Geological Analysis and Resource Assessment of selected Hydrocarbon systems

Deliverable 4.2

Description of the work done on EGDI, guidelines for uploading, updating and consulting information.

Authors and affiliation:
Uffe Larsen
GEUS – Geological Survey of Denmark and Greenland

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

Version: 7-12-2020

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	D4.2
Dissemination level	Public
Deliverable name	GARAH D.4.2: Description of the work done on EGDI, guidelines for uploading, updating and consulting information.
Work package	WP4, Knowledge database
Lead WP/Deliverable beneficiary	Uffe Larsen/GEUS

Deliverable status		
Submitted (Author(s))	7/12/2020	Uffe Larsen
Verified (WP leader)	7/12/2020	Uffe Larsen
Approved (Coordinator)	7/12/2020	Peter Britze

GENERAL INTRODUCTION

The aim of the Geological Analysis and Resource Assessment of selected Hydrocarbon systems (GARAH) is to develop a harmonized, scientific based, geological analysis and assessment conventional and unconventional hydrocarbon resources that will help member states to continue the transition to lower Carbon energy sources. This will contribute to climate commitments, and allow the planning for secure sources of affordable energy. The analysis and assessment of hydrocarbons will focus on two areas:

(i) in Europe's major petroleum province – the North Sea and include a "Geological analysis and resource assessment of North Sea petroleum systems", This research includes the assessment of conventional and unconventional oil and gas resources in the most important hydrocarbon basin in Europe. This will enable the remaining resource to be better understood and managed, and identify options for multiple and alternative uses of the subsurface as producing fields come off-line.

(ii) with a pan-European view, "Hydrate assessment in the European continental margin and related risks".
The assessment of gas-hydrates resources in the European continental margin represents an information gap of pan-European interest. This will improve the understanding of the potential role that gas-hydrates may play in the future EU energy mix, as it will constitute a base-line for future projects pertaining the improvement of the European model of the GHSZ, related hazards and potential for geological storage of CO_2.

A catalogue evaluating the multiple-use of hydrocarbon reservoirs, as integrated or alternative use of the subsurface, together with an appraisal on risks and safety, will be produced.

Data produced in the GARAH project (Geological Analysis and Resource Assessment of selected Hydrocarbon systems) must be disseminated using the infrastructure provided and developed by the European Geological Data Infrastructure (EGDI) and GeoERA GIP-P (the Geographic Information Platform Project).

The former report*, Preliminary data selection to provide relevant information in assessing hydrocarbon resources in subsurface (deliverable 4.1),* described the datasets, that were expected to be delivered at the start of the project. The report (deliverable 4.1) also listed the preliminary recommendations by GIP-P. Since then both the number of datasets, the structure of the datasets and recommendations by GIP-P have changed. Further changes must be anticipated in the future. This report describes the present status.

EXECUTIVE REPORT SUMMARY

Some key points from the GeoERA project GIP-P (the Geographic Information Platform Project) on how to upload and update datasets to the EGDI platform is described shortly.

An example of how a dataset is processed before it is uploaded to the EGDI platform is described in detail.

At last table with the current statuses for the GARAH datasets is given.

TABLE OF CONTENTS

1 GUIDELINES FOR UPLOAD AND UPDATING AND CONSULTING INFORMATION

It is GIP-P (the Geographic Information Platform Project) that specifies how data are uploaded and updated. They are also responsible for the portal where data can be accessed by the public. Some information from GIP-P is described shortly here. More information can be found on the GIP-P website (https://geoera.eu/projects/gip-p/). How the GARAH data will be presented on the EGDI portal is described in detail in the report that is deliverable 4.3.

The work of the GIP project takes place simultaneously with the work in the GARAH project. This means that not all upload facilities are ready yet. The GIP project has provided an administration module for uploading and configurations of maps and map layers, https://egditest01.geus.dk/egdiadmin/.

From the administration module it is possible to upload unstructured data as PDF, pictures and csv files. Spatial data are uploaded as GeoPackage or shape files. Uploads of raster data (GeoTIFF and NETCDF) and 3D models are planned but not implemented yet. For best performance the spatial reference system, SRID = 3034, is recommended.

Before data are uploaded, metadata must be added to the EGDI Metadata Catalogue (MIcKA).

It is recommended to use existing standards and code lists. In Work package 4 of the GIP project (Semantic harmonization issues) takes care of new vocabularies and code lists. New vocabularies are hosted at github. New code lists and extensions to existing ones are hosted at the European Geoscience Registry (https://data.geoscience.earth/ncl/geoera)

2 DESCRIPTION OF THE WORK DONE

The primary role of WP4(Knowledge database) is to manage the interaction between the GARAH work packages and other GeoERA projects and especially with GIP-P (the Geographic Information Platform Project). WP4 also takes care of the uploading of dataset to the EGDI portal, configuration of maps and adding extra functionality to the EGDI portal.

The present status of the GARAH project is that WP3 (Assessment of gas hydrates) has prepared complete datasets that can be uploaded to the portal. WP2 (North Sea Petroleum Systems) has not produced any datasets, which are ready for upload, yet, but - according to schedule – this is not expected either. It has shown to be rather difficult to produce harmonized maps that cover the whole of the North Sea. The main reason for this is that petroleum systems in the SW part of the North Sea are in other stratigraphic intervals than the petroleum systems in the NE part of the North Sea.

2.1 Standardization

When data are formatted, international standards will be followed where it makes sense. Many standards, for example the INSPIRE standard for geology, specifies complex xml structures, that are not supported by many GIS tools. To be able to support as many clients as possible the datasets in the GARAH project uses a flat data structure. Thus, the web feature services are not fully compatible with the INSPIRE standard.

For most of the datasets in the GARAH no obvious international standard exists. In theory one could use parts of other xml standards where the attributes resemble the attributes in GARAH. Unfortunately, picking of single attributes from a big selection of remotely related standards is not advisable either, since this makes the data definition very complicated and will deteriorate performance when data are parsed, reformatted, and validated against a multitude of XML schemas.

For the majority (probably all) of the datasets in GARAH project it will not be possible to validate the WFS xml against external XML schemas. Instead a compromise is used. The content of code lists will refer to international standards where possible. Also, the name of attributes will follow the naming used in international standards as much as possible.

2.2 Dataflow

An example of how data are manipulated/processed before data are uploaded to the EGDI test portal is given below for the layer, *Evidence and indicators of gas hydrates (point data)*.

The gas hydrate dataset originates from an ArcGIS project where the vector data were stored as shape files. The original structure of the file is shown below

Table 1. Definition of the dataset, GasHydrate_Site_Evidences&Indicators: Points as specified in deliverable GARAH.D.3.2: Hydrates GIS-dataset

FIELDNAME	FORMAT	DESCRIPCION
ID_IndiNa	Text, 254	Identification code of the evidence – PK
Lat_DD	Double14,6	Latitude in decimal degrees (WGS84)
Long_DD	Double14,6	Longitude in decimal degrees (WGS84)
WaterDepth	Double14,6	Seafloor depth (meters)
GeoSettin	Text, 50	Geographical / Geological Setting – Constrains NN, LV
LocalSite	Text, 50	Local site where the evidence is located
Data_Sourc	Text, 254	Institution/Company if Owner of Data. Project, database or publication where data have been collected
Cruise	Text, 50	Oceanographic Cruise where data have been recovered or observed
CName	Text, 50	Contact name
Email	Text, 50	Contact Email

FF_Type	Text, 50	Type of evidence o indicator – constrains: NN-LV_FF_Type
Descripti	Text, 254	Description of the evidence - free text
Sedi_Type	Text, 50	Sediment type – LV_
D_Indi_mtp	Double 10,4	Depth of the top of the evidence below seabed in meters
D_Indi_mbt	Double 10,4	Depth of the bottom of the evidence below seabed in meters
DOI	Text, 254	DOI of main data publication
Reference	Text, 254	References to data. Author, Year & Title. Link to PDF in data repository
Comments	Text, 254	Comments -- Free text

This data structure is changed in following ways

- The columns Sedi_Type and FF_Type now refer to external code lists.
- The data in the columns, Data_Source, CName and Email are metadata that now are stored in the MIcKA system.
- There is a many-to many between bibliographic references and the features. Also, features from different layers refer to the same bibliographic reference. For this reason, data are normalized and the information in DOI and Reference is put into another table.
- The columns Lat_DD and Long_DD are omitted because this information is already stored in the geometric object
- At last columns are renamed to give them more user-friendly names.

To be able to make these changes efficiently the shapefile was at first imported into a postgresql database using the QGIS database module.

2.3 Code lists

The column Sedi_Type refers to the lithology codelist in INSPIRE. Unfortunately, one value was missing, mud-breccia, so the lithology code list had to be extended. The extension is found here:
https://data.geoscience.earth/ncl/geoera/graph/lithology

FF_Type refers to a completely new code list, Fluid Flow Indicator, that is found here: https://data.geoscience.earth/ncl/geoera/graph/FluidFlowIndicator

2.4 Metadata

Data_Source, CName and Email refer to the name of the institution that have produced the data, contact name and contact email respectively. The persons in question have been asked if they allow that their contact information is visible on a public website.

For most dataset only one entry in the EGDI Metadata Catalogue (MIcKA) is needed but in this case the parts of the dataset with different contact information can be attributed to different metadatasets that all belong to the same parent metadataset. MIcKA, which is an implementation of ISO 19115, supports this parent-child relationship.

2.5 Bibliographic references

Bibliographic references were at first registered in the program, Endnote, to ensure a consistent formatting of the references. Endnote is able to export bibliographic references in many different formats, so the format can easily be changed if this is needed.

The formatted references were inserted into the database table, bibref, that have this structure

Table 2. Description of the database table, bibref.

column name	data type	description
bibref_id	integer	Internal database id
in_text_citation	varchar(254)	Citation in text. For example: Bourry et al., 2009
bibreference	text	The full bibliographic reference
doi	varchar(254)	Link to the article using DOI (the Digital Object Identifier)

The many-to-many relation between features and bibliographic references is resolved with the table feature_bibref

Table 3. Description of the database table, feature_bibref.

Column name	Data type	Description
feature_type	varchar(100)	name of the feature table
fid	integer	database id of the feature
bibref_id	integer	reference to the bibliographic reference in the table bibref
note	varchar(254)	examples are Figure 5, table II, etc

The tables bibref and feature bibref are imported into the postgresql database that are used by the EGDI portal. A REST webservice, which has been deployed in the EGDI environment, returns json that describes all bibliographic references for a specified feature and this result is shown on the EGDI portal when requested.

A more detailed description of how data is presented on the EGDI portal is described in the report deliverable 4.3

In this example the resulting new structure for the original dataset is show in Table 4. Where appropriate the other datasets in the GARAH project are treated in a similar way.

Table 4. Description of the database table, gas_hydrate_geology_point, including the mapping to the original dataset

Column name	Data type	The original column name	Description
fid	Integer		database id of the feature
geom	the geometry object		
originalidentifierid	varchar(254)	ID_IndiNa	The original identifier of the feature
waterdepth	double precision	WaterDepth	water depth in meters below sea level
geographicalarea	varchar(50)	GeoSettin	the name main geographical area where the feature is situated
localsite	varchar(50)	LocalSite	the name of the specific geographical site
metadataurl	varchar(254)	Data_Sourc, CName, Email	reference to MIcKA
cruise	varchar(254)	Cruise	Name of cruise
indicatortype	varchar(50)	FF_Type	A human readable name for the type of indicator or evidence of gashydrates.
indicatortypeurl	varchar(254)	FF_Type	The identifier of the indicator/evidence as defined in the external code list
description	varchar(254)	Descripti	description
topdepth	double precision	D_Indi_mtp	Depth to the top of the evidence below seabed in meters
bottomdepth	double precision	D_Indi_mbp	Depth to the bottom of the evidence below seabed in meters
lithology	varchar(50)	Sedi_Type	The lithology. For example clay or mud-breccia. This column is added because lithologyurl only contains a numeric identifier and not the actual name of the lithology

lithologyurl	varchar(254)	Sedi_Type	The identifier of the lithology in the external code list
comments	varchar(254)	Comments	comments

3 STATUS OF DATASETS IN THE GARAH PROJECT

The following tables show the current status of the different datasets in the GARAH project at the start of December 2020.

Table 5. Status for the data tables in the GARAH project. December 2020

dataset and workpackage	**Description**	**submitted files**	**expected dates (or time intervals) for delivery test data**
GasHydrate_Site_Evidences&Indicators, WP3	GasHydrate Site Evidences & Indicators shown as points, Geopackage format. Contains three tables: The table with spatial data, a table with bibliographic references and a many-to-many table that shows the relation between points and references. There might be another table with chemical analyses but it is not decided yet whether the chemical analyses should be delivered in an separat documents	one geopackage file	Delivery of test data: July 7. 2020. Feeback expected September 1. - September 15., 2020. Second test data deliverered 1 September 15. - October 1.
GasHydrate_Areal_Evidences. WP3	GasHydrate Areal Evidences presented as polygons. Delivered in geopackage format. Also contains af many-to-many relation to bibliographic references	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020 (done). 2. Test: October 15. - November 15. , 2020
GasHydrate_Local_Geophy_Indicators. WP3	geophysical indicators shown as points. Delivered in geopackage format. Contains af many-to-many relation to bibliographic references	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020. (done) 2. Test: October 15. - November 15. , 2020

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
GasHydrate_Profile_Geophy_Indicators, WP3	GasHydrate Profile Geophyical Indicators. Lines. Delivered in geopackage format. Contains af many-to-many relation to bibliographic references	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020. (done) 2. Test: October 15. - November 15. , 2020
GasHydrate_Areal_Geophy_Indicators, WP3	GasHydrate Areal Geophysical Indicators. Polygons. Delivered in geopackage format. Contains af many-to-many relation to bibliographic references	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020. (done) 2. Test: October 15. - November 15. , 2020
FluidFlow_Seafloor_Point_Features, WP3	FluidFlow Seafloor Areal Features. Points. Delivered in geopackage format. Contains af many-to-many relation to bibliographic references	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020. (done) 2. Test: October 15. - November 15. , 2020
FluidFlow_Seafloor_Areal_Features, WP3	FluidFlow Seafloor Areal Features. Polygons. Delivered in geopackage format. Contains af many-to-many relation to bibliographic references	one geopackage file	1. Test: Data delivered September 1. -

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
			October 1., 2020. (done) 2. Test: October 15. - November 15. , 2020
HeatFlow_Global, WP3	HeatFlow Global. Points. Delivered in geopackage format. Downloaded from The Global Heat Flow Database	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020.(done) 2. Test: October 15. - November 15. , 2020
Seafloor Temperature, WP3	Seafloor Temperature: Points. Delivered in geopackage format	one geopackage file	1. Test: Data delivered September 1. - October 1., 2020. (done) 2. Test: October 15. - November 15. , 2020
Exploration wells, WP2	Exploration wells. points. Probably delivered in geopackage format	It is not decided yet whether data are delivered as a	1. test: December 2020. 2. test. January 2021

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
		geopackage or whether we will use an existing service from one of the other projects in EGDI or EMODNET	
Hydrocarbon fields, WP2	Hydrocarbon fields. Polygons. oil, gas, condensate. Delivered in geopackage format	one geopackage file	1. test: December 2020. 2. test. January 2021
Project areas, WP2	A least one polygon showing the outline of the study area of conventional resources in the North Sea	one geopackage file	1. test: February 2021. 2. test: March 2021
HC Plays in DCG, WP2	HC Plays in DCG. Polygons. newly digitalised from GEU Assessment. Delivered in geopackage format.	one geopackage file	1. test: February 2021. 2. test: March 2021
Structural outlines, WP2	Basin outlines . Polygons. Delivered in geopackage format	one geopackage file	1. test: December 2020. 2. test. January 2021
Faults, WP2	Faults . Lines. Delivered in geopackage format	one geopackage file	1. test: February 2021. 2. test: March 2021

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
Salt diapirs outlines, WP2	Salt diapirs outlines. Delivered in geopackage format	one geopackage file	1. Test: December 2020. 2.Test. January 2021
TOC map, unconventional resources, WP2	TOC map. Polygons. Unconventional resources. Delivered in geopackage format	one geopackage file	1. test: February 2021. 2. test: March 2021
Maturity map, unconventional resources, WP2	Maturity map. Polygons. Unconventional resources. Delivered in geopackage format	one geopackage file	1. test: February 2021. 2. test: March 2021
Depth maps, unconventional resources, WP2	Depth maps. Polygons. Unconventional resources. Delivered in geopackage format	one geopackage file	1. test: February 2021. 2. test: March 2021
Thickness maps. Unconventional resourses, WP2	Thickness maps. Unconventional resourses. Polygons. Delivered in geopackage format	one geopackage file	1. test: February 2021. 2. test: March 2021
Identified CO2 storage sites, CSS, WP2	Identified CO2 storage sites, CSS. Polygons.	one geopackage file	1. test: March 2021. 2. test: April 2021
3D models, WP2	3D Volumes. Same type of 3D volumes as used in 3DGEOEU	?	1. test: March 2021. 2. test: April 2021
2D Horizon interpretations, WP2	The format is not decided yet	?	1. test: March 2021. 2. test: April 2021
3D surfaces, WP2	2.5D Grids. Problably deliverted in geopackage format	?	1. test: March 2021. 2. test: April 2021

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
Chemical analyses of gases. WP3.	Chemical analyses of gases from IODP drilling holes. Point Data	one geopackage file	1. Test: Data delivered October 15. - November 1., 2020. 2. Test: November 1. - December 1. , 2020
Base of hydrate stability zone for biogenic gas. NW Europe, WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of hydrate stability zone for biogenic gas. SW Europe. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of hydrate stability zone for 100% CO2. Celtic Sea & French EEZ. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of hydrate stability zone for 96% CO2. Celtic Sea & French EEZ. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
Base of negative bouyancy zone for 100% C02. Celtic Sea & French EEZ. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of negative bouyancy zone for 96% C02. Celtic Sea & French EEZ. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of hydrate stability zone for biogenic gas. From Piñero et al. 2013. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of hydrate stability zone for 96% CO2. Extended 200M in the FISU Area, Celtic Sea. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of hydrate stability zone for 96% CO2. South of Biscay Bay, Galicia Area. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP

dataset and workpackage	Description	submitted files	expected dates (or time intervals) for delivery test data
Base of negative bouyancy zone for 96% C02. Extended 200M in the FISU Area, Celtic Sea. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP
Base of negative bouyancy zone for 96% C02. South of Biscay Bay, Galicia Area. WP3	Gas Hydrate Stability. Raster data	Format not decided by GIP yet. No upload facilities	?. Depends on GIP

Table 6. External services that are used by the GARAH project, December 2020

Dataset (please, follow terminology used in GIP-P D2.2.2)	Service url and description of the data included in it.
Areas reserved/used for windmill	Areas with windmills. From EMODNET. https://ows.emodnet-humanactivities.eu/wfs?SERVICE=WFS&VERSION=1.1.0&request=GetFeature&typeName=emodnet:windfarmspoly&OUTPUTFORMAT=json
Fishing activities	Fishing activites From EMODNET. Several layes. This is an example. https://ows.emodnet-humanactivities.eu/wfs?SERVICE=WFS&VERSION=1.1.0&request=GetFeature&typeName=emodnet:fishingbeamtrawls&OUTPUTFORMAT=json
Licences	Licenses from EMODNET. Polygons. https://ows.emodnet-humanactivities.eu/wfs?SERVICE=WFS&VERSION=1.1.0&request=GetFeature&typeName=emodnet:activelicenses&OUTPUTFORMAT=json
Coastline + territorial boundary	this must be provided by GIP

Dataset (please, follow terminology used in GIP-P D2.2.2)	Service url and description of the data included in it.
Pipes and installations	Pipelines from EMODNET. https://ows.emodnet-humanactivities.eu/wfs?SERVICE=WFS&VERSION=1.1.0&request=GetFeature&typeName=emodnet:pipelines&OUTPUTFORMAT=json
Sediment_Thickness	https://gis.ngdc.noaa.gov/arcgis/services/web_mercator/sediment_thickness/MapServer/WMSServer
Sedimentation rates	https://drive.emodnet-geology.eu/geoserver/gtk/wms?service=WMS&version=1.1.0&request=GetMap&layers=gtk:seabed_accumulation_rates&styles=&bbox=-30.8698431209999,23.68047218,68.0297911930001,81.852091617&width=768&height=451&srs=EPSG:4326&format=image%2Fpng