



# Geological Analysis and Resource Assessment of selected Hydrocarbon systems

## Deliverable

### GARAH Data Management Plan (DMP)

Authors and affiliation:

**Name(s)**

**Peter Britze, GEUS**

**Niels Schovsbo, GEUS**

**Karen L. Anthonsen, GEUS**

**Stefan Ladage, BGR**

E-mail of lead author:

**Email**

**pbr@geus.dk**

Version: 10-12-2018

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.



<b>Deliverable Data</b>		
<b>Deliverable number</b>	D1.2	
<b>Dissemination level</b>	Public	
<b>Deliverable name</b>	Title GARAH Data Management Plan (DMP)	
<b>Work package</b>	WP1, Project Coordination	
<b>Lead WP/Deliverable beneficiary</b>	GeoERA	
<b>Deliverable status</b>		
<b>Submitted (Author(s))</b>	12/12/2018	Peter Britze, GEUS Niels Schovsbo, GEUS Karen L. Anthonsen, GEUS Stefan Ladage, BGR
<b>Verified (WP leader)</b>	14/12/2018	Peter Britze, GEUS
<b>Approved (Coordinator)</b>	17/12/2018	Peter Britze, GEUS



## LIST OF CONTENTS

1	AIM AND SCOPE OF THE DOCUMENT .....	4
2	INTRODUCTION .....	5
3	DATA MANAGEMENT PLAN.....	6
3.1	Data summary .....	6
3.1.1	What is the purpose of the data collection/generation and its relation to the objectives of the programme? .....	6
3.1.2	What types and formats of data will the programme generate/collect? .	6
3.1.3	Will you re-use any existing data and how? .....	7
3.1.4	What is the origin of the data?.....	7
3.1.5	What is the expected size of the data?.....	7
3.1.6	3. To whom might it be useful ('data utility')? .....	7
3.2	Fair data .....	7
3.2.1	Making data findable, inclusion provisions for metadata.....	7
3.2.2	Making data openly accessible.....	8
3.2.3	Making data interoperable.....	9
3.2.4	Increase data re-use (through clarifying licences) .....	10
3.3	Allocation of resources .....	10
3.3.1	What are the costs for making data FAIR in your project? .....	10
3.3.2	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)?.....	11
3.4	Data security .....	11
3.4.1	What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? .....	11
3.4.2	Is the data safely stored in certified repositories for long term preservation and curation?.....	11
3.5	Other issues .....	11
3.5.1	Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones? .....	11



## 1 AIM AND SCOPE OF THE DOCUMENT

The present document is the deliverable “Project **Data Management Plan** ” of the GARAH project, addressing the topic “Geological Analysis and Resource Assessment of selected Hydrocarbon systems” within the SRT GeoEnergy – GE1-Fossil energy, energy security and climate action.

The GARAH project aims to generate a harmonized, scientific based, geological analysis and assessment of conventional and unconventional hydrocarbon resources, and delivers key geoscientific subsurface data to stakeholders via a user-friendly web based information platform. This will help member states to continue the transition to lower carbon energy sources and will contribute to climate commitments, and allow the planning for secure sources of affordable energy.

Outcomes from the GARAH study will be the geological analysis and resource assessment of selected petroleum systems focussed on the North Sea, together with a pan-European view on gas hydrate assessment in the European continental margin. The study will also evaluate the multiple-use of hydrocarbon reservoirs, as integrated or alternative use of the subsurface, together with an appraisal on risks and safety.

This is the first version of the project Data Management Plan (DMP). It contains preliminary information about the data and metadata that the research project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The purpose of the DMP is to provide an analysis of the main elements of the data policy that will be used by the GARAH consortium with regard to all the datasets that will be generated by the project. The DMP is not a fixed document but will evolve during the lifespan of the project.



## 2 INTRODUCTION

This document is the GARAH data management plan (DMP). The DMP describes the data management life cycle for all datasets to be collected, analysed and/or generated by the partners. It defines, among other things: the handling of research data during and after the project period, the type of data that will be collected and analysed and what methodology and standards will be applied. The question of Intellectual Property Rights, ownership, diffusion and long term storage is also taken into consideration.

The purpose of the DMP deliverable is to provide relevant information concerning the data that will be collected and used by the partners of the GARAH project. The project aims to provide resource assessments of HC resources in the North Sea, together with an evaluation of the Europe's gas hydrates. Alternative use and hazards exploring and producing HC is included in the project. In large, the analytical geological data and products will be a part of the EGD compatible GeoERA geological data infrastructure. ~~(EGDI)~~. In order to achieve this, metadata and webgis platforms for harvesting, converting and sharing data between partners, stakeholders, and general public has to be implemented.

This goal entails the need for good documentation and implementation of standards, thesaurus, infrastructure, privacy settings and interoperability of data formats. The strength of having standards is that it should be relatively simple to upgrade with new specifications and implement conversion, distribution and privacy strategies.

The DMP will be refined in subsequent revisions of the present deliverables. This first version of the DMP mainly depicts what the project currently (M6) expects of the direction regarding the collection of the data.



## 3 DATA MANAGEMENT PLAN

### 3.1 Data summary

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

*The purpose is to collect and harmonize all geo-technical data aspects associated with the assessment of HC energy.*

*The assessed project data will be distributed in the following ways:*

- *GIS based datasets prepared for the assessment of the HC resources (maps, databases and geological 3D data).*
- *Data and reports without a spatial reference*

***Please note: due to possible limitations in data privacy rules, we will only publish processed output datasets (partially derived from unpublished input data) and analytic datasets generated during the GARAH project!***

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

***Type:*** *In general, all geological and subsurface management data related to a scientifically robust estimation of HC resources. In details the type of data is expected to consist of:*

##### ***North Sea data:***

*These data will mainly follow the datatypes generated in the EUOGA project e.g.: Boreholes, wells, outlines of formations, temperature maps, basin outlines, bathymetry, geothermal gradients, seafloor temperature, horizon and fault interpretations (linked to 3dGEO-EU)*

*In the pilot study area same functionality as in 3DGEO-EU is needed.*

##### ***Gas Hydrates data:***

*This will follow the structure and datatypes made in the MIGRATE project, e.g.: seafloor T heat flow, sedimentation rates in 4D,, gas hydrates below seafloor, gas stability map and faults.*

##### ***Format:***

*The format highly depends on the data itself but will be harmonized as much as feasible among the partners. It will also respect the common usage among the scientific community depending on the data type. Formats used include, Comma Separated values (CSV), Excel files, Shapefile & vectors, Grid & raster. A dedicated webgis platform is developed by the GeoERA Information Platform Project to support the provision of the results toward the public. Webservices such as*



---

*WMS, WMFS, Downloads services and metadata catalogue will be set up to assure the discovery and visualization of the data.*

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

*Yes, it is encouraged to make existing data available for research. GARAH will provide data templates and lists of pre-defined values, in order to be able to harmonize the different datasets that will be provided by the partners. On the other hand, the project will rely on data currently residing in existing data repository at national or regional level among the partners or publicly available.*

### **3.1.4 What is the origin of the data?**

*Data originates from national and regional data information repositories. GARAH will also collect new data on the study of gas hydrates.*

### **3.1.5 What is the expected size of the data?**

*As it is the first DMP at the start of the project the size is hard to predict. It is presumed that for documents storage space will be in the order of 10 GB or less. This will be defined in a further phase of the project and the DMP will be updated according to it.*

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

*GARAH consortium; European Commission services and European Agencies, Geological Surveys outside GARAH, Stakeholders and the general public in the addressed pilot areas, Research community inside and outside the pilot areas, third parties, consultancy agencies, software developers, regulators, policy makers.*

## **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)?*

*All published data must be discoverable with metadata. Many will also be identifiable and locatable by means of standard identification mechanisms.. Digital Object Identifiers (DOIs) will be generated in the central system to permanently and unambiguously identifies the data to which it is associated*

*3.2.1.2 What naming conventions do you follow?*

*Naming conventions have not yet been defined. There will be defined by the GIP-P after M6. The DMP will be updated according to it.*



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.*

*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.*

### **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.*

*Data provided to the GeoERA Information Platform for public distribution are open access.*

***due to possible limitations in data privacy rules, we will only publish processed output datasets (partially derived from unpublished input data) and analytic datasets generated during the GARAH project!***

***As the data collection is in progress and certain restrictions might apply to these datasets, the DMP will be updated accordingly.***

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

*Yes, the majority of the analytical data produced by GARAH will be made available in the IP. Considering the facts that, this is the first DMP and the GEOERA-IP is also under development during the project duration the direct accessibility of the data might be restricted by technical issues. For instance, the 3D data are not yet implemented in the IP.*

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

*documentation about the software needed to access the data included?*

*The data that will be produced by GARAH will follow the common good scientific practice in the field of HC energy resources assessments. Therefore, it does not require specific commercial software to access the data. Furthermore, the produced datasets will be made available on the GeoERA information platform developed by the GIP-P. IP data and metadata are available via any web browser application. As webservices such as WMS, WFS will also be produce as well as shapefiles, grids, csv to publish the datasets common GIS software are needed to visualise the data. The project will also produce word documents, excel sheets and binary data.*

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

GARAH does not require any specific software





*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.*

*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.*

### **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)?*

*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.*

*In GARAH, we will provide derived output datasets. The shown parameter themselves as well as the used physical units, parameter descriptions and metadata descriptions will follow harmonized standards.*

*3.2.3.2 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.*

*Furthermore, we would like to install project vocabularies in GARAH, which might be linked to a technical Glossary of Terms and to a knowledge repository. Our projects vocabularies will focus on technical terms used to describe resources, limitations and concepts for the use of HC energy. We will not focus on geological terms (e.g. formation names or rock types) in our project vocabulary but would like to link GARAH to such vocabularies already existing, if possible.*

*3.2.3.3 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. GARAH will introduce new vocabulary concerning technical terms and with the help of*



*the GIP-P WP3 and WP4 will try to map/link them into the commonly used standards vocabularies.*

### **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?**

*During the GeoERA projects period of time, the WP9 from the GIP-P will provide assistance and harvest for each projects the IPR of the produced datasets.*

*Nevertheless, for existing data, not generated with GeoERA co-fund, the Data Owner/Data Provider specifies exactly which data will be transferred. In case 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 agreement when projects members demand access to the data to meet the objective and deliverable of GARAH.*

*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.*

*There is no requirement to delay making the data available. Technical delays are possible during the duration of the project.*

*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.*

*The general rule will be that data produced in the projects shall be useable by third parties. 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.*

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

*For GARAH as the assessment HC map is highly dependent of new data, the long term reusability is variable. Five to ten years seems reasonable.*

## **3.3 Allocation of resources**

### **3.3.1 What are the costs for making data FAIR in your project?**

*One of the aims of GARAH consists in the integration of the produced data into the GeoERA IP which is managed by GIP-P. The data and metadata repository shall be stored after the project to meet the requirements of good scientific practice. A strategy for storage of the files after the project is being developed and will be included in an updated DMP. The project and the GIP-P will take care the data and supplementary materials have sufficiently rich metadata and a unique*



and persistent identifier to make them Findable. The Accessibility and Interoperability are assured by the fact that all the data and metadata will be stored according to the standards in the central database of GeoERA IP and Micka metadata catalog. Licensing and traceability issues of the data are also taken into consideration. Overall, the project will benefit from being included in the GeoERA framework for which a dedicated project concerning Information Technology has been designed. The centralisation of the storage and the publication of the data and metadata reduces the global cost.

**3.3.2 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 maintenance of the central information 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, will be documented in a further version of the DMP when the complete list of data provided or stored will be available.

**3.4 Data security**

**3.4.1 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 IP 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 GARA will document the data security issues when the complete list of data provided or stored will be available. Therefore, the DMP will be updated according to it.

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

See above

**3.5 Other issues**

**3.5.1 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.