**Establishing the European Geological
Surveys Research Area to deliver a
Geological Service for Europe**

Deliverable

DATA MANAGEMENT PLAN
VERSION 1

Authors and affiliation:
Jørgen Tulstrup
[GEUS]

E-mail of lead author:
jtu@geus.dk

Version: 28-06-2017

Deliverable Data		
Deliverable number	D1.3	
Dissemination level	Public	
Deliverable name	Data Management Plan	
Work package	WP1, Management	
Lead WP/Deliverable beneficiary	GEUS	
Deliverable status		
Submitted (Author(s))	09-06-2017	Jørgen Tulstrup
Verified (WP leader)	29-06-2017	Yvonne Schavemaker
Approved (Coordinator)	30-06-2017	Paul Bogaard

GENERAL INTRODUCTION

This report describes the Data Management Plan (DMP) for the whole of the GeoERA programme. The purpose of the plan is to ensure that research data generated in the GeoERA projects will be findable, accessible, interoperable and reusable (FAIR). This overall DMP explains how the central information system (the Information Platform) must deal with issues about data accessibility, interoperability, etc.

The individual projects, which will be funded by GeoERA, will all be required to write a specific Project Data Management Plan and in that describe how the research data generated in each project will be made FAIR.

This DMP is structured according to the Horizon 2020 FAIR Data Management Plan template. The specific Project Data Management Plans must use a template containing a subset of the Horizon 2020 FAIR Data Management Plan.

This DMP will be updated over the course of the GeoERA programme whenever new important information is available or significant changes arise.

TABLE OF CONTENTS

1   INTRODUCTION

The many projects that will run under GeoERA will generate very large and complex data sets and it is of utmost importance that these are made accessible to the researchers during the course of the projects and to all stakeholders after the end of the GeoERA. This is why GeoERA contains a specific theme (the Information Platform Theme, IPT) to deal with organising and disseminating the data resulting from the research.

In addition to organising the data management issues under the IPT, GeoERA will require that all data are made findable, accessible, interoperable and reusable (FAIR) as defined in the [Guidelines on FAIR Data Management in Horizon 2020](#).

These guidelines contain a template for describing how a project will ensure data are FAIR. In the case of the whole GeoERA programme, some general principles will be defined at an early stage (before the different research projects have been selected). This is what this report describes. But at this early stage the specific data, which will be generated in the projects, are not known, and specific requirements regarding those therefore cannot be defined yet. Instead, each project proposal for the Stage Two call will be required to describe how it will ensure data are FAIR, and each funded project will be required to make a specific data management plan in an early stage of the project. At the end of this report a first version of a template for the project specific data management plans can be found.

This overall Data Management Plan will be updated during the course of the GeoERA programme whenever new important information is available or significant changes arise. This will happen (at least) when the Stage 2 call texts are published, when projects to receive funding have been selected and in time with the periodic evaluations/ assessments of GeoERA.

2   THE OVERALL GEOERA DATA MANAGEMENT PLAN

The following is the overall GeoERA Data Management Plan. It is structured as recommended in the template for Data management Plans in [Guidelines on FAIR Data Management in Horizon 2020.](#)

2.1   Data summary

2.1.1.1 *What is the purpose of the data collection/generation and its relation to the objectives of the programme?*

Data harmonisation and provision are key aspects of the programme, and are specifically addressed though the dedicated Information Platform theme. The development of new and improved transnationally harmonized data and information services are also key objectives of the other three themes.

2.1.1.2 *What types and formats of data will the programme generate/collect?*

These will include derived information such as geological maps and models, resource assessments, risk and vulnerability assessments, and impact assessments, but also primary datasets such as borehole information and geophysical data.

2.1.1.3 *Will you re-use any existing data and how?*

The projects will to a high degree be based on data currently residing in existing national and regional data and information repositories to produce new compilations and other derived data sets.

2.1.1.4 *What is the origin of the data?*

Data originates from national and regional data and information repositories. A lot of compilations of data will be produced by the projects themselves, but to some degree the DMP must also handle data from other projects outside GeoERA.

2.1.1.5 *What is the expected size of the data?*

This is not possible to estimate before the projects start.

2.1.1.6 *To whom might it be useful ('data utility')?*

The Information Platform is intended to function as the data infrastructure for geoscience information and data for the GeoERA partners and stakeholders also beyond the duration of GeoERA. In this context, Information Platform projects will be required to explicitly address interfaces with other EU data infrastructures in the domains of geo-energy, groundwater and raw materials, for instance with EPOS, EMODNET, WISE, RMIS and GEOSS. The Information Platform will be based on the already existing concept of the European Geological Data Infrastructure. The platform will serve as a first access point for stakeholders, e.g. the European Commission, that are searching for pan-European data and information on the subsurface.

2.2 Fair data

2.2.1 Making data findable, inclusion provisions for metadata

2.2.1.1 Are the data produced and/or used in the programme discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

All data must be discoverable with metadata. Many will also be identifiable and locatable by means of standard identification mechanisms (for instance unique borehole numbers). The projects will be encouraged to consider Digital Object Identifiers (DOIs).

2.2.1.2 What naming conventions do you follow?

Naming conventions have not been defined yet. This will be one of the first tasks for the Information Platform Theme, when the projects are awarded.

2.2.1.3 Will search keywords be provided that optimize possibilities for re-use?

The metadata system of the Information Platform will provide facilities for tagging the datasets with keywords.

2.2.1.4 Do you provide clear version numbers?

End products, like maps, 3D models, etc. will have version numbers, but several of the more basic datasets (borehole data for instance) will be dynamic in nature and some providers of data may not define versions.

2.2.1.5 What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

A metadata system based on ISO 19115 will be part of the Information Platform. This will be used for basic datasets and services for maps. For 3D models and other data types there is currently no agreed, international standard for metadata. Defining this standard is part of the activities carried out within the Information Platform theme.

2.2.2 Making data openly accessible

2.2.2.1 Which data produced and/or used in the programme will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Some data (in particular primary or raw data) may be subject to access restrictions as the data may be subject to confidentiality and/or commercial restrictions at the national level. The specific Project Data Management Plans must explain and justify each case.

2.2.2.2 How will the data be made accessible (e.g. by deposition in a repository)?

The Information Platform will be the key tool to make the data and data products accessible to a broad range of Stakeholders, including policy makers, industry and academia. Access will be possible for both verification and re-use purposes and the Information Platform will contain view, search and visualization facilities alongside standardized view- and download-features.

2.2.2.3 What methods or software tools are needed to access the data?

Certain datasets may be accessible only using specific software, but this should be avoided as much as possible. Each case must be explained and justified in the data management plan for the particular project.

2.2.2.4 Is documentation about the software needed to access the data included?

It may be in certain cases. This will be requested for projects to define in their Project Data Management Plan.

2.2.2.5 Is it possible to include the relevant software (e.g. in open source code)?

This is not clear yet. This will be requested for projects to define in their Project Data Management Plan.

2.2.2.6 Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The major part of the data, metadata and documentation will be stored in the EGDI central database(s) which is intended to be sustained by the GSOs after the end of the GeoERA. This will ensure accessibility.

2.2.2.7 Have you explored appropriate arrangements with the identified repository?

Negotiations are currently on-going with EuroGeoSurveys regarding the long term operation and sustainability of the EGDI central system.

2.2.2.8 If there are restrictions on use, how will access be provided?

Access to operation of the EGDI will be clarified under the Information Platform Theme, when the projects are awarded.

2.2.2.9 Is there a need for a data access committee?

The GeoERA Executive Board will take the role as Data Access Committee once the Call for Project Proposals is launched.

2.2.2.10 Are there well described conditions for access (i.e. a machine readable license)?

This will be considered under the Information Platform Theme, when the projects are awarded.

2.2.2.11 How will the identity of the person accessing the data be ascertained?

Issues regarding authentication, authorization and accounting will be dealt with on general terms under the Information Platform Theme.

2.2.3 Making data interoperable

2.2.3.1 Are the data produced in the programme interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Very strong focus will be on interoperability of the data produced. The Information Platform Theme will deal specifically with that and all other projects within the other themes will be required to document how their data will be interoperable. All maps must

be made available for use by others as Open Geospatial Consortium (OGC) web services.

2.2.3.2 What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Projects will use established European and international standards for the storage, exchange and dissemination of project data. INSPIRE (the European Directive on Infrastructure for Spatial Information) compliance will be used wherever possible. Where this is not possible, Commission for the Management and Application of Geoscience Information (CGI) standards will be used. Nevertheless, for some data types (e.g. 3D/4D geological models) no accepted international standards currently exist. The GeoERA call will explicitly solicit projects that address the development of such standards.

2.2.3.3 Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Standard vocabularies will be used to the extent that they exist or will be developed in the projects.

2.2.3.4 In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

The projects will be required to document if they find it unavoidable to use uncommon of generate specific ontologies or vocabularies. If mapping to more commonly used ontologies will be possible the projects will be required to establish such a mapping.

2.2.4 Increase data re-use (through clarifying licences)

2.2.4.1 How will the data be licensed to permit the widest re-use possible?

The projects will be required to specify this in their Project Data Management Plan.

2.2.4.2 When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

The projects will be required to specify this in their specific Data Management Plan.

2.2.4.3 Are the data produced and/or used in the project useable by third parties, in particular after the end of the programme? If the re-use of some data is restricted, explain why.

The general rule will be that data produced in the projects shall be useable by third partied. If this for some reason will not be the case in a certain project this must be justified in the specific Project Data Management Plan.

2.2.4.4 How long is it intended that the data remains re-usable?

The projects will be required to specify this in their Project Data Management Plan.

2.2.4.5 Are data quality assurance processes described?

The projects will be required to specify this in their Project Data Management Plan.

2.3 Allocation of resources

2.3.1.1 What are the costs for making data FAIR in your project?
This will be specified in the Project Data Management Plans.

2.3.1.2 How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
Costs for making data FAIR must be covered by the individual projects.

2.3.1.3 Who will be responsible for data management in your programme?
The Theme Coordinator for the Information Platform Theme.

2.3.1.4 Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?
The costs for the maintenance of the central EGDI platform is under consideration (discussions with EGS). Costs for the long term preservation of data, which will not be integrated in the central Information Platform, must be documented in the Project Data Management Plan.

2.4 Data security

2.4.1.1 What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
For data stored in the central EGDI database the security is currently secured through the fact that it is operated by BRGM and included in their operational procedures. For the data which will not be included in this database a documentation for the data security issues must be part of the Project Data Management Plan for the project which will produce the data.

2.4.1.2 Is the data safely stored in certified repositories for long term preservation and curation?
See above.

2.5 Ethical aspects

2.5.1.1 Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
This is not likely to be the case.

2.5.1.2 Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?
Not applicable as the programme will not deal with personal data.

2.6 Other issues

2.6.1.1 Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?
The underlying data from the GSOs will in many cases be governed by national/regional or institutional rules. This must be described in the Project Data Management Plan for the individual project.

3   TEMPLATE FOR PROJECT DATA MANAGEMENT PLANS

The following is a first version of a template to be used in the project specific Data Management Plans for all funded projects under GeoERA. The questions are a subset of the questions of the full Data Management Plan Template used above for the whole of GeoERA with a few additions.

The funded projects must deliver a specific Project Data Management Plan within the first 6 months of the project and the Plan must be updated over the course of the project whenever significant changes arise. As a minimum the Plan should be updated in time with the periodic evaluations of the project.

3.1   Data summary

3.1.1.1   What is the purpose of the data collection/generation and its relation to the objectives of the programme?

3.1.1.2   What types and formats of data will the programme generate/collect?

3.1.1.3   Will you re-use any existing data and how?

3.1.1.4   What is the origin of the data?

3.1.1.5   What is the expected size of the data?

3.1.1.6   To whom might it be useful ('data utility')?

3.2   Fair data

3.2.1   Making data findable, inclusion provisions for metadata

3.2.1.1   Are the data produced and/or used in the programme discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

3.2.1.2   What naming conventions do you follow?

3.2.1.3   Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

3.2.2   Making data openly accessible

3.2.2.1   Which data produced and/or used in the programme will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

3.2.2.2   Will the data be made accessible through the Information Platform? If not, explain why.

3.2.2.3   What methods or software tools are needed to access the data?

3.2.2.4   Is documentation about the software needed to access the data included?

3.2.2.5   Is it possible to include the relevant software (e.g. in open source code)?

3.2.2.6   Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

3.2.3   Making data interoperable

3.2.3.1   Are the data produced in the programme interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

3.2.3.2   What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

3.2.3.3   Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

3.2.3.4   In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

3.2.4   Increase data re-use (through clarifying licences)

3.2.4.1   How will the data be licensed to permit the widest re-use possible?
The projects will be required to specify this in their Project Data Management Plan.

3.2.4.2   When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

3.2.4.3   Are the data produced and/or used in the project useable by third parties, in particular after the end of the programme? If the re-use of some data is restricted, explain why.

3.2.4.4   How long is it intended that the data remains re-usable?

3.2.4.5   Are data quality assurance processes described?

3.3   Allocation of resources

3.3.1.1   What are the costs for making data FAIR in your project?

3.3.1.2   Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

3.4   Data security

3.4.1.1   What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

3.4.1.2   Is the data safely stored in certified repositories for long term preservation and curation?

3.5   Other issues

3.5.1.1   Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?