Biomedical Research Database

**Data Protection Impact Assessment**

Controller details

|  |  |
| --- | --- |
| Name of controller |  |
| Subject/title of DPO |  |
| Name of controller contact /DPO  (delete as appropriate) |  |
|  |  |

# Contents

[1. Contents 1](#_Toc14447221)

[2. Summary 2](#_Toc14447222)

[3. The need for a DPIA 3](#_Toc14447223)

[4. Description of the processing 4](#_Toc14447224)

[4.1 The nature of the processing 4](#_Toc14447225)

[4.2 The scope of the processing 4](#_Toc14447226)

[4.3 The context of the processing 4](#_Toc14447227)

[4.4 Describe the purposes of the processing 4](#_Toc14447228)

[4.5 Provide a functional overview of the processing activities proposed 5](#_Toc14447229)

[4.6 What assets does your anticipated processing rely on? 5](#_Toc14447230)

[4.7 Database construction 5](#_Toc14447231)

[4.8 Data management and data ingestion 6](#_Toc14447232)

[5. Consultation with relevant stakeholders 7](#_Toc14447233)

[6. Necessity and proportionality of processing 8](#_Toc14447234)

[7. Measures to Protect Data Subject Rights 8](#_Toc14447235)

[7.1 Categories of data subject: 8](#_Toc14447236)

[7.2 Data subject rights (operative by law, operative by database policy, inoperative) 8](#_Toc14447237)

[8. Data Flow Map 9](#_Toc14447238)

[9. Identification and assessment of risks 9](#_Toc14447239)

[10. Measures to reduce risk 12](#_Toc14447240)

[11. Sign off and record outcomes 13](#_Toc14447241)

# Summary

# The need for a DPIA

* 1. Explain broadly what project aims to achieve and what type of processing it involves. You may find it helpful to refer or link to other documents, such as a project proposal. Summarise why you identified the need for a DPIA.

Identify the DPIA screening criteria that you flagged as high risk.

What elements of the DPIA are intended for publication? What is the intended timeline, scope, content and audience of the publicly released version (e.g. summary form, disclosure only to data contributors and data users, inclusion in the white paper, etc.)?

* 1. Stage of project at which the DPIA was carried out

2.3 Person(s) responsible for maintaining documentation of the progress of the DPIA

# Description of the processing

## 4.1 The nature of the processing

How will you collect, use, store and delete data? What is the source of the data?

Will you be sharing data with anyone? If so, identify the persons or categories of persons.

You might find it useful to refer to a flow diagram or another way of describing data flows.

What types of processing identified as likely high risk are involved?

Will you use any novel technologies? Will you use any novel methods of processing data?

Will you use any processors? If so, identify the persons or categories of persons and their anticipated roles.

## 4.2 The scope of the processing

What is the nature of the data, and does it include special category or criminal offence data?

How much data will you be collecting and using? How often? How long will you keep it?

How many individuals are affected?

What geographical area does it cover?

What is the anticipated volume and variety of data processed and of processing?

## 4.3 The context of the processing

What is the source of the data?

What is the nature of your relationship with the individuals?

How much control will they have over their data?

Would they expect you to use their data in this way?

Do they include children or other vulnerable groups?

Are there prior concerns regarding this type of processing? Have you or others performed this type of processing previously?

Is it novel in any way?

What is the current state of technology in this area? What is the current state of data security as regards your intended processing? Do you anticipate any advances therein?

Are there any current issues of public concern that you should factor in?

Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?

## 4.4 Describe the purposes of the processing

Is legitimate interest your s.6 legal basis for processing the data? If so, what are the legitimate interest(s) you pursue in processing the data?

What do you want to achieve?

What is the intended effect on individuals?

What are the benefits of the processing – for you, and more broadly?

## 4.5 Provide a functional overview of the processing activities proposed

4.6 What assets does your anticipated processing rely on? (Hardware, software, networks, people, paper or electronic transmission channels identified)

## 4.7 Database construction

1. Tiers of access (open access, controlled access, registered access) and role-based access (based on research purpose, job title, etc.)
2. Data de-identification measures contemplated (de-identification such as single or double coding, anonymization, pseudonymization, directly identifiable data)
3. Technical measures adopted to operationalize data subjects’ right to withdrawal.
4. Technical specifications regarding the retention of personal data necessary to the functioning of the database (e.g. metadata created in storage, upload, or other processing activities). For instance, personal data may be retained or otherwise processed as a byproduct of technical procedures involved in data upload.
5. Technical specifications as regards the transfer of legacy datasets from an earlier draft of the database. Specifications as regards the disposal of the legacy database, if applicable.
6. Offsite processing arrangements (processing activities, security measures anticipated, documentation to be signed by external processors, agreements to be entered into, list of the bodies/organizations responsible for processing)
7. Mechanisms used to effect offsite data-flows to and from the database [contributors, data ingestion, data storage, tertiary processors] both in the EEA and out of the EEA.
8. Disaster recovery and contingency planning. Can destroyed data be recovered (e.g. malicious intervention, human error, technological failure)? Further, are backups held? If so, where are they held?
9. Is mandatory training contemplated for staff, users, or contributors?
10. Do technical measures ensure that required disclosure have been made? That consent has been obtained (if necessary)? That other non-consensual processing bases’ requirements have been observed (if applicable)?
11. What technical measures are used to operationalize the established data retention periods?
12. Technical measures as regards data security (encryption, access controls, data destruction after the retention period, de-identification, physical sequestering of sensitive data).
13. Intended record-keeping as regards security measures’ testing – logs, records. Protocols regarding the frequency of testing, the retention of test records, the persons responsible for oversight and the disposal of such records.
14. Data users’ security obligations / responsibilities.

## 4.8 Data management and data ingestion

1. Does the project involve complex data controllership agreements?
2. Do different data collection policies apply to datasets?
3. Do different data quality policies apply to the datasets?
4. Does the project compel individuals to provide information about themselves? Does the project ensure that the research consents (or equivalent) used establish that consent to data provision is *not* required of the research participant? Does the project mitigate any pressure / compulsion on data subjects to
5. Does the database contain significant amounts of data about each data subject?
6. Does the database contain significant amounts of new data about each data subject?
7. Does the database significantly increase the population coverage for the kinds of data it retains?
8. Does the database involve or allow for new data linkages?
9. Does the database involve ambiguities as to the security of the data contained // the effectiveness of the security measures introduced?
10. Does the database involve novel data access / data disclosure arrangements? Does the database involve data access / data disclosure arrangements not contemplated by the original consent or other permissions obtained?
11. Does the database involve changing the medium of disclosure for publicly available information such that it becomes more readily accessible than before?
12. Could the database / the data it contains be used to make decisions concerning the data subject?

# International Data Transfers

Describe the intended international data transfers, including the source of the data being transferred and the destination of each such transfer.

Establish the legal justification for the transfer or transfers (i.e., adequacy decision, GDPR art. 46 appropriate safeguard, GDPR art. 49 derogation for a specific situation).

If the transfer is not being performed on the basis of an adequacy decision that benefits a specific law, country, or territory, determine if the law of the recipient jurisdiction and the practices of its authorities creates a risk for the fundamental rights of EU/EEA citizens.

These rights include the rights to respect for private and family life, the right to protection of personal data, and the right to an effective remedy and to a fair trial is enshrined at art. 47.

These assessments should be carried out in cooperation with experts or legal counsel in the recipient country or territory and should be revised from time to time.

If it is determined that the law and practice in the recipient jurisdiction cannot ensure respect for the fundamental rights of EU citizens, it is necessary to implement additional measures.

Such additional measures could include coding data prior to transfer (i.e., replacing all direct identifiers with a pseudonym) and retaining the re-identification key in the European Union. Such additional measures could also include certain forms of secure multi-party computation, or encryption.

It is a recommended best practice to document and justify both the analysis of the laws applicable to the destination of transfer, and the choice of additional safeguards.

# Consultation with relevant stakeholders

Describe when and how you will seek individuals’ views – or justify why it’s not appropriate to do so. Who else do you need to involve within your organisation?

Do you need to ask your processors to assist?

Do you plan to consult information security experts, or any other experts?

Internal stakeholders:

(Leaders, technical experts, and scientists)

External stakeholders:

Service providers (Processors, other contracted service providers) and data subjects (patients, research participants and vulnerable populations).

# Necessity and proportionality of processing

Describe compliance and proportionality measures, in particular:

What is your lawful basis for processing?

What purposes does your processing pursue? Are all processing activities relevant to your purpose and limited to what is necessary?

Does the processing actually achieve your purpose?

Is there another way to achieve the same outcome?

How will you prevent function creep?

How will you ensure data quality and data minimisation?

What information will you give individuals? What mechanisms will you use to provide it?

How will you help to support their rights?

What measures do you take to ensure processors comply? With GDPR? With other applicable privacy frameworks [e.g. Standard contractual clauses, if used in international transfer context]?

How do you safeguard any international transfers?

# Measures to Protect Data Subject Rights

## 7.1 Categories of data subject:

1. General
2. Vulnerable groups
3. Data subjects not concerned by GDPR / Privacy Legislation

## 7.2 Data subject rights (operative by law, operative by database policy, inoperative)

a) Information provided to the data subject

b) Right of access and data portability

c) Right of rectification

d) Right to object and restriction of processing

e) Relationships with processors

f) Safeguards surrounding international transfer

g) Notification of the designated point of contact

h) To complain to a supervisory authority

i) To know the legal basis for processing and how data is used

j) To know how long the data will be stored

To know the persons or categories of persons who are data recipients.

# Data Flow Map

Recommendations:

a) Distinguish data flows within the EEA from international data transfers

b) Identify points of contract / interaction across the parties sharing the data between themselves

c) Identify tangibly what data will be shared between one party and the next

d) For each input and outflow of data, establish:

1. Collection mechanisms

2. Parties responsible for proper collection and safeguards, location and nature of filing systems

3. Identity of direct end users, identity of potential indirect end users (such as IT support, suppliers, processors, etc.)

4. The relationship between the systems being used to host and transmit the data, and other systems (interfacing, interoperability, automatic data sharing, etc.)

# Identification and assessment of risks

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary.** | **Likelihood of harm** | | **Severity of harm** | | | **Overall risk** | |
|  | Remote, possible or probable | | Minimal, significant or severe | | | Low, medium or high | |
| **Inadequate disclosure controls or security measures:** | |  | |  |  | |
| 1. Due to the open access nature of the project | |  | |  |  | |
| 1. Due to unintended risks *at the design phase*, such as the inadvertent retention of metadata due to technical elements of database design | |  | |  |  | |
| 1. Due to unintended risks *at the use phase*, such as the erroneous introduction of identifiable data into the open-access or controlled-access database | |  | |  |  | |
| **Consent and ethics risks** | |  | |  |  | |
| 1. Vulnerable persons or persons whose data reveals prejudicial information may be exposed to a heightened risk | |  | |  |  | |
| 1. Vulnerable persons and other participants could be exposed to heightened psychological risks due to fear or anxiety concerning data misuse | |  | |  |  | |
| 1. Research participants might not understand the scope of their risk at the time of consent to inclusion in the database | |  | |  |  | |
| 1. Research participants might feel pressure or obligation to consent to inclusion in database | |  | |  |  | |
| **Privacy and rights risks** | |  | |  |  | |
| 1. Inability of the data subject to exercise rights (including, but not limited to privacy rights), access services or opportunities | |  | |  |  | |
| 1. Risk of economic or social prejudice caused to data subjects in the event of re-identification | |  | |  |  | |
| 1. Risk of physical harm to data subjects in the event of re-identification | |  | |  |  | |
| 1. Risk that authorities could request or access data and cause prejudice to data subjects | |  | |  |  | |
| 1. Inclusion of erroneous data could cause prejudice to data subjects | |  | |  |  | |
| 1. Illegitimate modification of data could cause prejudice to data subjects | |  | |  |  | |
| 1. Risk that the data subject will not be able to withdraw their data, or will not be able to compel researchers already having accessed/downloaded it from making further use of the data. | |  | |  |  | |
| **Technical Risks** | |  | |  |  | |
| Security risks | |  | |  |  | |
| 1. Data breach (unauthorized data modification of data in the database) | |  | |  |  | |
| 1. Data breach (unauthorized access to data in the database) | |  | |  |  | |
| 1. Data breach (unauthorized access to *or* modification of data hosted on servers of third-party processors) | |  | |  |  | |
| 1. Data breach (unauthorized access to *or* modification of data hosted on cloud servers) | |  | |  |  | |
| Technological risks | |  | |  |  | |
| 1. Heightened risk caused by the systematic processing of data | |  | |  |  | |
| 1. Heightened risk caused by the automatic processing of data | |  | |  |  | |
| 1. Heightened risk caused by the use of new technologies; namely the hosting of genetic / genomic data on an open-access database AND the use of novel visual and data representation technologies to view and interpret the data | |  | |  |  | |
| 1. Data could be contributed or stored in double, decreasing the accuracy of the database or exposing data subjects to heightened privacy/re-identification risk | |  | |  |  | |
| **Risks for pre-GDPR datasets** | |  | |  |  | |
| 1. Data subjects not informed of their rights at the time of collection (right to withdrawal, right to erasure, right to data portability, etc.) | |  | |  |  | |
| 1. Other GDPR requirements for data use may not have been respected | |  | |  |  | |
| **Future risks** | |  | |  |  | |
| 1. Data subjects may lose control over their personal data | |  | |  |  | |
| b) If a retention period is not established – information might be used for longer than necessary | |  | |  |  | |
| c) The risk that individuals are not aware that their data is being processed | |  | |  |  | |
| 1. New technologies could be developed that expose data subjects to new kinds of risks of which they were not made aware on collection of data | |  | |  |  | |
| e) New significance could be associated with the data contributed that exposes to new kinds of risks of which they were not made aware on collection of data | |  | |  |  | |
| f) Risk that data could be used to commercial ends or research ends objectionable to original data subject | |  | |  |  | |
| **Data matching risks** | |  | |  |  | |
| Collecting information from different sources and linking identifiers might mean an organisation is no longer using information which is safely de-identified *or* anonymized: | |  | |  |  | |
| 1. Merging of datasets within the database in order to re-identify data subjects | |  | |  |  | |
| 1. Merging of datasets with datasets outside the database to re-identify data subjects | |  | |  |  | |
| 1. Merging of datasets in the database with datasets in other health science databases allowing re-identification of data subjects | |  | |  |  | |
| 1. Cross referencing of identifiable information of known persons with database data in order to reveal identifiable information [aka jigsaw attack] | |  | |  |  | |
| 1. Use of data scraping or big data technologies to re-identify data subjects using metadata or data not otherwise obviously identifiable | |  | |  |  | |
| **Third-party / group risks** | |  | |  |  | |
| 1. Risk that family members of data subjects could be re-identified | |  | |  |  | |
| 1. Risk that negative inferences could be drawn concerning groups of which data subject forms part | |  | |  |  | |

# Measures to reduce risk

|  |  |  |  |
| --- | --- | --- | --- |
| **Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5** | | | |
| **Options to reduce or eliminate risk** | **Effect on risk** | **Residual risk** | **Measure approved** |
|  | Eliminated reduced accepted | Low medium high | Yes/no |

# Sign off and record outcomes

|  |  |  |
| --- | --- | --- |
| Item | Name/position/date | Notes |
| Measures approved by: |  | Integrate actions back into project plan, with date and responsibility for completion |
| Residual risks approved by: |  | If accepting any residual high risk, consult the ICO before going ahead |
| DPO advice provided: |  | DPO should advise on compliance, step 6 measures and whether processing can proceed |
| Summary of DPO advice: | | |
| DPO advice accepted or overruled by: |  | If overruled, you must explain your reasons |
| Comments: | | |
| Consultation responses reviewed by: |  | If your decision departs from individuals’ views, you must explain your reasons |
| Comments: | | |
| This DPIA will kept under review by: |  | The DPO should also review ongoing compliance with DPIA |