3D Geomodeling for Europe
Project number: GeoE.171.005

Deliverable 8.2

Project Data Management Plan

Authors and affiliation:
Stefan Knopf (BGR)

E-mail of lead author:
Stefan.Knopf@bgr.de

Version: 31-01-2019

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	D8.2	
Dissemination level	Public	
Deliverable name	Project Data Management Plan	
Work package	WP8 Project Management and Coordination	
Lead WP/Deliverable beneficiary	BGR	
Deliverable status		
Submitted (Author)	14/01/2019	Stefan Knopf
Verified (WP leader)	24/01/2019	Stefan Knopf
Approved (Project leader)	31/01/2019	Stefan Knopf

GENERAL INTRODUCTION

This report describes the Data Management Plan (DMP) for the GeoERA project 3DGEO-EU. The purpose of the plan is to ensure that research data generated in 3DGEO-EU will be findable, accessible, interoperable, and reusable (FAIR).

This DMP is structured according to guidelines of the Horizon 2020 FAIR Data Management Plan template. It will be updated over the course of the project whenever new important information is available or significant changes arise.

Version History

Version	Date	Description
V1.0	31.01.2019	First version of Data Management Plan (status M6)

TABLE OF CONTENTS

1 INTRODUCTION

The present document is the deliverable D8.2 "**Project Data Management Plan**" of the GeoERA project 3DGEO-EU. The purpose of this plan is to ensure that research data generated in 3DGEO-EU will be findable, accessible, interoperable, and reusable (FAIR).

This project relates to the GeoERA Specific Research Topic GE5 "Advancements in developing and using 3D transnational geomodels". The partners of the consortium aim to develop, test and improve methods for the harmonization of cross-border 3D geomodels in pilot areas of different geological situations and source data coverage. Additional work packages focus on selected aspects in regard of uncertainties of geological 3D models and their visualization, cross-border modeling of fault data, and optimized reconstruction and restoration workflows to reduce uncertainty of geomodels.

The principal outcomes of the 3DGEO-EU project will be methods and workflows for cross-border harmonization of 3D geomodels, harmonized cross-border 3D geomodels and maps in pilot areas, and example data sets which can be used to test visualization methods for model uncertainties. All technical results from the 3DGEO-EU project will be integrated in the European Geological Data Infrastructure (EGDI), through the GeoERA Information Platform project (GIP-P). The provided results will deliver an important input for further transnational harmonization projects in other regions in Europe.

This is the first version of the Project Data Management Plan (DMP). It contains preliminary information about what the project currently (M6) expects regarding the different data used and collected during the project. The DMP is not a fixed document but will be refined in subsequent revisions in the course of the project.

2 DATA MANAGEMENT PLAN

This DMP for the project 3DGEO-EU is structured as recommended in the Guidelines on FAIR Data Management in Horizon 2020. It provides relevant information concerning the data that will be collected, used, and generated by the partners of the project. This first edition of the DMP is based on the present state of knowledge and views of the partners regarding data management.

The question of intellectual property rights and ownership is important for all project partners and must be taken into account. Due to possible limitations in data privacy rules (variable among partners), we will only publish processed and interpreted output datasets (which are partially derived from unpublished input data) and datasets that will be measured in field campaigns during the project life.

2.1 Data summary

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

The main purpose is to work on the topic of cross-border harmonization of 3D geomodels and geodata. That involves tasks like harmonizing (seismo)stratigraphical interpretations and time-, depth-, or thickness-maps of certain litho-stratigraphical horizons across borders. For all pilot areas the inventory of existing data, maps, geomodels, model concepts, and geological interpretations on both sides of a border (with a focus on identifying cross-border differences) will be documented in status reports.

The project addresses testing and improving methods regarding harmonization of 3D geomodels, visualization of uncertainties, and optimized 2D/3D reconstruction and restoration workflows to reduce uncertainty of geomodels. Harmonized, consistent cross-border 3D geomodels will be prepared for the pilot areas.

The development of new and improved transnationally harmonized data and information services is a key objective of the GeoERA programme.

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

The project produces harmonized cross-border 3D geomodels and maps in pilot areas, example data sets which can be used to test visualization methods for model uncertainties, and reports on methods and workflows ([3D-] pdf-files). The results will mainly consist of derived information/data/models based on existing primary data (e.g. well data, seismics) and/or existing national/regional 3D geomodels.

Regarding WP6 (optimizing reconstruction and restoration workflows) however, gravity and petrophysics measurements and serial cross sections will be acquired/performed in Spain (Pyrenees case study). This data will honor FAIR rules and will be accessible from the EDGI and IGME servers.

Format:
The format highly depends on the data type but will be harmonized as much as possible among the partners. It will also respect the common usage among the scientific community depending on the data type. Format includes Gocad ASCII format, VTK ASCII and binary format (Visualization Toolkit, see www.vtk.org, open source), 2D grids in CPS3 format, shapefiles, properties in Excel format, and pdf-files.

A dedicated web-GIS platform is developed by the GeoERA Information Platform project (GIP-P) to support the provision of the results toward the public. Web-services such as WMS, WMFS and download services and metadata catalogues will be set up to assure the requirements of the guidelines on FAIR data management.

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

The project will considerably be based on data currently residing in existing national and regional data and information repositories to produce new or enhance existing derived data sets and compilations.

2.1.4 What is the origin of the data?

Data mainly originates from national and regional data information repositories of the involved countries/project partners. Partly, data is used which is not owned by the partners. Property rights must be taken into consideration and may vary among partners. The data not owned by partners will not be made available to the GeoERA community, let alone the public.

To a minor degree 3DGEO-EU will acquire new data from gravity and petrophysics field measurements in selected case study areas (e.g. Pyrenees).

2.1.5 What is the expected size of the data?

As it is the first DMP at an early stage of the project the size is hard to define. But, according to the expected deliverables and related formats (see above), it should be less than one gigabyte (GB) for documents to a few GB for 3D geomodel-datasets.

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

The results of 3DGEO-EU will be useful for the 3DGEO-EU consortium, for Geological Survey Organizations (GSO) in Europe, for the 3D geomodeling community, the research community, state institutes that deal with large underground facilities (e.g. traffic tunnels), and for stakeholders in the pilot areas.

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).

In the further course of the project the partners will decide if Digital Object Identifiers (DOIs) should be used (e.g. for reports).

2.2.1.2 What naming conventions do you follow?

Naming conventions have not been defined yet. They will be defined after M6 by the GIP-P, with contributions from the 3DGEO-EU project specific IP Interface Work Package 7. The DMP will be updated accordingly.

2.2.1.3 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 major part of the generated data, metadata, and documentation will be stored on the Information Platform and thereby 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.

In some cases, strict data privacy rules may also include derived and interpreted data, and physical data transfer to the EGDI central database is restricted by national laws. In these cases the GSOs/project partners will ensure the accessibility to national platforms via hyperlinks.

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.

The output data produced by 3DGEO-EU will be stored and published via the Information Platform (and thereby also EGDI) and is therefore openly available. Data used that is not owned by partners will not be made accessible, as it is subject to confidentiality and/or commercial restrictions or embargo rules at the national level. In later versions of the DMP each case of restricted data sets will be explained.

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

Yes, the output data produced by 3DGEO-EU will be made accessible through the Information Platform.

Considering the facts that this is the first DMP and EGDI is also under development during the project duration, the direct accessibility of the data might be restricted by technical issues. For instance, 3D data are not yet implemented in EGDI.

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

The data that will be originally produced by 3DGEO-EU will follow the common good scientific practice, mainly in the research field of geological 3D modeling. Access to the original 3D models/data requires specific commercial geomodeling software, open-source visualization tools or the development of specific conversion tools for provided ASCII data or other generic formats (i.e. *.grd, *.shp, *.dxf, *.pdf [3D]).

The produced datasets and/or 2D depictions/representations of these will be made available on the Information Platform, which is accessible via any web browser application.

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

Commercial specific software can most likely not be included. However, 2D depictions/representations of the generated 3D geomodels included in the Information Platform can be visualized with any web browser application.

2.2.2.5 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.

See 2.2.1.3.

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)?

The GeoERA Information Platform project (GIP-P) will deal specifically with this issue and will provide requirements to the other GeoERA projects.

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

As for 3D geomodels no accepted standards exist so far, 3DGEO-EU will cooperate with GIP-P and with other projects to address development of standards.

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

3DGEO-EU will follow 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. If 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.

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?

If such a case occurs, 3DGEO-EU will document it. If mapping to more commonly used ontologies is possible, we will provide 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 as possible?

During the GeoERA projects period of time, the WP9 from the GIP-P will provide assistance and will harvest the produced datasets from 3DGEO-EU.

Nevertheless, for existing data not generated with GeoERA co-fund, the Data Owner/Data Provider specifies exactly which data will be transferred. If the Data Owner/Data Provider indicates that the data are not directly accessible for use within the projects, the Data Owner/Data Provider will be asked for agreement when project partners demand access to the data to meet the objective and deliverable of 3DGEO-EU.

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.

There is no requirement to delay the release of produced data. Technical delays are possible during the duration of the project.

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 project shall be useable by third parties. If this for some reason will not be the case for some data produced during the project duration we will provide explanation/justification about the specific licensing in a later updated version of the Project Data Management Plan.

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

Data provided by 3DGEO-EU should in principle be reusable as long as the EGDI central database(s) exists.

2.2.4.5 Are data quality assurance processes described?

Quality assurance processes will be described in good practice guidances and protocols in different 3DGEO-EU deliverables.

2.3 Allocation of resources

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

One of the aims of 3DGEO-EU consists in the integration of the produced data into EGDI which is managed by the GeoERA Information Platform project (GIP-P). The data repository shall be maintained after the project to meet the requirements of good scientific practice. A strategy for storage of the files after the project is being developed by GIP-P. The related costs are a matter of GIP-P.

Data which will not be integrated in the Information Platform are stored at the proprietary databases of the national/regional GSO in charge. Its long term upkeep is secured by national/regional obligations.

2.3.2 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 long term maintenance of the central EGDI platform are under consideration (discussions with EGS).

2.4 Data security

2.4.1 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 ensured 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, 3DGEO-EU will document the data security issues when the complete list of provided or stored data will be available. Therefore, the DMP will be updated according to it.

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

See 2.4.1.

2.5 Other issues

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

The underlying data from the GSOs or partners will in many cases be governed by national/regional or institutional rules. This will be described in a later updated version of the DMP.