D2.8: Guidelines on Open Access and Restricted Data (final)

Author(s)	R Baxter, S Sacchi, H Tjalsma, V Cavalli, S Lambert, H Frew, C Inglis
Status	Final
Version	v1.0
Date	05/09/2017

Abstract:

This report is a full revision of the earlier D2.5 [Guidelines on Open Access and Restricted Data (draft)] on restricted data. It looks in particular at the new legal requirements for information service providers in storing and processing personal data under the EU General Data Protection Regulation (GDPR), the particular impacts of the GDPR on the EUDAT CDI and EUDAT service providers, and offers a number of supporting policies, technical approaches and guidelines for EUDAT in managing, storing and processing open and restricted data.

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

Document identifier: EUDAT2020-DEL-WP2-D2.8	
Deliverable lead	UEDIN
Related work package	WP2
Author(s)	R Baxter, S Sacchi, H Tjalsma, V Cavalli, S Lambert, H Frew, C Inglis
Contributor(s)	
Due date	31/08/2017
Actual submission date	05/09/2017
Reviewed by	D Foster, MF Iozzi
Approved by	PMO
Dissemination level	PUBLIC
Website	www.eudat.eu
Call	H2020-EINFRA-2014-2
Project Number	654065
Start date of Project	01/03/2015
Duration	36 months
License	Creative Commons CC-BY 4.0
Keywords	Policy, restricted data, personal data, GDPR, DataTags

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

PREFACE

This report is a revised and updated version of EUDAT2020 deliverable D2.5, *Guidelines on Open Access and Restricted Data (draft)*. For readers familiar with that earlier document we record the changes introduced in this version below.

D2.5 chapter	Notes on changes	D2.8 chapter
Executive summary	General updates to reflect new content.	Executive summary
1. Introduction – the openness of data	Updates to Sections 1.4, *Scope and approach of this report*, and 1.5, *Structure of this report*. Section 1.3, *Open data in European research projects*, deleted as now redundant.	1. Introduction – the Openness of Data
2. Established EUDAT policies for open data	Unchanged.	2. Established EUDAT Policies for Open Data
3. Management of personal data 5. Future legislation on personal data – the GDPR	Merged and renamed. Updated to reflect GDPR as "current legislation". New section on forthcoming ePrivacy Regulation	3. Management of Personal Data – GDPR and ePrivacy
4. The application of data protection in Europe: three case studies	Deleted as no longer current.	--
--	New commentary on research codes of practice and current legal position.	4. Research Use of Restricted Data – Current Positions
--	New chapter on broader ethical restrictions on open data, with a focus on temporal and spatial constraints.	5. Ethical Constraints on Openness
D2.5 Chapters 6, 7 and 8 have been re-organised, revised and updated as follows:		
7. Data controllers and data processors in EUDAT. 6.1 Administrative data 6.2 Content data 8.2 Considerations and issues for specific EUDAT Services	Revised and reorganised.	6. Potential Impacts on the EUDAT CDI
6. Classifying data in the EUDAT CDI and related services	Entirely revised to cover use of the DataTags system as an approach to classification. New sections added to consider a data-subject-centric view of EUDAT service design and issues of encryption.	7. Technical Policies for Restricted Data
8. Recommendations	Revised. Former Section 8.2 on EUDAT services moved into new Chapter 6.	8. Recommendations
9. Conclusion and plans for further work	Revised.	9. Conclusions and the Future
10. References	Updated.	10. References
Annex A. Country case studies (UK, Netherlands, Norway)	Deleted as no longer current.	--
Annex B. Template for privacy notice and disclaimer	Renumbered. Updated to strengthen privacy notice into privacy notice more aligned with GDPR.	Annex A. Template for privacy notice and disclaimer
Annex C. Template agreement for data controller-data processor	Renumbered.	Annex B. Template agreement for data controller-data processor
Annex D. Glossary	Renumbered. Minor updates.	Annex C. Glossary

Definitions

In this document we use the term 'EUDAT' to mean 'the EUDAT Collaborative Data Infrastructure', an ongoing collaboration between Service Providers and research communities, defined by the EUDAT CDI Agreement, working as part of a common framework for developing and operating an interoperable layer of common data services. In this context, EUDAT is not a legal person.

We use the term 'CDI' in isolation to refer to the underlying technical and service infrastructure of EUDAT.

We use the name 'EUDAT2020' to refer to the EU Horizon 2020 project which funded the second phase of the establishment of EUDAT between 2014 and 2018.

The structure and governance of EUDAT is described in EUDAT2020 D2.4 *Report on Governance Model* [1].

We use the term 'EUDAT management' to refer to the EUDAT CDI Board and Council, supported by the Secretariat, as defined in [1]. At time of writing, in contexts where 'EUDAT management' should refer to a legal person, this is to be interpreted as 'the hosting site of the EUDAT Secretariat'.

We use the term 'EUDAT service providers' to refer to individual member organisations of the EUDAT Collaborative Data Infrastructure, typically those who provide data services as part of the CDI service offering.

EXECUTIVE SUMMARY

The EUDAT Collaborative Data Infrastructure (CDI), a consortium of European research data service providers, was founded on the principle of supporting and promoting open access to research data. EUDAT encourages the open publication and sharing of research data under permissive licence conditions (we recommend CC BY 4.0) in line with the FAIR data principles of findability, accessibility, interoperability and reusability.

From this baseline of openness, EUDAT recognises that certain data cannot be fully open. The personal data of European citizens, as defined and codified under the General Data Protection Regulation (GDPR) of 2016, form a major class of restricted data. CDI services must be adapted to align with the new and strengthened rights of data subjects. EUDAT's legal basis for processing any personal data – and this includes storing such data – will be that of subject consent; gaining and recording such consent, along with subject rights to access, to update and to delete their personal data, must be addressed within the underlying CDI service infrastructure. Future service design must also follow the principles of (personal) data minimisation.

EUDAT's principal concern is the storing and process of research data, or data for research purposes. The GDPR derogates detailed handling of personal data in research contexts to Member State law and to competent community codes of conduct. Unfortunately, at the time of writing these national legal and community "soft-law" approaches are still being formulated. EUDAT is thus strongly advised to track development of research codes of conduct in particular, and work with community representatives accordingly, but also to "prepare for the worst" by following the letter of the GDPR. Compliance with the new European law can be enforced from May 2018.

Above the level of individual services and service providers, the EUDAT CDI Collaboration as an organisation must understand and implement the necessary legal arrangements between those service providers who will act as data controllers and those who will act as data processors. These agreements need to be formalised in legal agreements before the May 2018 deadline.

Beyond personal data, the issues of openly publishing data with potentially harmful impact on cultural, historical or scientific ethics is complex and best done in partnership with competent community ethics bodies.

The EUDAT CDI Collaboration supports, through common guidelines and suggested best practices, its members and service providers in devising and implementing appropriate technical solutions to make their infrastructure and services compliant with the GDPR. In this context, we suggest possible standard approaches for CDI service designers in handling data with time-constrained sensitivities (embargoed data; personal data of the deceased; children's data), and ways to support "degraded precision" in recording or reporting geographically restricted data.

We also examine the DataTags system, originally proposed by Harvard University as a way to map legal requirements onto suitable data handling patterns, and consider the adaptations needed to use it as a standardising tool under the GDPR. We also suggest one possible method of pseudonymisation for data subjects which could assist service builders in complying with requests from data subjects exercising, for example, their right to be forgotten.

We recommend the EUDAT CDI take steps now to re-orient data services towards supporting the rights of data subjects, and to adopting a standard internal metadata scheme like DataTags that is able to categorise sensitive or otherwise restricted data (with support for changing levels of sensitivity over time) and assist underlying data processing systems to manage both open and restricted data appropriately and automatically.

1. INTRODUCTION – THE OPENNESS OF DATA

In the context of EUDAT, it hardly needs reiterating that there is a powerful and universal trend for openness of data—including but not restricted to research data. The G8 Open Data Charter [2] declares that "open data are an untapped resource with huge potential to encourage the building of stronger, more interconnected societies that better meet the needs of our citizens and allow innovation and prosperity to flourish", and sets out five principles that will be the foundation for access to, and the release and re-use of, data made available by G8 governments. Science and research is recognised as one of the areas of high-value data. The European Commission published in 2011 a Communication entitled "Open data: An engine for innovation, growth and transparent governance" [3] that singles out the acceleration of scientific progress as one of the reasons why open data is crucial for Europe.

Funding agencies increasingly require open access to research data in the investigations that they support. For example, Research Councils UK has a set of principles [4] starting with "Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner."

Such statements and initiatives are not mere aspirations or impositions from on high. The Research Data Alliance has mobilised over 4,000 individuals in pursuit of its mission to build "the social and technical bridges that enable open sharing of data" [5]. EUDAT itself aims to support sharing and reuse of open data through its services, while recognising that not all data will be completely unrestricted.

The push for openness has been refined and synthesized into the concept of FAIR data: Findable, Accessible, Reusable and Interoperable [10], emphasizing not only the openness of data but the principles enabling its effective reuse. The FAIR ideal is gaining wide acceptance and uptake, for example in the context of the European Open Science Cloud.

1.1. Constraints on openness

There are some restrictions on the general openness of data that must be acknowledged. Many stakeholders have an interest in controlling or restricting access to some digital material, and not only for selfish reasons. The enforcement of restrictions might be underpinned by legislation, or by policies and practices of particular organisations. In any case the aim is to prevent unauthorised access to digital material that might cause harm of some kind, whether to national security, the lives of individuals, or commercial or scientific interests.

At the highest level, there are laws in place concerning disclosure of official secrets. In the United Kingdom, for example, the Official Secrets Acts 1911–1989 provide the main legal protection against espionage and the unauthorised disclosure of information. Their scope is information concerned with national security, defence, international relations, criminal activities and the like. The protection of personal data is also enshrined in legislation all over the world, respecting the rights of the individual on the collection and processing of data that is (or can be) linked to them.

Within the world of scientific data, not all is necessarily openly and instantly available. The Common Principles on Data Policy of Research Councils UK recognises that "there are legal, ethical and commercial constraints on release of research data" and acknowledges that "those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research." Even when there is no embargo period, user registration may be required simply in order to track who has accessed data. Legal constraints include the acquisition and processing of personal data, while ethical issues might arise where release of data might have unwanted consequences: for example, through revealing the location of archaeological sites or of rare animal or plant species.

In the commercial world, it goes without saying that there is much digital material that its owners wish to keep secret since its release would give competitors an advantage.

1.2. Why EUDAT must be prepared

Such considerations will certainly become relevant in EUDAT when some communities will wish to use EUDAT services to store and manage restricted data as well as open data, with even more complications when open data and restricted data are interwoven into individual datasets. The EUDAT Collaborative Data Infrastructure (CDI) must be able to handle these situations, and the mechanism chosen is to produce consistent guidelines for restricted data access to be adopted in the CDI. In particular, data protection legislation in Europe sees a strengthening of emphasis on giving citizens control of their personal data, requiring explicit consent for their storage, processing and use and placing more formal restrictions on the data controllers and data processors who handle them.

In assessing the impact of data restrictions on EUDAT, and in particular EUDAT's management of personal data, we draw a distinction between data that is submitted by external parties to the CDI, designated *content data*, and the data that EUDAT itself gathers and processes as part of its operations, *administrative data*, which might be personal in nature, for example about individual users of the CDI. EUDAT needs to be able to handle both of these in compliance with the data protection legislation in Europe. In handling administrative data such as emails or login names EUDAT service providers will become "data controllers"; for storing content data they are clearly data processors. We address these distinctions in EUDAT's case in Chapter 7. When the word "data" is used on its own, unless the context indicates otherwise, it is understood to mean "content data".

Another guiding principle for handling restricted data that EUDAT must keep in mind when designing data services is that one person's data may be another person's metadata, and vice versa. Any approaches to minimising undue exposure of restricted data within the CDI must be applied to metadata services as well as data services.

1.3. Scope, approach and structure of this report

As in the first version of this report [6] our focus is largely on personal data and the legislative environment that protects it. The principal reason is the incoming General Data Protection Regulation (GDPR) which enters into force in 2018. Its impact is both significant and far-reaching; in particular it has triggered ongoing reviews of, and revisions to, numerous scientific codes of conduct of particular relevance to the handling of sensitive data in infrastructures like the EUDAT CDI. We review the current state of some of these "soft law" approaches to restricted data in research contexts in Chapter 4, and examine some practical considerations around ethical restrictions in Chapter 5. Preparing EUDAT for the GDPR, however, is our main driver.

The target audience for the guidelines in this final version of the report are the EUDAT management and service providers. They will wish to be sure that their policies and implementations are well founded, defensible, and coherent with the policies of other service providers—particularly as EUDAT services may be distributed across multiple providers. The aim is to tease out some of the questions, the risks and impacts and to make recommendations and identify where further clarification and decisions are needed. Ultimately the aim is to produce a consistent and acceptable position for EUDAT as a whole on what restricted data may be stored and how; who may access it, when and how; and assurance that indeed it is secure.

In considering policies for open and restricted data, EUDAT is not starting from scratch: it already has established policies and guidelines, and **Chapter 2** summarises these. **Chapters 3 and 4** review the new European legislative framework around personal data, the General Data Protection Regulation, and discuss the current state of related legislation and "soft law" codes of conduct with respect to research. **Chapter 5** looks at some broader ethical considerations and restrictions on the openness of data in EUDAT. **Chapter 6** reviews possible organisational and infrastructural impacts of the legislative framework on service design within the CDI, and **Chapter 7** touches on some possible technical solutions for classifying data and designing services to meet legal requirements on restricted data. **Chapter 8** outlines a set of recommendations for EUDAT, and **Chapter 9** concludes. **Annexes** provide a number of practical templates for different policies or agreements between parties that relate to handling personal data.

2. ESTABLISHED EUDAT POLICIES FOR OPEN DATA

Since its inception under the Framework 7 'EUDAT' project, the EUDAT Collaborative Data Infrastructure consortium has believed fundamentally in open access to data. By "open access to data" we mean the free availability of data on the public Internet, permitting any user to reproduce and redistribute them for any purpose, and in particular for the purpose of non-commercial research, with no (or limited) financial, legal or technical barriers that might impede their meaningful reuse. The only allowable constraint on reproduction and redistribution should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

One of the prime motivations for the CDI is to create a single domain of registered, well-described, cross-disciplinary data, connecting collections and data centres across Europe and harmonising access to them – harmonising access not just in the technical sense but in the policy sense. In this, EUDAT subscribes to the ideas of intelligent openness as described in the 2012 Royal Society report "Science as an Open Enterprise" [9] and summarised as accessible, useable, assessable and intelligible. To this, we add the desirable property of discoverable. Consequently, all sites joining the CDI under the 2016 Collaboration Agreement are strongly encouraged to adopt open access policies towards their collections in return for the benefits of EUDAT replication and management services.

It is not accidental that EUDAT's position aligns very closely with the FAIR agenda for open data[1]. EUDAT is a fundamental supporter of the FAIR agenda for increased 'findability', 'accessibility', 'interoperability' and 'reusability' of research data.

In the final version of the Sustainability Plan from the Framework 7 'EUDAT' project [11] a number of common policies for EUDAT sites were defined which have helped to steer the ongoing development of the common data infrastructure. In particular, EUDAT has adopted the following policies and principles that are directly relevant to its open data agenda.

2.1. Openly discoverable

All data objects deposited in the CDI will be assigned a unique, persistent identifier (a "CDI-assigned PID") at a suitable level of granularity, and these PIDs will be communicated to the data depositor[2]. EUDAT adopts globally unique Handles to identify digital objects within the CDI. The Handle System [12], the system behind DOI [13] and other well-known identification mechanisms, is administered by the Digital Object Naming Authority (DONA) and is used worldwide. EUDAT works with the European PID Consortium (EPIC) [14] to ensure all data objects registered in the CDI receive a unique, persistent Handle.

EUDAT sites will ensure that resolution of a CDI-assigned PID results in common, defined and stable behaviour. A CDI-assigned PID should be all a user of EUDAT services needs to retrieve the associated metadata record and (where authorised) data object from any EUDAT site.

EUDAT sites will ensure that data deposited in the CDI are documented with an agreed common metadata baseline to support discovery, citation and provenance. EUDAT strongly encourages adoption of the OpenAIRE application of the DataCite version 3.1 mandatory metadata schema [16]. EUDAT sites that store data will ensure that all metadata records are discoverable by (at least) the EUDAT metadata catalogue B2FIND. Metadata collected by EUDAT services are made available using the OAI-PMH publication standard as recommended by the OpenAIRE Guidelines for Data Archives [17].

[1] See the FAIR principles, FORCE11 website (https://www.force11.org/group/fairgroup/fairprinciples) and Wilkinson et al [10].

[2] In this context, the assignment of a PID to a collection of related data can be said to define the term 'data object'.

2.2. Openly accessible

All data in the CDI should, in time, become full open access. Open access is the norm for EUDAT data. Nevertheless, where necessary or required, embargo periods for original producers are fully supported, on condition that such data become openly accessible when the embargo period expires.

EUDAT sites will ensure that metadata and (where authorised) data are accessible by users of EUDAT services over the Internet through common, defined and stable access methods.

EUDAT sites that store data will be agnostic to any particular data format or set of formats. Users of EUDAT services will be encouraged to deposit data in open (i.e. non-proprietary) formats appropriate to the content, but no format will be proscribed.

EUDAT's notion of open access follows closely the "Open Definition" [18].

2.3. Openly (re) useable

EUDAT sites will encourage depositors of data in the CDI to licence their data for open access under the Creative Commons Version 4.0 Attribution licence scheme (CC BY 4.0) [19]. EUDAT sites that store data will ensure that all rights associated with data objects within the CDI are respected and that access to data objects not openly licensed is subject to appropriate authorisation checks. The reason for recommending adoption of the OpenAIRE Guidelines for metadata is because of their requirement for including a rights statement in the DataCite metadata record for each accessible data object [16]. EUDAT recommends that such rights statements adopt the standard machine-readable forms defined by Creative Commons, thus facilitating automation of processing at data service level.

3. MANAGEMENT OF PERSONAL DATA – GDPR AND EPRIVACY

A major focus of these guidelines is on personal data protection legislation, not least because of the recently adopted EU-wide General Data Protection Regulation (GDPR) [20] on personal data which will come into force in May 2018. It is important that EUDAT service providers comply with the changes brought about by the GDPR. Our aim in this report is to prepare service providers in good time for the legislation in its current form, and review ways in which EUDAT already complies and the areas which will need to be updated. Quotations in this chapter are taken from [20] unless attributed otherwise.

It is worth nothing that, given the pan-European applicability of the GDPR, the correct and efficient handling of personal data will become a competitive advantage as it will become a major decision factor for those looking for service providers to handle personal data. We hope this will help motivate EUDAT service providers to work on the advice and recommendations in the document.

3.1. Data protection in the European Union

The General Data Protection Regulation replaces the provisions of the earlier 95/46/EC Data Protection Directive [21] and sets out the principles and conditions for processing personal data across the EU, as well as the rights of data subjects and the obligations of data controllers and data processors. From 25 May 2018 all organisations within the EU (and in other countries "where Member State law applies by virtue of public international law").

The GDPR identifies and defines certain key terms (Article 4 *Definitions*):

'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

'processing of personal data' ('processing') means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

'personal data filing system' ('filing system') means any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis;

'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;

'processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

'third party' means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data;

'recipient' means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not. However, public authorities which may receive personal data in the framework of a particular inquiry in accordance with Union or Member State law shall not be regarded as recipients; the processing of those data by those public authorities shall be in compliance with the applicable data protection rules according to the purposes of the processing;

'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

It is important to note that EUDAT activities do fall under the processing definition. However, personal data processing is not a large part of the work carried out in our data centres. We therefore have policy in place to ensure compliance with the limited data processing of this type that we do carry out as well as to prepare us for future instances where such data might be held.

Within the framework of the 1995 Directive it was for Member States to determine more precisely the conditions under which the processing of personal data was lawful. Under the 2016 Regulation these conditions have now been harmonised. Some, indeed many, areas of application are still derogated to Member States – notably, and relevant for EUDAT, the processing of data for research purposes (see Chapter 4) – but for most purposes the laws on personal data protection and privacy are now common across the EU, and we examine these below.

3.2. The rights of data subjects

Protection of personal data is recognised as a fundamental right under Article 8 of the Charter of Fundamental Rights of the European Union. The GDPR fully harmonizes EU data protection law in order to give individuals greater control over their personal data including in the following ways:

- The right to be forgotten (Article 17);
- Better control over who holds one's private data (Article 7);
- The right to switch one's personal data to another service provider (Article 20);
- The right to be informed in clear and plain language (Articles 12, 13, 14);
- The right to know if your data has been hacked (Articles 33 and 34);
- Clear limits on the use of profiling (Article 21);
- Special protection for children (Article 8);
- Privacy as the norm ('privacy by default' for users of information services, and 'privacy by design' as a principle for service providers).

3.3. Personal data and processing

Two key concepts trigger the data protection regime, namely personal data and processing.

3.3.1. Personal data

When processing data, it will be important for EUDAT to be aware of whether a person is identified or identifiable. There are some grey areas here: for example an IP address might be a marker of identification to an internet service provider but not to the layman. Secondly, data that may not constitute personal data by itself, might in combination with other data, enable identification. Other media such as photographs may also constitute personal data. Given the above it may be safer for EUDAT to assume that data processed in these contexts may constitute personal data and to follow steps to determine whether GDPR rules apply.

3.3.2. Special categories

As noted above, a stricter regime applies to special categories of data revealing, for instance, that constitute sensitive data; according to Article 9, these include personal data revealing:

- racial or ethnic origin;
- political opinions;
- religious or philosophical beliefs;
- trade-union membership, and;
- data concerning health or sex life.

Generally, the processing of such data are prohibited under Article 9 (1). Possible exceptions to this blanket prohibition are noted in Article 9 (2). EUDAT, as a general-purpose data infrastructure, cannot assume that it will not encounter data under these categories: archive recordings of data subjects discussing unusual lifestyles, for instance, might form part of a social science archive of which EUDAT preserves a replica copy. The handling of such data (provided this is allowed under one or more of the exceptions in Article 9 (2)) must be left as a matter for careful definition in a formal data handling agreement between the data controller in question and the relevant EUDAT service providers in their role as data processors (see below).

3.3.3. Processing

It is the processing of personal data that triggers the application of the GDPR. The definition of processing has been interpreted fairly broadly as noted above; consequently, it is safe to regard all EUDAT services as data processing services, and the analysis of the use of personal data within EUDAT can proceed accordingly.

3.4. Data controller and data processor

The GDPR emphasises that when dealing with personal data, organisations must identify the data controller and data processor. This relates to the obligations and liabilities under the Regulation which are primarily aimed at the data controller with some responsibilities for the processor. The roles of controller and processor are more formal than under the 1995 Directive; under the GDPR their relationship must be covered by a contract or other legal instrument. (This follows the current practice in Norway, for example.)

According to Article 4 (7), the controller is “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law”.

Article 4 (8) regards the data processor to be the “natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller”. As follows from the definition of controller (‘alone or jointly’), multiple parties might be considered to be the controller of the data, which are then regarded as “co-controllers”.

When a controller chooses a processor to process data on his behalf, this does not discharge the controller from obligations relating to the security of the data. Article 28 (1) GDPR provides that the controller must “use only processors providing sufficient guarantees” relating to the technical and organisational security of the processing of the personal data and he must also ensure that these measures are complied with.

Chapter IV of the GDPR sets out the general obligations of the controller and processor and their implementation of the appropriate technical and organisational measures such as making records of processing activities, security against risk, data breach communication and the designation of a data protection officer. The Regulation also encourages organisations to draw up a code of conduct.

Section 6.2 of this report examines the implications of the controller–processor relationship for EUDAT in greater detail.

3.5. Applicable law

An additional point brought in by the Regulation is the territorial scope of the data protection. According to Article 3, the Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not. The protection also applies to the processing of personal data of data subjects who are in the Union, by a controller or processor not established in the Union under specified conditions.

For EUDAT this will apply to service providers in two countries with data centres outside of the EU: Switzerland, Norway, (in time) the UK and certain international organisations. In both of these cases the service providers will need to operate under GDPR standards.

3.6. Principles and obligations

Any processing of personal data must comply with the six main principles provided for by Article 5(1) GDPR and shall be:

a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89 (1), not be considered to be incompatible with the initial purposes ('purpose limitation');
c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89 (1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

In addition, Articles 13 and 14 provide that the data controller must inform the data subject of, inter alia, his identity, the purposes of the processing of data, the recipients and categories of personal data, and the existence of the data subject's right of access and right to rectify his data.

For EUDAT it will be important to show that the service providers know of the principles and obligations mentioned above and that there are provisions to inform data subjects where Articles 13 and 14 apply. The classification scheme sketched in Section 7.1 is designed to assist service providers in this.

3.7. Legal grounds and consent

Any processing of personal data requires a legal ground. Article 6 provides a limitative list of six grounds that legitimise the processing of personal data:

a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
c) processing is necessary for compliance with a legal obligation to which the controller is subject;
d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

The key considerations for EUDAT here hinge on consent.

3.7.1. Consent – recording, managing, withdrawal

Article 4 (11) of the GDPR defines 'consent' of the data subject as meaning any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her. Conditions for consent are explained in Article 7.

An important consideration for EUDAT is that this informed and unambiguous consent for the use of personal data must now be sought by service providers, *and must be recorded* so that compliance can be demonstrated at a later date. An equally significant consideration is that a data subject may *withdraw* their consent at any time. This will have an impact on any future processing of the data in question, where, it is worth reiterating, storage is regarded as a form of processing. Take, for example, the case of the data subject of a voice recording stored in a social science archive which itself stores replicas with two EUDAT service providers. Should that data subject withdraw processing consent for the recordings, the social science archive must delete them and the EUDAT service providers must also delete the replicas they hold. Ensuring this is both possible and straightforward will require additional infrastructure to be put in place within the CDI. This infrastructure must, at the very least, be able to do the following:

- **for each** data subject within the collection of data objects processed by a service provider:
  - **for each** data object for which the data subject is a subject:
    - create a record of the data subject's consent for processing, where processing has been explicitly defined to mean, for instance, "storage for safe replication purposes";
    - delete both the record of consent **and** the object itself where consent has been withdrawn.

This may require extension to metadata records in service provider catalogues to ensure that, for a given data subject, all data objects in which they are a subject can be located, and deleted if necessary.

Note that deleting the record of consent ('the metadata') as well as the data object in question ('the data') follows the principle of a data subject's right to erasure (the 'right to be forgotten') of Article 17. Service providers will need to make clear, perhaps in the service's privacy notice, *how* data subjects can exercise their rights: an online form or a clear email address, for example.

3.8. The ePrivacy Regulation

The ePrivacy Regulation, currently in proposal form[3], aims to do for the 2002/58/EC ePrivacy Directive what the GDPR does for the 95/46/EC Directive. The Regulation follows from evaluations of the impact of the ePrivacy Directive, most recently in 2009, and seeks to update protections of citizens' privacy in light of both lessons learned and of technological developments in modes of online communication that currently fall outside existing privacy rules within the EU.

In essence, the ePrivacy Regulation extends EU privacy laws to all machine-to-machine and Internet-based electronic communication methods, including the newer "over-the-top" technologies that Internet firms offer on top of telecoms providers' infrastructure covered by the earlier Directive. At time of writing the proposed regulation has the following principal features:

- it covers all new players providing digital communications services in the Digital Single Market;
- it provides one common set of rules across the whole EU, with less Member State variation;
- it guarantees privacy of both communication content and metadata. In particular, a user's communication metadata may not be saved without the user's consent;
- it revises the (now infamous) "cookie consent" rules, aiming to streamline users' browsing experiences and remove the need for users to consent to "non-privacy invading" cookies;
- it outlaws "spam", unsolicited email, SMS and automated calling machines;

[3] Available from https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications

- it hands the enforcement of the confidentiality aspects of these new rules to the same data protection regulatory bodies responsible for overseeing the GDPR.

As a "new player" in the Digital Single Market, EUDAT will need to be aware of the provisions of the ePrivacy Regulation but the impact should be low. While EUDAT is not an electronic communications service provider it does use electronic channels, particularly email, to interact with users. EUDAT maintains an internal database of user requests, feedback, bug reports etc. – a "trouble ticket system" – which does store both content and metadata of individual communications. It would be wise to note this in the privacy notices associated with each relevant EUDAT service, and perhaps request and record a user's consent for this storage at the time of ticket submission.

Streamlining of cookie consent is only to be welcomed.

3.9. Article 29 Working Party clarifications

The Article 29 Working Party of European information commissioners ("WP29") has provided, and continues to provide, valuable clarifications of some of the terms and definitions noted here. In particular, we would highlight the following additional sources of relevance not only to the 1995 Directive but also to the new 2016 GDPR:

- WP29's opinion 4/2007 on the concept of personal data:
  - http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf
- WP29's opinion 15/2011 on the definition of consent:
  - http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf
- WP29's opinion 03/2013 on purpose limitation:
  - http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf
- WP29's opinion 06/2014 on the notion of legitimate interests…:
  - http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf
- WP29's opinion 05/2014 on Anonymisation Techniques:
  - http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

4. RESEARCH USE OF RESTRICTED DATA – CURRENT PERSPECTIVES

The use of personal data for scientific or historical research purposes is recognised in the GDPR as a special case that is best served in law by particular legislation at Member State level, and by "soft law" codes of conduct within individual research disciplines. The GDPR's Article 89 states:

Article 89 (2): "Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21…" ('right of access', 'right to rectification', 'right to restriction of processing' and the 'right to object').

Article 89 (3): "Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21…" ('right of access', 'right to rectification', 'right to restriction of processing', 'notification obligation [on the data controller]', 'right to data portability' and the 'right to object').

There are two consequences here for data processing research infrastructures like EUDAT. The first is that the final picture of how a research data subject might request changes to the way their data are processed is unclear; the second, and more insidious, is that the final picture may *remain* unclear should Member State legislation introduce differences to the way data subjects' rights are handled in this particular context.

Further, Article 40 notes:

Article 40 (1): "The Member States, the supervisory authorities, the Board and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper application of this Regulation, taking account of the specific features of the various processing sectors and the specific needs of micro, small and medium-sized enterprises."

Article 40 (2): "Associations and other bodies representing categories of controllers or processors may prepare codes of conduct, or amend or extend such codes, for the purpose of specifying the application of this Regulation…"

Further, Recital (98) notes: "In particular, such codes of conduct could calibrate the obligations of controllers and processors, taking into account the risk likely to result from the processing for the rights and freedoms of natural persons." This emphasises the principle of "balance-of-risk" prevalent throughout the GDPR.

In the European research context a number of the major research infrastructures, particularly those formally incorporated as ERICs or other legal personas, and particularly those in life-science domains, have naturally taken on the roles of "… bodies representing categories of controllers and processors" and have initiated forums for the drawing up of "GDPR-compliant" codes of conduct[4]. The significant challenge here is one of timing. At time of writing, these particular derogations are (generally speaking) still in formulation at Member State level, and discussions on subject-specific codes of conduct are, not unnaturally, somewhat in abeyance awaiting the outcomes of Member State legislation. It is thus difficult to draw up any significant recommendations at this stage.

Where detailed rules for the management of research data are currently missing, EUDAT can and must recognise that its services will need to assume an environment governed by restrictions on the processing of personal data. Engaging with key community stakeholders in the co-design of "compliant" services is a natural way forward. It is not EUDAT's role to enforce personal data privacy across the varied European research landscape but rather to provide a data infrastructure able to support codes of research conduct and patterns of information governance across a broad stakeholder base.

[4] See, for example, the work-in-progress on the BBMRI code of conduct:
http://www.bbmri-eric.eu/news-events/code-of-conduct-for-using-personal-data-in-health-research/

To use an example with which the authors are familiar, SHIP, the Scottish Informatics Programme[5] (originally the Scottish Health Informatics Partnership), has developed a proportional, balance-of-risk approach to the use of personal data in research, anticipating the approach taken by the GDPR. The GDPR highlights the "principle of proportionality" in applying rules on personal data, noting "the right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights" (Recital (4)). SHIP's original remit was to design a governance framework around this balance-of-risk principle to manage in particular the safe *linkage* of multiple personal datasets to support research in public health and social policy.

The SHIP governance model is based around the assessment of five benchmarks: public interest; safe people; safe data; safe environment; and, consideration of relative risk (across the preceding four). These five headings are useful touchstones for EUDAT's future design decisions; while EUDAT provides an infrastructure, not a full-blown information governance body, the ideas of 'safe data', 'safe people' and 'safe environment' nevertheless offer useful pointers.

'Safe data' requires that data be handled within the CDI with due regard for the protection of the privacy of any data subjects. EUDAT needs to design services which are able to manage personal data accordingly. The DataTags approach to "tagging" sensitive data objects objects at a designated level of granularity within the CDI is one suggested approach; we examine this other organisational and infrastructure considerations in Chapter 7.

'Safe people' suggests at the very least a basic level of training for CDI service operators is both appropriate and necessary. EUDAT service providers need to be aware that personal data may find their way onto their systems, and accordingly how to handle them.

'Safe environments' is something to consider in designing additional services that might "sit on top" of the CDI data layer. Analytics, data joining or query services, for example, if operating on personal data, might increase the risk of identification or personal data breach. EUDAT may wish to consider the introduction of services following the "safe haven" approach of secured environments for research on sensitive data[6].

At bottom, though, assuming EUDAT intends to support a CDI that is able to host "non-public domain" data it must continue to engage with community partners as research codes of conduct develop. These codes of conduct, constrained necessarily by the GDPR but encompassing the appropriate proportional balance needed by research, will form the most relevant data governance frameworks for the future CDI.

[5] See http://www.scot-ship.ac.uk/publications.html

[6] See, for example, https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens for a list of data safe heavens operated by the UK National Health Service in Scotland. The Scottish National Safe Haven is operated by EPCC at the University of Edinburgh under contract to the eDRIS unit of NHS Scotland.

5. ETHICAL CONSTRAINTS ON OPENNESS

Personal data are by no means the only data which need sensitive handling. A 2014 report from the EU RECODE project [8] on *Legal and ethical issues in open access and data dissemination and preservation* classifies ethical concerns about the openness of data under five headings: unintended secondary uses and misappropriation; dual use; violations of privacy and confidentiality; unequal distribution of research results; commercialization, and; restriction of scientific freedom. Chapter 3 of that report offers good examples of each of these areas of concern, and notes that the principal mechanisms for handling data in such circumstances arise from the associated scientific or research communities and published codes of conduct (cf. Chapter 4). In this regard, EUDAT's approach of building collaborative data infrastructure involving both generic service providers and community-oriented data repositories is well placed to incorporate the necessary fine detail of research sensitivity into the design of data services.

This might suggest that the management of sensitive data must by necessity be dealt with on a case by case basis through, for example, research ethics committees. While it is certainly true that ethics committees and codes of conduct will play a defining role, there are two broad areas in which the CDI infrastructure and services could potentially be developed to support more easily the management (including the automated management) of sensitive data. These we can usefully term "temporal" and "spatial".

5.1. Temporal restrictions on openness

Support for open and restricted data cannot be static. There are temporal dimensions and concerns that should be considered in order to implement, effectively and appropriately, the GDPR (by, for example, adopting a GDPR-harmonized DataTags system) and, more broadly, to ensure a comprehensive support policy for managing open and restricted data access in the CDI, now and in the long term.

While promoting open access to data as a default position (where possible), EUDAT does support restricted access to data, both in terms of administrative data (as defined in Section 6.1.1) and content data (Section 6.1.2), and enforces this in specific cases. Data access restrictions typically fall under three categories:

1. access restrictions based on (admissible) user requests (e.g. embargo);
2. access restrictions required by laws and regulations (in particular regarding personal data);
3. access restrictions based on ethical considerations.

As a collaborative infrastructure, EUDAT is ready to accommodate, apply, and enforce access restrictions upon submitted data where appropriate. EUDAT is not a "dark archive" but an infrastructure of "living data". Further, following from EUDAT's stated policy that "*all data in the CDI should, in time, become full open access*"[7], EUDAT should assume that at a certain time in the future, data submitted to the CDI might change their status from "restricted access" to "open access". Consequently, considerations about "when" and "for how long" access restrictions shall be permitted and enforced must necessarily be reflected explicitly in EUDAT's regulations and policies. These policies need also to be actionable at the machine level to enable, enforce, and monitor access constraints and their change over time.

We go on discuss more specific considerations according to the three categories above.

5.1.1. User requested embargo

When it comes to research data there are admissible constraints on openness, to support, for example, an appropriate, time-limited, privileged exploitation of datasets so that contributing researchers can publish the results of their research (this type of constraint is typically called "embargo"). Even when admissible, embargo durations may vary according to specific policy provisions (e.g. allowed by a funding agency) but usually take the form of a relative time span (e.g. 18 months) with a start date that, again, varies according to policy provisions.

[7] c.f. Section 4.2 of [11].

As of version 2.0 the B2SHARE simple repository service supports an "open-after" embargo date field in submitted metadata[8], and design work is underway to propagate this concept into the lower-level infra-structure services B2SAFE and B2STAGE. EUDAT service providers will, ideally, need to implement means of capturing such embargo information in order to provide (possibly automatic) mechanisms to release the embargo when data are supposed and/or required to become openly available. These policies and procedures must also be explained, made available to, and explicitly accepted by the agents submitting data to the CDI.

Consideration on how to implement embargo and embargo-release mechanisms are left to individual implementations, but a number of common features should be enabled:

1. it should be possible to accept embargo requests either as time spans (e.g. 18 months) and absolute (future) dates and times (e.g. 31 December 2017, midnight);
2. time spans and dates must be recorded in machine processable formats (e.g. following RFC 3339 and ISO 8601). This might imply translating time spans into absolute dates;
3. services should support automatic embargo release mechanisms, with notifications sent to the user that data have been released as open access;
4. if adopting the DataTags system for managing embargo periods, tag implementation shall always be timestamped (embargo, by definition, is never indefinite).

5.1.2. Personal data through time

As discussed throughout this report, personal data must be stored and processed in the CDI according to the current GDPR regulations. The application of the DataTags system, harmonized with the GDPR, to classify data within the EUDAT CDI provides an efficient solution to translate natural language statements into formal, decision-tree driven, and machine-actionable tags that reflect a finite set of precisely defined categories of data restrictions and their characteristics. The DataTags system though, does not inherently provide an explicit and declarative way to represent "for how long" a certain restriction or constraint shall be enforced, nor any means to release the data if and when certain restriction cease to apply. Changes in status shall be admitted, enabled and monitored within the CDI according to the GDPR regulations.

The most explicit case of an admissible change in status of a dataset that include personal data is when a data subject dies. In this case, GDPR restrictions cease to apply[9]. While we cannot predict any future timeframe in this case, there needs to be means to record (e.g. in the associated metadata) when and why changes can be admitted and actually applied. Another related example is that categories of personal data collected and processed for research without explicit consent might include children's data and that will entail special considerations in management. Children, of course, grow up; a person ceases to be a child at some point in time and therefore the additional processing restrictions arising from special considerations cease to be required by law. Further complication might arise from national legislation and the legal concept of "child" which might be differently conceived (e.g. different legal positions on when a person ceases to be a child).

In this case again, consideration on how to implement and provide these mechanism are left to individual implementations, but a number of common features should be enabled (as above):

1. the indication of time spans and dates must be recorded in machine processable formats (e.g. following RFC 3339 and ISO 8601). This might imply translating time spans into absolute dates;
2. services should support automatic embargo release mechanisms, with notifications sent to the user that data have been released as open access;

[8] See B2SHARE user documentation at https://eudat.eu/services/userdoc/b2share-usage

[9] Although GDPR Recital (27) allows for the additional regulation of post-mortem personal data at Member State level.

3. if adopting the DataTags system for managing embargo periods, tag implementation shall always be timestamped (embargo, by definition, is never indefinite).

5.1.3. Ethical issues in the time-based release of restricted data

Personal data of a deceased person, even if not falling any longer under the GDPR, might need to remain under access constraints due to impact on family and/or future generations, or on other members of society. In certain world cultures there can be restrictions on viewing the likeness of the deceased, or even speaking their name. The indigenous Yolngu of Elcho Island, Australia, for instance, hold any images of the deceased as taboo for a certain period of time determined often by the deceased's family. (For a good discussion about the sensitivities around video data collection among Elcho Islanders, see [25].) This points towards a possible "reverse embargo" approach; certain data may be accessible before the death of a data subject, then subject to embargo for a period of time, then openly available again. Whether the management of data under such conditions can be fully automated or whether it must be handled on a case-by-case basis, it suggests that the time-stamping data for embargo purposes needs to be flexible enough to accommodate multiple possible patterns.

5.2. "Spatial" restrictions on openness

A significant number of the examples quoted under RECODE's heading 'unintended secondary uses and misappropriation' involve geographical sensitivities of one form or another: sites, for example, of cultural, archaeological, zoological or botanical importance. Recording and publishing data openly on the precise location of the few families of mountain gorilla in Uganda, for instance, increases the risk to those endangered animals of death by poaching.

One important point to recognise here is that geographical sensitivity may appear in a dataset's *metadata* rather than its data content. A series of observations of the hunting times of a nesting pair of ospreys, for example, poses no risk to the nest; a metadata record of *where that nest is* does. This underlines the notion that "one person's data is another person's metadata" noted in Section 1.2.

Whether or not to publish data like these is not a decision for EUDAT CDI infrastructure providers; clearly this is a research-ethical decision. Nevertheless, there are a number of points for EUDAT service designers to consider:

1. where data or metadata may be "spatially sensitive", ensure no overly-precise location data leaks into publically available records such as Handles or metadata catalogues. The definition of "overly precise" is likely to be case or community specific (unfortunately);
2. consider the introduction of standard ways of recording spatial location that are easy to render less precise, for example by numerical filtering on metadata queries, or by recording location information at different granularities protected by different levels of data tag (cf. Section 7.1). A simple significant figure filter on (latitude, longitude) pairs could be applied to all metadata queries, for instance, or a mechanism like geohashing could be adopted[10].

[10] Geohashes offer a way to encode locations on Earth into short URLs in a way that supports "gradual precision degradation". See https://en.wikipedia.org/wiki/Geohash and http://geohash.org/.

6. POTENTIAL IMPACTS ON THE EUDAT CDI

6.1. Impacts on general data handling

In assessing the impact of the GDPR on EUDAT data services and data stored across multiple sites, we begin by classifying data in the CDI to be either *content* – data uploaded by EUDAT users into a service for storage or other processing – or *administrative* – data collected (perhaps automatically) by a service as part of its normal operation.

6.1.1. Administrative data

Administrative data are those which are collected directly from users of EUDAT services, including PIDs, email addresses, accounting data, login names, IP addresses, file checksums and other file attributes, and so on. They are under the direct responsibility of – and the direct control of – EUDAT service providers and are, in many ways, the easier type to deal with.

Our primary concern here is with personally identifiable administrative data from EUDAT service users. We assume that an EUDAT service user is not deceased for these purposes (a reasonable assumption). Whether these data are automatically collected (e.g. by logging IP addresses from service requests) or user-supplied (e.g. a login name or email address), key considerations come down to three points.

Legal basis – whatever personal data EUDAT services are collecting, the service provider must be able to demonstrate a legal basis for doing so, and this must be reflected in a clear, public *privacy notice* for each service and each service provider: "we collect the following pieces of personal data for the following reasons". Certainly none of the personal data that EUDAT service providers might require should fall into the *special categories of personal data* (so-called "sensitive data") – race, religion, sexuality etc. In almost all cases, an EUDAT service provider's legal basis for processing personal data is likely to be *consent*.

Consent – the key to compliance with the GDPR is seeking consent from users for the specific use of their personal data. This can be achieved by obtaining a user's consent to the privacy notice, thus defining the legal basis for data collection. Another key point is that consent cannot be assumed from use; consent must be actively sought (e.g. by a tick-box or active click), and must be recorded in order to demonstrate compliance with the GDPR. A mechanism to record and act upon the withdrawal of consent must also be enabled for each service.

Age – children, meaning a person up to an age of between 13 and 16, depending on jurisdiction, cannot give consent under the GDPR. If a child creates a login account on the B2ACCESS service, the EUDAT service provider must take "reasonable steps" to ensure that they (the service provider) have consent from that child's *legal guardian* for them to do so lawfully. This raises a significant issue for EUDAT services (see later).

All administrative data collection is also subject to the blanket *personal rights* of a user: the right of access (to any personal data stored); the right to be forgotten (to have personal data deleted); the right to move elsewhere; the right to be informed of hacking.

The challenges of complying with some aspects of this framework will colour the way EUDAT services evolve in the future. This is, of course, in line with the GDPR's *principle of minimisation*: data services should be designed to collect or process the minimum required set of personal data to deliver their service (see below).

6.1.2. Content data

Where content data has no data subject and no personal data records as part of it, or if the data subject in question is deceased, then storage or other processing is not an issue under the GDPR, even if, in the latter case, ethical considerations might still apply and need to be taken in to consideration within the EUDAT CDI.

Personal content data face similar challenges to administrative data but add a few more of their own. To begin with can we divide content data into two classes: those with a data subject, and those without. Our definition of content data here includes user-supplied metadata of any kind; metadata are data too.

Thus, previously impersonal data (measurements of rainfall, for example) may be rendered personal by the addition of relevant metadata (measurements of rainfall made on 17 June 2016 by Dr John Smith, submitted for archival on 28 July 2016 by Dr Jane Doe). The contributor may add their own name and contact details when they create or upload a content file (e.g. as DataCite *creatorName* and *nameIdentifier*). These personal data will then appear in the metadata associated with the initial impersonal data, and consent for the lawful storage of personal content metadata like these must be obtained in the same way as for administrative metadata noted above.

Content data with a data subject can raise additional issues on top of the three principal issues – legal basis, consent and age – discussed in the previous section. These are: special categories of personal data; and processing for research or statistical purposes. They are related.

Special categories – special categories of personal data cannot easily be processed by general-purpose data services like those in the EUDAT CDI without important specialisation. Where a community data provider (e.g. in the social or medical sciences) has data of this nature and wishes to use the CDI for data storage and management, a specific solution will need to be designed between them and a designated service provider, following the principle of minimisation. Even where service providers have obtained specific consent for the storage or processing of data of this nature, strong arguments can be made against storing or processing them on any Internet-connected system such as the CDI; the impact of leakage or unauthorised access is extremely high.

Research use – the principle of minimisation also applies to the storage of personal data for historical, statistical or research purposes (we use the term "research" to cover these specific adjectives from the GDPR). Storing data for a particular data subject (e.g. a voice recording) for research purposes covers the fact that specific consent for all possible future reuse scenarios for those data cannot be obtained at the time the data are recorded; it does provide a get-out clause against minimisation principles. This means that EUDAT may need to design specific services to store personal data of this nature (e.g. end-to-end encryption with a user's public key, requiring no shared secret between user and service provider).

It is difficult to offer more specific guidance on these topics here. As noted in Chapter 4 the GPDR is non-specific; detailed provision of data-use in research is derogated to national lawmakers and community guidelines. There is, of course, as yet no case law to which to refer. Also, confidentialisation techniques for different kinds of data quickly become technologically specific[11]: de-identifying medical image data stored in DICOM formats, for instance, is both a topic of active research and an expert area for sophisticated specialist companies. There is no easy answer, but EUDAT's open culture of service building with community drive lends itself well to future co-design of suitable services.

6.2. Impacts on organisation: Data Controllers and Data Processors in the CDI

In assessing the impact of the GDPR on EUDAT CDI, the roles of the EUDAT data service providers with regard to the data handled must be clearly identified. While the ownership of the data is unequivocally assigned to the data subject, discerning the role of a service provider in term of Data Controller or Data Processor is not always simple, but it is crucial in order to adopt the correct measures to comply with the GDPR.

In the simplified scenario shown in Figure 1, the data controller is the person, organisation, authority or agency who determines the purposes for which, and the manner in which, any personal data are processed. Prior to collecting the data, the data subject must give his or her consent to the collection of a given set of data, for a given purpose and a given time. The data controller might process the data or engage other service to process the data on its behalf. The data processor is the person, organisation, authority or agency who processes the data on behalf of the data controller. The data agreement signed by both parties (data controller and data processor) ensures that the data are processed according to certain standard and states the relative roles and responsibilities of the two parties. An example of a data agreement is given in Annex

[11] We use the term "confidentialised data" after the 2016 ANDS report *Publishing and sharing sensitive data* [24]: "when data has been modified to remove or reduce the risk that people or subjects of the data can be identified".

B. Often the data controller and data processor are the same institution, and therefore a data agreement is not needed. Furthermore a service provider that acts as data controller for a given set of data might be the data processor for other data.

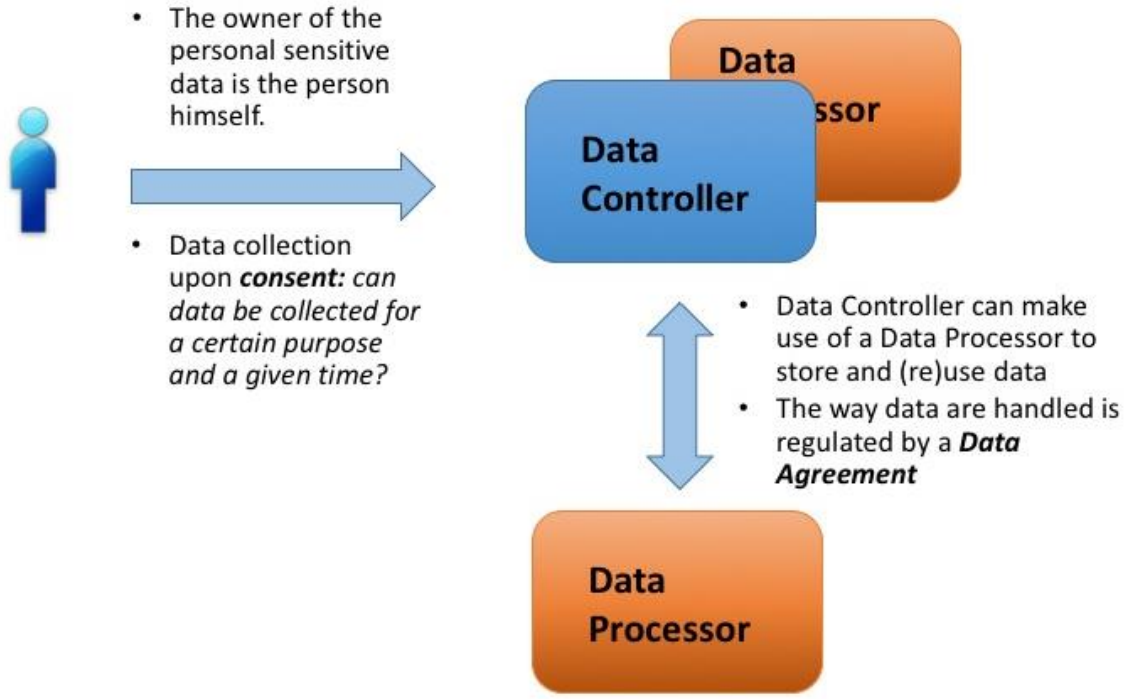

Figure 1: Simplified picture of the relationship between data subject, data controller and data processor.

The main significance of the GDPR for EUDAT at an organisational level is the requirement to formalise in law agreements between data controllers and data processors. Under the 1995 Directive, controllers and processors had to have a data agreement; under the GDPR this agreement must now be in the form of a contract or equivalent legal undertaking.

As noted in the general case above, service providers in the EUDAT CDI will take one, or both, of these roles. Figure 2 indicates potential flows of personal data (both administrative and content-specific) between principal EUDAT service providers, and the likely roles this will require from each.

In general, it is safe to assume that non-user-facing services like B2SAFE and B2HANDLE will operate solely as data processors. User-facing services that require user authentication – B2ACCESS and B2SHARE – will almost certainly operate in the role of data controller through their handling of user identity data. B2FIND, while user-facing, does not require authentication and can be used anonymously, although administrative data such as IP addresses can, in principal, be collected for statistical purposes. However, since B2FIND does not transmit non-confidentialised data onwards for further processing, we do not categorise it as a data controller.

The requirement for formal contracts between these service providers needs to be included in the future legal framework of the EUDAT CDI.

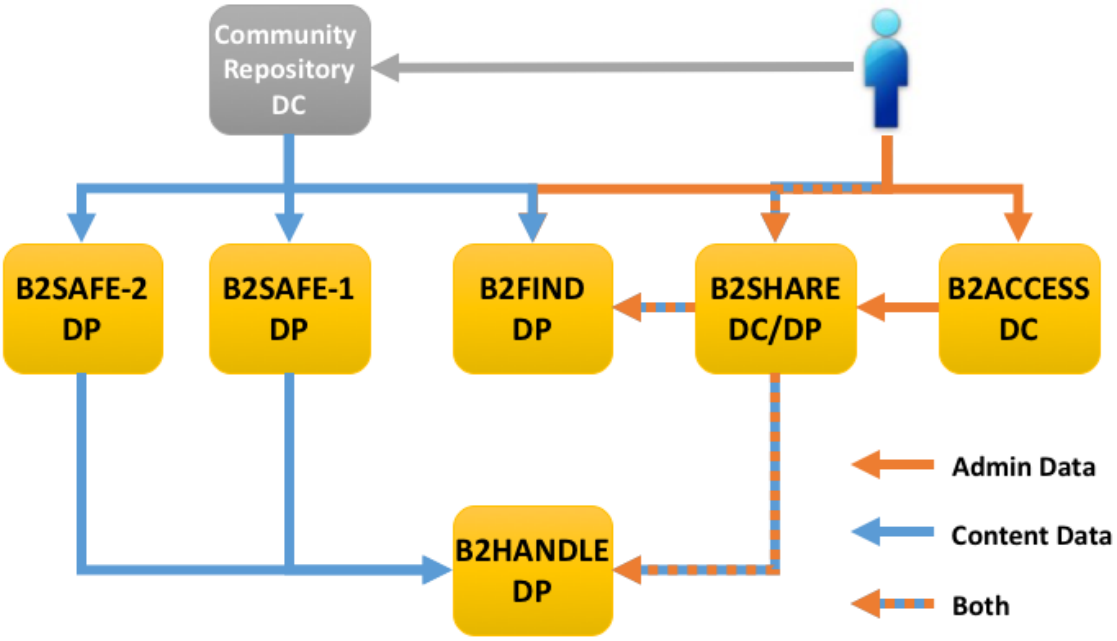

Figure 2: Potential flows of personal data between EUDAT service providers (arrows indicate "from-to"). Orange denotes administrative data, as collected directly by EUDAT services; blue indicates content data, which could include personal data of either the illustrated user or another data subject. Gold boxes indicate EUDAT services (properly, service providers). The grey box illustrates a community data repository making use of EUDAT services. Service providers are labelled DP if they take the role of data processor, DC if data controller.

6.3. Impacts on specific EUDAT Services

The GDPR will impact generally the way EUDAT CDI services are designed, implemented and operated. In this section we summarise the probable impacts on each of the B2- services specifically, and offer some initial guidelines on future service design. Note that these are initial assessments and guidelines, not points of law; the actual impact of a number of these issues may not be known until the GDPR has been tested in the courts – hopefully by entities outwith the EUDAT CDI Partnership!

6.3.1. B2ACCESS

Consent for user identity management: B2ACCESS, the common authentication service, is likely to see the largest impact, given that it manages personal data as a matter of design. Consent must be obtained from each and every user of B2ACCESS, and must be recorded. It must be made clear what personal data are collected and for what purpose; this must be explained in a clear *privacy notice*.

Age: As noted, children are unable to consent to use of their personal data. While it's unlikely that a child would be interested in creating either a federated or indeed simple identity on B2ACCESS, nevertheless some form of age verification for the service may be inescapable. How this could or should be done is unclear. As a guard, a disclaimer could be included in the privacy notice stating that EUDAT services are not intended for use by children.

Consideration of identity federation: B2ACCESS can combine user identities from multiple sources. This almost certainly brings it under the scrutiny of the minimisation principle: a data breach of such multiple identities is more serious than that of a single identity, and so thought must be given to how user identities are recorded in the service databases. A security assessment should be carried out to help understand the risks.

Data controller – data processor: B2ACCESS both federates user identities from external providers and allows users directly to create an "EUDAT" identity. This almost certainly places the B2ACCESS service provider in the role of both data processor to the identity providers' data controllers, and data controller to

other EUDAT sites making use of the service for single sign-on. These controller-processor relationships will need to be codified in contracts or other legal instruments.

6.3.2. B2SHARE, B2DROP

Consent: B2SHARE and B2DROP offer user-facing web-sites requiring authentication, either via B2ACCESS or directly. In either case, consent for use of personal data (logins) must be obtained and recorded. A clear *privacy notice* must be in place for each site (ideally a common policy with B2ACCESS).

Age: As noted in B2ACCESS, age verification for consent is a potential issue. Given that B2SHARE is designed for "long-tail" science and "citizen science", and one potential group of citizen scientists is classes of schoolchildren, the likelihood of this occurring is perhaps higher than we might expect. Devolving all authentication issues to B2ACCESS is one way to collect the problems – and solutions – in one place.

Content: Users can upload anything into B2SHARE; some of this could be personal data, potentially even special categories of personal data. There is no policing of content, nor is there likely to be in the future. B2SHARE should publish a clear *Disclaimer* to this effect as part of the privacy notice (cf. Annex A).

6.3.3. B2FIND

Content: As with B2SHARE, the B2FIND service provider has little or no control over personal data that may find their way into the service. Consequently, a clear *Disclaimer*, possibly common with other services like B2SHARE, needs to be in place on the site. The current service provider, based in Germany, provides a disclaimer in German, noting that the ruling law is that of the Bundesrepublik Deutschland; a non-binding English translation should be provided alongside.

6.3.4. B2SAFE, B2STAGE

Content: As a multi-site service, the operation of B2SAFE between, say, a community site and one or more data centres is or will be covered by service level agreements (SLAs). Where data to be replicated or otherwise processed might be personal data, these agreements will need to reflect appropriate minimisation principles, probably on a case-by-case basis. Any data transfer mechanisms should be secure (ssh; sftp; https); data may need to be stored encrypted. There is almost certainly no one-size-fits-all solution here.

Data controller – data processor: If a community site has collections which include personal data (e.g. interviews with data subjects) then they automatically take the role of data controller for those data. An EUDAT CDI site receiving those data by onward transmission through B2SAFE, for instance, automatically takes the role of data processor; the B2SAFE agreement between controller and processor (between community site and data centre) must now be codified in a contract or other legal instrument; a sub-legal agreement is no longer sufficient.

6.3.5. B2HANDLE

Data controller – data processor: B2HANDLE is currently used behind the scenes by a number of other services for PID creation and has no user-facing interface. However, *if* the B2HANDLE PID schema records any personal data (e.g. *creatorName* from the DataCite schema) *then* this would potentially place the B2HANDLE service provider in the role of data processor to a PID-requesting data controller, requiring a contractual agreement between the two parties.

Consent: Again, *if* the B2HANDLE PID schema records any personal data, explicit consent must be obtained from the relevant data creator. The request for consent must make clear that this personal data will propagate into the global Handle System.

6.3.6. Future Services

The minimisation principle must be applied to all future service design, not only as necessary to comply with the law but also as a defensive mechanism. Consider, for example, the propagation of (*creatorName, nameIdentifier*) into 1,000 or more Handles and the impact of a subsequent invocation by that creator of their right to be forgotten.

7. TECHNICAL POLICIES FOR RESTRICED DATA

The implications of broadening the remit of the CDI beyond purely open access data are significant but not insurmountable. In this chapter we suggest three practical approaches to handling personal, sensitive or otherwise restricted data in the CDI that can help in particular streamline data management under the GDPR and support service providers in meeting their new obligations.

7.1. Classification: using DataTags to classify CDI data

GDPR Article 4 (3) defines 'restriction of processing' as the marking of stored personal data with the aim of limiting their processing in the future. In the first draft of this report [6] we introduced an *ad hoc* data classification scheme as a first step towards meeting this requirement. In this report we build upon this idea of a classification scheme and explore adoption of the Harvard DataTags system for data within the EUDAT CDI.

In the DataTags system, a *data tag* is a label indicating the level of protection to which a data object should be subject within a repository (or elsewhere). This system was originally developed at Harvard University [26], and the notion of a *DataTags repository* was introduced by its creators Sweeney, Crosas and Bar-Sinai as a facility which stores and shares data files in accordance with different levels of security, access requirements and auto-generated data use agreements. The system basically defines security features and access requirements for handling sensitive data. The original American system uses six levels of access from blue (public data) to crimson (highest level of restriction) and is modelled on the various U.S. privacy laws. Effectively, the DataTags system informs the technical infrastructure of handling requirements that are needed for a given data object by attaching a data tag to it, taking into account the specific legal obligations for processing these data.

Sweeney, Crosas and Bar-Sinai define a DataTags repository as a repository of files held for data sharing that satisfies the following conditions:

1. A data tag is a set of security features and access requirements for file handling. A DataTags repository has a finite, partially ordered set of data tags, where the strictness and strength of data tags' security features and access requirements dictate the ordering. A repository must have more than one data tag.
2. All files in the repository must have a data tag, and each file in the repository has one and only one data tag. A file may optionally have additional handling requirements, such as an audit trail log or an expiration date. A file may optionally require additional terms for a data use agreement or additional terms of access by a recipient of the file from the repository. A file may have attributes that further describe it for reporting purposes. None of the optional requirements may weaken or replace the security requirements for the file's assigned data tag, and none may adjust a data tag's security requirements to be the same as another data tag or stronger than a more restrictive data tag.
3. A recipient who receives a file from the repository must satisfy the file's associated access requirements, produce sufficient credentials as requested, and agree to any terms of use required to acquire a copy of the file.
4. Technological guarantees exist that the requirements in 1 and 2 are satisfied for all files in the repository and for all accesses to those files from the repository. This imposes auditing obligations on transactions in the repository.

The Harvard authors discuss the notion of using a flow chart or decision tree approach to arriving at a given tag for a particular file or data object (we use the terms interchangeably here), thus describing a way to generalise the DataTags system for other privacy and personal data frameworks. In assessing the use of DataTags for EUDAT we have, of course, based our analysis on the framework of the GDPR. Our analysis is based on a pilot study carried out by DANS during summer 2017.

The first issue to decide on was how many tags would be appropriate for this tool based on the requirements of European law. Building on the GDPR it proved possible to identify the need for four DataTag levels: blue

(public access) for non-personal data, green (basic access) for personal data which is not sensitive, yellow (restricted access) for personal data which contains sensitive information[12], and red (selected access) for personal data which contains highly sensitive information[13]. Additional requirements under each tag can be seen in the table below.

Table 1: DataTags based on the GDPR classification.

Tag type:	Authentication	When transmitted	When stored	Reading/downloading rights
0. Public access (non-personal data)	None needed	Without encryption, with checksum	Standard - clear storage	Everyone (with or without registration)
1. Basic access (non-confidential personal data)	Registration necessary	Without encryption, with checksum	Standard - clear storage	All registered users
2. Restricted access (sensitive personal data)	Registration via repository and approval of depositor	With encryption, with checksum	Standard - clear storage	All registered users, *after approval of depositor*
3. Selected access (highly sensitive data)	Registration via repository and mandatory further identification	Multi-encryption, with checksum	Not accessible via the internet and with encryption	NOT via repository, checked users only

The next step in the pilot was to codify the GDPR law into a set of questions which users can answer in order to establish the level of privacy protection their personal data need. The question and answer path of this tool results in the creation of a whole decision tree leading to a recommended data protection level. Added to this is a description for the user of what that protection level implies for the storage and sharing of the data. This recommendation will finally be displayed as a data tag at the end (see Figure 3).

The pilot leads us to conclude that a classification scheme based on DataTags for types of data, personal and otherwise, could be used as a tool in assessing the risks associated with processing certain data within the CDI. We have to bear in mind here the principal feature of the CDI as an Internet-connected distributed platform for research data with a mandate for open access.

This means that the use of a uniform system containing simple, clearly defined tags, could be used throughout all EUDAT sites and services. The model is certainly not yet final: some choices are open for debate. A more operational point in designing the decision tree is from which perspective the tree should start: the researcher or the site/repository. This is however not a fundamental obstacle. A more complex issue is how

[12] The word sensitive is used here in the meaning of "special categories of personal data" as defined in Article 9 (1).

[13] "Highly sensitive" data are data which should be protected on a higher level than "normal" sensitive data, because of the possible vulnerability of the data subjects. This is NOT a legal category, contrary to sensitive (= special categories of) personal data. Examples are oral history interviews with people who were engaged in recent warfare, former psychiatric patients etc. The key qualifying point is often that people do want to run the risk of being recognised, except by academic researchers under certain conditions and warranties.

to handle the exceptions and derogations for processing personal data for historical and statistical purposes as well as research, highlighted in Chapter 4. The problem here is that the exact formulation of these derogations is not yet clear as this is left to national legislation and not yet decided upon. This will call for further work on extending the CDI DataTags system as these issues are clarified over the next 12 months.

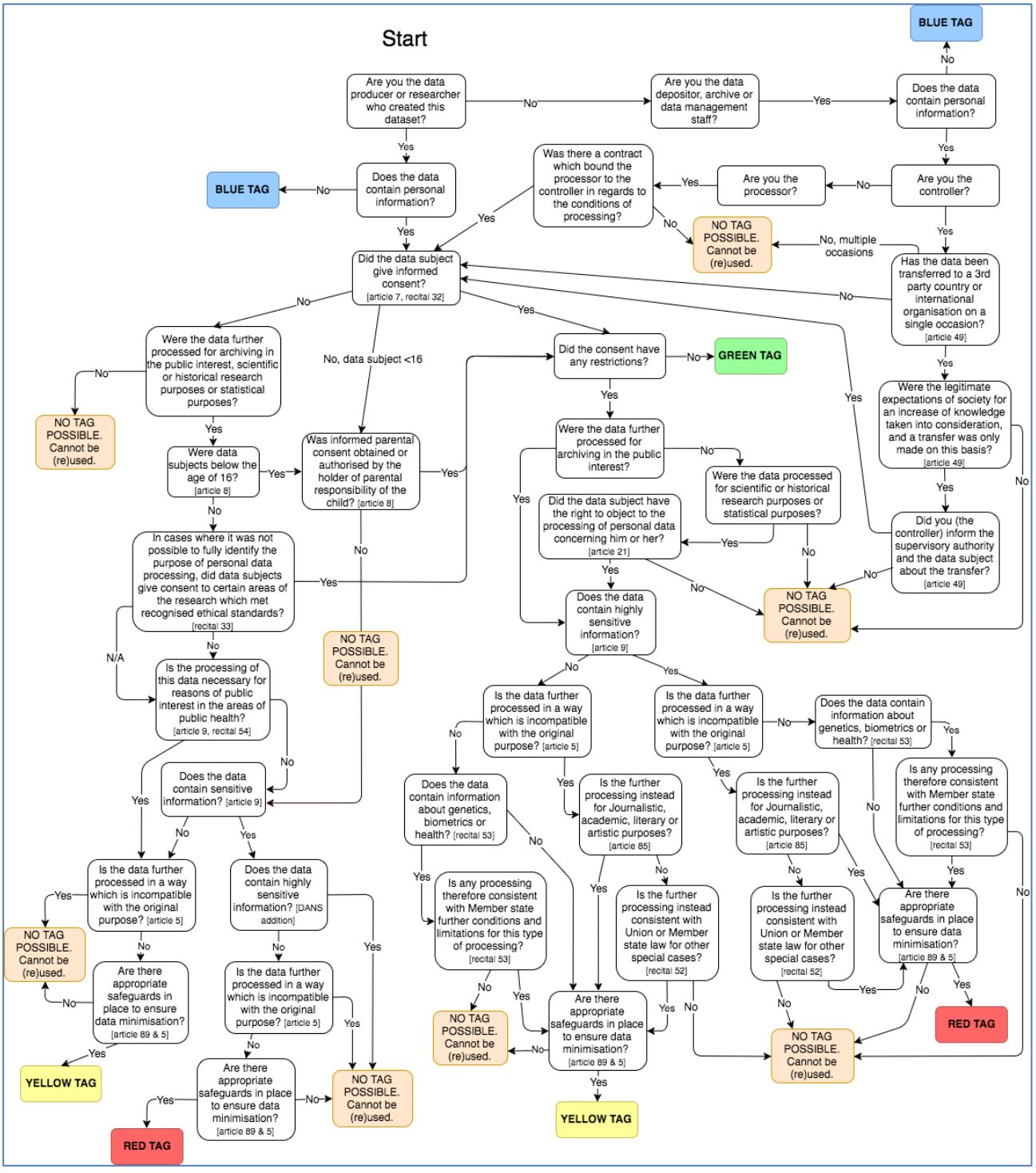

Figure 3: A DataTags flowchart based on the GDPR.

It should be noted that this classification only covers personal data as defined under the GDPR; it does not yet assess data which may be restricted or sensitive for other reasons, nor does it address copyright or other intellectual property rights. However, extending the DataTags system to other areas of sensitivity, like data

which have to be kept secret or confidential for commercial or state-security reasons, is straightforward. A data tag is an indicator to the data processing infrastructure on how to handle the accompanying data object; the DataTags systems separates the concerns of *how the tag was decided upon* from *how the object should be managed*. In this sense it forms a canonical interface between policy and implementation, and one that already has broader use beyond the EUDAT CDI.

7.2. Pseudonymisation: designing services around the rights of data subjects

As the final picture of the rights of data subjects whose data are stored in the CDI for "scientific or historical research purposes" is still to emerge (see Chapter 4), EUDAT would be wise to consider the new rights of data subjects as defined in Chapter III of the GDPR and consider the impact on EUDAT services and infrastructures of being asked the following questions:

"I am a potential data subject. Does the CDI hold any personal data about me? If so:

- what do you hold, and for what purpose?
- where is your record of my consenting to this?
- how can I access it?
- how can I request it be changed?
- how will you delete it should I withdraw my consent?"

This suggests that, as an adjunct to the CDI's indexing and cataloguing processes already in place to harvest metadata across all sites, EUDAT should build an internal catalogue of all data objects which have data subjects, *indexed by data subject*.

Building an index by data subject is a practical necessity in order to be able to answer most of the posed questions, but runs the risk itself of propagating personal information (the data subject's name, for example) into indexes and metadata records unnecessarily. One way to avoid this might be to introduce anonymous identifiers for data subjects, record corresponding identifying information in a secure index file, and propagate the anonymous identifiers into metadata records. The index file can be held offline by an EUDAT data privacy office function. If a data subject wishes to exercise their rights across data concerning them stored within the CDI, they are first looked up in the index file to find their anonymous identifier, then that identifier is used to index the usual metadata records stored within CDI metadata services.

This indirect cataloguing is a form of *pseudonymisation* as recommended in the GDPR as a suitable technical approach to addressing the principle of data minimisation.

Drawing together a number of threads from the preceding discussions, EUDAT should revise its required common metadata record for data objects to include a number of "sensitivity-related" fields. As noted in Section 2.1 EUDAT already follows the OpenAIRE guidelines in advocating a minimal metadata set. OpenAIRE's guidelines [16] are based on the DataCite version 3.1 "recommended" metadata set [15], with the addition of a rights-statement to cover intellectual property licensing. We recommend that EUDAT adds fields to record:

- a sensitivity-related DataTag as described above;
- a defined-time "embargo-release" date;
- a way to record event-based embargo release (e.g. upon the deaths of all data subjects);
- anonymous identifiers for each data subject.

All these fields must be machine-processable to enable automated data management services like B2SAFE to read them and trigger appropriate data handling rules (e.g. encrypt on transmission).

7.3. Encryption: who holds the keys?

The current interpretation of the Harvard DataTags system is largely a technical one, advising data processors how they should handle a data object with a particular tag, principally in terms of whether it should be encrypted or not.

Encryption of personal, or otherwise sensitive, data at rest within the CDI is an element of good practice which follows from the principles of Article 5 (1) [e] and [f], 'storage limitation' and 'integrity and confidentiality'. The main question with data encryption is always "who holds the keys?". To best support the storage and archiving of personal data for long-term scientific and historical research purposes EUDAT would be well-advised to begin designing encrypted data services, but the key question of encryption keys, their storage, use and access to them, is almost certainly a governance issue that should be addressed by EUDAT management (as defined in the Preface, p. 6).

8. RECOMMENDATIONS

The EUDAT CDI, built as it is from existing data repositories, already complies with national data protection legislation. In order to comply with the GDPR some areas will require renewed attention. These relate to **consent**, **age**, responsibility (**data processor/controller relationship**) and **content audit**. We summarise our general guidelines here.

Overall, to ensure compliance with existing laws and with the future requirements of the GDPR, EUDAT service providers will have to make informed decisions on the following:

For protection of EUDAT users and consent:

1. EUDAT service providers *must* ensure that service users are made aware of their rights in a clear understandable format. This can be done through a *privacy notice* for each service (cf. Annex A).
2. EUDAT services *must* have a method to gather and record consent from users – "freely given, specific, informed and unambiguous" – over the use of EUDAT services, and in particular acceptance of the privacy notice – the equivalent of "I accept cookies" and/or "I have read and agree to the terms and conditions".
3. EUDAT service providers *must* ensure that there are procedures on how to handle requests about personal data (e.g. consent withdrawal). An appropriate contact should be noted in the privacy notice, and clear lines of action should be created.
4. EUDAT service providers *must* ensure privacy notices are reviewed regularly.
5. EUDAT service providers *must* ensure that changes in applicable and enforced data restrictions are tracked and, when appropriate, provenance and timestamped information recorded to handle temporal and spatial dimensions.
6. EUDAT service providers *must* put in place procedures to detect, report and investigate personal data breaches. This could be added to the responsibilities of the EUDAT security officer, or form part of the new role of data protection officer.

On processing and identifying processors and controllers:

7. EUDAT service providers *must* be aware of the change in EU legislation and their responsibilities as data processors, data controllers, or both. This report serves as a foundation document here.
8. EUDAT service providers in the role of data controllers *must* arrange contracts or other legal forms with any and all data processors to whom they transmit personal data.
9. EUDAT service providers *must* ensure that there are measures in place to uphold the principles and obligations of the GDPR in Article 5, vis: 'lawfulness, fairness and transparency'; 'purpose limitation'; 'data minimisation'; 'accuracy'; 'storage limitation'; and 'integrity and confidentiality'.
10. EUDAT management *should* incorporate relevant recommendations into the CDI Collaboration Agreement.

On data held (both administrative and content data):

11. EUDAT service providers *must* ensure that there is a record and understanding of personal data held, and that this is in line with the stated privacy notice.
12. EUDAT service providers *must* ensure that there is a mechanism in place to identify personal data including the special categories of personal data (cf. point 16).
13. EUDAT management *should* instigate a documented review of the various types of data processing EUDAT carries out and the legal basis for carrying it out.

Service design:

14. EUDAT service designers *must* be able to demonstrate that there is *data protection by design*, ideally through impact assessments, and that data services are designed around the *principle of (personal) data minimisation*.
15. EUDAT service designers *should* introduce a DataTags-like system for marking data objects within the CDI with sensitivity levels. We recommend the introduction of a data tag into the standard metadata of EUDAT Handles created through B2HANDLE.
16. EUDAT service designers *should* introduce pseudonymised identifiers for data subjects in metadata and other internal records. A master index of identifiers should then be created and held off-line by the relevant data controller(s) (cf. points 3 and 12 above).
17. EUDAT service designers *could* put in place mechanisms to change or update data tags according to time- or event-based rules.

Protecting children:

18. EUDAT service providers *must* ensure that systems are in place to verify individuals' ages and to gather parental or guardian consent for the data processing activity if required. A paragraph in the *Disclaimer* section of the privacy notice is necessary but possibly insufficient.

Other:

19. EUDAT service providers *should* individually appoint Data Protection Officers.
20. EUDAT *should* appoint a Data Protection Officer to coordinate compliance across the members of the EUDAT CDI.

9. CONCLUSIONS AND THE FUTURE

EUDAT deals with open data, some of which is covered by data protection regulation. EUDAT has already developed national solutions under existing legislation but updates are now required to comply with the EU wide GDPR as it comes into force in 2018, with (in due course) derogated Member State legislation covering the particular case of research data, and with in-progress codes of conduct created in response to the GDPR by various discipline-specific authoritative bodies.

EUDAT will be drawing on established best practice as a source of recommendation for our service providers. Key to our approach is ensuring that there is the awareness of our obligations following the principles of the GDPR, knowing where responsibility lies and upholding user rights including that there is explicit consent given for the processing of personal data. This report has therefore identified the areas in which EUDAT already complies with the GDPR and set out further recommendations for our service providers. There is work to be done at the infrastructure level to support the new rights of data subjects, and future policy work around research codes of conduct should be conducted in close collaboration with key EUDAT community partners.

10. REFERENCES

[1] M. Dovey et al, *Report on Governance Model*, EUDAT-DEL-WP2-D2.4 v1.0, October 2016. http://doi.org/10.23728/b2share.2f585488d0be4f9591f65cccf1b588c0

[2] G8 *Open Data Charter*, 2013. https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

[3] European Commission, *Open data: An engine for innovation, growth and transparent governance,* 2011, COM/2011/0882 final. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52011DC0882

[4] RCUK *Common Principles on Data Policy*, 2011 (revised 2015). http://www.rcuk.ac.uk/research/datapolicy/

[5] Research Data Alliance website, http://www.rd-alliance.org

[6] H. Frew et al, *Guidelines on Open Access and Restricted Data (draft)*, EUDAT-DEL-WP2-D2.5 v1.0, August 2016. http://doi.org/10.23728/b2share.5ae5027c74d649f0961946d9e9887803

[7] OpenAIRE website on the EU Open Data Pilot. https://www.openaire.eu/opendatapilot

[8] RECODE project website. http://recodeproject.eu/

[9] The Royal Society, *Science as an Open Enterprise*, The Royal Society Science Policy Centre report 02/12, ISBN: 978-0-85403-962-3.

[10] M.D. Wilkinson et al, *The FAIR Guiding Principles for scientific data management and stewardship*, Scientific Data 3, March 2016, doi:10.1038/sdata.2016.18.

[11] R. Baxter et al, *EUDAT Sustainability Plan (final)*, EUDAT-DEL-WP2-D2.1.3 v1.0, May 2015.

[12] Handle System website. http://www.handle.net/

[13] Digital Object Identifiers website. http://www.doi.org/

[14] European PID Consortium website. http://www.pidconsortium.eu/

[15] The DataCite Consortium, *DataCite Metadata Schema for the Publication and Citation of Research Data*, version 3.1, October 2014, doi:10.5438/0010

[16] OpenAIRE, *Guidelines for DataCite*, 2015. https://guidelines.openaire.eu/en/latest/data/use_of_datacite.html

[17] OpenAIRE, *Guidelines for use of OAI-PMH*, 2015. https://guidelines.openaire.eu/en/latest/data/use_of_oai_pmh.html

[18] *Open Definition*, version 2.1. http://opendefinition.org/od/2.1/en/

[19] Creative Commons website. http://creativecommons.org/

[20] EU, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, 2016. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

[21] EU, *Directive 95/46/EC, Protection of personal data*, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV:l14012

[22] ECJ, *The Court of Justice declares that the Commission's US Safe Harbour Decision is invalid*, October 2015, http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf

[23] Article 29 Working Party, *Statement on the decision of the European Commission on the EU-U.S. Privacy Shield*, 2016. http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29_press_material/2016/20160726_wp29_wp_statement_eu_us_privacy_shield_en.pdf

[24] ANDS, *Publishing and sharing sensitive data*, 2016. http://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2016.pdf

[25] A. Bauer, *The Use of Signing Space in a Shared Sign Language of Australia*, Walter de Gruyter GmbH & Co KG, 11 Sep 2014, ISBN 978-1-61451-733-7.

[26] L. Sweeney, M. Crosas, M. Bar-Sinai, *Sharing Sensitive Data with Confidence: The Datatags System.* Technology Science [Internet], 2015. http://techscience.org/a/2015101601/

ANNEX A. TEMPLATE FOR PRIVACY NOTICE AND DISCLAIMER

The following text can be used by service providers as a basis for a common privacy notice and disclaimer. It is derived from existing EUDAT statements (notably B2FIND and B2SHARE) and has been strengthened where necessary to align with the new requirements of the GDPR[14].

Privacy Notice

The service … (hereafter "the Service") is operated by … (hereafter "We", "Us", "the Service Provider", "Our"). To make use of the Service you must consent to processing by Us of your personal data as described below.

The careful and lawful handling of your data is important to Us. This privacy notice explains what personal data the Service collects, for what purposes and how they are processed.

Please read the privacy notice thoroughly before you provide Us with any personal data, or use the Service in any other way. If you do not consent to these terms, you may not use the Service. The following data privacy provisions for the Service are the current version.

Your Consent

By using this Service, you agree to the terms of this Privacy Policy and Our Terms of Use. Whenever you submit information to this Service, you consent to the collection, use, and disclosure of that information in accordance with those Terms of Use and this Privacy Policy.

Your Rights

You have the right to withdraw your consent to the collection, processing and use of personal data at any time with effect for the future.

You have the right to receive free information about the personal data stored about you. On request We will inform you in writing in accordance with applicable law whether and what personal data is stored by Us.

You have the right to update or correct inaccuracies in the personal data We hold about you.

You have the right to delete your personal data from the Service (your "right to be forgotten").

You have the right to know if your personal data have been hacked or compromised.

To exercise these rights, please send requests to: *[……]*

Information We Collect

When you use the Service the following data may be collected if you choose to supply them in content you upload to the Service ('content data'):

- contact information, such as your name, email address;
- your username and password;
- other personal information in content you provide to the Service;
- institutional or organisational affiliations.

[14] A number of major European websites with privacy policies updated at the beginning of 2017 were cross-checked against this template. These included Figshare (https://figshare.com/privacy), Mumsnet (https://www.mumsnet.com/info/privacy-policy), Box.com (https://www.box.com/en-gb/legal/privacypolicy) and the BBC (http://www.bbc.co.uk/usingthebbc/privacy/).

The Service also creates information about service usage and status ('administrative data'). This information (in the form of server log files) is automatically and manually monitored and processed. We may collect information about the Service by analysing types and kinds of data stored in the Service and how these data are accessed.

When you use the Service the following administrative data may be collected automatically to create anonymous statistics; for the purpose of monitoring data protection; and to ensure the proper operation of our data processing systems:

- the IP address of your computer;
- the date and time of visit;
- the operating system and browser on your computer;
- the amount of data transmitted;
- the internet address of the website from which you have accessed this site.

We may also combine information with other EUDAT service providers. Information created by collecting and analysing administrative and service usage information will only be used to check that the relevant service Terms of Use are being followed and for service development purposes.

What We Do with the Information We Collect

We may, with your consent, share your personal information with other EUDAT service providers in order to ensure the correct operation of the Service and our underlying data processing systems.

We do not share your personal information with any other third party, unless you specifically consent to Us doing so or We are specifically required to by law.

We only use the information you give Us to understand your needs and provide you with a better service, and in particular:

- for internal record keeping.
- to improve Our Services.
- to respond to service queries you have reported.

Any personal data that you give Us will be retained by Us for as long as you make use of the Service.

Children's Privacy

Our Service is not aimed at children under the age of *[13..16]* and We do not knowingly collect personal information from children under the age of *[13..16]* through the Service. If We become aware that We have inadvertently received personal information through the Service from a child under the age of *[13..16]*, We will delete the information from our records.

Changes to the Contents of this Privacy Notice

The Service Provider reserves the right to change the content of this privacy notice from time to time in accordance with legal data protection regulations. Changes to this privacy notice will become effective when those changes are posted to the Service.

Disclaimer

Service Website

The Service Provider endeavours to provide accurate and up-to-date information on the Service website. However, errors cannot be ruled out, and the Service Provider accepts no responsibility for the correctness or completeness of the information provided.

The Service Provider reserves the right to change the website in part or in whole without prior notice.

The Service website includes links to external sites. The Service Provider accepts no liability or responsibility for the accuracy, completeness or legality of the content of any linked external websites.

Third-Party Content

The Service Provider will not monitor third-party content for completeness, accuracy, or compliance with binding rules. This holds for all data accessible through the Service, regardless of whether they are stored on servers owned by the Service Provider or others. Accordingly, the Service Provider accepts no liability or responsibility for the completeness, accuracy or legality of third-party content.

Unauthorised Access by Third Parties

Despite all best efforts, no method of transmission over the Internet and no method of electronic storage can be guaranteed to be absolutely secure. The Service Provider accepts no liability for unauthorised access by third parties or for any possible transmission of computer viruses, Trojan horses or other malicious programs. The user is responsible for making arrangements to protect their own computer from such malicious programs, in particular by installing antivirus software and using the latest antivirus definitions.

If the Service Provider learns of a security breach, affected users will be notified immediately by email and via the Service website, so that they can take appropriate protective steps.

ANNEX B. TEMPLATE AGREEMENT FOR DATA CONTROLLER-DATA PROCESSOR

The following text can be used as a template for a legal agreement between the Data Controller and the Data Processor, where the latter processes personal data on behalf of the former. The draft is an adaptation of what it is presently in use in Norway. The legal formalisation of the agreement between controller and processor now present in the GDPR stems, at least prima facia, from the Norwegian model, hence our suggestion of this as a starting point. In the agreement the articles/act that the partners need to comply with have to be specify and normally consist of the national Personal Data Act and/or national Personal Health Data Act, but in the near future also the GDPR articles/acts might be invoked.

AGREEMENT ON THE PROCESSING OF PERSONAL DATA

Storage and Processing of research data in the *<Insert the name of the Data Processor>*

The text in blue italics must be removed and replaced with relevant text, in some cases by selecting one of several alternatives.

1. Parties to the Agreement

1.1 Parties

The Agreement is entered into between the party responsible for the data processing: *<Insert the name of the project/institution>* (Org. no. ……) (hereafter referred to as the Data Controller) and the party that processes the data: *<Insert the name of the service/Institution>* (Org. no. ……) (hereafter referred to as the Data Processor).

1.2 Contact persons

Contact person for the Data Controller: *<name, contact information, role>,……..*

Contact person for the Data Processor: *<name, contact information, role>,……..*

2. Purpose of the Agreement

The Data Processor offers storage services for researchers who conduct research on person-sensitive data, including health data.

The purpose of the Agreement is to regulate rights and duties pursuant to:

- *<list of the articles/act/regulations of the GDPR regarding the personal data and personal health data>*

The Agreement regulates the Data Processor's processing and securing of personal data and health data that have been made available by the Data Controller. It must be clearly stated whether the Data Processor is permitted to surrender data to other parties for storage, processing or other use.

The purpose of the processing shall not be changed by either of the parties without a new agreement being signed.

3. The parties' area of responsibility pursuant to *<list of the articles /act/ regulations of the GDPR regarding the personal data and personal health data>*

The Data Controller is to be deemed the unit responsible for the data processing pursuant to *<list of the articles of the GDPR to which the agreement aims at complying…>*

The Data Controller is responsible for ensuring the fulfilment of the requirements laid down in the *<list of the articles of the GDPR to which the agreement aims at complying…>*, including those relating to security. This entails the Data Controller being charged with ensuring that the requirements relating to the storage and use of health data and sensitive personal data are complied with by the Data Processor.

The Data Processor can only process health data and personal data that have been made available by the Data Controller in accordance with this Agreement. Any other use of health data and personal data shall be agreed with the Data Controller in advance and in writing.

The Data Processor shall ensure that health data and personal data made available by the Data Controller are kept separate from its own and others' data and services.

4. Description of the purpose of the use of the Data Processor

The Data Processor can only process personal data in accordance with the purposes that have been specified by the Data Controller and pursuant to the terms stated in this Agreement.

<This point MUST be filled in, and it must be stated clearly and precisely what the data are to be used for. Any link with other data sets must be approved by the Data Controller. An exception from this is when the links are made anonymous.

State what the data is to be used for:>

5. Specification of the data that is to be processed

<Must be filled in, and must indicate the type of data that is to be processed, and whether these data are directly identifiable or have been made unidentifiable (i.e. whether the data appear as anonymous, but where it is actually possible to go back and find out who the date/information concerns).

If the Data Controller finds it necessary to change the data that are to be processed, or to add a new type of data to those that are to be processed, he is under the obligation to make a new security assessment. If a material change is involved, the change cannot take place without a new data processor agreement being signed.

6. Requirements regarding data security

Pursuant to the provisions stated in the *<list of the articles of the GDPR to which the agreement aims at complying…>*, both parties shall at all times meet the requirements regarding data security and internal control, as well as those relating to access control.

The Data Processor shall ensure that all processing of health data and personal data encompassed by this Agreement is carried out in accordance with the acceptable level of risk defined by the Data Controller. As part of this the Data Processor shall submit risk assessments of its own security.

With regard to security, the Data Processor is required to have defined its objectives, strategy, organization and responsibility in accordance with the *<list of the articles of the GDPR to which the agreement aims at complying…>*, and is required to ensure that these are followed up by the necessary internal control system.

Any breach of security or any suspected breach of security shall immediately be reported to the Data Controller.

The Data Processor shall have clear procedures for logging errors and nonconformities in systems that are used to handle health data and personal data and that are included in this Agreement. If such errors or nonconformities are detected, the Data Processor shall notify the Data Controller of this as soon as possible and at the latest within 24 hours (48 hours if the incident arises at the weekend or on a public holiday). In such an event the Data Processor shall immediately take steps to minimize possible damage to the Data Controller.

The Data Controller can at any time demand documentation from the Data Processor as reassurance that the Data Processor is complying with all relevant requirements concerning data security stated in the *<list of the articles of the GDPR to which the agreement aims at complying…>*. The Data Controller can request access to the Data Processor's reports etc. on periodic audits of its procedures and routines.

The Data Processor shall be able to demonstrate good routines concerning data security, including in particular technical security, access control and physical security.

The Data Controller is responsible for adequate security at the units that are used for remote access to the Data Processor. With regard to updating and virus control this will in many cases mean that the units must be in the Data Controller's operating regime or that of parties closely related to the Data Controller.

7. The Data Controller's right to access, inspection and testing

The Data Controller shall have the right to access the solution and to verify how it is secured. In this context 'access' means documentation, interviews, meetings and any other forms of verification that may be appropriate. The Data Processor accepts that access can be exercised by the Data Controller or by the third party the Data Controller may select to carry this out as long as the access extends only to the area designated to the processing of the Data Controller's data. The right to access applies to all technical, organizational and administrative aspects that are relevant for security in the services that are delivered to the Data Controller.

The Data Processor is obliged at four weeks' notice to surrender security documentation relevant for the Data Controller, or otherwise to ensure access to such documentation.

If the Data Controller makes use of the right to access, and nonconformities are detected in the security of the Data Processor's systems, the Data Processor shall remedy the nonconformity as quickly as possible. The Data Processor shall give a written description of the remedial measures and the plan for implementing them.

8. Confidentiality obligation

The parties shall observe professional secrecy on all confidential information, people's personal circumstances, security and business matters, and information that may cause harm to one of the parties or that may be utilized by a third party.

The confidentiality obligation applies to the parties' employees and to others who act on behalf of the parties in connection with the implementation of the contract. All employees must have signed a non-disclosure declaration.

The parties are under the obligation to take the necessary precautions to ensure that others do not come into possession of material or information in conflict with this clause. Employees and others who resign from the service of one of the data processors shall be subject to confidentiality on the matters mentioned above also after their resignation.

This provision also applies after the termination of the Agreement.

9. Entry into force, duration and termination

9.1 Entry into force and duration

The Agreement comes into force when it has been signed by both parties.

<alt. 1>

The Agreement applies as long as the Data Processor processes personal data on behalf of the Data Controller in accordance with the purpose stated in this Agreement.

<alt.2>

The Agreement comes into force on ..... and lasts until ..... . The Agreement can be terminated with ..... months' written notice.

9.2 Termination

Unless otherwise agreed with the Data Controller, on termination of this Agreement the Data Processor undertakes to return all health data and personal data that have been received on behalf of the Data Controller and that are included in this Agreement.

The Data Processor shall delete all documents, data, hard disks, CDs and other storage media that contain information that is included in the Agreement. The deletion shall be carried out in a way that prevents the data being retrieved. This also applies to any back-up copies.

10. Breach of contract

Breach of contract occurs if one of the parties does not fulfil its duties according to this Agreement and when this is not due to circumstances for which the other party bears responsibility or risk. If one of the parties wishes to invoke breach of contract, the other party must be notified of this in writing without undue delay.

In the event of breach of contract, the injured party may withhold payment in return, although the amount withheld shall not be clearly higher than what seems necessary to remedy the effects of the breach, and only until the matter has been brought into accordance with the Agreement.

Should a material breach occur, the other party may – after having given written notice and a reasonable deadline for remedying the matter – terminate all or parts of the Agreement with immediate effect, and may demand compensation for any loss this has caused.

11. Transfer of rights and obligations

The Data Controller may, completely or partially, transfer its rights and obligations pursuant to this Agreement to another body, which is then entitled to equivalent terms and conditions. The Data Processor can require any additional expenses incurred by the transfer to be covered.

The Data Processor can transfer its rights and obligations pursuant to the Agreement with the written consent of the Data Controller. Such consent cannot be refused without reasonable grounds. The right to remuneration according to the Agreement can be freely transferred, but the transfer does not exempt the Data Processor from its obligations and responsibilities.

12. Governing law

The parties' rights and duties pursuant to this Agreement are determined in their entirety by the law of *<usually the country of the Data Controller>*.

13. Signing

This Agreement has been signed in 2 – two – copies, each party retaining 1 – one – copy.

<Place>, on	*<Place>*, on .....
<Data Controller>	*<Data Processor>*
(signature)	(signature)
……	……
Name: …….	Name: …….
(in block capitals)	(in block capitals)
Position:……	Position: ……

ANNEX C. GLOSSARY

Term	Explanation
AAA	Authentication, Authorisation and Accounting
AAI	Authentication and Authorization Infrastructure
ANDS	Australian National Data Service
APARSEN	Alliance for Permanent Access to the Records of Science in Europe Network
CC	Creative Commons (an IPR licensing scheme)
CDI	Collaborative Data Infrastructure
CERT	Computer Emergency Response Team
DataCite	An organisation which standardises and assigns PIDs for data (cf. DOI)
DCC	Digital Curation Centre
DMP	Data Management Plan (or Planning)
DOI	Digital Object Identifier (a de facto standard PID for data)
DSA	Data Seal of Approval (a repository certification scheme)
EC	European Commission
IPR	Intellectual Property Right
ISO	International Organization for Standardization
ITSM	IT Service Management
NDS	National Data Service (USA)
OAI-PMH	Open Archives Initiative Process for Metadata Harvesting (a metadata exchange protocol)
OAIS	Open Archival Information System
OpenAIRE	Open Access Infrastructure for Research in Europe
OSCT	Operational Security Coordination Team
PID	Persistent Identifier
QA	Quality Assurance
QC	Quality Control
QoS	Quality of Service
RDA	Research Data Alliance
RI	Research Infrastructure
RP	Resource Provider
SLA	Service Level Agreement
SP	Service Provider
ToU	Terms of Use

