



## Deliverable

### Development of Questionnaire with Scientific Support. Version 2

Authors and affiliation:

**Name(s)**

Christopher Luton

E-mail of lead author:

**Email : CHTO@BGS.AC.UK**

Version: 16-05-2019

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		
<b>Deliverable number</b>	D10.1	
<b>Dissemination level</b>	Public	
<b>Deliverable name</b>	Development of Questionnaire with Scientific Support	
<b>Work package</b>	WP10, IPR & Data Policy Issues	
<b>Lead WP/Deliverable beneficiary</b>	UKRI/BGS	
Deliverable status		
<b>Submitted (Author(s))</b>	16/05/2019	Christopher Luton
<b>Verified (WP leader)</b>	16/05/2019	Christopher Luton
<b>Approved (Coordinator)</b>	28/06/2019	Jørgen Tulstrup

# GeoERA Materials\* Questionnaire



## BACKGROUND

As has been pointed out previously, the overall aim of GeoERA is to integrate information and knowledge to support sustainable use of Earth's geological subsurface. Three groups of geoscientists involved in each of the three overall GeoERA themes intend to identify, use and produce large amounts of geological data and information which will be managed by a GeoERA Information Platform Project (GIP-P). This, in turn, will create a common platform for organising, disseminating and sustaining the digital outputs and results of those projects.

Work Package 10 and this deliverable were created to facilitate the investigation of intellectual property and data policy issues to support the release of project results of GeoERA. To assist this vision, at an early stage it was recognised that for the GIP-P to operate effectively and, more importantly, legally, the main requirement is that information and knowledge used by the science teams and projects would need to be free and unencumbered, making future access and use non-constrained. Consideration was also given to derived data (as was defined in the Questionnaire) and finally an issue about the use of personal data and data protection issues, which have placed an additional burden on users of personal information since the emergence of General Data Protection Regulations.

With the Questionnaire (Deliverable D10.1) in place, it is important to look at the feedback from the science teams about use and usefulness of the questionnaire. For information, a copy of the questionnaire is attached.

# GeoERA Materials\* Questionnaire



## TABLE OF CONTENTS

1	Background.....	2
2	Science Group Feedback.....	<b>Fejl! Bogmærke er ikke defineret.</b>
3	Conclusion .....	6
4	Questionnaire .....	7

\* Materials covers data, datasets, data products, & software.



## 1 FEEDBACK ON USE OF THE QUESTIONNAIRE

The science groups have been actively using the questionnaire since it was developed and, when questioned, they have very little feedback. In their opinion the questionnaire is useful and “understandable” and they have discovered that some of the contents of the questionnaire have already been covered in their own data management plans – such as that utilized in the MUSE project.

Further enquiries were made about additional issues which have come to light when sourcing and using geological information and datasets and other third party materials, and which were not covered in the questionnaire; the science groups felt that there would be other matters which would arise in the future concerning the knowledge-based approach, especially as some of the groups want to build knowledge repositories.

The subject of derived data was raised with the groups and most, such as MUSE, do not intend to publish background or raw data, much of the work involves datasets that are freely available to use/publish (e.g. in registers), derived and interpreted raw data, as well as data which is provided by partners to the science groups.

### **Contaminated Water/Land**

An interesting point was raised by science staff in the water area, where it is recognised that many private companies (and, in fact, Government bodies) in Europe are very reluctant to make raw and interpreted data public, because of concerns around security, GDPR and potential devaluing of land because of contamination. Security issues are recognized in various European countries (Environment Agency-UK) because of potential terrorist attacks on water supplies in the UK. This is probably similar across other countries in Europe. Therefore, the publication of exact grid references and locations is a problem. This also applies to land where contamination is identified and the potential “harm” it can bring to landowners; publicising exact locations has been found to be a breach of personal data/information under data protection legislation. With GDPR now in full swing in all European countries, publishing personal data is becoming a litigious area meaning publishing exact locations need to be managed.

As a follow on, the inclusion of information on drinking water supplies also poses a potential problem because of the effect they can have on the value of land. So again discretion needs to be paramount when publishing or using data/information.



## **Changing Circumstances and Legislation**

An issue was raised about the way in which answers may vary with time on the questionnaire; a matter covered off in the questionnaire at one time, might have a different answer if considered at another time, due to such things as legislation changes, etc. The matter of changing legislation with relevance to R&D projects was given as an example where it was considered there would be “knowns” and “unknowns” about changes during the period a project was running. It was felt maybe there should be a review of some type to ensure continuity.

## **Review**

Because of the potential for changes in circumstances, including changes to relevant legislation, it was suggested the questionnaires should be reviewed after around 18 months, to ensure the original answers are still valid. Of course, this then raised the matter of where the collected data/information was being stored and also the questionnaires themselves. Should these be stored in case of an audit? An audit may be necessary to examine the trail of permissions concerning a particular dataset.

## **Acknowledgement**

The final point raised was that of acknowledgement. Some owners of datasets permitted access to the dataset and free use, *subject* to acknowledgment. This then begged the question – how do we adequately acknowledge? How do we ensure that future users of the materials also provide adequate acknowledgement and, in the cases where they don't, who takes responsibility/liability?

The subject of acknowledgement is a common problem across the whole academic publishing field. The digital object identifier (DOI) is the recognized system for identifying content objects in the digital environment. DOIs are names assigned to any entity for use on Internet digital networks and provide persistent identification together with current information about the object. The system is managed by the International DOI Foundation (founded in 1998) - an open membership consortium including both commercial and non-commercial partners, and DOIs are accepted for standardization within International Standards Organization (ISO).

Presently the DOI is in widespread use and DOIs are being accepted for use in Government documents (such as EU, OECD, UK Government, etc.). Most importantly DOIs are built on a set of standards and business practices that provide a self-sustaining model which is being

## GeoERA Materials\* Questionnaire



widely adopted. At this stage of the GeoERA project it is not intended to dwell on DOIs, but, as such a crucial element or tool in the “permissions” chain of obtaining consent to use and publish datasets and information, it will be addressed in later deliverables for WP10.

### Conclusion

The conclusion from WP10 is that the questionnaire is being well used and there are very few issues to report. It was drafted by administrators involved in managing geological materials on a daily basis, so many of the common issues were raised. It probably does need some slight adaption to cover the points raised, and the question of potential review of the questionnaire (after a set period of maybe 18 months) merits consideration.

The question of water/land contamination is one which regularly need to be addressed in geological surveys across Europe: each has its own set of particular circumstances and each should be handled differently, depending on the circumstances.

It is intended, when the next head-to-head meetings take place in June 2019, that further discussions will be held with the science groups to determine if there are other matters which need to be addressed. The whole question of “acknowledgement” is slightly outside the remit of WP10, but it is likely to be addressed in later WP10 deliverables concerning open access.

## GeoERA Materials\* Questionnaire

WP10

	Question	Comment	Answer
1	Have you identified materials not owned by your organisation that you wish to use in connection with your work? (Note: if the materials are from your own organisation you may also need permission).	<p>Third party information and data from non-Government Sector Organisations (GSOs) that is intended to be used for GeoERA projects should be identified.</p> <p>GSOs may have restrictions on the use of their own data and information (e.g. only after publication has taken place, embargo periods, etc.). Such limitations need to be added (as at point 9 below)</p>	(this response can be a “yes” or “no” or might result in a list of materials).
2	What do the materials comprise?	<p>A short description is required.</p> <p>Please use an additional page if further explanation is required</p>	
3	Where are the materials available from and how did you find them – known repository, Government link, university, open website?	<p>Were possible please provide a reference number and identify any DOI or acknowledgement required</p>	
3	From the location of the materials, was there any mention of who owned the materials?	<p>We probably mean here the copyright ownership – however some companies/organisations licence in materials so they don’t actually own them. Bear in mind that if you have obtained materials from an organisation that has itself licensed them in from elsewhere, it may be that the owner of the materials, rather than the organisation you have obtained them from, will need to be consulted.</p>	

\* Materials covers data, datasets, data products, & software.

## GeoERA Materials\* Questionnaire

5	What is your intended use of the materials and will the finished product/dataset include a substantial part of the materials? Could the materials be reverse engineered from your product?	Copyright law allows one to derive materials as long as you don't copy a substantial part of the original. If the source materials are not visible in what you create and if you cannot reverse engineer from what you have created to the original material, this is termed <i>non-copy derived</i> material. Further information regarding <i>copy derived</i> and <i>non-copy derived</i> is contained at Schedule 1.	
6	Do you intend to modify the format of the materials in any way?	I.e. are you planning to convert the materials to a different delivery format; and/or are you converting from a digital to a 'flat' format (or vice versa)	
7	Do you see a need to share the materials, or anything that you derive from the materials, with anyone else in connection with this project?		
8	Do you see any future commercial applications for what you have derived from the materials?		
9	Were the materials accompanied in any way with terms and conditions concerning use or are they freely available (without any cost) to everybody using FAIR Principles (Schedule 2) both during and post the GeoERA project?	This could take many forms: Creative Commons licences, Open Access, conditions concerning commercial and non-commercial use, free use but improvements made open, etc.	

\* Materials covers data, datasets, data products, & software.



## GeoERA Materials\* Questionnaire

		If terms and conditions apply please provide links or describe these in detail for all restricted Materials.	
10	Are you sure that these terms and conditions of use cover all of your proposed usage of the materials (i.e. questions 5-8 above)?	If you are not sure, or if you are sure that some (or all) of your proposed usage of the materials is outside of what is permitted under the terms of use you have seen, you need to contact the owner of the materials to get full permission. You also need to check not only that you have permission, but whether or not the owner of the materials may have a claim of IPR ownership over anything you derive from their materials (this is more likely if you are producing <i>copy derived</i> materials – see Schedule 1). This may have implications for whatever you have derived: i.e. could the owner request a royalty/revenue share in relation to any commercial use that you maybe intending to make?	
11	Processing personal data/information. Does any of the materials you wish to use contain “Personal Data”? Examples might be where personal names have been left in the digital content. Please see Schedule 2 that covers the obligations that might be on you to ensure this matter is clearly accounted for and that where you pass on personal data instructions are given as to what and cannot be processed by the user.	Obligations on you will relate to whether you are “controlling” or “processing” personal data.	

\* Materials covers data, datasets, data products, & software.

### GeoERA Materials\* Questionnaire

12	When passing on data that holds Personal information you will need to consider further you and your users' position. Are you passing on personal data?	If you transfer or pass on the personal data, you may have a requirement to ensure that anyone you pass the personal data to also complies with the Data Protection Laws/Regulations. See Schedule 3.	
----	--	---	--

\* Materials covers data, datasets, data products, & software.



## Schedule 1

**Non-copy derived** means that the derived data set does not include a copy of the whole or any substantial part of the original information (as defined by Copyright legislation and case law) and the derived data set cannot be reverse engineered to create a copy of the original information, or any substantial part of it.

**Copy derived** means that the derived data set includes a copy of the whole or any substantial part of the original information or that the derived data set can be reverse engineered to create a copy of the original information or any substantial part of it.

If an external body owns the copyright in the materials/information and you are engaged in copying that information, you are *prima facie* breaching copyright. However as a general rule what the law states is that:-

- Firstly, there must be sufficient objective similarity between the infringing work and the copyright work, or a substantial part thereof, for the former to be properly described, not necessarily as identical with, but as a reproduction or adaption of the latter; and
- Secondly, the copyright work must be the source from which the infringing work is derived.

So, a key consideration is substantiality/quality. Not every act of copying is infringement: one must have copied the whole work or a substantial part of it. When we think about a "substantial part" we need to think of quality and not just quantity, as a single figure or image if sufficiently important (whatever its size) as a part of the copyright work could be considered substantive!



## Schedule 2

### FAIR Data: Background and Rationale

Good data stewardship is the key to knowledge discovery and innovation. To generate value for a research community beyond the initial researchers, funding agencies are increasingly setting requirements for proper data stewardship of research data. Beyond proper collection, annotation, and archival, data stewardship includes the 'long-term care' of research data, with the goal that they can be found and re-used in downstream studies and research. To facilitate good data stewardship, a broad community of international stakeholders have developed the FAIR Data principles, which have been embraced by both the [European Commission](#) and the [G20](#). The first formal [publication of the FAIR Principles](#) further describes the rationale behind them.

### FAIR Data

The paper '[The FAIR Guiding Principles for scientific data management and stewardship](#)' was written in 2016 and is the first formal publication of the FAIR principles. In short, the FAIR Data Principles propose that all scholarly output should be:

- **Findable:** easy to identify and find for both humans and computers, with metadata that facilitate searching for specific datasets,
- **Accessible:** stored for long term so that they can easily be accessed and/or downloaded with well-defined access conditions, whether at the level of metadata, or at the level of the actual data,
- **Interoperable:** ready to be combined with other datasets by humans or computers, without ambiguities in the meanings of terms and values,
- **Reusable:** ready to be used for future research and to be further processed using computational methods. This requires adequate information about how the data were obtained and processed (provenance) and an appropriate license

Please consult '[The FAIR Data principles explained](#)' for a more detailed description.

Of interest note there is no "O" for "Open" in FAIR. Proponents of FAIR data often also stress that data should be as open as possible, access only being restricted where necessary.

Further details are at:-

<https://www.go-fair.org/fair-principles/>



## Schedule 3

As background, Data Protection Laws means the General Data Protection Regulation (EU) 2016/679 (GDPR) and any other law applicable to Europe relating to the protection of personal data and the privacy of individuals, including where applicable guidance and codes of practice issued by any European Authority.

Understanding whether you are processing personal data is critical to understanding whether the law applies to your activities. "*Personal Data*" means personal data as defined in the legislation, which is information that relates to a *living* identified or identifiable individual. What identifies an individual could be as simple as a name or a number or could include other identifiers such as an IP address or a cookie identifier, or other factors.

If it is possible to identify an individual directly from the information you are processing, then that information may be *personal data*. If you cannot directly identify an individual from that information, then you need to consider whether the individual is still identifiable. You should take into account the information you are processing together with all the means reasonably likely to be used by either you or any other person to identify that individual.

Even if an individual is identified or identifiable, directly or indirectly, from the data you are processing, it is **not** personal data unless it 'relates to' the individual.

When considering whether information 'relates to' an individual, you need to take into account a range of factors, including the content of the information, the purpose or purposes for which you are processing it and the likely impact or effect of that processing on the individual. It is possible that the same information is personal data for one controller's purposes but is not personal data for the purposes of another controller.

Information which has had identifiers removed or replaced in order to pseudonymise the data is still personal data for the purposes of the law (GDPR).

Information which is truly anonymous is not covered by the law (GDPR). If information that seems to relate to a particular individual is inaccurate (i.e. it is factually incorrect or is about a different individual), the information is still personal data, as it relates to that individual.

For the purposes of this Questionnaire, "**controller**", "**data subject**" and "**personal data**" have the meanings given under the Regulation; "**processing**" has the meaning given under the Regulation (and "**process**", "**processed**" and "**processes**" shall be construed accordingly); and "**Regulation**" means the General Data Protection Regulation (EU) 2016/679, as defined as part of the Data Protection Laws.

To the extent that you provide another ("user") with *personal data* in connection with the performance of any agreement or arrangement with the user, the user must agree that, unless the parties specifically agree otherwise, each (you and the user) shall be independent controllers of the *personal data* in their own right. The user must agree they shall comply with all applicable data protection and privacy laws in relation to the *personal data*, including the Regulation. If required, you may wish to ensure a potential user provides details of any *personal data* it intends to process in accordance with the questions below:-



## PROCESSING OF PERSONAL DATA

This series of instructions for your user sets out the scope, nature and purpose of the processing of Personal Data by the user on your behalf and constitutes your written instructions.

### 1. THE SCOPE, NATURE AND PURPOSE OF PROCESSING

1.1 User may process Personal Data in order to provide a service to its users.

### 2. THE DURATION OF PROCESSING

2.1 User may process Personal Data as permitted by you for a duration (to be determined).

### 3. THE TYPES OF PERSONAL DATA

3.1 User may process Personal Data of the following types:

- (a) [personal details]
- (b) [financial details]
- (c) [employment and education details]
- (d) [goods or services provided]

3.2 User may not process sensitive Personal Data of the following types:

- (a) [physical or mental health details]
- (b) [racial or ethnic origin]
- (c) [religious or philosophical beliefs]
- (d) [trade union membership]
- (e) [political opinions]
- (f) [genetic data]
- (g) [biometric data]
- (h) [sex life and sexual orientation]]

### 4. THE CATEGORIES OF DATA SUBJECT

4.1 User may process Personal Data relating your

- (a) [employees]
- (b) customers and clients]
- (c) [suppliers and service providers]
- (d) [advisors, consultants and other professional experts]
- (e) [complainants and enquirers].