



Project acronym and title:  
SECURE – Subsurface Evaluation of Carbon capture  
and storage and Unconventional risks

## SECURE DATA MANAGEMENT PLAN

Authors and affiliation:  
**M Mowat**

British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK

Email of lead author:  
mmow@bgs.ac.uk

D7.8  
Revision:2

### Disclaimer

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 764531.

The content of this report reflects only the authors' view. The *Innovation and Networks Executive Agency (INEA)* is not responsible for any use that may be made of the information it contains.



**Project funded by the European Commission within the Horizon 2020 Programme**

**Dissemination Level**

<b>PU</b>	Public	X
<b>CO</b>	Confidential, only for members of the consortium (incl. the Commission Services)	
<b>CL</b>	Classified, as referred to in Commission decision 2001/844/EC	

<b>Deliverable number:</b>	7.8
<b>Deliverable name:</b>	SECURE Data Management Plan
<b>Work package:</b>	WP7 [MANAGEMENT]
<b>Lead WP/deliverable beneficiary:</b>	UKRI-BGS

Status of deliverable		
	By	Date
Submitted (Author(s))	M Mowat	24/07/2018
Verified (WP leader)	JM Pearce	27/07/2018
Approved (EB member)	DI Schofield	26/07/2018
Approved (Coordinator)	JM Pearce (pp E Hough)	27/07/2018

Author(s)		
Name	Organisation	E-mail
M Mowat	UKRI-BGS	mmow@bgs.ac.uk



## Public introduction

Subsurface Evaluation of CCS and Unconventional Risks (SECURE) is gathering unbiased, impartial scientific evidence for risk mitigation and monitoring for environmental protection to underpin subsurface geoenergy development. The main outputs of SECURE comprise recommendations for best practice for unconventional hydrocarbon production and geological CO<sub>2</sub> storage. The project is funded from June 2018–May 2021.

The project is developing monitoring and mitigation strategies for the full geoenergy project lifecycle; by assessing plausible hazards and monitoring associated environmental risks. This is achieved through a program of experimental research and advanced technology development that includes demonstration at commercial and research facilities to formulate best practice. We will meet stakeholder needs; from the design of monitoring and mitigation strategies relevant to operators and regulators, to developing communication strategies to provide a greater level of understanding of the potential impacts.

The SECURE partnership comprises major research and commercial organisations from countries that host shale gas and CCS industries at different stages of operation (from permitted to closed). We are forming a durable international partnership with non-European groups; providing international access to study sites, creating links between projects and increasing our collective capability through exchange of scientific staff.

## Executive report summary

*This deliverable comprises the Data Management Plan for SECURE. It provides a summary of data and information on making the data FAIR (findable, accessible, interoperable and re-usable).*

*SECURE is committed to open data access, long-term archiving and availability after the funding period of the project has finished.*



# Contents

<b>Public introduction</b> .....	<b>ii</b>
<b>Executive report summary</b> .....	<b>ii</b>
<b>Contents</b> .....	<b>iii</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>2 Data summary</b> .....	<b>1</b>
2.1 Summary of data types.....	2
2.2 Summary of datasets.....	1
2.3 Reuse of existing data .....	1
2.4 Data utility .....	1
<b>3 Fair data</b> .....	<b>1</b>
3.1 Making data findable, including provisions for metadata .....	1
3.2 Making data openly accessible.....	3
3.3 Making data interoperable .....	5
3.4 Increasing data re-use.....	6
<b>4 Allocation of resources</b> .....	<b>6</b>
<b>5 Data security</b> .....	<b>7</b>
5.1 Disaster planning.....	7
5.2 Persistent access to data .....	7
<b>6 Ethical aspects</b> .....	<b>7</b>
<b>7 Existing procedures for data management</b> .....	<b>8</b>
<b>Appendix 1 Data Management Questionnaire</b> .....	<b>9</b>
<b>Glossary</b> .....	<b>12</b>
<b>8 References</b> .....	<b>13</b>

## TABLES

Table 1: Summary of Data Types.....	2
Table 2: Summary of WP datasets (not complete).....	1
Table 3: Work Package 6 data tasks/deliverables .....	<b>Error! Bookmark not defined.</b>
Table 4: Resuse of existing data .....	1



# 1 Introduction

SECURE is committed to open data access, long-term archiving and availability after the funding period of the project has finished. The project is participating in the Pilot on Open Research Data in Horizon 2020, which aims to improve and maximise access to, and re-use of, research data generated by projects.

Partners are required to

1. deposit the data in a recognised research data repository.
2. as far as practicable, take measures to enable third parties to access, mine, exploit, reproduce and disseminate this research data.

Open data is data that is free to access, reuse, repurpose, and redistribute. The Open Research Data Pilot aims to make the research data generated by Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access. This Data Management Plan (DMP) defines certain datasets to remain closed according to the principle "as open as possible, as closed as necessary".

As part of making data findable, accessible, interoperable and re-usable (FAIR), this Data Management Plan includes information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access
- how data will be curated and preserved (including after the end of the project)

Overall data management for the project will be undertaken by BGS (UKRI) as the project co-ordinator with co-ordination of data generated by individual work packages resting with individual work package lead organisations. BGS (UKRI) has appointed a data manager for SECURE who has responsibility for coordinating and managing the collation and archiving of SECURE's data to ensure long-term data management complies with current best practice to allow continued data availability.

This Data Management Plan is an active document and will be updated over the course of the project as required. The data management plan will be discussed annually at General Assembly meetings, following which it will be revised as necessary to ensure it remains representative of the data management strategy for the project. The data manager may also attend work package meetings and Project Management Board meetings as necessary to discuss data management requirements with partners.

For data management support and assistance with archiving data, contact the SECURE data manager, Mary Mowat [secure.data@bgs.ac.uk](mailto:secure.data@bgs.ac.uk).

## 2 Data summary

The purpose of the data collection/generation in SECURE is to provide unbiased data, information and recommendations regarding the potential environmental impacts associated with geological carbon dioxide storage (CCS) and shale gas production, how these impacts are identified, monitored, mitigated and remediated.

Much of the data generated within work packages will be newly generated and describe subsurface and surface processes and impacts related to shale gas and CCS, and the understanding of those issues by stakeholder groups including the general public. Interpreted datasets will form the scientific evidence base from which recommendations for best practice will be made to the EU.



The data collected/generated by SECURE will be varied, with a range of data types and formats. These will include field and analytical data (including geochemical, geophysical and biological), data produced via computer simulations and data on participants in citizen science projects as part of SECURE's participatory monitoring.

Data management is imbedded within each work package. A Data Management Questionnaire (DMQ), Appendix 1, will be sent to WP leaders to gather more information on the data outputs. WP leaders can also forward to task and sub-task leaders as required. The DMQ is a data management planning tool to help identify data of long-term interest and also assists with data management requirements (e.g. expected high volume datasets). This should be completed with details of the expected data sets, formats, etc. The status of data should also be flagged as open access or confidential with any restrictions specified.

The overall size of the data generated by the project is unknown at this stage. Given previous experience with establishment and management of environmental datasets, we expect there to be a large amount of data, and it is expected to be of quite high volume.

A summary of data types is presented below. More information/detail can be added to the Data Summary as the project progresses and more information on specific data within the work packages is known. Sections can be added for each WP with specific WP datasets.

## 2.1 SUMMARY OF DATA TYPES

**Table 1: Summary of Data Types**

Data Types	Data Formats	Open/Restricted/ Confidential?	Embargo Period
Remote sensing data			
Georeferencing data	.shp, .mdb		
Geochemical data	.xlsx, .csv	Embargo until published (then open)	
Petrophysical data	.csv		
Gas monitoring data	.csv		
Microbiological data	.csv		
Geophysical data, seismic	segy, .dat		
Geophysical data: well logging	.las, csv		
Seismological data	SEED		
Seismological bulletins with results	Nordic		
Simulation results	VTK		
Interpreted datasets			
Numerical models (inputs/code)	FLAC3D DIANA/PFC Tough (react) BBN PHREEQC EQ3/6 MRST MOVE	Input and output data are open, codes will be restricted	
Software code necessary for validating results	Matlab C++ Python Matlab C-sharp Java-Eclipse	Can be shared with constortium if developed within the project, source code confidential if	



		developed outside the project	
Social science data collected from questionnaires, in-depth interviews and group sessions (anonymized before analyzing)	.pdf .xlsx	Confidential	
Teaching e-resources	Webpages, downloads of above datasets	open	
Public project deliverables	pdf, .xls, .jpg	open	
Other additional potentially useful outputs, such as posters, presentations	pdf, ppt		



## 2.2 SUMMARY OF DATASETS

There are a number of data related tasks from the various partners within different work packages which should be reviewed and may require data management support or coordination. A summary of these datasets are presented in table 2, with further details below.

**Table 2: Summary of datasets (not complete)**

WP/Task/ Deliverable	Dataset	Volume	Format	Access	Contact	Organisation	Archive	When
T2.3	Stress-permeability datasets	1GB	.xlsx	Open access	Andeas Busch	HWU	NGDC	Oct 2020
T2.3	3D models of fracture networks	10GB	Move, MRST	Open access	Andeas Busch	HWU	NGDC	Oct 2020
2.1.2	Laboratory and field test and measurements		.csv, .jpg, .pdf		Mirosław Wojnicki	INiG - PIB		
2.1.2, 4.2.3, 4.3.3	Existing out-of-date model of Borzecin site		.txt, Schlumberger code (Petrel, Eclipse and Visage)	Restricted. All the commercial software use their specific binary type input/output files as well as a general Rescue type. However, for an effective information exchange text type input/output files available for the commercial software as well as in-house software may be used.	Piotr Letkowski	INiG - PIB	INiG - PIB	
2.1.2, 4.2.3, 4.3.3	Production and injection data of Borzecin Site (including geological, geophysical, petrophysical, production, etc. data)		.txt, .xlsx	Restricted – All data regarding the Borzecin Site owned by the Polish Oil and Gas Company - PGNG are restricted. INiG has a restricted access to them. The exact status of data relating to Borzecin site will be determined after obtaining the official permission for the use of data from Polish Oil and Gas Company.	Piotr Letkowski	INiG - PIB	INiG - PIB	
2.1.2, 4.2.3, 4.3.3	Results of analyzes of reservoir fluids		.txt, .xlsx	Restricted	Piotr Letkowski	INiG - PIB	INiG - PIB	
T2.1.3	Bowtie risk framework	<20MB	MS Office	Open access	Matt Beeson	Risktec Solutions Ltd	NGDC	2020
WP2/5	Model software and codes (new model technology will be developed for coupled modelling and optimization rather than collecting new data)		FLAC3D, DIANA, TOUGH	Restricted. Numerical models and software codes developed within the project can be made available to the consortium if it is of interest to link/integrate modelling activities or exchange information. If developed outside of the project (prior knowledge) may be used by TNO within the project but use will be restricted, and source codes cannot be made available to the consortium or be made open access.	Jens Wollenweber	TNO	TNO Data Repository	





WP2/5	Input and output of model simulations		.doc, .pdf	Open Access if performed within project. The models will be applied to existing data available in the open domain, or data that is made available by partners in the SECURE consortium.	Jens Wollenweber	TNO	Publications	
WP2/5	Input or run files		txt, exe	Available to consortium	Jens Wollenweber	TNO		
WP2/5	Modelling results		.doc, .pdf	Open access	Jens Wollenweber	TNO	Reports, papers	
6.3	Social science data		.pdf, .xlsx	Restricted	Hanneke Puts	TNO		
D6.5	Training software and dataset					TNO		
4.1.4	Seismological data acquired at Stenlille		Seismology standard Seed format	Open Access	Trine Dahl-Jensen	GEUS	?	
4.1.4	Seismological bulletins		Nordic (see the SEISAN manual appendix A: <a href="http://seisan.info/">http://seisan.info/</a> ).	Open Access	Trine Dahl-Jensen	GEUS	?	
T3.4/D3.9	Integrated platform for multi-source multiscale sensor data - Simulation model code and associated data assimilation algorithms	<100MB	Matlab or C++	Open Access.  The objective of task 3.4 (deliverable 3.9) is to demonstrate how to efficiently incorporate heterogeneous data to subsurface flow models. This might include logging and production data as well as seismic and lab-data to better characterise transport and reaction properties. Starting from simple (e.g. homogeneous) best estimates for the parameters, a few different data sets will be assimilated to give calibrated model.  This will read a few geological and experimental/monitoring data (of heterogeneous type and different formats) made available by other partners (this will rely on what other partners will upload). This should be reviewed in month 24, once the code and assimilation algorithms are ready. This task and the proposed example will demonstrate how to appropriately convert and pre-process the available data to use it for calibration.	Matteo Icardi	UNOTT	meta-database at UNOTT	
T3.4/D3.9	Simulation results(prior and after the calibration)		VTK	?	Matteo Icardi	UNOTT		
D6.4	Online e-resources for training	< 200MB		Open Access	Bagus Muljadi	UNOTT		
4.1.2	Geochemical data linked to the well remediation aspects. Data are likely to be from lab experiments or analyses of synthetic or natural materials.		.xlsx, .csv	Open Access. Data are most likely to be generated by BGS during the project, or generated by BGS in the past, or from the open literature, so will be open file	Chris Rochelle	BGS	NGDC	



4.1.2	Geochemical models (there will probably not be much modelling)		PHREEQC or EQ3/6	?	Chris Rochelle	BGS	NGDC	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Operational datasets, 2008-2013 (injection history,p,T)	10 MB	Excel, .xlsx	Open	Knut Behrends	GFZ Potsdam	GFZ Archive	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Environmental Monitoring Datasets (CO2, Weather Data) from Ketzin	10 MB	.xlsx, .csv	(Open?). Needs to be clarified. Can be shared within project.	Knut Behrends	GFZ Potsdam	GFZ Archive	
	Reservoir modelling and simulation data at Ketzin			Confidential	Knut Behrends	GFZ Potsdam	NA	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Borehole Measurements (Geophysical Well Logs)	10 GB	.las, .lis	On Demand	Knut Behrends	GFZ Potsdam	GFZ Archive	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Video Data (Camera Traversal of Cased Holes)	10 GB	.mp4, .flv	On Demand	Knut Behrends	GFZ Potsdam	GFZ Archive	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Geological Field Data (Stratigraphy, core sample descriptions)	10 GB	.csv, .jpeg	On Demand	Knut Behrends	GFZ Potsdam	GFZ Archive	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Geophysical Monitoring Data	100 GB	.zip	Confidential	Knut Behrends	GFZ Potsdam	GFZ Archive	
3.1.3, 3.3.1, 3.5.1, 4.2.2, 6.3, 6.4.2	Documents and reports	1 GB	.docx, .pdf	Confidential	Knut Behrends	GFZ Potsdam	GFZ Archive	
D6.10	Educational talks			Open Access	Simon Shackley	UEDIN		
T6.4.1	International platform for environmental monitoring. The platform will include a data-sharing infrastructure (M10, Month 32).							



### 2.3 REUSE OF EXISTING DATA

Existing data will be brought in to use for the project within the various work packages. This will include data from project partners and third parties. Further work will be done to categorise the data which will be used in the project (i.e. new data, existing data from partners, external data from third parties).

**Table 3: Reuse of existing data**

WP/Task	Reuse of existing data	
<b>T1.3</b>	There is some overlap with the DETECT project, funded through the ERA NET program	HWU
<b>WP2</b>	We maybe acquire well data from (an operator) outside the consortium. This can include but is not limited to: well reports, well logs, well schematics geological and lithological data.	TNO
<b>WP2/4</b>	Simulation model, data concerning production and injection operations as well as results of analyzes of reservoir fluids from Borzecin site provided by Polish Oil and Gas Company.	INiG-PIB
<b>WP3</b>	Data from University of Calgary and Duke University.	

### 2.4 DATA UTILITY

The datasets from the project are useful to future collaboration and research and will be useful to the international CCS and shale gas communities to share best practice as widely as possible and to avoid duplication of effort in future. The data will also be useful to regulators in member states in order to establish and amend an appropriate regulatory regime to ensure the appropriate development of the subsurface with respect to Shale Gas and CCS technologies.

## 3 Fair data

In general, data from SECURE should be 'FAIR', that is findable, accessible, interoperable and re-usable. [Guidelines on FAIR Data Management in Horizon 2020](#) are available in the Online Manual.

More information about FAIR:

[FAIR data principles \(FORCE11 discussion forum\)](#)

[FAIR principles \(article in Nature\)](#)

### 3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

Data should be discoverable, with fully searchable metadata to inform prospective users of the data prepared to recognised data management standards and published in data repositories.

A listing of datasets will be maintained on the SECURE website with links to the data, giving a central source of information describing data associated with the project. This will be coordinated between the data manager and project dissemination team at the Scottish Carbon Capture and Storage (SCCS). Release of datasets will be made publically known through the project website and social media accounts.

#### 3.1.1 Metadata

Datasets generated will be INSPIRE compliant, with full metadata conforming to Directive 95/46/EC of the European Parliament and DOI (Digital Object Identifiers) where appropriate. Metadata containing details of the dataset will be captured in a standardised discovery metadata format which complies with ISO standard 19115.

The Research Data Alliance provides a [Metadata Standards Directory](#).



Partners completing metadata should ensure that this is of high quality enabling users in future to find a dataset and determine if they wish to use it. Metadata must include a good explanatory title and an accurate concise description (e.g. what, where, when, how, why, who).

Metadata examples: <http://www.bgs.ac.uk/services/NGDC/metadata/examples.html>.

### 3.1.2 Digital Object Identifiers (DOIs)

It is recommended that Digital Object Identifiers (DOIs) are applied to archived datasets where appropriate to enable citation of the information, particularly when data are referenced in a publication. This is a pre-requisite of leading science journals.

The DOI:

- allows data to be cited in the same manner as a scientific journal article
- enables credit to be assigned the dataset creators
- recognises the value of the data
- and the effort that has gone into its creation
- ensures the discoverability, permanence and stability of the dataset

For more information, see <http://www.bgs.ac.uk/services/ngdc/citedData>.

DOIs can be issued for datasets which are archived in the National Geoscience Data Centre (NGDC). Any other data repositories used should also offer this service.

A DOI can be assigned before the dataset is released so that it can be referenced in the associated publication and the dataset can be released, when notified, at the same time as the publication. Datasets can be cross-linked back to the article.

Data can be archived without a DOI as not all data are appropriate for a DOI. For a dataset to be assigned a DOI, it must be provided to the data repository in good condition, with appropriate metadata and of a suitable level of technical quality. The data depositor will be responsible for ensuring the data meets the required level of quality.

A DOI gives assurance to future users that the dataset is:

- stable — it is not going to be modified
- complete — it is not going to be updated
- permanent — the data repository is committing to make the dataset available for the foreseeable future
- of good technical quality — the data repository is giving its stamp of approval, saying that the dataset is complete and that all the necessary metadata are available

### 3.1.3 Data Access Statement

Partners must include a statement in their publication(s) describing how to access the data (or a statement explaining why access to underlying data has been restricted). If data are openly available, the name(s) of the data repositories should be provided, as well as any persistent identifiers (e.g. DOI) for the dataset. Some examples are provided below.

- All data created during this research can be accessed from the National Geoscience Data Centre at <http://dx.doi.org/10.5285/xxxx>, and used under the Creative Commons Attribution licence (CC-BY).
- Supporting data will be available from the National Geoscience Data Centre at <http://dx.doi.org/10.5285/xxxx> after a 6-month embargo from the date of publication to allow for commercialisation of research findings.
- Due to the sensitive nature of the research, no interviewees consented to their data being retained or shared. Additional details relating to other aspects of the data are available from the National Geoscience Data Centre archive at <http://dx.doi.org/10.5285/xxxx>.
- The data underlying this article are not available by agreement with our partners to protect their commercial confidentiality.

All written project deliverables should also include a similar 'data statement' specifying how to access the data.



### 3.2 MAKING DATA OPENLY ACCESSIBLE

Due to the high level of public and industry interest in the potential impacts of shale gas and CCS, the default position for the project will be for all finalised datasets to be open access. It is a requirement that all open data are accessible for the long-term. This makes the research process more robust by enabling validation of results and maximising the value obtained from publically-funded data. All public (written) deliverables should also be also archived.

Data should be archived as open access which:

- underpins a publication
- has long-term interest with potential for re-use (including currently unforeseen uses)
- validates research findings
- is worth keeping

Benefits of open access:

- Accelerations of the research and discovery process
- Avoidance of the duplication of research efforts
- Enhanced opportunities for collaborations
- Broader and faster opportunities for the adoption and commercialisation of research findings

#### 3.2.1 Categorisation of data access

However, not all data generated by SECURE must be open. The need to balance openness and protection of scientific information should be taken into account and certain datasets may need to remain closed according to the principle "as open as possible, as closed as necessary".

Project datasets should therefore be categorised. It should be carefully considered which data can be made public (open access) from the onset, which should be placed under temporary embargo (< 2 years) before open release and which must remain confidential. WP leaders should discuss with their work package participants/task leaders to determine this.

There is not a need for a separate data access committee, but this will be an item for discussion on Project Management Board meetings.

Examples of data which could be closed or restricted:

- Confidential information
- External industry data
- Commercial sensitivity/interest (e.g. new tools being developed with potential for patenting)
- Data with IPR issues
- Sensitive data containing personal information

Participant consent may also need to be obtained. This should be agreed during early stages of the project.

IPR and innovation for the project should be considered to ensure there are no conflicts between data sharing and these.

If certain datasets cannot be shared (or need to be shared under restrictions) the restrictions associated with the data must be valid/reasonable. Metadata should include a statement specifying any restrictions.

For confidential data, it is recommended that a discovery metadata record is published to signpost that the data exists without necessarily archiving the data. This should contain a brief description, which directs any potential user to the data owner contact if more information is required or to discuss the possibility of data access, which could lead to future collaborations. It should be discussed with the participants if they wish to advertise their (confidential) data in this way or if they prefer to make no information available in some cases.

#### 3.2.2 Archiving data in a recognised data repository

Data from the project should be made accessible by deposition in a suitable recognised data repository. A listing of datasets will be maintained on the SECURE website with links to the data. BGS (UKRI)'s proven data archiving service will be used as the basis for data archiving and provision for the SECURE project though other recognised repositories may also be used. Appropriate arrangements with the main data repositories,



the National Geoscience Data Centre (NGDC) and the UKCCSRC Data and Information Archive, which are managed by BGS (UKRI) have been explored.

Partners are expected to archive their own data in their own recognised data repository provided they are openly accessible. Where partners do not have access to their own recognised data repository, data may also be archived in the UK National Geoscience Data Centre (NGDC) based at the British Geological Survey <http://www.bgs.ac.uk/services/NGDC>, following discussions with the SECURE project data manager.

### 3.2.2.1 NATIONAL GEOSCIENCE DATA CENTRE (NGDC)

Data from SECURE can be deposited in the UK National Geoscience Data Centre (NGDC) based at the British Geological Survey <http://www.bgs.ac.uk/services/NGDC>. NGDC collects and preserves geoscientific data and information, making them available to a wide range of users and communities and is committed to ensuring data will remain usable and shareable in the long-term. It is a recognised trustworthy data repository committed to supporting open access data publishing and long-term storage and was awarded the Core Trust Seal in December 2017. This is an international core level certification based on the [DSA-WDS Core Trustworthy Data Repositories Requirements](#) catalogue and procedures. NGDC is also registered in the re3data Registry of Research Data Repositories <https://www.re3data.org/repository/r3d100010189>. The data repository is maintained and kept up-to-date with latest data management and web delivery developments.

There is a streamlined archiving process making it as easy as possible to deposit data. Data can be deposited online along with metadata describing the data via the NGDC data deposit portal -

<http://www.bgs.ac.uk/services/ngdc/guidelines.html>

Data archive staff will check and validate the information before adding it to the Archive.

Data can be searched and downloaded –

<http://www.bgs.ac.uk/services/NGDC/dataDeposited.html>

Larger datasets may be sent by alternative methods such as portable hard drive or file transfer application. In this case, the data portal should still be used to enter the metadata and a word or text file should be uploaded stating how the data will be transferred.

Support of users is also provided by NGDC, including:

- Initial contact with project participants
- Provision of guidance to data depositors
- Guidance to prepare and check metadata

As data archived in the NGDC will be made available via open access online with no login required, it is not possible to identify the individuals accessing the data other than web summary statistics.

### 3.2.2.2 UK CCS RESEARCH CENTRE (UKCCSRC) DATA AND INFORMATION ARCHIVE

Additionally, project data related to CO<sub>2</sub> storage will also be made available via the UKCCSRC Data and Information Archive [www.bgs.ac.uk/ukccs](http://www.bgs.ac.uk/ukccs). This is the subject-specific recognised data repository for Carbon Capture and Storage data in the UK and provides data archiving services to various projects. It uses the same underlying infrastructure as the NGDC with a separate front-end interface.

It holds data and other outputs generated by UKCCSRC research projects, UK's Engineering and Physical Sciences Research Council (EPSRC) funded CCS projects and other CCS projects. The data and other project outputs are archived for the long term making them easily accessible and more widely available for future reuse. This provides a unique and useful resource for Carbon Capture and Storage research and applications.

Data can be deposited using the UKCCSRC Data Deposit Application –

<http://transfer.bgs.ac.uk/carbon>

Data can be searched and downloaded at -

<http://www.bgs.ac.uk/ukccs/accessions/index.html>

Datasets are also listed by project -



<http://www.bgs.ac.uk/ukccs/accessions/projects.html>.

### 3.2.2.3 OTHER DATA REPOSITORIES

Data may also be archived in other suitable recognised data repositories provided these are openly accessible. National Geological Survey Organisations are more likely to already have data archiving procedures in place than some smaller partner organisations, and this can be achieved by lodging datasets in their respective national data repository.

See the [Registry of Research Data Repositories](#) for appropriate repositories.

Partners must inform the SECURE data manager if data are archived in another repository and a link to the data should be provided. Metadata for these datasets will also be published in the NGDC with a link to the data, giving a central source of information describing data associated with the project. A listing of datasets will be maintained on the SECURE website with links to the data.

## 3.3 MAKING DATA INTEROPERABLE

Data produced in the project should be interoperable and standard open formats should be used to allow data re-use. Data should be usable without the need for communication with the data creator.

### 3.3.1 Data formats

- The format must be well documented and conform to widely accepted standards.
- The format must be readable by tools that are freely available now and are likely to remain freely available in the future.
- List of NGDC acceptable formats:  
<http://www.bgs.ac.uk/services/ngdc/preferredDigitalFormats.html>

### 3.3.2 Data files

- Parameters in data files should either be labelled using an internationally recognised standard, or by local labels that are accompanied by clear, unambiguous plain text descriptions.
- Data must be accompanied by sufficient usage metadata to enable its reliable reuse. Some of this may be embedded within the data files. If not it should be included as additional documents.
- Data should be quality controlled by the data creator before archiving.

### 3.3.3 Naming conventions

- Data files should be named in a clear and consistent manner throughout the dataset.
- Filenames (rather than pathnames) should reflect the contents and uniquely identify the file.
- Filename extensions should conform to appropriate extensions for the file type.
- Filenames should be constructed from lower case letters, numbers, dashes and underscores
- Filenames should be no longer than 64 characters.

### 3.3.4 Data and metadata vocabularies, standards or methodologies

Partners are encouraged to use data and metadata vocabularies, standards and methodologies where these exist to make data more interoperable. Data should conform to [INSPIRE](#) where appropriate or other appropriate international standards.

Geoscience vocabularies used within BGS/NGDC -

<http://www.bgs.ac.uk/services/NGDC/management/geology/vocabulary.html>.

The Commission for the Management and Application of Geoscience Information promote geoscience information standards and best practice <http://www.cgi-iugs.org/home.html>.



### 3.4 INCREASING DATA RE-USE

#### 3.4.1 Data licencing

Data will be licenced to permit the widest re-use possible. The Creative Commons Attribution (CC-BY) licence with the appropriate acknowledgement is recommended for maximum dissemination and use of open access data. This licence lets others use the data for any purpose, as long as the data creator and the SECURE project is acknowledged.

The EU funding acknowledgement: "European Union (EU)" & "Horizon 2020" should be included. An example acknowledgment statement is provided below.

This data set is available under CC-BY Licence, subject to the following acknowledgement: "Data supplied by permission of Edinburgh University and funding provided by "European Union (EU)" and "Horizon 2020" under the SECURE project.

The appropriate licence must be specified when data is deposited. See <https://creativecommons.org/licenses/> for more information on the CC-BY license and other types of Creative Commons licences.

For more information on data licences, refer to Ball, A. (2014). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>.

#### 3.4.2 Embargos

Research data will be made available as soon as possible. Data received by a data repository (e.g. NGDC) as open access will be made available for re-use without delay once the data and metadata have been verified and archived in the system.

Datasets may be deposited in a data repository with an embargo if necessary in order to exploit data, publish results or seek patents, after which the data will be released as open access. The duration of the embargo or a release date must be specified. This should be no longer than 2 years in order to make the data available as soon as possible. If necessary, a review date can be set rather than automatic release and the depositor will be contacted at that time. Please notify the data manager if data can be released before the embargo has passed (e.g. when the related paper is published). During the embargo a metadata record will be visible but not the data.

#### 3.4.3 Ensuring long-term usability

Data will be archived for the long-term and it is intended that it remains re-usable for as long as possible. Data repositories used must have measures in place to ensure data does not become obsolete or unusable. For example, NGDC has a Digital Preservation Policy, <http://www.bgs.ac.uk/downloads/start.cfm?id=3173>, to ensure the longevity of the digital information assets in a sustainable way by addressing the factors which risk making them unusable and inaccessible.

#### 3.4.4 Quality Assurance

Quality assurance processes should be part of the metadata. Any laboratory data should meet appropriate QA standards.

## 4 Allocation of resources

Data management costs as covered as part of the grant. The cost of making data FAIR is €19,250 (Data Manager Allocation), though this does not include costs of archiving by partners.

Overall data management for the project will be undertaken by BGS (UKRI) as the project co-ordinator, with co-ordination of data generated by individual work packages resting with individual work package lead organisations. BGS (UKRI) have allocated resources and appointed a data manager to ensure long-term data management complies with current best practice to allow continued data availability.





Beyond the end of the project, data archived in the NGDC (or other repositories) will be preserved and maintained for the long term using the data repository resources. The costs associated with this will not be substantial.

## 5 Data security

Data stored in NGDC is covered by the BGS IT Security Policy which follows the guidance of the NERC Information Security Policy. The ongoing activities of the British Geological Survey are dependent on the availability and integrity of a range of IT systems and services. In using these technologies, the BGS has a duty to protect its valuable scientific information from damage and loss, whether accidental or deliberate. The BGS network is protected by firewalls and related security systems.

Sensitive data should be encrypted when transferring.

BGS maintains an Information Risk Register. This is a formal system for recording potential risks to either datasets or storage facilities including the probability of the risk occurring, who is the responsible person for ensuring that the risk is investigated and what actions are recommended/progress made towards implementing the solution to minimise the risk.

### 5.1 DISASTER PLANNING

For BGS, disaster planning is both site specific and data type specific and there are disaster plans in place.

For example:

- digital data on the SAN backed up daily to tape so that there is a near time back-up and an offsite backup of all data
- replication of 'critical' digital data between the main BGS sites
- replication of the BGS Oracle database between the main BGS sites
- installation of standby generators to power the computer suites in the event of power failure to the buildings
- installation of environmental monitoring and fire suppression systems
- critical components in the network are clustered so that they are more resilient to failure.

Any other certified data repositories used should have similar provisions in place.

### 5.2 PERSISTENT ACCESS TO DATA

Data will be stored in certified repositories for long-term preservation and curation. NGDC is a certified repository and has a Digital Preservation Policy, <http://www.bgs.ac.uk/downloads/start.cfm?id=3173>, to ensure data remains usable and accessible. Persistent access to datasets are guaranteed for the long-term. Any other data repositories used must also ensure this.

## 6 Ethical aspects

All data must conform to the EU General Data Protection Regulation (GDPR) when personal information is involved. Data can be anonymised if needed.

There may be sensitives with some data which means it is unsuitable for sharing. This should be decided by the data generator in discussion with the WP lead and Coordinator.

The TNO approach for GDPR/WP6 will be the same as that developed for the H2020 project ENOS: when collaborating with citizens for participatory monitoring via meetings and questionnaires, the citizens will receive a privacy information sheet that informs them of the aim of the research and the process of the data collection and storage. This privacy information sheet was developed in close cooperation with the TNO privacy officer.



## 7 Existing procedures for data management

The project makes use of NGDC and UK Research Councils (including NERC and EPSRC) procedures, policies and guidelines for data management.

- NGDC: <http://www.bgs.ac.uk/services/ngdc>.
- EPSRC expectations for research data management: <https://epsrc.ukri.org/about/standards/researchdata/expectations/>.
- NERC Data Centres: <https://nerc.ukri.org/research/sites/data/>.
- NERC data policy: <https://nerc.ukri.org/research/sites/data/policy/data-policy/>.
- Digital Curation Centre guide, Five steps to decide what data to keep: <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>



# Appendix 1 Data Management Questionnaire

In order to refine the SECURE Data Management Plan it would assist if you could complete this questionnaire and return to the SECURE data manager [secure.data@bgs.ac.uk](mailto:secure.data@bgs.ac.uk). This will help identify and categorise project data and any potential problem areas.

1. Work Package

2. Task/subtask

3. Contact for data management Name/Position

Organisation

Email

4. Projected timescale of data collection/production  to

5. Where will you deposit data for long-term archiving?

6. When do you expect to deposit data?

7. Specific datasets (add more rows as necessary)

Title of dataset	Estimated total volume	Formats	Open Access/Embargo/Confidential/Restricted?
a)			
b)			
c)			



d)			
----	--	--	--

8. All data will not necessarily be appropriate for long-term preservation with a data repository. Which of the above data sets do you consider may be not be appropriate to deposit?

9. Will you be providing any software? Y/N

If yes please provide details. Are there likely to be any licensing requirements?

10. Are you using any existing data from project partners? Y/N

If yes, please provide details:

11. Are you using any external third party data? Y/N

If yes, please provide details:

12. Does data require an embargo (< 2 years) before open access release? Y/N

If yes, please provide details:

13. Is any data confidential or restricted? Y/N

If yes, please provide details:



14. Are there any IPR or commercial sensitivity issues that will restrict access to data? Y/N

If yes, please provide details:

15. Any other comments or issues



## Glossary

<i>DMQ</i>	Data Management Questionnaire
<i>DOI</i>	Digital Object Identifier
<i>FAIR</i>	Findable, Accessible, Interoperable, Re-usable
<i>NGDC</i>	National Geoscience Data Centre



## 8 References

- H2020 PROGRAMME. 2016. Guidelines on FAIR Data Management in Horizon 2020. Available online: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- H2020 PROGRAMME. 2017. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Available online at: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- BALL, A. 2014. How to License Research Data. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>.
- DCC. 2014. Five steps to decide what data to keep: a checklist for appraising research data v.1. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>.
- WILKINSON, M. D. ET AL. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3:160018, doi: 10.1038/sdata.2016.18