Excel, or PDF formats. The numerical data to be generated should be in the range of several gigabytes, since small and simple test cases will be chosen for the education purposes.

WP7 – Project management

The main objectives of the WP7 are to carry out an effective technical, scientific, legal, financial and administrative coordination, establishing the mechanisms and management procedures to that end; an appropriate governance structure and internal communication methods; the project

monitoring & risk management; the project QP, DMP, Gender Action Plan; the interactions with the Advisory and User Group; to enforce gender equality.

Project documentation will be generated and collected in MS Word, MS Excel, or PDF formats. The size of the data should be within several hundreds of megabytes.

4. FAIR Data Management in GO-VIKING

In compliance with applicable rules, every Horizon Europe project is required to draft a DMP in order to make the data Findable (1), Accessible (2), Interoperable (3) and available for Re-use (4) (FAIR principles).

4.1 Making data findable

This section will provide insight into how GO-VIKING intends to make it easier to find data collected or produced by the consortium. The way to proceed in order to achieve this goal is to describe properly the content of the data sets.

Name and Reference code of datasets

In order to imbue the names of datasets with easily identifiable meaning that conveys important information, the following naming convention shall apply:

CountryCode.DataOwner.Openness.Title

CountryCode: this string identifies the country to which the data pertains/where the data was collected using the ISO 3166 Alpha-2 coding system.

DataOwner: this string identifies the project partner in GO-VIKING that is associated with the dataset (data collector/custodian) using the official abbreviated partner names.

Openness: this string determines whether a given dataset is intended to be shared with the public as Open Data. It may take the following values:

- Open: can be accessed, used and shared by anyone without limitations, accessible on the internet in a machine-readable format, free of restrictions on use in its licensing)
- Shared: available to use, but not under an open data license. Restrictions on its use or reproduction may apply (limited to a given group of people or organisations, may not be reproduced without authorisation, etc.)
- Closed: can only be accessed by its subject, owner or holder

Title: a short and descriptive string to identify the contents of the data

Using these strings, the name of a dataset would look like this:

DE.GRS.Open.CFDCodeSurvey

A dataset with this name would describe a survey on CFD code preferences, conducted in Germany and curated by GRS.

Description of the data

Each data set that will be collected, processed, or generated within the project will be accompanied by a brief table description (see an example in Table 1). The following detailed information sheet will be produced for every dataset to be produced/collected/curated in the project.

Name of the data set[i]	Complete title of the data set
Description	• **A brief, easy to understand description of what the dataset contains and what it will be used for in the project** • **A list of institutions to whom the data set could be useful outside the project** • **Whether the dataset has been/will be used for a scientific publication (if yes, brief details about the content and journal)** If the dataset is collected, a brief description of its origin and how it was collected will be provided
Media Type	The physical medium of the content representation, e.g., video, image, text, numerical data, n-grams, etc.
Language(s)	The language(s) of the resource content
Use & reuse	Foreseen use of the resource for which it has been produced
Size	Size of the resource with regard to a specific size unit measurement in the form of a number
Format/license	The format in which the data will be available (e.g. .xls, .csv, .txt) will be provided. The license to be used will also be provided.
Version Number	Specify the version number of the document

Table 1: Table specifying the content of a dataset

If a dataset is directly collected, the origin of the data set will also be provided.

4.2 Making data openly accessible

Open accessibility of the data is the second key aspect for making data FAIR. The strive in GO-VIKING is to make as many deliverables public as possible: 33 out of 34 deliverables will be public; only the deliverable on the member working area will be restricted to the GO-VIKING partners only.

Data licensing

Data licensing standards are used to lay out the openness of data sets in concrete terms. There are many types of licenses to choose from, and this document will not cover them in depth. Table 2 provides a summary of common data licenses that will be considered for use in the project (based on definitions from opendefinition.org):

Data storage

After collection, data will be generally organised in Excel files and Word documents. GO-VIKING will use Zenodo to systematically publish open data, open access presentations and public deliverables in order to maximise reuse and promote the project results. Prior to any upload, open publications on Zenodo will have to be approved by the Project Management Board of GO-VIKING. If requested, LGI may provide its support and advice to the partners prior to the publication on Zenodo. Storing data on Zenodo is free of charge and has no expiry date.

Name	Domain	Attribution	Share-alike*	Notes
Creative Commons CCZero (CC0)	Content, data	N	N	All rights (including those of attribution) waived
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	All rights (including those of attribution) waived
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, data	Y	N	Credit must be given, a link to the license must be provided, changes made must be indicated. If these terms are not followed, license may be revoked
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Credit must be given, share-alike must be assured, data may be redistributed using DRM as long as a DRM-free version is also released

Table 2: Example of licenses

*Share-alike is the requirement that any materials created using the given dataset must be redistributed under the same license

4.3 Making data interoperable & increasing re-use

Making data interoperable

Standard vocabulary may be used on a case-by-case basis to make the data interoperable between researchers, institutions, organisations, countries, etc. Further, a list of acronyms and/or abbreviations will be provided at the beginning of every report. Data will be stored using file formats in widespread use, if possible, to maximise interoperability between software solutions, operating systems, etc.

Restrictions for reuse

GO-VIKING will be compliant with the General Data Protection Regulation (GDPR). To allow reuse, respect privacy and avoid loss of research data, two different techniques could be used to disseminate its data, while abiding by regulations on privacy.

1) Anonymization of data

"Anonymization" of data means processing it with the aim of irreversibly preventing the identification of the individual to whom it relates. Data can be considered anonymised when it does not allow identification of the individuals it is related to, and no individuals can be identified from the data by any further processing of that data or by processing it together with other information which is available or likely to be available.

There are different anonymization techniques like:

- Generalisation: generalising data means removing its specificity. For example, in the case of a table containing household income levels, with 4 figures mentioned: $164,000, $58,543, $90,893, and $232,234. One way of generalising this numbers would be to write that the values are "more than $150,000, less than $60,000, between $90,000 and $100,000, and more than $225,000" respectively. Essentially it means taking exact figures, establishing a baseline category, and then obfuscating the data by assigning it to one of the categories in order to remove any sense of specificity from it.

- K-anonymity; A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from the other individuals whose information also appear in the release. For instance, in a table composed of six attributes (Name, Age, Gender, State of Domicile, Religion and Disease), removing the name and the religion column while generalising the age is a way to effectively k-anonymise the data.

- Other techniques, such as "masking" or "pseudonymisation", which are aimed solely at removing certain identifiers, may also play a role in reducing the risk of identification. In many cases, these techniques work best when used together.

2) Pseudonymisation of data

"Pseudonymisation" of data means replacing any identifying characteristics of data with a pseudonym, or, in other words, a value which does not allow the data subject to be directly identified.

Although pseudonymisation has many uses, it should be distinguished from anonymization, as it only provides limited protection for the identity of data subjects in many cases as it still allows identification using indirect means. Where a pseudonym is used, it is possible to identify the data subject by analysing the underlying or related data.

Archiving and preservation

It is of utmost importance for GO-VIKING to keep the data available for partners after the end of the project. To ensure medium-term preservation of the datasets, anonymised data will be stored on Zenodo, which is a multi-functional open platform recognised by OpenAIRE and the European Commission.

5. GO-VIKING and Open Science

GO-VIKING will fully embrace the open access policy of Horizon Europe (European Commission[2], 2022). Public scientific deliverables and associated supplementary data will be available on an EU-endorsed trusted repository (e. g. OpenAIRE on Zenodo) under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights in addition to the project website. Publications of GO-VIKING results in scientific journals will use, whenever possible, the highest level of open access. Similarly, publications in conference proceedings will be gold open access, whenever possible. Beneficiaries (or authors) shall retain sufficient intellectual property rights to comply with the open access requirements. The selection of journals and conferences as part of dissemination activities will prefer the ones enabling open access. Metadata of deposited publications shall be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles. Where applicable and possible, the metadata shall include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication. Open access requirements will not apply to background by partners as identified in the Consortium Agreement.

6. Data Management Plan implementation

Every WP Leader in GO-VIKING will be responsible for the data management within his own WP in accordance with the present GO-VIKING DMP. Data will be curated by individual partners overseen by the WP Leader. The overall implementation of the DMP in the project will be controlled by the Coordinator. In case of deviations, appropriate measures shall be identified and implemented together with the PMB members.

7. Ethical aspects

7.1 General Data Protection Rules

This DMP was drafted and updated taking into account the GDPR for the collection, storage and re-use of the data, in line with the following general principles.

Personal data shall be:

1. processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
2. 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 not be considered to be incompatible with the initial purposes ('purpose limitation');
3. adequate, relevant and limited to what is necessary for relation to the purposes for which they are processed ('data minimisation');

4. 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');
5. 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 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');
6. 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').

Conclusion

This report collects information about the Data Management Plan to be implemented within the GO-VIKING project. It provides information on the data that will be generated and collected in the different work packages, as well on the underlying principles that will be used for the generation, collection, storage, dissemination, and curation of the GO-VIKING data.

Bibliography

European Commission[1]. Directorate-General for Research & Innovation (August 2022). HORIZON EUROPE Data Management Plan – General Definition. Retrieved from: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

European Commission[2]. Directorate-General for Research & Innovation (August 2022). HORIZON EUROPE Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Retrieved from: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm