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

Deliverable 10.5

SUPPORTING DOCUMENTS

Authors and affiliation:
Mary Carter
[GSI]

E-mail of lead author:
mary.carter@gsi.ie

Version: 26-03-2019

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 731166.

Deliverable Data		
Deliverable number	D10.5	
Dissemination level	Public	
Deliverable name	Data Management Plan GIP-P	
Work package	WP10	
Lead WP/Deliverable beneficiary	BGS	
Deliverable status		
Submitted (Author(s))	26/03/2019	Mary Carter
Verified (WP leader)	27/03/2019	Christopher Luton
Approved (Coordinator)	30/04/2019	Jørgen Tulstrup

GENERAL INTRODUCTION

This report is the Data Management Plan(DMP) for the Information Platform Project (GIP-P). It follows from 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.

TABLE OF CONTENTS

1 GIP-P DATA MANAGEMENT PLAN

The following is a first version of the GIP-P specific Data Management Plans.

1.1 Data summary

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

The purpose of data collection is to provide trans nationally harmonised data and information services in a dedicated Information Platform theme. The GIP-P will liaise with the other projects trough dedicated interfaces.

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

The programme is expected to both raw data and derived information such as geological maps, 2-d and 3-D models, fault databases, groundwater data, urban data and geotechnical data, raw materials data. It will generate a harvesting functionality for some data, also visualization tools, and information knowledge and Intelligence

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

The projects will use data from existing national and regional repositories, but will produce new compilations and derived data sets

1.1.1.4 What is the origin of the data?

Data exists in national and regional repositories. Data will also be generated during compilation and harmonization exercises

1.1.1.5 What is the expected size of the data?

This is not possible to estimate at this time.

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

It is likely to be useful as a data infrastructure to GeoERA partners and stakeholders for the duration of GeoERA and beyond it. It will interface with other EU data infrastructures, and will be based on the already existing concept of the European Geological Data Infrastructure (EGDI). Stakeholders, policy makers and academia will use the GIP-P to access pan European data and information on the subsurface.

1.2 Fair data

1.2.1 *Making data findable, inclusion provisions for metadata*

1.2.1.1 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 will be discoverable with metadata. Many will be locatable through visualization using coordinate systems. The projects will be encouraged to use Digital Object Identifiers.

1.2.1.2 What naming conventions do you follow?

Naming conventions have not been defined yet

1.2.1.3 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.

The data and information will be distributed but centrally coordinated in EGDI by the GIP-P

1.2.2 *Making data openly accessible*

1.2.2.1 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 may be such as raw data or primary data may be subject to restrictions because it is subject to confidentiality or commercial restrictions at the national level. Each case will be explained in the Data management plan

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

All data will be available through the Information Platform except some raw or primary data and that data that is subject to restrictions for commercial or confidentially reasons.

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

Many datasets will be openly accessible although some datasets will be only available using specific software.

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

It may be in certain cases

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

This is not clear yet

1.2.2.6 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 will be stored in EGDI central databases which is intended to be maintained by EuroGeoSurveys after the end of GeoERA. Some data and information may be distributed.

1.2.3 *Making data interoperable*

1.2.3.1 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)?

Yes, the data will be interoperable. EGDI will focus on interoperability, and will require all other projects to demonstrate their interoperability. All maps will be made available as Open Spatial Consortium (OGC) web services.

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

We will use established European and international standards for the storage, exchange, and dissemination of 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. For some data types (e.g. 3D/4D geological models) no accepted international standards currently exist.

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

Yes standard vocabularies will be used where they exist, or they will be developed during the project.

1.2.3.4 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?

If it is unavoidable to use uncommon or generate specific project ontologies or vocabularies, this will be documented. If a mapping to more common ontologies is possible

1.2.4 ***Increase data re-use (through clarifying licenses)***

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

This will be specified in the project data management plans at a later stage.

1.2.4.2 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.

This will be specified in the data management plan at a later stage.

1.2.4.3 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.

As a general rule the data produced will be available by third parties. If for some reason it is not, then this will be justified in the Data Management Plan.

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

This will be stated in the Data Management plan at a later stage

1.2.4.5 Are data quality assurance processes described?

These will be described in the Data Management Plan at a later stage.

1.3 Allocation of resources

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

This will be specified in the data Management plan at a later stage.

1.3.1.2 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)?

The costs for the maintenance of the central EGDI platform are 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 Plans for the individual geoscientific GeoERA projects.

1.4 Data security

1.4.1.1 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 databases the security is currently secured through the fact that it is operated by BRGM, GeoZS and GEUS and included in their operational procedures. For the data which will not be included in this database documentation for the data security issues must be part of the Project Data Management Plan for the project which will produce the data.

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

See above.

1.5 Other issues

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

Yes, the underlying data from the Geological Surveys will in many cases be governed by national, regional, or institutional rules.