IMI2 821520 -
ConcePTION

ConcePTION

WP7 – Information and data governance, ethics, technology and data catalogue and quality support

D7.4 Report on initial information and research governance for WP1-5

Lead contributor	6. The European Institute for Innovation through Health Data (i~HD)
Other contributors	41. Glaxosmithkline Research and Development Ltd. (GSK) 1. Universitair Medisch Centrum Utrecht (UMCU)

Document History

Version	Date	Description
V0.3	06/03/2020	Early draft adapted into house style
V0.4	11/04/2020	Second draft prior to Managing Board review
V0.5	14/04/2020	Third draft for Managing Board review
v.06	02/05/2020	Final draft after comments from Managing Board
V0.7	04/05/2020	Final version

Abstract/Executive Summary

The advent of the General Data Protection Regulation (GDPR) in May 2018 has not fundamentally altered the data governance arrangements for undertaking research. It has placed more emphasis on the pre-production of documentation, such as Data Protection Impact Assessments (previously described as Privacy Impact Assessments) as well as prescribing explicitly what information needs to be provided to potential data subjects about possible data processing, including the legal bases to support such processing.

This document lays out these aspects in more detail. In particular it notes (on page 16) that a consortium, such as ConcePTION, cannot be a 'controller' as it is not a legal entity, so all data protection obligations fall to member organisations to apply and ensure compliance with both EU and Member State law.

The ConcePTION project has chosen to follow the ENCePP Code for WPs 1 & 2 and RD-Connect (as pertinent to biobanks) for WP4.

The ConcePTION project has undertaken the following actions:

- a survey of likely information assets to be developed across all Work Packages to establish their nature and which partner organisation will actually be holding and managing the asset and under what code of practice it will be managed
- production of a Data Management Plan for the consortium (additional more detailed plans may be required in respect of particular assets – e.g. WP4 bio-bank)

The following recommendations have been made to the Managing Board to consider these actions for coordinated implementation across the consortium:

- to provide a structured list of information assets (derived from the results of the survey) to partner DPOs to ensure they are aware of these assets, can arrange appropriate DPIAs (if required), and add them to any information asset registers (including Article 30 registers) they may have;
- to provide generic instructions to consortium staff about privacy, security, and legal requirements in respect of such assets
- to encourage WP leads to ensure that their own WP and Task members are aware of the generic instructions and also any additional information required in respect of assets used in their own WP or Task
- to provide appropriate GDPR awareness training at consortium group meetings

It is expected that this document will be updated later in the project to reflect possible changes within the ConcePTION project itself and also within legal and regulatory guidance, including any possible new codes of practice or codes of conduct.

Glossary

This covers terms that are used throughout the document. Other terms or abbreviations may be used within a particular section but are elaborated there.

Term or abbreviation	Full text or description
DP	Data Protection
DPO	Data Protection Officer
DPIA	Data Protection Impact Assessment
EC	European Commission
EEA	European Economic Area
EFPIA	European Federation of Pharmaceutical Industries & Associations
EU	European Union
EUDRACT	
GDPR	General Data Protection Regulation
IMI	Innovative Medicines Initiative – a joint initiative between EFPIA and EC
ISO	International Standards Organisation
MS	Member State (of EU)
PMI	Patient Master Index
REC	Research Ethics Committee
SME	Small/Medium Entreprise

Contents

Purpose & scope of this document

The purpose of Work-Package 7 in the ConcePTION project is to provide ethical and governance and quality assessment support for the conduct of distributed data collection and analyses which will support generation of high quality real world evidence on the effects of drugs during pregnancy and lactation.

The purpose of this document, Deliverable D7.4, is to cover the activities of Task 7.2 as part of the wider Work Package 7 (WP7) in terms of appraising the requirements in terms of information and research governance for the anticipated data processing across the ConcePTION project.

D7.4 definition from project proposal: *'Report on initial information and research governance for WP1-5 - task 7.2 (M12)'.*

T7.2 Task description

From the project proposal, ConcePTION Technical Annex Section 1-3 (p48):

The objective of this task is to ensure, and assure, data access providers and research users that data and samples are treated in full compliance to the GDPR (for personal data), the IMI Secondary Use Code and other codes of practice that protect the privacy of data subjects (especially EHR4CR, EMIF, i~HD and BBMRI-ERIC), that research is conducted in ways that accord with codes of conduct from ENCePP and the ethics policies of ELIXIR, as well as other relevant standards.

In achieving this, the partners involved in this task will reuse these existing instruments – some of which they have themselves developed in other initiatives – and carefully combine these and fill any gaps that are relevant to the project, while including the results from task 7.1. [Defining the rules and collaboration models for data reuse]

We will describe the procedures for a common trusted data management and research ecosystem. The main elements of such trusted ecosystem, and the areas of coverage of some of the instruments we will combine and reuse. After the initial development of instruments and mapping to data access providers, this task will monitor their adoption and maintain the instruments in the light of new policies and legislation (such as national GDPR implementation laws). For instance, it will be important to consider how each EFPIA company is interpreting the grey areas of GDPR in the scope of real world data use for evidence generation.

ConcePTION will (re)-use human health/specimen data from various sources (WP1, 2, 3 & 4), that has been consented for research purposes (i.e. biobank, cohorts, some registries, reports) or was collected for routine healthcare (e.g. healthcare data, national statistic registries).

The GDPR requires data privacy impact assessments prior to conducting any research on the data. In this task we will provide clear guidance documents to adhere to GDPR and provide templates for data privacy impact assessments and will provide guidance and support for DPIA locally as well as for the central data platform. It will collect and monitor the results of the DPIAS and be the reference for the local data privacy officers. This task will liaise with Ethics and Governance specialists in the Independent Scientific Advisory Board.

In this task we will develop information and consent procedures and templates for biobanking. Development and implementation of information and consent procedures and templates for collection, storage, analysis of samples in different European countries and transfer of data across borders will be based on:

- *i) the legal premises of the GDPR;*
- *ii) the National laws of participating countries;*
- *iii) The Article 29 Working Party Guidelines on consent;*

iv) The BBMRI-Code of Conduct;
v) The Recommendation of the OECD Council of Health Data Governance (of 17 January 2017);
vi) The CIOMS International Ethical Guidelines for Health-Related research involving Humans (2016) and;
vii) 20 years of published research on information and consent procedures by the Uppsala partner Centre for Research Ethics & Bioethics.

The purpose of sampling will be described in general terms while the processes will be described in more detail, e.g. that

- *samples and data will be sent across borders for research that may include both academic and commercial partners;*
- *means for privacy protection;*
- *how to withdraw consent;*
- *kind of analyses to be made;*
- *linkage to other data sources;*
- *re-consent of children when they reach age of maturity; and*
- *how to access to information about projects.*

Templates will be formulated in the relevant language and ethics approval will be obtained by legitimate ethics committees in each country.

Additional material

Technical Annex 5 - Ethics (5.2) states: *A ConcePTION code of conduct specifying the rules of data access and use in this collaborative project will be created in task 7.2-7.3 to ensure organizations will follow this code*.

Also 5.4 'Personal Data' states: *Compliance with the GDPR will include undertaking Data Protection Impact Assessments, having transparency to data subjects including subject rights, identifying the Data Protection Officers for each data source and the project as a whole, about the data processing that will be undertaken, third country transfers etc. and ensuring that any new patient/subject consent is fully informed and non-coercive*.

This rightly notes that DPIAs would likely be required under GDPR, but not that these are necessarily done within T7.2. As noted in the main body of this report, that is actually the responsibility of each participating institution as a legal entity, unless it considers itself a 'processor' for another organisation – in which case there should be an explicit controller-processor contract or agreement detailing the relationship and what processing the processor should undertake for the controller.

Further: *Given the complex nature of the rights of data subjects when it comes to maternal, paternal, familial and neonatal data, it is possible that ethical (in addition to governance) data handling issues will arise. Work Package 8 (Task 8.6) will run an External Advisory Board with ethics experts who will provide peer oversight of the policies and measures and any compliance issues that arise with respect to data processing. This Board will also advise on any ethical issues that arise in relation to the data*.

Again, this does not directly entail T7.2 in developing any materials or other activities. It is, however, an area unaddressed by GDPR where there is a presumption that a 'data subject' is clearly identified (there are parallels with the possible ambiguity about de-identified data and whether it can be determined to be within the scope of GDPR or not). Family members may not be the 'subject' of a record but may be associated with its content through their familial (or other) relationship with the 'primary' data subject.

Interpreted scope of T7.2 and D7.4

The scope is interpreted as being strictly about:

- 'information and research governance' issues, that is issues relating to either 'information governance' or 'research governance' or possibly both insofar as they relate to the ConcePTION project, though some aspects may apply more generally

- An 'initial' report, so that it may need updating through the life of the project to reflect changes in external aspects (e.g. legislation or regulatory guidance) or internal aspects (such as modification of research protocols, changes in relevant project partners, etc.)

- Focused on the project work in WPs 1-5, so not administrative or stakeholder engagement elements, such as may be handled by WP6 and WP8. Equally, it does not include issues arising from patient engagement activities within WP7 itself (e.g. in T7.3)

It should be noted that there are aspects where 'information governance' and 'research governance' overlap and others where they do not. So the information governance recommendations in this document does not clash in any sense with the 'research governance' decision to follow the ENCePP Code of Practice.

However, it is clear from the task description for T7.2 that wider issues need to be considered and some of these issues have arisen during the first year of the ConcePTION project. This report seems an appropriate place in which to consider these (e.g. Appendix D on GDPR-compliance issues for similar IMI/EC-funded projects).

This document is intended for a general audience, though some elements may be of a more technical/legal nature where more precise terminology and detailed argument is required. The detailed discussion has been covered in some ancillary internal documents as intermediate products towards the development of this Deliverable:

- GDPR Compliance Position paper
- Data Survey collation and comparison
- Data Survey responses review

Introduction

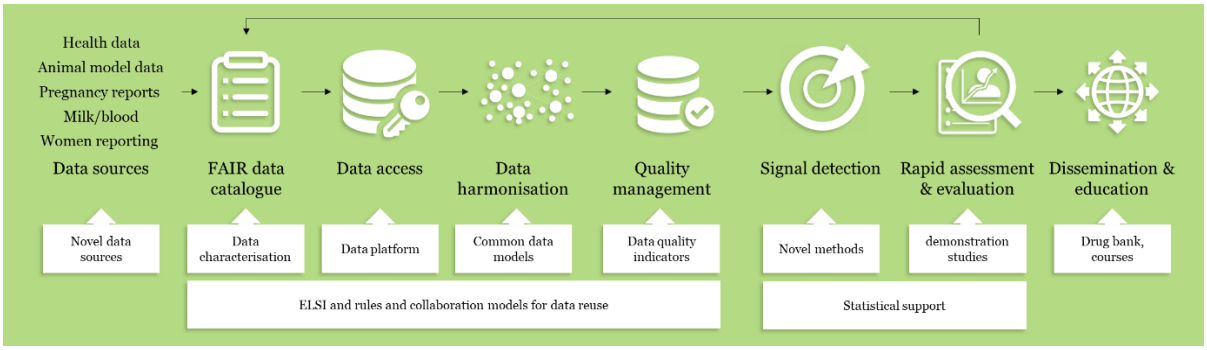
ConcePTION is an IMI-funded research project involving a consortium of 51 partners and a number of associated organisations (88 in total - see Appendix D for more detail).

Summary

Women who are pregnant or breastfeeding are traditionally excluded from medical research due to safety concerns. As a result of this, only 5 % of medications have adequate safety information on their use in pregnant or breastfeeding women, and this makes it very hard for doctors and women to make informed decisions about their treatment. Nonetheless, some 90 % of women are exposed to a prescription medication at some point during their pregnancy. The ultimate goal of ConcePTION is to create a trusted biomedical ecosystem capable of providing evidence-based information on the safety of medications during pregnancy and breastfeeding in an efficient, systematic and ethically responsible way. The information will be provided in a form that is usable by both healthcare providers and patients alike. The project will achieve this in a number of ways. Firstly, it will improve and unify existing approaches to data collection in this area by re-using existing, de-identified data generated during routine patient care. The project also aims to deliver procedures and tools for the collection of digital data and samples directly from pregnant women. They will also create the first Europe-wide breast milk biobank for research purposes, and develop tools to predict which drugs are likely to be transferred to breast milk. Finally, the team will establish a web-based drug information knowledge bank.

Its overall objective(s) are to build various dimensions of the ConcePTION ecosystem:

- the societal and ethical dimension
- the technical infrastructure dimension
- the data dimension
- the evidence generation dimension
- the communication & evidence dissemination dimension
- the sustainability dimension



ConcePTION will (re)-use human health/specimen data from various sources (WP1, 2, 3 & 4), that have been consented for research purposes (i.e. biobank, cohorts, some registries, reports) or were collected for routine healthcare (e.g. healthcare data, national statistic registries). The full detail is available in section 5 of the project proposal.

WP5 may undertake focus groups, interviews, and/or surveys to identify what information is available to and needed by pregnant and breast-feeding women – while much will be done through anonymous surveys online, there may be some collection of 'personal data'.

ConcePTION data collection

The main aim of the ConcePTION project is to facilitate research access to relevant medical data about pregnant and breast-feeding women.

ConcePTION will collect data about other relevant data sources (to develop a data catalogue with associated meta-data) in order to help promote research into the health of pregnant and breast-feeding women.

Its purpose is not generally to collect data about medical subjects *per se*, though it will seek to develop a distributed analysis system or 'gateway' so that data can be analysed in a harmonized manner from data access providers (if appropriate) to create aggregated data from the different data-sources. ConcePTION is not expected to provide facilities to create linkage between different data collections itself, these are supposed to be conducted at the level of the data access provider.

Where demonstration projects collect 'personal data' then this data will be held only by the original research institutions involved in those studies. Wider sharing within the consortium will be limited to meta-data about the data being held or aggregated data based on queries provided through agreed mechanisms.

Summary of data collections used in ConcePTION WPs 1-5:

WP	Data type	Data collection
1	Meta-data and aggregate data	Details about EUROmediCAT registries (registries of congenital anomaly with information on medication exposure in the first trimester of pregnancy).
1	Meta-data only	Details on sources for electronic health record (EHR) data, education data and civil registration data.
1	Meta-data only	Details on sources for prospective cohort data collected for research purposes
1	Meta-data only	Details on Population-based registries other than EUROCAT (e.g. cerebral palsy registers, cancer registers)
2	Meta-data & aggregate data	Aggregate data from and details about publicly available international spontaneous reporting systems: EUDRAVIGILANCE, FAERS, VAERS, Vigibase
2	Meta-data & aggregate data	Available spontaneous reporting data from pharmacovigilance centers locally
2	Meta-data & aggregate data	Available Pregnancy registry data & prospective cohort datasets
2	Primary study data, meta-data, and aggregate data	Newly collected data in ConcePTION on neurodevelopmental outcomes
3	Primary data and data from other sources	Not personal data, as relating to *in silico* or *in vivo* animal data
4	Anonymised data	Anonymised medical data and primary data from anonymised tissue samples from existing biobank facilities
5	Consented & anonymised data	Interview/focus group data (subject to consent) and anonymous online survey responses

Approach/Methods used in T7.2

Initial GDPR appraisal for ConcePTION project

An internal deliverable, ConcePTION - GDPR Compliance Position Paper, was developed to clarify how GDPR applied in the context of a multi-party project such as ConcePTION. The paper covered points such as:

- A brief summary of GDPR for project participants
- Discussion of 'controllership', particular where there is no lead legal entity
- The role of Data Protection Impact Assessments (DPIAs) as explicitly mentioned in terms of work)
- Need for 'Personal data' registers and link to WP7 data survey
- 'Transparency' requirements
- Transfers outside EU/EEA (as were overseas participants)
- Anonymisation & pseudonymisation considerations
- A preliminary list of expected data holdings based on the project proposal
- A particular focus on internal 'personal data' holding in relation to individual participants in the projects

This paper has informed a number of elements in this deliverable.

Outline of Data-sharing Agreement terms

Any data release by a project like ConcePTION will require a data-sharing agreement (DSA) or equivalent licence terms. It was not part of this Task's brief to develop a template for such an agreement and, in any event, the actual agreement would need to be framed by the actual organisation acting as 'controller' for the resource in question.

A list of the likely 'heads of agreement' (main topics to be covered) was drawn up and is hoped to be useful later in the project.

Questions to ConcePTION Managing Board

From the Position paper four points were submitted to the Managing Board:

- To what degree did Managing Board feel it should undertake GDPR compliance checking across project activities and project partners?
- To what degree should WPs 7 & 8 ensure that partners understood the legal position re controllership
- Should the consortium undertake direct communications with individual project participants re likely project use of their personal data (mainly contact details)
- Should the ConcePTION web-site include a GDPR statement making clear that any data subject should address themselves to relevant consortium members rather than the consortium itself

These points are likely to be relevant to similar projects in the future

Data survey

A survey was sent out by WP7 leads to all work packages in order to identify what datasets (human/personal, animal, or other) each work package expected to collect as part of its project activities. This allowed an initial view of overall project data holdings and how these might differ from the original indications in the project proposal to IMI for funding.

The format of the survey was based on the structure required by the Horizon 2020 Data Management Plan template

The results from the survey were collated into a spreadsheet, both as a register of the planned data holdings and also to be able to compare the results of the effectiveness of the survey. An outline review of the responses to the survey was also completed to inform possible updates or extensions of the survey in the future.

Collation of requirements across Codes of Practice and data protection law

While not an explicit requirement of this Task, it was originally envisaged that some detailed comparison between the various codes suggested in the project proposal would be undertaken. However, given the differences both in intended scope and the level of detail at which different codes were formulated, this proved to be unworkable except at a very broad level (see Appendix A for more detail.

The comparison of requirements did not seek to detail variations in data protection legislation between member states, particularly as permitted by derogations in the GDPR, as these were still becoming clear during the life of the project, and especially during the main phase of T7.2 (April 2019-March 2020).

The various codes and their possible relevance are considered in *Appendix A – Reflections on Codes of Conduct/Practice*.

	Research Regulation	Primary Data Use	Secondary Data Use
General	ECCRI		FAIR
Human	Clinical Trials Directive GCP, GVP, GEP, GPP, GLP, etc. EUDRACT	General Data Protection Directive (GDPR)	GDPR EMIF, ENCePP, ADVANCE
Animal	Animal Trials Directive PREPARE, ARRIVE		

It should be noted that the EMIF Code of practice is not currently widely adopted, as it is not publicly accessible.

The ConcePTION project has chosen to follow the ENCePP Code for WPs 1 & 2 and RD-Connect (as pertinent to biobanks) for WP4.

Human Research Governance

The legal and ethical requirements for studies are laid down through the Clinical Trials Directive 2010 and the Good Clinical Practice (GCP). Further guidance is normally provided by Member State regulators (e.g. Health Research Authority (HRA) in UK).

Clinical trials must be registered through EUDRACT.

There are generic research integrity codes such as ECCRI and UKRIO.

Human Research Data Governance

Clearly all data processing needs to be GDPR-compliant and conform to any MS law (e.g. Data Protection Act 2018, DPA2018, in the UK). GDPR includes some exemptions for research, as well as possible derogations for member states.

The position as to when Research Ethics review is or is not required for observational research studies (involving only re-use of patient data) is not always been clear. The need for ethics review varies across member states (and even within them):

For example, in the UK, the guidance document, Governance Arrangements for Research Ethics Committees (GAfREC 2018) is clear (1.3.3(b)) that research involving anonymised information does not of itself require REC approval, though the situation was not so clear prior to this edition – and there are still exceptions to this exemption (as detailed in 2.3.5 of that document).

There are also initiatives, such as the BMJ Open Data Campaign, which seek to have anonymised data from clinical trials available to others for verification or different methods of analysis (including meta-analyses). These are echoed in the ICMJE (international Committee of Medical Journal Editors) policy on data-sharing, though this only calls for a clear declaration of the study's position on data-sharing.

The ConcePTION study clearly supports the spirit of these open-data initiatives, though the position for any of the demonstration projects will remain with the institution undertaking the actual research study. ConcePTION aims to make access to aggregated data much easier.

It should also be noted that even aggregated data may pose potential re-identification risks through reconstruction attacks, so statistical disclosure controls, such as small cell-counts would be applied, even where not explicitly required by data protection law.

Animal Research Governance

European Directive governing the use of animals in scientific procedures, EU Directive 2010/63/EU

Codes of practice such as PREPARE and ARRIVE have been developed.

There is also now www.animalstudyregistry.org for the registration of all scientific studies involving animals conducted around the world. It is run by the German Centre for the Protection of Laboratory Animals (Bf3R). Registration is purely voluntary.

Animal Research Data Governance

There is no explicit data governance (beyond the need to keep good and effective records as required above), viz. any further use, re-use, or sharing is not subject to specific regulation (beyond being available for audit).

There is no parallel to the wide range of laws, regulation, and literature concerning data-sharing of human health records.

Ensuring GDPR compliance across consortium and projects

The ConcePTION consortium is not a legal entity, so can have no legal responsibilities under GDPR – it can be neither a 'controller' nor a 'processor' as such.

Other projects may have an associated legal entity or a single lead partner which may contract with other parties to fulfil aspects of the work involved in the project. The necessary arrangements for this type of project are not considered here.

While the ConcePTION project may have no legal requirements, there are perhaps moral or at least administrative requirements that it should fulfil in order to ensure that the project as a whole is managed correctly and appropriately under relevant laws.

It would seem appropriate therefore that such a consortium should at least;

- Understand what information assets it will create or hold as part of the project, whether as primary data or to support administrative operations
- Ensure that it is understood across all partners which partner is the legal 'controller' for any information assets that include 'personal data' of whatever sort
- Support partners in ensuring that these information assets are appropriately protected against corruption, loss, or misuse
- Ensure that all staff who may have access to such information assets (particularly those which include 'personal data'):
    - understand their obligations to protect privacy as well as respect any intellectual property rights or other constraints on use of the data (if only to avoid unnecessary or inappropriate use of email addresses)
    - know whom to contact in their own organisation (or the organisation which is the 'controller'/owner of the asset) in order to seek guidance over legal compliance and to whom to report any possible incidents concerning those assets.

To this end, the ConcePTION has undertaken the following actions:

- a survey of likely information assets to be developed across all Work Packages to establish their nature and which partner organisation will actually be holding and managing the asset and under what code of practice it will be managed
- production of a Data Management Plan for the consortium (additional more detailed plans may be required in respect of particular assets – e.g. WP4 bio-bank)

It is recommended that the Managing Board also consider these actions:

- to provide a structured list of information assets (derived from the results of the survey) to partner DPOs to ensure they are aware of these assets, can arrange appropriate DPIAs (if required), and add them to any information asset registers (including Article 30 registers) they may have;
- to provide generic instructions to consortium staff about privacy, security, and legal requirements in respect of such assets
- to encourage WP leads to ensure that their own WP and Task members are aware of the generic instructions and also any additional information required in respect of assets used in their own WP or Task
- to provide appropriate GDPR awareness training at consortium group meetings

Implications of GDPR for research projects

The GDPR concerns only 'personal data', so information derived from non-personal subjects (e.g. *in vivo* or *in vitro* animal studies) are not subject to GDPR; similarly, information about deceased persons is not subject to GDPR (but may be under ethical confidentiality restrictions).

Data on patients or participants in clinical studies will generally be 'personal data' under GDPR, though may be 'anonymised' to a degree that it is no longer identifiable (see section on Anonymisation & Pseudonymisation).

One main change from the previous Data Protection Directive is that personal email addresses (even with institutional domain names) are considered 'personal data' (unless of the form 'postmaster@ xyz.com') – even 'rgz12@cam.ac.uk' would be considered personal if 'rgz' are the person's initials as it would likely permit them to be identified.

It is anticipated that, while most consortium participants would be aware that any data relating to patients, whether from clinical trials or EHR systems, are normally 'personal data', they may be unaware of this minor, but key, change in the law (and its interpretation) and the need to register such 'administrative' holdings of 'personal data' with their Data Protection Officers (DPOs).

GDPR Principles

These 'principles' are laid out in Article 5 and require that 'Personal data shall be:'

Principle title	Description
lawfulness, fairness and transparency	Processed lawfully, fairly and in a transparent manner in relation to the data subject
purpose limitation'	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
data minimisation	adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed
accuracy	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
storage limitation	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

Principle title	Description
integrity and confidentiality	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
accountability	The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 [viz. all the preceding principles]

The implications for EU-wide projects are developed further in *Appendix D – Implications of GDPR for projects such as ConcePTION*.

Controllership

GDPR concerns 'controllers' of 'personal data' and 'processors' as well as the 'data subjects' themselves and their individual rights and freedoms. Article 4 defines many of the relevant terms, though the precise interpretation may be elaborated in other Articles or in the Recitals:

Term	Definition	Notes
controller	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	See Article 24 on 'Responsibility of the controller'; also Article 26 on 'Joint Controllers'
processor	a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;	See Article 28 for more details as well as Article 29 'Processing under the authority of the controller or processor'
processing	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	

Term	Definition	Notes
recipient	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	
third party	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	

ConcePTION is a project but not a legal entity, so that it cannot be a 'controller' under GDPR. The obvious contenders for controllership are the institutions involved in the consortiums and the individual project participants themselves (who may, deliberately or inadvertently), create their own holdings of 'personal data').

At one level, this could be interpreted to mean that the project need undertake no GDPR compliance checks as all responsibility lies with the component organisations to undertake all such checks themselves in accordance with their own organisational standards, policies, and procedures and their various countries' precise legal codes. In strict legal terms, this is indeed the case and ConcePTION cannot change this.

However, it would be clearly problematic to the consortium if one or more of its organisations were found in breach of GDPR through any misunderstanding as to which organisation was the controller and hence some handling of 'personal data' was mis-managed and hence non-compliant.

Note: There is differing guidance as to whether the institution running a study is the controller or whether it is the sponsor of the study that is the controller (with the institution(s) running the study as 'processors'). This may have implications for some of the ConcePTION demonstration projects, especially where run across different legislations or regulatory settings.

SME exemption

Recital 13 notes: '*To take account of the specific situation of micro, small and medium-sized enterprises, this Regulation includes a derogation for organisations with fewer than 250 employees with regard to record-keeping*'. This is detailed in Article 30(1-4) but relates solely to keeping a register of personal data processing activities, so does not carry any exemption from other records such as DPIAs.

'Filing system' concept

One other exemption is offered by Recital 14: 'Files or sets of files, as well as their cover pages, which are not structured according to specific criteria should not fall within the scope of this Regulation'. Generally, this should mean that occasional or casual references to people would not be considered to make a record 'personal data', but if there were a ready way to locate information about a person, then it is likely to be considered a 'filing system'. An email system would likely be considered a 'filing system' in this context. Hence the concern about contact lists.

Strictly 'personal' use

Recital 18 makes clear that individual use is not restricted by GDPR; however, its phrasing:

This Regulation does not apply to the processing of personal data by a natural person in the course of a purely personal or household activity and thus with no connection to a professional or commercial activity.

so use by members of the consortium in their capacity as members of the consortium through their organisation would not be exempt under this Recital and would be considered a 'personal data' holding or processing for their organisation.

'Personal Data' register

Article 30 of the GDPR requires every controller to keep a record of its 'personal data' holdings.

To this end, it would be wise at least for the consortium to undertake a review of what 'personal data' holdings are anticipated to be created for the consortium and ensure that each 'controller' organisation is aware of this holding and has it recorded in their own Article 30 register and so that it is properly managed and clearly that organisation's responsibility.

The WP7-T7.2 task force has already started a preliminary list of such holdings (see Appendix 1) and would look to WP8 to take this forward as part of its responsibilities in administering the ConcePTION project.

GDPR Transparency requirements

Transparency requirements (Articles 13 & 14) fall mainly on the 'controller' of any personal data (PD) processing. However, there is both a general ethical obligation on the project to be transparent as well as self-interest in promoting the activities and progress of the project. This last falls into the ambit of WP5, WP6 and WP8.

It also makes sense to have a single central explanation of what ConcePTIOn does to which other 'controllers' can refer and direct users for more detail, then require each site to have detailed (and possibly conflicting) descriptions of the project, the PD processing, and what individual data subjects may do to exercise their rights. This would include a list of participants, likely to hold relevant PD, and their supervising authorities.

The individual organisations would still need to provide the minimum legal details, but could rely on the main ConcePTION description for the broader perspective as well as finer detail, unless it makes more sense to describe on the 'controller' site (e.g. for demonstration studies).

[need to review 'Privacy Statement' on imi-conception.eu as appears to have been taken from a commercial site]

Data Protection Impact Assessments (DPIAs)

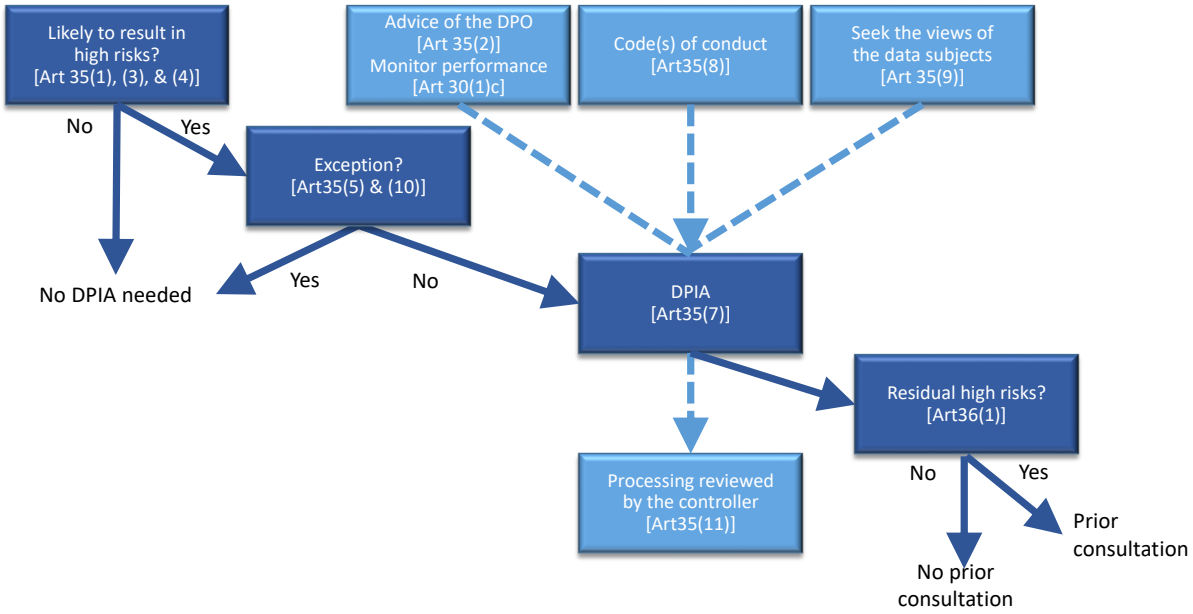
The GDPR Article 35 introduces a legal requirement to perform DPIAs in certain circumstances – normally where there may be 'high risk' to the rights and freedoms of data subjects through the proposed processing. This builds on the earlier 'good practice' of Privacy Impact Assessments recommended by the Article 29 Data Protection Working Group.

The actual requirement falls to individual consortium members to do such assessments on any 'personal data' holdings or processing that they may have or perform, but it is considered advisable that the consortium as a whole perform a form of DPIA to advise its members as to the overall processing that is expected to take place. However, this cannot replace each consortium member's legal obligation to perform a DPIA where necessary and to do its own due diligence over the detail of such a DPIA. Any document produced as part of the project can be at most indicative and 'advisory' only.

The Article 29 Working Group (now European Data Protection Board (EDPB)) produced a Working Paper (WP248) detailing how a DPIA should be performed.

A DPIA is only required where processing of 'personal data' is likely to present a 'high risk' to the rights and freedoms of the individual data subjects. If the DPIA shows that here is indeed a 'high risk' to the rights and freedoms of the individual data subjects, then it must be referred to the supervising Data Protection Authority (DPA) before any processing takes place. The DPA may approve or reject the proposed processing.

WP248 includes this diagram to aid decision-making about having to do a DPIA or not:



Article 35(3)b makes clear that for 'special category' data, e.g. health data, then a DPIA is required. This would apply to any patient records created or gathered for the demonstration studies or other purposes. It may be that where suitable controls are applied (e.g. anonymisation or aggregation of results) then any DPIA might be relatively brief.

Article 35(7) notes that a DPIA must contain the following elements:

- a systematic description of the envisaged processing operations;
- the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
- an assessment of the necessity and proportionality of the processing operations in relation to the purposes;

- an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and
- the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

Individual organisations should have their own templates and processes for completing and approving DPIAs, so there is no need to develop a ConcePTION template or approach. Data Protection Authorities may also provide exemplar templates[1].

However, for simple administrative use of contact list details for non-patient contacts, especially if used within a controlled and secure environment (e.g. an institutional IT environment) then it is hard to see how these could constitute a 'high-risk' as long as the controlling institution properly upholds data subject rights (e.g. of access, correction, erasure, and to prevent further processing).

Contact details for 'patient representatives' or representatives from patient groups (e.g. EIWP) would be no different as long as the data is used in the same way as for other ConcePTION contacts (who may also be patients, but in a distinctly different context).

Transfers outside EU/EEA[2]

GDPR Chapter V (Articles 44-50) lays down requirements for 'personal data' transfers outside the EEA (or to international organisations).

The ConcePTION project does not expect any personal data concerning patients to be transferred outside the EEA[3], except in aggregated format, so that the restriction on overseas transfers would not apply as the data would not constitute 'personal data'.

The importing of patient data from outside the EU will depend on the nature of the data (whether identifiable, anonymised, or aggregate – the last would make such transfers very straightforward) and the privacy laws of that external country. GDPR does not place restrictions on the importing of 'personal data', though once within the EU it would be subject to EU data protection laws, if applicable.

[1] For example, in the UK: https://ico.org.uk/media/for-organisations/documents/2553993/dpia-template.docx

[2] At the time of writing the UK was no longer a member of the EU, but was in a 'transition period' during which EU legislation still applied; at the end of the 'transition period' (31st December 2020), the UK might fully exit on a 'no-deal' basis (so be outside the EEA) or under some negotiated deal which may include it with the EEA or equivalent arrangement. Its data protection legislation may be deemed to be 'adequate' and so join the 'white-list' of countries/jurisdictions which are deemed to have equivalent privacy controls to the EU/EEA

[3] Similarly for transfers from the UK (when outside the EEA) to locations within the EEA.

Anonymisation & Pseudonymisation

The definition of 'personal data' is quite broad, but there is no definition of 'anonymised data' – basically if there is some potential privacy risk, then the data should be protected – only where there is no risk at all could we expect that the data could be shared freely, e.g. by publication on the Internet. Indeed, all too often such datasets openly published have proved to have weaknesses such that some of the data (not necessarily all) can be re-identified.

Article 4 defines 'pseudonymisation' as 'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'. However, that is not to say that such data cannot possibly be re-identified.

Recital 26 elaborates on this point: The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

This text can be developed into the following table:

Category of data	Description	GDPR applies
Identified data	Data which identifies a person	✓
Identifiable (but not identified) data	Person is not obviously identified, but by linkage with other data or other means can be re-identified	✓
Data which cannot be identified by means 'reasonably likely' to be used	Identifiers have been removed; data has been blurred; there are contractual and organisational constraints on its use or availability; a risk assessment has been made concerning 'means reasonably likely to be used'	✘
Data which can never be re-identified or was never personal data	Note: non-personal data might become personal in some contexts, e.g. car registration or IP address; aggregate data may still be subject to reconstruction and other attacks	✘

Notes from Recital 26:

- *The principles of data protection should apply to any information concerning an identified or identifiable natural person* – this is covered by the first two categories of data
- *Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person* – this is occasionally précised into 'pseudonymised data is personal data' which is rather too simplistic: just pseudonymising data may not ensure that the data is anonymous, so a better precis would be 'pseudonymised data may well still be personal data, unless assured to be unidentifiable by means reasonably likely to be used'.
- *To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments* – this has two aspects: the 'singling out' or locating unique records in the dataset (so k-anonymity would prevent this) and then the actual possible identification of the person to which that record relates. What is not covered in this is the question of assured identification: might the attacker locate a record which they think is a person's, but they are mistaken (as the person may be in the dataset) or they have located a similar record but not the actual target person's record. The key element here, though, is the 'means reasonably likely to be used' – this suggests not just a theoretical risk (e.g. NSA could crack it in a century) and a recognition of the context (e.g. who may actually have access)
- *The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes* – this last phrase is just 'for the avoidance of doubt'. However, the general premise of GDPR is that if the processing is on the basis of the data being anonymous, then there must have been an assessment of sorts and this should be documented – unless it is self-evident (in the case of non-personal data).

Conclusions

It is recommended that the Managing Board consider these actions:

- to provide a structured list of information assets (derived from the results of the survey) to partner DPOs to ensure they are aware of these assets, can arrange appropriate DPIAs (if required), and add them to any information asset registers (including Article 30 registers) they may have;
- to provide generic instructions to consortium staff about privacy, security, and legal requirements in respect of such assets
- to encourage WP leads to ensure that their own WP and Task members are aware of the generic instructions and also any additional information required in respect of assets used in their own WP or Task
- to provide appropriate GDPR awareness training at consortium group meetings

It is expected that this document will be updated later in the project to reflect possible changes within the ConcePTION project itself and also within legal and regulatory guidance, including any possible new codes of practice or codes of conduct.

Appendix A – Reflections on Codes of Conduct/Practice

The definitions usually used are:

- Code of Conduct: a set of rules outlining the norms, rules, and responsibilities of, and proper practices for, an individual or organisation when performing particular activities within an defined arena [Wikipedia – adapted by PS]
- Code of Practice: A code of practice can be a document that complements laws and regulations to provide detailed practical guidance on how to comply with legal obligations, and should be followed unless another solution with the same or better standard is in place, or may be a document for the same purpose published by a self-regulating body to be followed by member organisations [Wikipedia – adapted by PS] – for example:
    - the 'Confidentiality: NHS Code of Practice' (2003) sets out what UK health and care organisations have to do to meet their responsibilities around confidentiality and patients' consent to use their health records.
    - 'Anonymisation: managing data protection risk code of practice' (2012) by the UK Information Commissioner's Office (ICO) - The code will help all [UK] organisations that need to anonymise personal data, for whatever purpose. Note: produced pre-GDPR.

Source	Code of Conduct	Code of Practice
The International Federation of Accountants	Principles, values, standards, or rules of behaviour that guide the decisions, procedures and systems of an organization in a way that: (a) contributes to the welfare of its key stakeholders, and (b) respects the rights of all constituents affected by its operations.	[not defined]
Collins Dictionary	an agreement on rules of behaviour for the members of that group or organisation	a set of written rules which explains how people working in a particular profession should behave.
YourDictionary	a collection of rules and regulations that include what is and is not acceptable or expected behaviour	
Cambridge Dictionary	a set of rules about how to behave and do business with other people	a set of standards agreed on by a group of people who do a particular job
Chambers Dictionary	An established method or set of rules for dealing with, behaving in, etc. a particular situation [considers synonymous]	
Business Dictionary		Written guidelines issued by an official body or a professional association to its members to help them comply with its ethical standards.

Source	Code of Conduct	Code of Practice
Wikipedia		A code of practice can be a document that complements [occupational health and safety] laws and regulations to provide detailed practical guidance on how to comply with legal obligations, and should be followed unless another solution with the same or better [health and safety] standard is in place, or may be a document for the same purpose published by a self-regulating body to be followed by member organisations

GDPR Articles 40 and 41 provide significant detail on the requirements for a 'code of conduct' at least in relation to data processing (or particular aspects of processing as detailed in Article 40(2)) and the approval of a particular code:

- There must be mechanisms to monitor compliance to the code – it does not make a requirement on any particular body to effect this, but the mechanisms must exist (and presumably be robust) as detailed in Article 41 – unless the code applies to public bodies (who are presumed to have to comply and do not need to be monitored).
- A code may be limited to a single Member State (codes covering several MSs require a more complex approval process)

Effectively, a 'code of conduct' sets minimum obligatory requirements (and so needs a mechanism for ensuring compliance/adherence), whereas a 'code of practice' (which may well be more detailed), while identifying minimum standards may also promote 'good practice' above the minimum as well as allowing possible variation where several approaches may be possible or pragmatic considerations may apply.

Selecting the appropriate Code of Conduct/Practice

A number of different 'codes of practice' or 'codes of conduct' have been developed by past EU projects or international collaborative bodies in relation to information processing and use – often focused on a particular aspect of information use in research.

This paper explores some of the history of such codes and seeks to identify which codes are most appropriate for certain types of data use, as well as where these may need updating (often to reflect the introduction of GDPR in May 2018) or perhaps consolidation.

Comparison of the various codes was one of the anticipated activities for Task 7.2. In practice this proved difficult except at the most general level:

Governance instrument profiles for consideration in the IMI Conception IG Policies and CoC

	ADVANCE CoC	ENCePP CoC	EMIF CoP	RD-Connect CoC	BBMRI CoC	Potential role of SAB
Research governance: ensuring appropriate conduct of the research and the handling of results						
Over-arching principles of the purposes of the research	■	■	▒			
Transparency of how data sharing/research requests are assessed	■		▒			
Whether data access providers have privileged / first research use						
Handling overlapping research requests						
Transparency about research sponsors	■	■				
Restriction on the role of research sponsors		■				
Restrictions on kinds of research organisation						
Restrictions on kinds of research purpose	▒					
Requirements for declarations of interest	■	■				
Requirements regarding independence (e.g. of persons involved)		■				
Handling potential conflicts of interest or risks of bias		■				
Requirements for transparency of the study, public registers	■	■				
Restrictions regarding development of the research protocol		■				
Restrictions regarding data processing and analysis		■				
Requirements regarding transparency of the analysis		■				
Scientific engagement of the Data Access Provider						
Analysis engagement of the Data Access Provider						
Requirements regarding transparency and timeliness of the results	■	■	■	■		
Handling of negative findings	■	■				
Dissemination of findings	■	■				
Publication constraints and inclusion of the Data Access Provider				■		
Information governance: privacy protection and safeguarding the interests of data sharing parties						
Differentiating the processing of identifiable, pseudonymous and anonymous data			■	■	■	
Transfers of data between jurisdictions			■	■	■	
Protecting data subject privacy			■	■	■	
Complying with applicable ethics and consent			■	■		
Capture of consent, withdrawal of consent				■	■	
Interactions with data subjects, trial participants				■		
Obligations on data documentation, metadata, provenance data, catalogue entries			■	■		
Formalising and adhering to the purpose of agreed data sharing			■	■		
Information security policies and obligations			■	■	■	
Obligations to prevent re-identification			■		■	
Arrangements for data transfer, remote access, remote querying			■		■	
Data cleaning and enhancement			■			
Methods for obtaining additional data and samples			■			
Remuneration and recognition of the Data Access Provider			■	■		
Handling of incidental findings			■	■		
Accountability, audit, sanctions			■	■	■	
Notification of issues, breaches			■		■	
Possible topics for the Scientific Advisory Board						
Relevance of the research topic to knowledge gaps						■
Scientific validity and quality of the research question						■
Statistical validity and quality of the research question						■
Applicability of the source data to the research question						■
Potential value of the results						■
Legend: Deep colour indicates that the instrument has clauses that cover this topic, Light colour indicates that the instrument touches on this topic.						
NOTE: This table only indicates the areas of scope overlap. It does not seek to detail what these clauses contain or to compare clause details between instruments. This will be later work.						

The difficulty was that each code was structured very differently, used different terminology (partly, of course, to reflect the different area of processing/use), and often worked to a different level of granularity, so that possibly comparable requirements would appear as detailed but distinct requirements across one document, but as a single quite general requirement in another. This made it hard to determine whether the requirements actually aligned or not or were intended to address the same fundamental need or obligation.

To this end, this document suggests a possible overarching structure for such codes, including particularly a separation of legal/regulatory requirements.

In particular, it is recommended that such codes are called 'code of practice' to avoid any confusion with possible Article 40/41 'code of conduct' which would have regulatory power once approved by appropriate SDAs or EDPB, especially as they are usually about promoting 'good practice' rather than just legal conformance (which is not to preclude detailing practical solutions to legal compliance).

Background to existing Codes of Conduct

Some of the Codes of Conduct have been designed with particular data distribution models in mind, often based on particular projects or types of study or trial.

FAIR Data principles

This is a core requirement on the ConcePTION project: its data catalogue should meet the FAIR data principles as completely as possible. This is, in effect, a broad 'code of practice' for research data catalogues.

These principles are detailed in Wilkinson et al 2016:

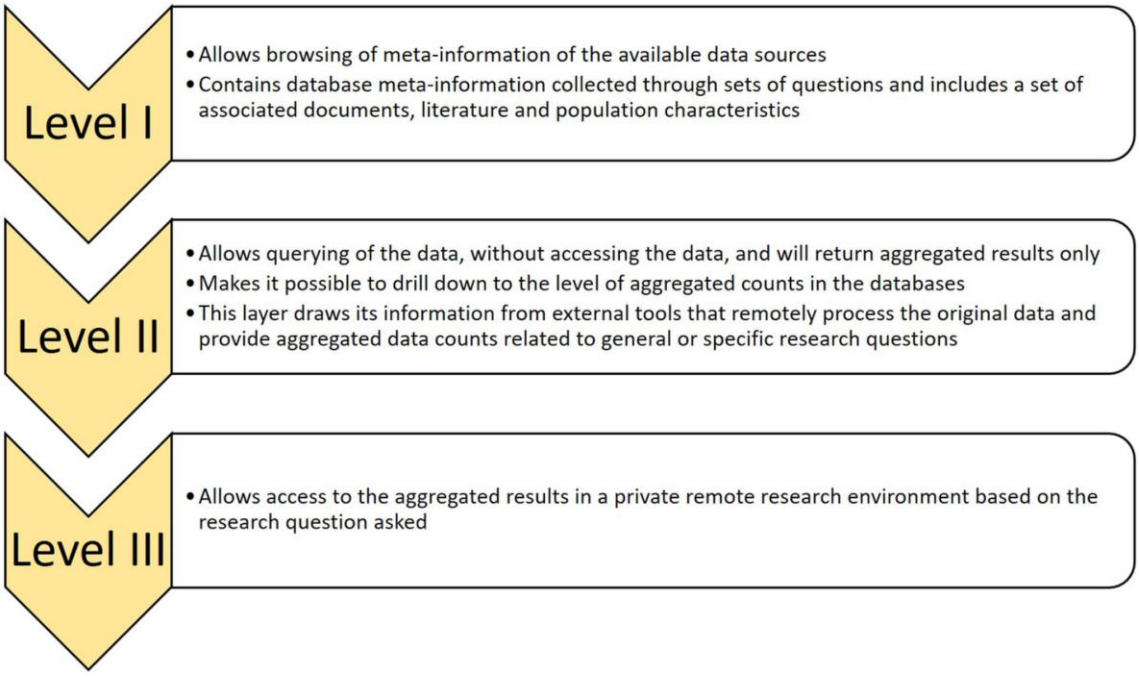
	Principle[4]	
FINDABLE	F1.	(meta)data are assigned a globally unique and persistent identifier
	F2.	data are described with rich metadata (defined by R1 below)
	F3.	metadata clearly and explicitly include the identifier of the data it describes
	F4.	(meta)data are registered or indexed in a searchable resource
ACCESSIBLE	A1.	(meta)data are retrievable by their identifier using a standardized communications protocol
	A1.1	the protocol is open, free, and universally implementable
	A1.2	the protocol allows for an authentication and authorization procedure, where necessary
	A2.	metadata are accessible, even when the data are no longer available
INTER-OPERABLE	I1.	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
	I2.	(meta)data use vocabularies that follow FAIR principles
	I3.	(meta)data include qualified references to other (meta)data
RE-SUABLE	R1.	meta(data) are richly described with a plurality of accurate and relevant attributes
	R1.1.	(meta)data are released with a clear and accessible data usage license
	R1.2.	(meta)data are associated with detailed provenance
	R1.3.	(meta)data meet domain-relevant community standards

[4] from www.nature.com/articles/sdata201618 ; Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

The principles may be applied to either the meta-data about a digital object or its actual content – hence the use of '(meta)data' above where a principle may be applied to both the data and the meta-data.

It is worth noting that in the context of the data itself, 'interoperable' is used in the sense of 'processible', not necessarily that the data is appropriate for some other use in a different context. This is a key consideration when considering re-use of data recorded for a different (though related) purpose.

This distinction between application of the FAIR principles to both data and metadata is reflected in the EMIF approach of levels for data catalogues:



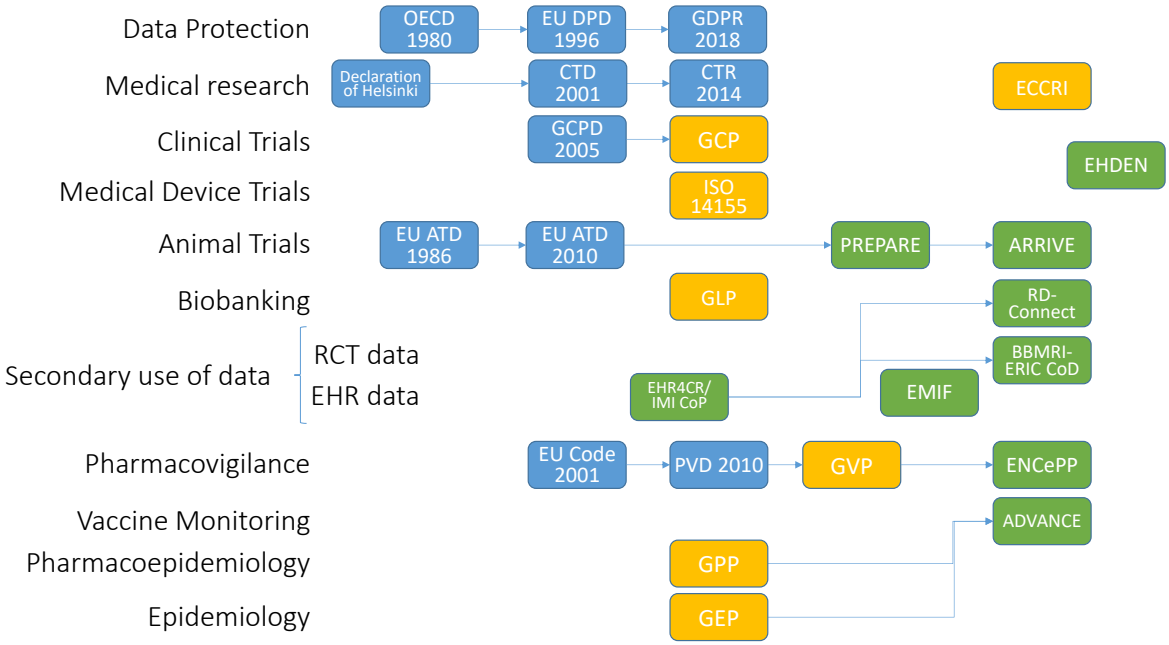
From 'The European medical information framework [EMIF]: A novel ecosystem for sharing healthcare data across Europe'; Learning Health Systems, First published: 25 December 2019, DOI: (10.1002/lrh2.10214)

Relevant research codes of practice

Acronym	Organisation	Focus	Notes/comments
ENCePP	encepp.eu	Pharmacoepidemiology and Pharmacovigilance	Part of GVP (Good Pharmacovigilance Practice)
ADVANCE	vac4eu.org	Best practice and code of conduct for benefit-risk monitoring vaccines	By Miriam Sturkenboom; designed as an alternative to ENCePP to address conflicts of interest Builds on Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology (ISPE) and the Good Epidemiology Practice (GEP)
EMIF	emif.eu	Data catalogues, data access, and distributed querying	Built on ENCePP & EHR4CR SOP (not EHR4CR CoP)
ELIXIR ELSI Policy	elixir-europe.org	Omics data re-use	ELIXIR framework for secure archiving, dissemination and analysis of human access-controlled data
RD-Connect	rd-connect.eu	Biobanking, including genomics	Produced by Mats Hansson; specific to RD-Connect GPAP, but adapted from EHR4CR CoP
EHR4CR	imi.europa.eu	Secondary Use of Medical Data in Scientific Research Projects	Bahr & Schlünder – developed in parallel with EHR4CR but as part of integration across IMI projects
BBMRI	code-of-conduct-for-health-research.eu		Building on EHR4CR CoP? Since 2015 No information currently available
SUMMIT	www.imi-summit.eu		A set of principles [no document found!] referenced in EMIF documentation as source
PREPARE	Norecopa	Planning Research and Experimental Procedures on Animals	Checklist.
ARRIVE	NC3Rs	Animal Research: Reporting of In Vivo Experiments	
CORBEL?	corbel-project.eu		Code of Conduct for Health Research (Mayrhofer)= BBMRI above Mats Hansson is on BBMRI ELSI

Acronym	Organisation	Focus	Notes/comments
GEANT	geant.net	Data Protection Code of Conduct for identity providers	Pre-GDPR; minimal requirements and relates to identity of end-users
EU Directive 2010/63/EU	European Commission eur-lex.europa.eu	Protection of animals used for scientific purposes	Article 30: Animal Records Article 31: Information on dogs, cats and non-human primates Article 32: Marking and identification of dogs, cats and non-human primates
LASA 2015	RSPCA/LASA		Royal Society for the Prevention of Cruelty to Animals (RSPCA) and Laboratory Animal Science Association (LASA) – seems to be UK-specific based on Home Office requirements through Animal Welfare and Ethical Review Bodies (AWERBs)
ECCRI	Allea.org		European Code of Conduct for Research Integrity – first published 2011
UKRIO	www.ukrio.org		UK Research Integrity Office Published 2009
GLP	oecd.org	Good Laboratory Practice	Chemical safety & testing
FAIR		Research data re-use	Principles: findable, accessible, interoperable and reusable(FAIR)

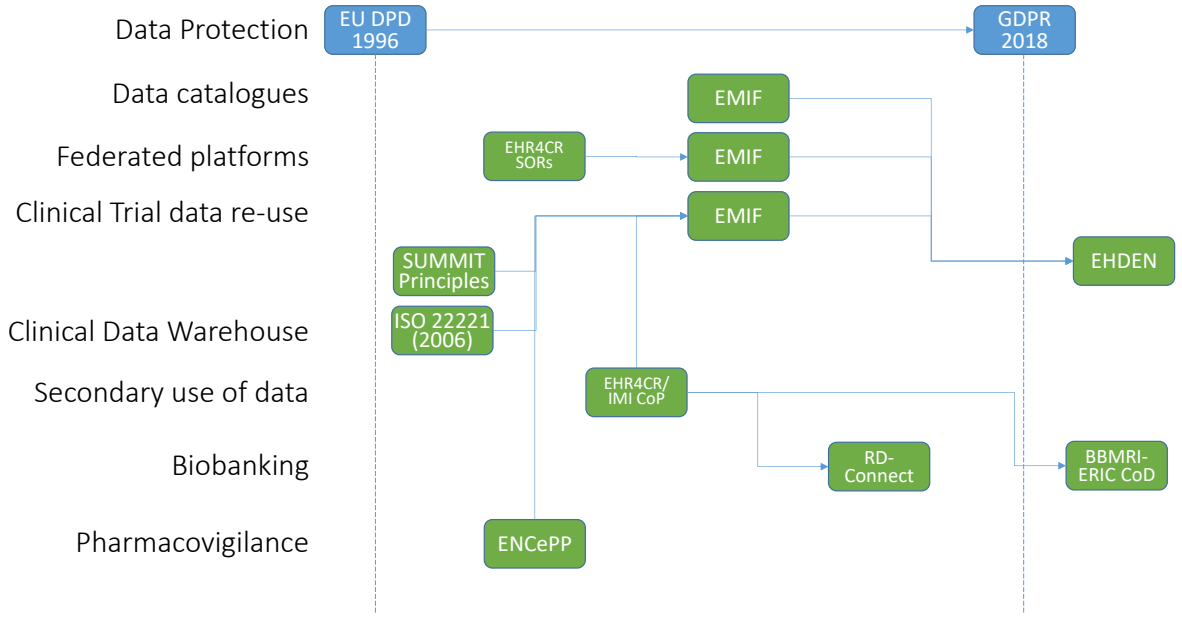
Several of these codes are inter-related, having developed from a number of broader initiatives, including general principles which were then adopted into law, e.g. data protection and privacy, but applied in particular areas of research:



Timeline of Legal requirements & Codes of Conduct for research

Blue is legislative; yellow – codes of conduct, and green codes of (best) practice

In particular for Codes of Conduct relating to data use:



Timeline of Legal requirements & Codes of Conduct for data use

The diagram omits any reference to member state legislation, either providing specific health-related law or implementing derogations permitted or other aspects required by GDPR itself. Equally, EHDEN has only been operational for a year and is not expected to produce any pertinent deliverables yet.

Appendix B – Data Governance for Research projects

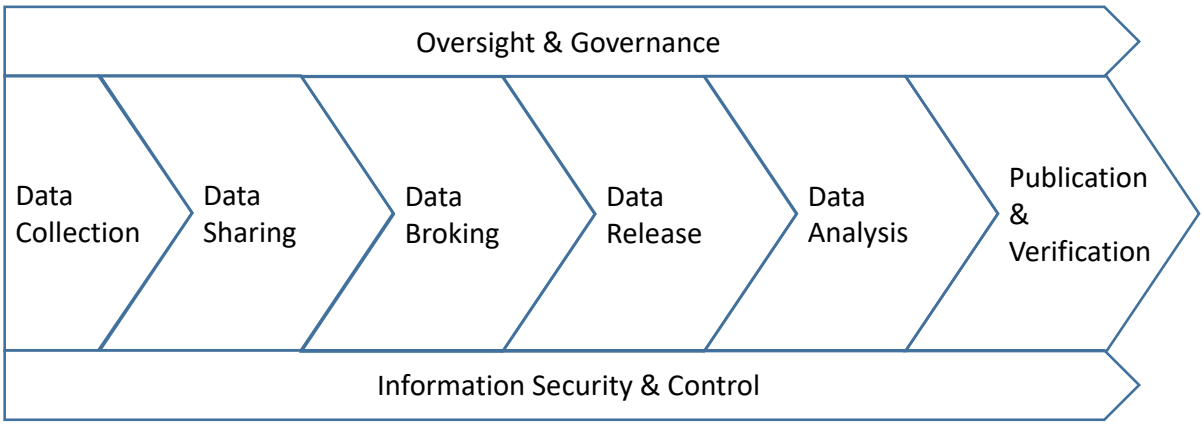


Figure 1- Data Governance aspects

All research projects will require some measure of Oversight and Governance as well as adequate Information Security; some straightforward projects will only cover Data Collection, Data Analysis and Publication steps, though will still have to address some questions over data-sharing in terms of the Open Data initiative to ensure that the analytic data is open to scrutiny and re-use.

Description of steps

Step	Description
Data Collection	Original data collection from patients, care provision, or sample analysis; includes requirements for transparency (including GCP and GDPR Article 13 transparency requirements)
Data Sharing	Covers regular data-sharing with other data controllers for generic purposes, possibly including routine collection for research, e.g. longitudinal studies
Data Broking	Covers the collection of data from other sources, curating, and making available to studies; usually involves a rigorous assessment process, often with independent oversight or advice; inc. GDPR Article 14 transparency requirements
Data Servicing	[Not in diagram] Provision of data processing services, including acting as a 'trusted third-party' or 'safe haven' for handling 'personal data' as in intermediary between other parties. May collate, clean, and/or de-identify. May overlap significantly with preceding or other steps, but acting as a 'processor' rather than a controller, but can be a significant aspect of any governance arrangement by providing an assured service
Data Release	Covers specific data releases to research institutions for specific research studies for health-related purposes, mainly assurance and establishing Data-sharing Agreement (DSA), but may also include risk assessment to ensure meets data minimisation standards (=anonymisation)

Step	Description
Data Analysis	The core of data research; may include some data cleansing or normalisation. Main current issue is machine learning, though many of the more straightforward approaches introduce no new issues versus statistical techniques; may include GDPR Article 14 transparency requirements, depending on status of data received
Publication & Verification	May cover two sets of requirements: limiting identifiability of results in publications and allowing appropriate data to be available for re-use or result verification (possibly using different analytic methods); ICMJE requirements
Oversight & Governance	Any arrangements need to have independent oversight as well as governance structures to ensure that policies and procedures are respected and regularly reviewed to maintain legal compliance and best practice
Information Security & Control	Covers not only basic IT security, preventing unauthorised access or loss, but also access control and authorisation, as well as pro-active auditing and monitoring with reporting to oversight bodies
Standards Assurance and Certification	[not in diagram] Helps minimise overheads of vetting other parties if they are assured or certified as meeting certain standards as to information security and/or governance

Data Collection requirements

Requirement	Description
GDPR transparency	Need to define purposes, legal bases, likely recipients of data [Article 13]
GCP consent	Need to provide specific information to patients concerning the proposed research study as well as data use.
General transparency	There should also be public materials about the specific research study and data use, both for the general public, those wishing to gauge the research study or institution, or for potential participants seeking further information. Note: all the communication material need to be consistent and in line with consent forms
Determine Purpose(s)	Both for immediate purposes (provision of healthcare, involvement in a research study, quality, safety, and governance requirements) and for further use (e.g. provision to other studies in anonymised form, to other approved studies to apply different analytical techniques to the data, for validations studies, etc.)
Legal bases	Article 6; and likely Article 9 where data collected relates to human participants, their tissues, or medical history

Requirement	Description
Consent Model	Need to anticipate likely uses in order to cover under consent (=set purpose); may need to separate choices to ensure each choice has valid consent and is not coercive (=respect patient autonomy); ensure consent properly captured in dataset and coding is flexible to allow for changes in consent model; equally, be able to handle withdrawal appropriately (complete withdrawal from intervention and data use or just from intervention or just supplementary data uses – or whatever)
Data Management Plan	Various funders provide templates for DMPs; some put more emphasis on data-sharing, others on data security and assurance.
FAIR principles	Findable, Accessible, Interoperable, Reusable (Wilkinson et al., 2016[5]).
ICMJE	International Committee of Medical Journal Editors (ICJME) and 'Open Data' initiatives by BMJ Clinical Trials (post Jan 2019) must have a DSP Data-sharing statement required in any publication: *Data sharing statements must indicate the following:* • *whether individual de-identified participant data (including data dictionaries) will be shared;* • *what data in particular will be shared; whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.);* • *when the data will become available and for how long;* • *by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism).*

The EDPB Opinion 3/2019 is particularly helpful in resolving some of the issues around consent under GCP and under GDPR.

One key aspect is to determine what options or choices can be supported within the remit of the proposed study – to what degree can the candidate have a free choice over further processing and at what point does it become ineffective for a study to include participants who impose significant restrictions on data use.

There is no doubt that seeking consent to further use for research purposes is critical here, though requires clarity about how further release will be controlled. Part of the difficulty is the general GDPR requirement for 'specific consent' which tends to preclude seeking consent for a broad range of future research – one way around this is de-identification and the use of the research exemption, but this can limit options for data linkage and enrichment.

It will be difficult for each research study to establish its own governance programme for reviewing and approving data release requests for research. However, it is probably equally difficult to establish a national or European body to receive diverse datasets and manage a range of different consent models (as well as the technical difficulties of managing the diverse datasets and met-data required).

[5] The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. doi:10.1038/sdata.2016.1

Dryad, the digital repository
While cited by BMJ as possible repository, it is mainly used by non-human studies, where there are far fewer issues around data re-use. Its mission includes three aims: 'discoverable, freely reusable, and citable' – regrettably not the 'interoperable' aspect of FAIR – mainly because it is 'flexible about data format', so does not require any form of meta-data to cover structure.

Data Sharing requirements

Requirement	Description
Agreements	Need to define uses as well as all procedural aspects (termination, data destruction, incident management, process variation, etc.)
Controller/ processor relationships	This will determine a number of legal obligations under GDPR (e.g. transparency, processing registers, etc.) as well as the relationships for contractual agreements
Legal bases & consents	Need to ensure that legal bases for sharing are consistent with original collection and that onward sharing is covered by any original consents (or that further consent is sought if possible/practical)
Data sharing policy	Need to determine what categories of sharing will be supported (both operationally and from a consent perspective) and with what categories of institution (including how eligible institutions will be vetted); what restrictions on data shared (data minimisation) and any subsequent data-sharing; process (=DPIA)
Risk assessment	Viz. DPIA, to include data minimisation approach
Data-sharing Plan	ICMJE requirement (see above)
Data-sharing mechanism	A mechanism whereby a research institution may apply for access to the data, either for a one-off release, a repeated release (for longitudinal studies) or to permit queries to be run through a portal and aggregate results returned

The core of these requirements is a Data-Sharing Plan (DSP) which determines how (if at all) data will be onwardly shared, either as an active option or merely to meet 'Open Data' requirements.

Data Broking requirements

Requirement	Description
Data gathering	As part of data 'broking' it is assumed that other parties will be actually doing the original data collection; the data broker arranges or coordinates access to other parties' datasets, either through distributed queries or by holding curated (and possibly reduced) copies of the original data
Data cleansing	Including reduction or re-coding;
Dataset catalogue	Details of actual collections and pertinent meta-data, including legal or other restrictions on access or use

Requirement	Description
Data linkage	May or may not be available or may only be available between certain datasets or at certain population levels (e.g. at postal area or region, but not at individual-level) May make use of a trusted third-party or 'data safe haven' to effect the actual linkage process at the individual level
Data selection	It is normal only to release an appropriate sub-set or sample of the population, both for operational efficiency and also to reduce data attribution risks, while permitting effective analytical power to any study. Very small samples or cohorts may not be permitted as re-identification risks may be too high.
Data extraction	The production of the requisite linked datasets pertinent to the actual request – this may be released directly or processed further (see next) so that only aggregate data is released
Data aggregation	Rather than releasing micro-data, analytic summaries may be provided. This may be managed through an online portal with user access limited in the range of data available or the types of query that can be performed. Reconstruction attacks can be a particular difficulty for this data distribution model.

Data Servicing requirements

Requirement	Description
Identity management	For linkage purposes; may require access to an authoritative master patient index to match diverse identifier sets
Data security certification	If acting as a TTP, then needs a very high level of security and governance certification
Pseudonym key management	Multiple release to the same or different recipients requires both that distinct pseudonym schemes are used for different applicants (or studies) to ensure that data cannot easily be cross-compared, but also that longitudinal releases have persistent pseudonyms [to avoid having to release very large volumes of data]
Consent/dissent management	There may need to be mechanisms to be able to uphold data subject rights, particularly for objections to processing

Data Release requirements

Requirement	Description
Vetting & approval mechanism	Almost certain that some form of vetting will be required before release of health-related data, even if in aggregate form only. Often, includes some form of independent advisory committee to avoid possible bias or discrimination, but can represent a significant overhead.
Data retention policy	If releasing microdata, then usually there will be a requirement to delete the original data after some period; aggregated derived data is usually considered as controlled by the recipient

Requirement	Description
Data Release agreement template(s)	Requirements to be covered (Heads of Agreement) – see Appendix
Data access information	Usually on a web-site and needs to include an application form (online/ downloadable) as well as indication of approval criteria and likely timescales.

Data Analysis requirements

The actual requirements may depend on the approach taken:

1) data analysed in-house and released through a portal or sent to recipient via secure communication, perhaps as a chargeable service
2) microdata released for recipient to perform further analysis, in which case some appraisal of the security of the system environment would be needed, along with auditing and security arrangements to ensure compliance

Requirement	Description
Analytic service	Usually using a standard package, e.g. Stata, SAS, etc.
Client system appraisal	Where microdata is released, assurance is needed that data will be held securely and only used appropriately in line with contractual restrictions. May be no more than seeking ISO27001 or equivalent security – most likely covered as part of data-sharing arrangements and contracts above.

Publication and Verification requirements

The ICJME (International Committee of Medical Journal Editors) have developed a number of recommendations concerning the reporting of research in journals. As these are quite detailed, so only a relevant precis is included here:

Requirement	Description
Accountability	It should be clear who the authors are and where further information can be sought
OpenData/ Reproducability	Ideally the analytic data should be available for others to verify the results or to apply different methods to validate the conclusions of a study; at least there should be a statement about data availability
Acknowledgement	Data sources, and any IP rights, should be acknowledged
Dissemination/ Benefits	This may be a requirement of data provision or by Ethics Committees that studies should publish their results and seek to promote the knowledge gained as widely as possible
Transparency	Particularly over any possible conflicts of interests or potential biases in the data or results

Oversight and Governance requirements

These will cover both internal and external (client and processor) governance assurance – and many of the requirements have been addressed under separate elements above (e.g. vetting, auditing, system assurance, etc.)

Requirement	Description
Oversight	Many of the organisations involved will have the equivalent of an Ethics Board or independent board to provide assurance to the public that their information is only being used appropriately
Incident reporting	Required under GDPR, but should include the internal reporting of lesser incidents and their resolution to provide an overall picture of the security and resilience of manual and computer systems and processes. Ideally there should ab at least an annual report to the oversight board about system security, incidents, staff training, and organisational resilience.

Information security and Control requirements

These can be extensive (as anyone who has prepared for ISO27001 certification can attest)

Requirement	Description
External certification	E.g. ISO27001 or similar (e.g. Cyber Essential Plus in UK). This provides external assurance of meeting at least minimum requirements
Access control	How new users are recorded on the system, privileges managed and reviewed, actual access monitored to detect intrusion or potential misuse, and obsolete account frozen or removed.
Staff training	Ensuring users understand the 'appropriate use' requirements, including contractual clauses and general confidentiality and data protection restrictions.
Pro-active Auditing & monitoring	Systems often have audit trails, but no ready facilities for analysis and are not routinely used by administrators to check for signs of poor use or mis-use. Effective procedures and practical tools need to be in place so that early signs of poor training or external attacks can be detected and addressed; redundant accounts can also be suspended or revoked to reduce the possibility of misuse.

Standards Assurance and Certification requirements

Requirement	Description
Certification criteria	[Note: these would only apply to a formal 'Code of Conduct' which would need certification, etc.]
Certification process	
Accreditation criteria	
Accreditation audit process	
Standards setting	

Appendix C – Data protection concepts

Note: the discussion here is intended to inform a general audience, in particular the likely audience for this deliverable, which requires some simplification at various points in the interests of brevity and readability. This should not be taken as a definitive interpretation of the law.

In legal terms[6] there are two distinct rights:

- right to a private life ('privacy') from the European Convention of Human Rights (Article 8)[7] and the European Charter of Fundamental Rights (Article 7 – Respect for private and family life); and
- 'data protection' rights from the European Charter of Fundamental Rights (Article 8 – the protection of personal data) and the Lisbon Treaty of 2009 which gave the Charter legal force.

The former may be interpreted as protection from 'intrusion' by state or other bodies, particularly the protection of correspondence, whereas the latter is about protecting a person's broader interests in terms of data collection and use.

Personal data

A concept fundamental to 'data protection' is what is (or is not) 'personal data' which should be protected. This is defined in Article 4 of the General Data Protection Regulation (GDPR) as:

> *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.*

This is very similar to the previous definition of 'personal data' under the EU Data Protection Directive (95/46/EC), which was explored in detail by the Article 29 Data Protection Working Party (now the European Data Protection Board (EDPB)) in WP136 'Opinion 4/2007 on the concept of personal data'. However, this paper has not been endorsed by the EDPB as being applicable under GDPR, but nor has the EDPB provided a replacement opinion (though individual DPAs may provide their own guidance).

While much focuses on 'relating to', so that mere ancillary information about individuals may not count as 'personal data' if the information is not indexed in reference to this information – however, electronic information is nearly always capable of being searched by text, so it would be hard to claim the appearance of a person's name was not their 'personal information' as might be the case in paper format[8].

Equally, items closely associated with a person may allow data that is about an object rather than a person to be 'related to' a person, e.g. a car registration number may allow the driver's location to be tracked, or a computer IP Address to be used to track a person's browsing habits.

These are interpretations that pre-existed GDPR, and still apply post-GDPR.

[6] See https://edps.europa.eu/data-protection/data-protection_en

[7] Based on the United Nations' Universal Declaration of Human Rights (Article 12)

[8] Such an argument might be more successfully used in terms of copies of electronic data held in archive format, so not readily searchable except through retained index files.

Company email addresses as 'personal data'

One key change is guidance about email addresses[9]. Previously, a business email address, fred.bloggs@company.com (as against gmail or Hotmail, etc.), was normally considered as 'relating to' a person's work activities rather than 'personal use' (indeed for many organisations there are general prohibitions on using a work email address for 'personal use').

However, with the advent of GDPR, guidance has made clear that this should also be considered as potentially 'personal data' as relating to that person's employment and business activities. This is not a specific change brought about by GDPR itself, but rather a coincidental clarification[10], which can have some significant effects in how organisations should control email-related information. A generic 'post box' address, e.g. info@company.com, would not be considered 'personal data', though specific emails to such an address might be 'personal data' depending on the context and the content.

Special categories of 'personal data'

Under the previous EU DP Directive, the term, 'sensitive personal data', was used to describe more sensitive categories of personal data, such as health-related data – the phrasing is that these are now 'special categories' of 'personal data' – to some degree avoiding the question of what makes them more 'sensitive', though deeming them to be self-evidently 'special':

- racial or ethnic origin,
- political opinions,
- religious or philosophical beliefs, or
- trade union membership, and
- the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person,
- data concerning health or
- data concerning a natural person's sex life or sexual orientation

It is notable that biometric and genetic data when processed other than for identification purposes would probably be treated as 'special category' as 'data concerning health', though perhaps not if used for purely genealogical purposes (e.g. ethnic mix).

Controller and Processor

Article 4 defines both these concepts:

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

Articles 24 & 25 further elaborate the obligations on a controller. Articles 13 & 14 cover the transparency obligations on the controller to inform data subjects about the processing.

[9] See https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en

[10] Perhaps from Privacy and Electronic Communications Regulations (PECR)

Article 26 notes the possibility of 'Joint controllers', but rather leaves the resolution of responsibilities to the controllers to decide between them.

Article 28 covers the responsibilities of the processor, particularly that there should be a clear contract establishing the responsibilities and the level of control to be exercised.

Legal basis for processing

Under the previous EU DP Directive, an organisation needed to consider the legal basis for processing of 'personal data' (Article 7) or the legal basis for processing 'special categories of data' (Article 8). Under GDPR, these are replaced with:

Article 6(1) – processing 'personal data'

- consent to purpose(s)
- performance of or entering into a contract to which the data subject is party
- legal obligation of the controller
- vital interests of the data subject
- a task carried out in the public interest or in the exercise of official authority of the controller
- legitimate interests pursued by the controller[11] – subject to the individual rights and freedoms of the data subjects
-

Article 9(2) – processing 'special category of personal data'

- explicit consent
- obligations or rights in respect of employment and social security and social protection law
- vital interests of the data subject
- legitimate activities of a foundation, association or any other not-for-profit body concerning its members
- data manifestly made public by the data subject
- necessary for the establishment, exercise or defence of legal claims or courts
- a task carried out in the substantial public interest
- for medical purposes, including public health – subject to professional secrecy or equivalent
- for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes – subject to Article 89(1)

These are very similar to the original Article 7 & 8 legal bases, however, under GDPR the phrasing requires the selection of one or more bases under **both** Article 6(1) and Article 9(2) when processing 'special category personal data'

Pseudonymisation and Anonymisation

Only 'pseudonymisation' is actually defined within Article 4 of the GDPR, though a form of data attack, 'profiling' is also defined:

[11] This does not apply to processing carried out by public authorities in the performance of their [public] tasks. They may have 'legitimate interests' simply as organisations needing to use email, keep proper accounts, etc.

Pseudonymisation	the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person
Profiling	any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements

This last might be deemed to include 'risk profiling' or 'screening' within healthcare or social care services. It echoes the use of the term 'singling out' which is used in Recital 26 as part of discussing 'anonymous data' which would be outside the scope of GDPR.

This perhaps illustrates the need at times to separate out what are the legal bases for different steps of processing, so for 'screening' there would be at least these elements:

- generic segmental analysis: what types of person are likely to be at risk from what sort of harm, and how best might they be detected/located (e.g. from particular events or levels of results) – this would be Article 9(2)j – scientific research in the public interest (the purely analytical/epidemiological processing) or perhaps Article 9(2)h – the management of health or social care systems and service (the planning & costing or health economics aspects)
- From the screening criteria determined, health data would be processed to identify specific individuals at risk and invite them to participate in the screening programme – this would be on the basis of Article 9(2)h – the provision of health or social care or treatment

Data subject rights

Articles 15-23 cover data subject rights (in addition to Articles 13 & 14 over the right to be informed):

- Right of access (Article 15)
- Right to rectification (Article 16)
- Right to erasure (Article 17)
- Right to restrict processing (Article 18)
- Right to know data recipients/obligation to inform data recipients (Article 19)
- Right to data portability (Article 20)
- Right to object (Article 21)
- Right not to be subject to automated individual decision-making, including profiling (Article 22)

These rights are subject to some qualification: e.g. the right to portability only applies where the data subject has provided the information; other rights may not apply if the data is anonymised. Articles 23 includes specific restrictions on these rights for purposes such as national security.

Article 25 - Data protection by design and by default – is quite broad, but would generally be deemed to include an obligation to ensure that these data subject rights can easily be upheld through appropriate system design. Equally, should a consortium, such as ConcePTION, receive a data subject rights request, then it should be able to direct the individual to the appropriate controller of any relevant dataset – hence the need for the data survey.

Codes of Conduct

Articles 40-43 introduce the idea of a 'code of conduct' as an instrument to bring together good practice which can be approved by a DPA (including the consistency requirements of Articles 60-67) and hence promote a standard of practice which is deemed consistent with the GDPR.

So far none have been approved, though BBMRI-ERIC has been working on a draft code for secondary use of healthcare data for some years. A potential code also needs a certification mechanism to validate that organisations claiming to abide by the code actually do so.

This requirement is considered the main difference between a 'code of conduct' and a 'code of practice' in this paper. Approval of both the code and the certification mechanism by a supervisory authority is part of the difficulty of getting such a code established.

Appendix D – Implications of GDPR for projects such as ConcePTION

Background to GDPR compliance

GDPR support for health data collection and further use

Article 4 defines '**data concerning health**' as 'personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status'. Recital 35 dwells on this a bit further: 'Personal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject. This includes information about the natural person collected in the course of the registration for, or the provision of, health care services as referred to in Directive 2011/24/EU of the European Parliament and of the Council (1) to that natural person; a number, symbol or particular assigned to a natural person to uniquely identify the natural person for health purposes; information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test.'

It also defines '**genetic data**' as 'personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question'. This last is clearly relevant for biobanking – or perhaps most medical records in the future. Recital 34 develops this slightly more: 'Genetic data should be defined as personal data relating to the inherited or acquired genetic characteristics of a natural person which result from the analysis of a biological sample from the natural person in question, in particular chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis, or from the analysis of another element enabling equivalent information to be obtained.'

The GDPR did not seek to address any of the complex issues around the familial implications of genetic material, preferring to deal only its implications as a potential identifier.

Further, it covers '**biometric data**' as 'personal data resulting from specific technical processing relating to the physical, physio-logical or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'. The difficulty is that it may include many medical data recordings (e.g. ECGs) which are not intended as a unique identifier but might be capable of being used in a 'biometric' sense for identification. A good example might be retinal scans by opticians – not particularly practical as a biometric measure (in comparison to an iris scan) but could easily be used in that way. Recital 51 does have some sense of relaxation of interpretation, at least in respect of the general processing of photographs

These three categories of 'personal data' are all considered 'special category', so subject to Article 9 requirements (as well as Article 6 requirements) for processing, so for the purposes of this document they may be considered essentially the same – and referred to generically as simply 'health data'.

Recital 52 allows member state derogations 'for health security, monitoring and alert purposes, the prevention or control of communicable diseases and other serious threats to health. Such a derogation may be made for health purposes, including public health and the management of health-care services, especially in order to ensure the quality and cost-effectiveness of the

procedures used for settling claims for benefits and services in the health insurance system, or for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes'.

Article 89 expands on the possibilities for the last element (see section below).

Recital 53 is less clear-cut in its implications, though introduces the question of obligations of 'professional secrecy': 'Special categories of personal data which merit higher protection should be processed for health-related purposes only where necessary to achieve those purposes for the benefit of natural persons and society as a whole, in particular in the context of the management of health or social care services and systems, including processing by the management and central national health authorities of such data for the purpose of quality control, management information and the general national and local supervision of the health or social care system, and ensuring continuity of health or social care and cross-border healthcare or health security, monitoring and alert purposes, or for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, based on Union or Member State law which has to meet an objective of public interest, as well as for studies conducted in the public interest in the area of public health. Therefore, this Regulation should provide for harmonised conditions for the processing of special categories of personal data concerning health, in respect of specific needs, in particular where the processing of such data is carried out for certain health-related purposes by persons subject to a legal obligation of **professional secrecy**. Union or Member State law should provide for specific and suitable measures so as to protect the fundamental rights and the personal data of natural persons. Member States should be allowed to maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health. However, this should not hamper the free flow of personal data within the Union when those conditions apply to cross-border processing of such data.'

Recital 54 makes it clear that '**consent**' is not a prerequisite for processing health data: 'The processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject. Such processing should be subject to suitable and specific measures so as to protect the rights and freedoms of natural persons. In that context, 'public health' should be interpreted as defined in Regulation (EC) No 1338/2008 of the European Parliament and of the Council (1), namely all elements related to health, namely health status, including morbidity and disability, the determinants having an effect on that health status, health care needs, resources allocated to health care, the provision of, and universal access to, health care as well as health care expenditure and financing, and the causes of mortality. Such processing of data concerning health for reasons of public interest should not result in personal data being processed for other purposes by third parties such as employers or insurance and banking companies.'. The last sentence does not wholly preclude the use of health data by such organisations, but would require a very clear justification that the processing was for 'public health benefit' in 'the public interest'.

Article 89(1): Safeguards for research

This Article requires that 'Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' (from Article 9(2)j) 'shall be subject to appropriate safeguards … for the rights and freedoms of the data subject'.

It explains that 'Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner'. This expands on the 'data minimisation' principle by requiring that the data is as de-identified as possible, possibly anonymous (though without using that term).

No precise detail is given for the level of de-identification required, but is clearly to be established within a DPIA and by having a specific 'risk to rights & freedoms' assessment in such a document.

Article 89(2) and (3) allow for Member State law to provide for derogations to (most) of the data subject rights detailed in Articles 15-21, where this might prevent or impair the fulfilment of the research purpose.

Background to ConcePTION project

ConcePTION is an IMI-funded project, so part-funded by each of EFPIA and the European Commission. It consists of a consortium of 88 organisations.

Legal structure

ConcePTION may be unusual in the number of members in the consortium as well as the number of associated organisation, but the fact that there is no legal entity associated with the consortium is far from uncommon in academic projects, particularly where there is a blend of academic and commercial partners.

There is, of course, a 'Consortium Agreement' which binds all the members of the consortium and determines the governance for the consortium and how disputes will be resolved, but this binds all of the partners to a common enterprise rather than creating a distinct legal entity.

Note that the third-party organisations are not actually bound to the project nor do they form part of the consortium itself. They may be involved through other agreements, such as data-sharing agreements, but these will be with consortium partner organisations directly rather than with the consortium *per se*.

Implications under GDPR

This means that for GDPR purposes, the ConcePTION project cannot be a 'controller' (or 'processor') for any personal data that may be used within project activities, so any 'personal data' processed will be processed by one of the partner organisations as 'controller' – or possibly by several of the partners as 'joint controller' and/or 'processors'.

Use of 'personal data' by third-party organisation will be their responsibility as distinct controllers and their use of such personal data justified under their own purposes and legal bases for such processing, even if in support of ConcePTION project objectives.

Implications for consortium partners

The main implication is that any personal data collection (which might just be email addresses) needs to be associated with at least one consortium member as 'controller' in order to fulfil accountability requirements under GDPR.

Principal Investigators (PIs) need to be aware of this and ensure that they and their staff follow their own organisation's policies and procedures for GDPR compliance (e.g. when and how to perform a DPIA, registering any personal data collection – or at least ensuring that it is covered under an existing processing category in their organisation's Article 30 register of processing).

Where a PI is leading a particular study and so creating datasets concerning patients, then they will need to ensure that it is clear whether their organisation is sole controller or whether other members of the consortium should be considered as 'joint controllers' with them.

Implications for consortium administration

Clearly, it is important that all individual participants are aware of the legal position and do not blithely assume that all processing of any personal data for the project is covered by the consortium itself, but take steps to ensure that it is covered under their own policies & procedures. A briefing note to this effect should be distribute as soon as possible into the project (perhaps at the project kick-off meeting (or meetings within each organisation).

Usually, there will be one organisation (the 'lead' organisation) which will handle the administration of the consortium as a whole. This organisation will need to be mindful of information assets such as any project web-site and mailing lists, which may contain personal data and will need to be managed properly under GDPR – and it is likely that their organisation would be the 'controller' as it will need to generate the facilities to support them.

An alternative view is that these information assets would be under the 'joint controllership' of all consortium members and the 'lead' organisation is merely the 'processor' (as well as one of the 'joint controllers').

Implications for other consortium projects in the future

These are the lessons learned for future consortium-style European-wide projects:

1. Establish the necessary data protection governance arrangements within the project proposal
   The original ConcePTION proposal was developed prior to GDPR becoming effective and participant organisations becoming familiar with the necessary arrangements under GDPR. This meant that it was assumed that the consortium would need to do a range of activities, such as develop DPIAs, whereas the responsibility for these fall to the consortium members, though the consortium may wish to take steps to ensure that all members understand their responsibilities. This may be detailed more fully in the consortium agreement.

2. Ensure that all individual project participants are briefed on GDPR governance arrangements, in particular to abide by their own organisation's GDPR policies and procedures
   A briefing note and privacy notice comparable to the one developed and distributed by ConcePTION to its participants may be a useful template.

3. In particular, establish from the outset which consortium member (or members) will be the controller for consortium-level assets (e.g. web-site and/or distribution lists)
   While not a legal requirement, it may be as well to develop a listing of such assets and who is the controller and /or processor of such assets for clarity and so that the legal position is clear and can be reflected in each consortium member's own processing register appropriately.

4. While most organisations will have a generic entry for customer relationship management (CRM) activities (e.g. contact list, enquiry management, workflow, etc.), it will be important to ensure that it is understood that consortium distribution lists are to be maintained separately (or at least marked distinctly) except where individuals have agreed to be included for general communication purposes

5. It would be helpful if funders such as the European Commission or IMI had guidance for project consortia on GDPR compliance and the specific requirements that they may have for proposals in terms of overall information governance compliance (including GDPR).

This would help avoid the initial confusion within a project as to what GDPR-compliance activities were needed, given that project proposals are usually produced by scientists rather than governance experts.

6. Each consortium should know what significant information assets may be held on behalf of the consortium by member organisation, so that it can support individuals in locating their data
   An individual may know that they were involved in a consortium-led study and may wish to locate whatever personal data may be held about them. While the consortium is not a legal entity and its organisation will not persist beyond the lifetime of the project itself, it should be able to direct enquiries to the appropriate controller or controller's