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.

EIS

Exploration
Information
System

Data Management Plan

Version 1.0

Lead Beneficiary: LGI
November 2022

Dr. Gabor Szendro[1]

[1]LGI

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	101057357
Project Title	Exploration Information System
Project Acronym	EIS
Scientific Coordinator	Vesa Nykänen (vesa.nykanen@gtk.fi) – GTK
Project starting date (duration)	1 May 2022 (36 months)
Related Work Package	WP 6 Dissemination, exploitation, communication and improving the awareness of the general public
Related Task(s)	Task 6.5 Data and Knowledge Management
Lead Organisation	LGI
Contributing Partner(s)	All
Due Date	9 November 2022
Submission Date	9 November 2022
Dissemination level	Public

History

Date	Version	Submitted by	Reviewed by	Comments
23 June 2022	0.1	Dr. Gabor Szendro		Initial version
10 November 2022	1.0	Hafsa Munia		

Table of contents

List of figures

List of tables

Summary

This is deliverable "D6.2 Data Management Plan" of the EIS project.

This document is structured in three distinct parts. The first part provides some basic contextualisation of a data management plan in Horizon Europe projects and provides key definitions necessary to understand the deliverable. The second part aims to explain the scope of EIS Data Management Plan, in other words, the purpose of data collection in the frame of EIS, the type, format and the origin of the data to be collected and the use and re-use of the data. The third part details the various actions EIS will implement to make its data findable, accessible, interoperable and available for re-use.

According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated – if appropriate – during the project lifetime (in the form of deliverables). DMPs should, therefore, have a clear version number and include a timetable for updates.

Keywords

Data, Knowledge, Information, Data Management

1. INTRODUCTION

1.1 Background of the Data Management Plan in Horizon Europe

This document is the EIS Data Management Plan (DMP), a deliverable that has been required by the European Commission for every project participating in the Open Research Data (ORD) Pilot. Since 2017, the European Commission has extended this pilot project to all thematic areas of H2020 and then Horizon Europe. According to the European Commission[i], 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 in itself but a key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. It follows from this that projects must aim to improve and maximise access to and re-use of research data generated while balancing openness and protection of scientific information, commercialisation and Intellectual Property Rights, privacy concerns, etc[ii].

Data Management Plans (DMPs) are a key element of good data management. As part of making research data findable, accessible, interoperable and re-usable (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.

The EIS DMP was written in reference to the task Task 6.5 on Data and Knowledge Management.

1.2 Definition

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 s/he 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[iii].
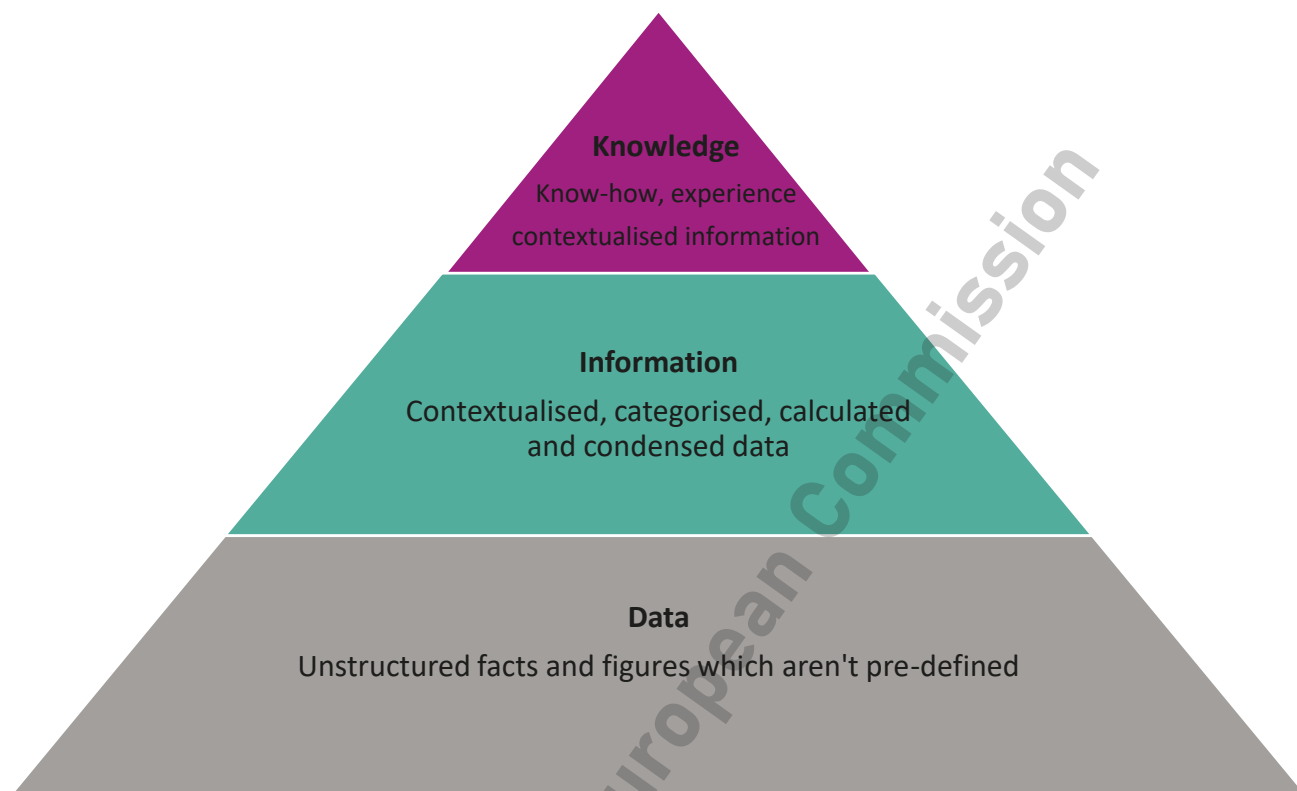

Figure 1: Knowledge Management – Definitions and hierarchy

This present DMP will mainly deal with how the data will be managed and will mention superficially the links with knowledge.

- 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[iv].

- Data set: a data set 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 data set. The data set may comprise data for one or more members, corresponding to the number of rows[v].

1.3 Versions

According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated - if appropriate - during the project lifetime (in the form of deliverables). The minimum requirement is that the DMP be updated for each periodic evaluation of the project. If there are none, such an update must be made in time for the final review at the latest.

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.

DMPs should, therefore, have a clear version number and include a timetable for updates. The timetable of updates for the present DMP will be aligned with the Progress Reports of the EIS Project.

2. Scope and lifecycle of the EIS data

2.1 General framework for data collection

In this section, the data to be collected in each WP of the EIS project will be presented and described in order to define the purpose of the collection as well as to previously define the type, format and origin.

2.1.1 WP1 – Coordination

WP1 ensures the achievement of the project's objectives, in terms of quality, timely delivery, and contribution to the expected impact of the project.

No specific research data is planned to be collected from WP1, as it is devoted to management. Personal data from project partners, advisory boards, will be collected.

2.1.2 WP2 – Mineral systems and geology for CRM deposits

The main aim of WP2 is to develop models for mineral systems hosting critical commodities and opening for new areas for exploration as well as widening the targeted commodities beyond the up-to-date exploited ones.

Data will be collected in the field and from archive data. Field data comprises rock samples from outcrops and drill core, information on lithologies, mineralogy, structures, textures. Collected data covers geochemical and geophysical datasets. Archive data includes geological data sets from national surveys and companies.

2.1.3 WP3 – Innovative digital exploration toolbox

The main objective of WP3 is development of a GIS (Geographical Information System)-based Exploration Information System (EIS) for predictive mapping of mineral resources. The backbone of EIS will be the "EIS Toolkit", comprising a comprehensive set of software tools for performing specialized tasks involved in mineral prospectivity analysis.

2.1.4 WP4 – Test case studies and CRM potential in Europe

WP4 is essentially dedicated to the operational testing and application of the mineral systems models integrated in the EIS tools developed by WP2 and WP3, respectively, using "real life" datasets. The main objective of WP4 is to produce mineral prospectivity maps with the EIS tools for the test sites that represent mineral systems investigated in WP2.

Data will be provided by the partners at various scales in order to be processed using the EIS toolboxes. Collected data will be composed of geochemical, geophysical, geological and structural features. As for WP2, the archive data includes geological data sets from national surveys and involved mining companies.

2.1.5 WP5 – Engaging with stakeholders

The main aim of WP5 is to perform clustering activities with Horizon 2020, Horizon Europe, EIT Raw Materials, national and international projects and initiatives related to EIS. No particular data collection is foreseen apart from what is necessary for events and stakeholder exchanges. These will be handled in accordance with GDPR.

2.1.6 WP6 – Dissemination, exploitation, communication and improving the awareness of the general public

During the project a GDPR compliant list of interested stakeholders will be compiled for the public EIS Scope Engagement Forum (ESEF). There are no further processing of the personal data and it is not planned to export personal data from the EU to non-EU countries or vice versa.

The main objectives of this WP are to disseminate the results of the EIS project to key stakeholders; to communicate on the activities and results of the EIS project to target audiences, while raising the awareness of the general public on raw materials; and to provide an exploitation strategy for the EIS technology to meet its market and have a sustainable economic model.

2.2 Use and Re-Use of the data

The data collected and generated by the consortium will be useful to the development of further activities related to other WPs within the project, to specific end-users and to some EU regulators.

As part of WP6, data will be re-usable only within the consortium, and for EIS related activities, such as newsletter dissemination, posters presentation, invitation to events or specific news on the website.

WP5 handles the exchange of data related to stakeholder exchanges.

Data collected within WP2 and WP4 will be re-used within the consortium. Parts of the data will be also made public in order to be used by exploration and mining companies.

3. FAIR Data Management in EIS

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).

3.1 Making data findable

This section will provide insight into how EIS 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.

3.1.1 Storing the data with datasets

- **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 EIS 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:

1. **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)**
2. **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.)**
3. **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:

FR.LGI.Open.CommuteHouseholdSurvey

A dataset with this name would describe a household survey on commuting preferences conducted in France and curated by LGI.

- **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. The following detailed information sheet will be produced for every dataset to be produced/collected/curated in the project.

Name of the data set[vi]	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 & re-use	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.

3.2 Making data openly accessible

Open accessibility of the data is the second key aspect for making data FAIR. This section will describe the type of data to be made available, its location and the procedure to obtain it.

Several degrees of accessibility are identified below, including both open access and restricted access.

3.2.1 Data licensing

Data licensing standards are used to layout 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. The table below provides a summary of common data licenses that will be considered for use in the project (based on definitions from opendefinition.org):

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

3.2.2 Datasets that could be made openly accessible in EIS

The table below contains already identified datasets that can be made openly accessible. The list will be regularly updated.

Data producer	Brief description of dataset	Foreseen use & re-use	Possibility to share the data beyond the consortium
BRGM	Test site for Li Mineralization in the Armorican in France at regional scale	Dataset is public and can be used by partners.	YES

Table 3: Datasets that could be made openly accessible in EIS

3.2.3 Datasets to remain confidential

Most of the data produced by EIS will be public. Exceptions will be added as part of updates to the DMP over time.

3.2.4 Data storage in EIS

After collection, data will be generally organised in Excel files and Word documents.

EIS will use Zenodo to systematically publish open data, open access presentations and public deliverables in order to maximise re-use and promote the project results. Prior to any upload, open publications on Zenodo will have to be approved by the Steering committee of EIS. 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.

3.3 Making data interoperable & increasing re-use

3.3.1 Making data interoperable

As described on 3.1, Standard vocabulary may be used on a case-by-case basis to make the data interoperable between researchers, institutions, organisations, countries, etc.

- 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 to maximise interoperability between software solutions, operating systems, etc.

- For surveys, standard definitions for entities such as trips, trip chains etc. will be adopted.

3.3.2 Restrictions for re-use

EIS will be compliant with the General Data Protection Regulation. To allow re-use, 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[vii]

"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. Here are the two most relevant:

- 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.

3.3.3 Archiving and preservation

It is of utmost importance for EIS to keep the data available for partners after the end of the project.

As already mentioned in 3.2, to ensure medium-term preservation of the datasets, anonymised data will be publicly shared by the consortium will be stored on Zenodo, which is a multi-functional open platform recognised by OpenAIRE and the European Commission.

4. Human resources

Every Work Package Leader in EIS will be responsible for the data management within its own Work-Package. LGI is responsible for authoring the present Data Management Plan in collaboration with project partners, and to make updates to it during the project as necessary.

5. Ethical aspects

1.1 General principles

The general principles of ethical standards and guidelines defined for the Horizon Europe projects as well as relevant national and EU guidance, directives and legislation will be followed in the project. The most important articles from the EIS Grant Agreement t are listed below. Relevant EU directives and regulations are also listed.

- EIS Grant Agreement ARTICLE 15 — DATA PROTECTION
    - 15.1 Data processing by the granting authority
    - 15.2 Data processing by the beneficiaries
- Data Protection Directive 95/46/EC
- General Data Protection Regulation 2016/679 (Enforcement date: 25 May 2018)

Within the EIS project there will be no collection and processing of sensitive personal data that could reveal political opinions, religious or philosophical beliefs, trade-union membership, nor data concerning health or sexual orientation.

5.1 GDPR

This Data Management Plan (DMP) was drafted and updated taking into account the General Data Protection Rules (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, in accordance with Article 89(1), 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 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');

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')[viii].

5.2 Informed consent

EIS will follow the rules defined in GDPR. As a rule, under the GDPR processing of personal data is lawful only if and to the extent that at least one of the grounds for processing set forth in GDPR Article 6 applies. In the EIS case, the only applicable ground would be the consent given by the data subject (e.g. the contact person in question). This does not apply to the organizations and their general contact data as such, i.e. the name of the organization, its address, telephone number to the reception or operator, general e-mail address, as such information does not constitute personal data.

But processing (i.e. collecting, publishing, etc.) of contact data of persons representing the organization, including their names, e-mails, direct dials, address etc. requires consent as constituting personal data. The consent should be explicit and the party collecting and processing the data must be able to demonstrate that the data subject has consented to processing of his or her personal data.

In EIS, the Lyyti registration system is used to collect personal information to GTK´s register of participants of events. It can be used for registering participants of stakeholder events and internal meetings. It is not possible to submit Lyyti registration forms without accepting the privacy policy statement, which describes in detail how the submitted personal information will be managed. All information on the participants will be removed from the register at the latest after one month after the end of each event. In addition, the privacy policy statement explains that names and organisational information of participants may be permanently stored in reports describing each event.

All stakeholders attending EIS events or answering the questionnaires will be recruited into these activities fully on a voluntary basis. Stakeholders will be given the EIS Project basic information material and they are kept aware that they can exercise their rights to protect their personal data. Participants of the EIS interviews and other activities will be asked to sign the informed consent form which clearly describes the intention of the activity, what data is collected, how it is used by whom, and how the participant can update or delete his/her information.

Conclusion

This is the initial version of the Data Management Plan of the EIS project. It has been drafted to provide initial information on the data collected, stored, analysed and published as part of the work. It will receive updates over time, notably after each periodic report, as necessary.

Annex 1 – detailed dataset descriptions

Name of the data set[ix]	Armorican Dataset from BRGM (FRANCE)

Description	The dataset includes geological, structural, geophysical, geographic and hard rock lithium mineralization in the Armorican Massif (Brittany, NW France). This note also briefly describes the geological and geodynamic history of the region. These data layers and information constitute a test dataset, for hard rock lithium mineralization in the Armorican massif, that is provided by BRGM and could be used to test the EIS wizard and toolkits that will be developed by WP3 of the project.
	
	- Mask_Brittany: study area
	- Faults_1M.shp: faults extracted from the 1 million scale geological map of France – mean orientation is stored in attribute ORI180. Other attributes: Nature_ENG : Nature of the fault, Important: Importance of the fault,
	- Faults_1M_Plz.shp: faults extracted from the 1 million scale geological map of France; faults oriented between 150° and 170° related to Tertiary events have been eliminated of the dataset,
	- Main_faults.shp: synthetic major faults,
	- Geology.shp : geological map synthetized from the 1:50 000th scale geological maps of France. Attributes are: Area (in sq km) Lithology, System (from Paleoproterozoic until Quaternary), RMK (remarks), Veins (1 if true) and Code (coded concatenation of lithology and system as described in the tables) – see Geological coding,
	- BD Lithium-Brittany_WGS84.xls and BD_Lithium_Brittany.shp: all occurrences of Li in Brittany; long/lat; TYPOLOGY: metallogenic type, COMMODITIES: main known metallic commodities,
	- BD Lithium-FRANCE_WGS84.xls and BD_Lithium_France.shp: all occurrences of Li in France; long/lat; TYPOLOGY: metallogenic type, COMMODITIES: main known metallic commodities,
	- main-streams.shp: main rivers of Brittany,
	- Main_cities : a selection of main cities of Brittany,
	- Grids:
	  o bougano1000m.tif: 1000m grid Bouguer anomaly,
	  o dem100m.tif: 100m Digital Elevation Model from SRTM,
	  o RTP.tif: 250m grid of mag survey, magnetic field reduced to the pole,
	  o RTPVG.tif: vertical gradient of the reduced to the pole magnetic field.].

Media Type	Image and tables (excel files)
Language(s)	English

Use & re-use	Public
Size	N/A
Format/license	Vector data layers are in ESRI Shape format (.shp) and all raster data layers are in .tif format. Projection system of the whole dataset is [UTM WGS84], zone [30N]. Coordinates mentioned in excel tables are in WGS84. ArcGis License is needed
Version Number	V0.1

Name of the data set[x]	List of interested stakeholders – EIS Scope Engagement Forum (ESEF)
Description	List of interested stakeholders is the main part of the Task 5.2 Development and launching of public EIS Scope Engagement Forum (ESEF). The list will be used in WP5 to perform clustering activities with Horizon 2020, Horizon Europe, EIT RawMaterials, national and international projects and initiatives related to EIS. WP5 will engage also with the policy makers and with several initiatives at EU, international, national and regional level, including DG GROW, S3 Industrial Modernisation partnership - Mining industry, EIT RawMaterials, European Raw Materials Alliance, European Battery Alliance, IPCEI project European Battery Innovation (EuBatIn), OECD´s Mining regions and Cities initiative, Council of Mining and Metallurgy Regions EU - CoMMER, European Clusters Alliance, European Cluster Collaboration Platform. The end goal is to build a critical mass of interested stakeholders to stimulate stakeholder engagement and to enhance societal acceptance and uptake of new innovations. This dataset will not be used for a scientific publication.
Media Type	Text, excel file
Language(s)	English
Use & re-use	Confidential. It will be used to invite stakeholders e.g. to clustering events.
Size	N/A
Format/license	xls/no license
Version Number	V0.1

Bibliography

i European Commission. Directorate-General for Research & Innovation (March 2017). HORIZON
EUROPE Programme Guidelines to the Rules on Open Access to Scientific Publications and
Open Access to Research Data in Horizon 2020. Retrieved from:
http://ec.europa.eu/research/participants/data/ref/Horizon
Europe/grants_manual/hi/oa_pilot/Horizon Europe-hi-oa-pilot-guide_en.pdf

ii European Commission. Directorate-General for Research & Innovation. (July 2016). HORIZON
EUROPE Programme – Guidelines on FAIR Data Management in Horizon 2020. Retrieved from:
http://ec.europa.eu/research/participants/data/ref/Horizon
Europe/grants_manual/hi/oa_pilot/Horizon Europe-hi-oa-data-mgt_en.pdf

iii Alan Frost. Defining Knowledge, Information, Data. (2017). Retrieved from: http://www.knowledge-
management-tools.net/knowledge-information-data.html

iv DDI Alliance. Create a codebook. Retrieved from: http://www.ddialliance.org/training/getting-
started-new-content/create-a-codebook

v Rajit Dasgupta. 19 Free Public Data Sets for your first data science project. Springboard.com.
(October 2015). Retrieved from : https://www.springboard.com/blog/free-public-data-sets-
data-science-project/

vi Maria Koutsombogera & Stelios Piperidis (Athena RC). D3.1 Data Management Plan. Cracker,
Cracking the language barrier. (June 2015). Retrieved from:
https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=08016
6e5a0203500&appId=PPGMS

vii Data Protection Commissioner. Anonymization and Pseudonymization. Retrieved from:
https://www.dataprotection.ie/docs/Anonymisation-and-pseudonymisation/1594.htm

viii Intersoft Consulting. GDPR. Article 5. (2018) Retrieved from : https://gdpr-info.eu/art-5-gdpr/

ix Maria Koutsombogera & Stelios Piperidis (Athena RC). D3.1 Data Management Plan. Cracker,
Cracking the language barrier. (June 2015). Retrieved from:
https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=08016
6e5a0203500&appId=PPGMS

x Maria Koutsombogera & Stelios Piperidis (Athena RC). D3.1 Data Management Plan. Cracker,
Cracking the language barrier. (June 2015). Retrieved from:
https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=08016
6e5a0203500&appId=PPGMS