Research Data Management Protocol UvA Economics and Business

version 2020-02-17

Summary

As of March 1, 2020, in line with UvA guidelines the faculty board of UvA Economics and Business (EB) expects all researchers to adhere to UvA EB Research Data Management (RDM) protocol to safeguard data for research for a period of at least 10 years. UvA provides an easy-to-use environment (Figshare) where data can be stored, and, if desired, shared. All projects in which data is collected or used should have a research data management plan, which contains a description of e.g. what data is collected, the way in which and where it is collected and how the data will be stored. Appendix 6 provides a quick overview of the steps to be taken when starting a research project in which data is used. In general, the place to start is the comprehensive EBEC (Economics and Business Ethical Committee) form.

0. Introduction

This protocol is a formal registration of the responsibilities of those involved with academic research at UvA EB. It is meant to support ethical management of data associated with research projects at UvA EB. The protocol starts with a description of responsibility for data and definitions, followed by a general protocol for the storage of research data, exceptions to this protocol, and appendices. This protocol is effective per March 1, 2020 and in accordance with the UvA guidelines on Research Data Management and the Netherlands Code of Conduct for Research Integrity.

0.1 UvA guidelines on data

UvA EB assumes adherence to UvA guidelines concerning data (as can be found on the RDM website). UvA requires that data is handled with integrity and methodically, and in such a way that it can at all times be tracked and accessed during the complete 'academic life cycle'. UvA adheres to the FAIR principles: data should be Findable, Accessible, Interoperable and Reusable. All researchers involved should be aware of what has been done with the data and where it is located. According to the UvA guidelines, data (including edited/modified data) should be

a. Accurate, complete, authentic and reliable;
b. Recognisable, traceable and accessible;
c. Stored with the proper meta-data
d. Where possible use open standards for files and meta-data;
e. Have a persistent object identifier;
f. Follow and adhere to the relevant laws, the rules and regulations of funding agencies and the university and academic code of conduct;
g. Open and accessible where possible, in line with the current open access principles and if not, an explanation should be provided;

h. Stored securely, guarded against unauthorised access;
i. Handled and stored in compliance with the relevant privacy laws and regulation, where personal data is involved

UvA EB in principle embraces the Open Access movement and the need for replicability in research, including the need to share data for meta-analyses, as well as the demands set by funding organisations (such as NWO), whilst understanding the need to help researchers to safeguard and not share data of proprietary or confidential nature. It also is aware of the additional administrative burden created.

0.2 Formal RDM Responsibilities of UvA EB and Research Staff

The faculty, the research institutes as well as the individual researcher bear responsibility for the storage of research data. Here, we outline who bears responsibility for which parts of the RDM process.

0.3 Research institutes

First and foremost, the Amsterdam School of Economics Research Institute (ASE-RI) and the Amsterdam Business School Research Institute (ABS-RI) are responsible for providing and managing the resources needed for adequate data storage. In that sense, the research institutes should provide researchers with an infrastructure that allows them to store data in accordance with this protocol. The institutes are responsible for ensuring this storage is safe (protected from threats, such as theft and technological malfunction) and managed properly by stewards. UvA has opted for Figshare as the default storage for data and UvA EB follows suit. Figshare is user-friendly, useable for sensitive data, allows data to be shared and made public easily. Furthermore, the research institute is responsible for the distribution of the protocol among research conducting employees. In the future RDM should be part of the education of students and when engaging in research projects, students should adhere to the same principles. Dissemination of the protocol amongst students, and having students save their data in Figshare (or another environment) will be part of future policy, in line with that of UvA. A revised protocol will be disseminated when the time arrives.

0.4 Research faculty

The research faculty of UvA EB is responsible for the practical adherence to the protocol in their research. During the storage of data at the UvA, they are responsible for providing the data stewards of their data with accurate contact information so that they can be contacted when necessary (e.g. after leaving UvA). The researcher is also responsible for the notification of third parties regarding this protocol and the associated data storage. Finally, the researcher is responsible for a research data management plan being present for each research project.

1 Definitions

1.1 Data

Data is anything that is collected for and provides the basis for analysis in research projects. It can either be in raw form (as collected) or in the form of edited data (as analyzed). All types of data are considered under this definition.

1.2 Research project

A research project is any effort undertaken by a researcher to collect data to be analyzed in order to share a conclusion with an audience. In that sense, any endeavor that results in conclusions presented at conferences, published in any form or made available to others in any other sense, is considered a research project here as it connects the data to the researcher, the research institute as well as UvA EB.

1.3 Data owner(ship)

The principal owner of the data is the (group of) researcher(s) responsible for the data collection within the research project as defined above. The secondary owner of the data is usually the UvA, as the collection of data happens in commission of the UvA and often bears the name of the UvA in public. However, it is important to state that collected data includes the intellectual property of the researcher(s) as the idea for a study is based on is consequential to the researcher's career, also before and after working for the UvA. It is also possible for a researcher or external party not affiliated with UvA to be the owner of the data. See the section exceptions for guidelines for this situation (for example, when external partners are involved in financing the research or researchers at other institutes are the primary data owner).

1.4 Data steward

On behalf of the UvA EB, ASE-RI or ABS-RI, the data steward is responsible for the practical management of data storage, information dispersion among research staff on RDM, and the management of access to the data. The default data steward is the Research Director (in accordance with UvA policy). However, within EB this responsibility will be delegated to the research institute officers (see Appendix 1).

1.4.1 Data-steward Role

The data-steward role consists of two components. On the one hand (s)he is the representative of a faculty, a research institute or group. The data steward needs to have accurate knowledge of the protocol and should be able to advise the faculty/research institute/group on data management issues. On the other hand, the data steward manages the research data management storage system on behalf of the faculty/research institute/group with regard to monitoring the use of the system and reporting on the use of the system. For EB the data steward can upload data in Figshare.

1.4.2. Data steward Tasks

Protocol representative:

- Making new employees aware of the protocol
- Answering questions from employees and (in the future) students on the protocol

Data management system representative:

- Reporting on the number of projects active and/or published in the research group
- Assessing the validity of and facilitating requests for extra or special storage space when requested by members of the research institute/group

1.5 Research Data Management Plan (RDMP)

A research data management plan (RDMP) is obligatory for a research project and contains a description of what data is collected, the way in which and where it is collected for the project, the status of approval by the ethics committee, what the main research question is and how the data will be stored. Ideally, it should be written and stored in Figshare before data collection commences. In any case a final version is stored together with the data. The RDMP states the researchers who are owner of the data and the name of the data steward for this research project. In case of high risks related to personal data the RDMP contains a Data Protection Impact Assessment (DPIA). The researcher is responsible for a RDMP being present for each research project and the RDMP being stored in Figshare (see below). The data steward is responsible for the periodical assessments of the RDMPs in terms of adherence to this protocol and the storage of the RDMP. See the appendix for a template, but a first RDMP will be generated after filling in the EBEC form.

1.6 Editing data

For edited data (data that has in any way been transformed by the researcher) a description of the editing process needs to be clarified in the (appendices of the) RDMP. This does not imply that all transformations should be reversible, but it does imply that researchers need provide insight in their editing process. Examples of this would be the data cleaning process, the generation of variables from items, or the anonymization of data, that would captured by for example SPSS syntax files or R-scripts. Alternatively, a codebook/syntax can be added to fulfill this role (see below)

2 General protocol for the storage of data (simple cases)

For a research project resulting in a publication (edited and raw) data needs to be stored in Figshare for a minimum of 10 years (in accordance with UvA policy), but the period should also meet the standards of the journal that the research based on this data is published in. The standard term of storage when putting items such as data in Figshare will be 10 years. Data is not deleted automatically afterwards, but the researcher may delete it manually and will be notified when the 10-year period has passed. For research projects involving a PhD project, the data needs to be stored indefinitely (in accordance with UvA policy). For other research projects, it is recommended that the data is stored for at least 10 years (after storage)

When the researcher responsible for the collection of the data leaves the UvA, the data steward takes over the role of the owner, however the researcher remains the holder of the intellectual property rights and still has the freedom to decide on the usage of the data. The researcher is obliged to be available for communication about the data for the remainder of the time applicable to the storage of the data. The storage of the RDMP remains with the EB.

Data should be in storage at the latest when the conclusions of the research project are made public in any way and thus impacting the name of the researcher and that of the research institute. All stored data should be stored with the matching RDMP.

PhD candidates are responsible for storing their research data correctly and on time. Next to the check on plagiarism, there will be a check on the correct storage of research data as a precondition for admission to the defense. The check on research data storage is compulsory and without approval of the data steward and the Dean, the candidate will not be admitted to the defense.

2.1 RDMP and Projects in Figshare

Per project that meets the requirements specified above that make storage necessary, a project needs to be generated in Figshare. Where possible, this project needs to include

a. the raw data and/or
b. the edited data set on which the analyses was done
c. syntax/codebook (key/manual to the dataset)
d. the EBEC approval (if applicable)
e. the RDMP
f. other relevant documents

2.2 Storage of raw data

For each research project, the raw data file needs to be stored in Figshare in accordance with the availability principles mentioned above. With this data, it should be possible to

replicate the analyses done in the research project. This means a syntax/codebook needs to be present to reconstruct the final dataset from the raw data with the help of the researcher. Where this information needs to be added manually (instead of, e.g., a codebook extracted from SPSS) a short description will suffice if replicating every step taken will take unreasonable amounts of time. If the data is to be shared with others, or made public, the researcher makes sure the version that will be used for those purposes does not contain any personal data. Shared data should be anonymized to protect the respondents. The quality of the data needs to be in line with the policy on RDM of the UvA. It is mandatory to also store a data file that includes the data as it was analyzed (the last edited version).

2.3 Access rights

If the data has not been made available publicly, it is available for inspection by the data steward without the consent of the researcher *only* in case of suspected fraud or unethical behavior (in this document: when a formal complaint at the UvA committee *Wetenschappelijke Integriteit* has been made and found admissible). In all other cases, the researcher needs to give consent for access to the data stored for a research project. The data steward of the research project is the first contact to access the data. The data steward is responsible for contacting the researcher with a request for access. When the data steward is unable to contact the researcher, access is denied. The only exception to this being the case of suspected fraud or unethical behavior as mentioned above, in that case data stewards can access the data. A few data specialists from the University Library also have access, as they are involved in the overall maintenance of the system, but will only access data after being ordered by the University Board.

2.4 Deletion of unpublished data

In principle, the data steward will delete the data after the retention period of availability of the data has expired, or the responsibility for the practical storage of the data is transferred to another party (e.g. when the researcher leaves the UvA without a substitute present). However, this will be done manually by the data steward and only after contacting the researcher. If the researcher needs to store the data longer, for example, to meet journal standards this is possible. In case a researcher leaves UvA access to Figshare is terminated after a 90 day grace period.

3 Publishing data

Figshare allows the researcher to publish data online. Generally, the UvA wants publishing data of finished projects of which results have been made public to be the rule where possible, and not publishing the data the exception (see section 4) as UvA RDM policy aims for openness when possible. Published data can be available for all others to access and use, depending on the access granted by the researcher who publishes the data. The EB protocol does allow for data to not be openly published when researchers have reasons for this as is also explained below.

3.1 Consequences of publication

If the researcher decides to publish the data, it will mean that a permanent identifier is attached to the dataset and that it is made publicly available for anyone to access and use it (within limits of the chosen license, see point 3.2). It is recommended to add a dataset specifically for publication to the project in Figshare, in which all identifying information has been deleted and no intellectual property beyond the scope of the related published work is discernable. All authors and owners of the data should be made aware of and have signed off on publication of the data. All researchers should be aware that publication of the data is irreversible.

3.2 Licenses

When publishing data, the researcher needs to select a license for publication. The license lets others know what kind of data is uploaded and determines what others may do with the data. More information can be found in the Figshare metadata manual. Default license is CC BY 4.0 (data can be used for other purposes by others, as long as an attribution is given, but others can be selected).

4 Protocol for complex cases and exceptions

4.1 Sensitive Data

When research projects include sensitive data that cannot be anonymized to an acceptable level (to any party involved) and both the researcher and the data steward agree that the data should not be stored within Figshare, UvA EB will provide local storage. A form stating this exception should be uploaded by the researcher in Figshare within the relevant project(s) (see appendix). The data steward assesses (in concurrence with the Research Director) whether the request for the exception is grounded.

When collecting, using and/or processing personal data, from May 25, 2018 onwards, researchers need to adhere to the new *Algemene Verordening Gegevensbescherming (AVG,* or General Data Protection Regulation, GDPR). This law stipulates that when personal data is collected for research the following principles apply:

a. Only collect data necessary for the research
b. Use data for the purposes they were collected for (there are guidelines for the re-use of existing data for research)
c. Be transparent and clear (in the RDMP) on which data is collected, used and processed and whom the data is shared with.
d. Make sure the personal data collected are accurate and truthful and participants were informed on all aspects of the research and the use of the data and have agreed on participating. In the case of special personal data, such as sexual preferences, medical information, ethnicity, political views and religious beliefs, explicit informed consent is mandatory.
e. Make sure the data is stored safely and privacy is safeguarded.
f. Make sure the UvA Data Protection Officer (*Functionaris voor Gegevensbescherming*) is informed [within UvA this will be done with an automated register]
g. In case of a data-leak the UvA Data Protection Officer will have to be notified immediately, (s)he will inform the Dutch Data Protection Authority (DPA, *Autoriteit Persoonsgegevens, AP*).
h. Note that participants have the following rights regarding their personal data: information, rectification, limitation in use and processing, receiving a portable copy of the data, to be forgotten (removal of the data on request), resistance against profiling (automated decisions based on the research that effect the participant).

It is advised that for all projects where personal data is collected, used and/or processed a Data Protection Impact Assessment (DPIA) is made. In these cases we expect that the raw data will be stored without being made accessible and an anonymized version is shared (unless there are other reason not to make the data public, see 4.4. for instance).

4.2 Big/un-storable Data files

When research projects include large amounts of data or data types that cannot be stored within the facilities offered by UvA EB, both researcher and the data steward will look for suitable alternatives for the practical storage of the data. When asked by the data steward,

the researcher must provide access to the data in the cause of questions on suspected fraud or unethical behavior. A form stating this exception should be uploaded by the researcher in Figshare within the relevant project(s) (see appendix). The data steward assesses (in concurrence with the Research Director) whether the request for the exception is grounded.

4.3 External Data Owners

When research projects include data that is not principally owned by a researcher affiliated to the UvA the responsibility for the practical storage of data is transferred from the research institute to the external owner of the data. When asked by the data steward, the researcher must provide access to the data in the case of questions on suspected fraud or unethical behavior. The external owner is responsible to ensure that the data steward has correct contact information of the external data owner so the data steward can contact them with requests for access to the data for the duration the data needs to be available for the research project. A form stating this transfer of responsibility must be signed and stored by the data steward (see appendix). The data steward assesses (in concurrence with another data steward) whether the request for the exception is grounded. The data steward is responsible for the storage of the form for the research project for the duration of time the data needs to be available for the project.

When the data is an excerpt from a database on which UvA EB has a subscription/license, this excerpt can be stored in Figshare but cannot be made available to the public. In that case, the status (confidential) must be selected.

4.4 Proprietary data

When the data is considered proprietary by the researcher (for example, (s)he expects to publish (additional) material based on the data), this is considered a valid reason not to make the data public. In that case, the status (confidential or embargoed) can be selected.

4.5 Exceptions

When a researcher is planning to do research under very specific circumstances and has valid reasons why the protocol will hamper the research or make it impossible, (s)he can ask the research director for to make an exception for a specific project. The decision of the research director including the request and the argumentation supporting the decision will be stored at the research institute.

Amsterdam Business School Research Institute

Director: prof. dr. D.N. den Hartog
Contact: B.C. Bouten
Tel: 06 4552 1079
E-mail: ABS-ri@uva.nl
Address: Plantage Muidergracht 12, 1018 WB Amsterdam

Amsterdam School of Economics Research Institute

Director: prof. dr. F.R. Kleibergen
Contact: L. Koks
Tel: 020 525 4276
E-mail: ASE-ri@uva.nl
Address: Roetersstraat 11, 1018 WB Amsterdam

APPENDICES

1: Data steward UvA EB

UvA EB general	B.C. Bouten

1.1 Data stewards ABS sections

ABS general	B.C. Bouten
Accounting	B.C. Bouten
Entrepreneurship & Innovation	B.C. Bouten
Finance	B.C. Bouten
Leadership & Management	B.C. Bouten
Strategy & International Business	B.C. Bouten
Marketing	B.C. Bouten
Operations Management	B.C. Bouten

1.2 Data stewards ASE sections

ASE general	L. Koks
Quantative Economics	L. Koks
Marco & International Economics	L. Koks
Microeconomics	L. Koks
Economics of Taxation	L. Koks

2: Links

University Library (UB) RDM and Figshare website
UvA-Economics & Business Ethics Committee website
UvA GDPR and research website
UvA Research Support Portal
VSNU Code of Conduct on the use of Personal Data in Research
UvA website on Academic Integrity
Wettekst Algemene Verordening Gegevensbescherming (Dutch)

3: RDMP template

4: External ownership of the data Form

5: Sensitive and big data exceptions Form

6: Process overview

3: RDMP template

After completing the EBEC approval online, a first draft RDMP is generated automatically

1. Project name:

2. Lead researcher:

3. Data steward:

4. Research question(s):

5. Data to be gathered (including location):

6. Method of data collection (in case of personal data indicate the basis (*grondslag*)):
Basis being either *informed consent* or *legitimate interest (academic research)*

7. Individuals involved in data gathering, data manipulation/editing and with access to the data:

8. Data Protection Impact Assessment
required when a processing operation "*is likely to result in a high risk to the rights and freedoms of natural person*".

- a systematic description of the envisaged processing operations and 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 that are likely to result from the processing (and in particular the origin, nature, particularity and severity of such risks); and
- the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and demonstrate compliance with the GDPR.

Practically speaking a DPIA is needed when two or more of the casus below are applicable:

- Assessing people based on personal characteristics
- Automated decision making
- Structured and large-scale monitoring
- Sensitive personal data
- Large-scale data processing
- Linked databases
- Data on vulnerable persons
- Use of new technologies
- Blocking of (a) right(s), service(s) or contract(s)

When data on genetics and/or health is concerned, a DPIA is mandatory. Please contact the data steward when drafting a DPIA so we can guide you in the process.

9. Data editing/manipulation steps (e.g. SPSS Syntax files, R scripts).

10. Where and how will the data be stored (including temporary storage for research use) and security measures applied:

11. Approval EBEC (Economics & Business Ethics Committee) obtained: approval yes/no

12. Intellectual property, copyright and ownership of the data:

☐ The researcher [name] hereby states that the data will be stored will be in line with the UvA guidelines and UvA EB protocol on RDM.

4: External ownership of the data Form

1. Project name:

2. Lead researcher:

3. Data steward UvA EB/ASE-RI/ABS-RI:

4. External owner of the data:

The external owner is responsible to ensure that the data steward of UvA EB has correct contact information of the data owner so the data steward can contact the owner with requests for access to the data in case of suspected fraud or unethical behavior (when a formal complaint at the UvA committee *Wetenschappelijke Integriteit* has been made and found admissible). Access can be denied for other purposes. If desired, both parties can jointly appoint a third party to access and check the data.

The external owner is responsible to ensure that the data is stored safely in accordance with the Dutch law for at least 10 years (and at least 5 years after publication of an academic paper based on the data).

Signatures

Name: External owner	**Name:** Data Steward UvA EB/ASE-RI/ABS-RI
Date: Place:	Date: Place:

5: Sensitive & big data exception Form

1. Project name:

2. Lead researcher:

3. Data steward UvA EB/ASE-RI/ABS-RI:

For the above-mentioned project, UvA EB will ensure that the data is stored safely and in accordance with the law (especially the *Algemene Verordening Persoonsgegevens*), the UvA guidelines and the EB RDM protocol outside Figshare at:

☐ a local storage within UvA EB and that the data steward of UvA EB is the only person able to access the data.

☐ a storage facility elsewhere (information on location and access are known to the Data Steward). Location:

The storage is done outside Figshare for the following reasons:

Signatures

Name:
External data storage party

Name:
Data Steward UvA EB/ASE-RI/ABS-RI

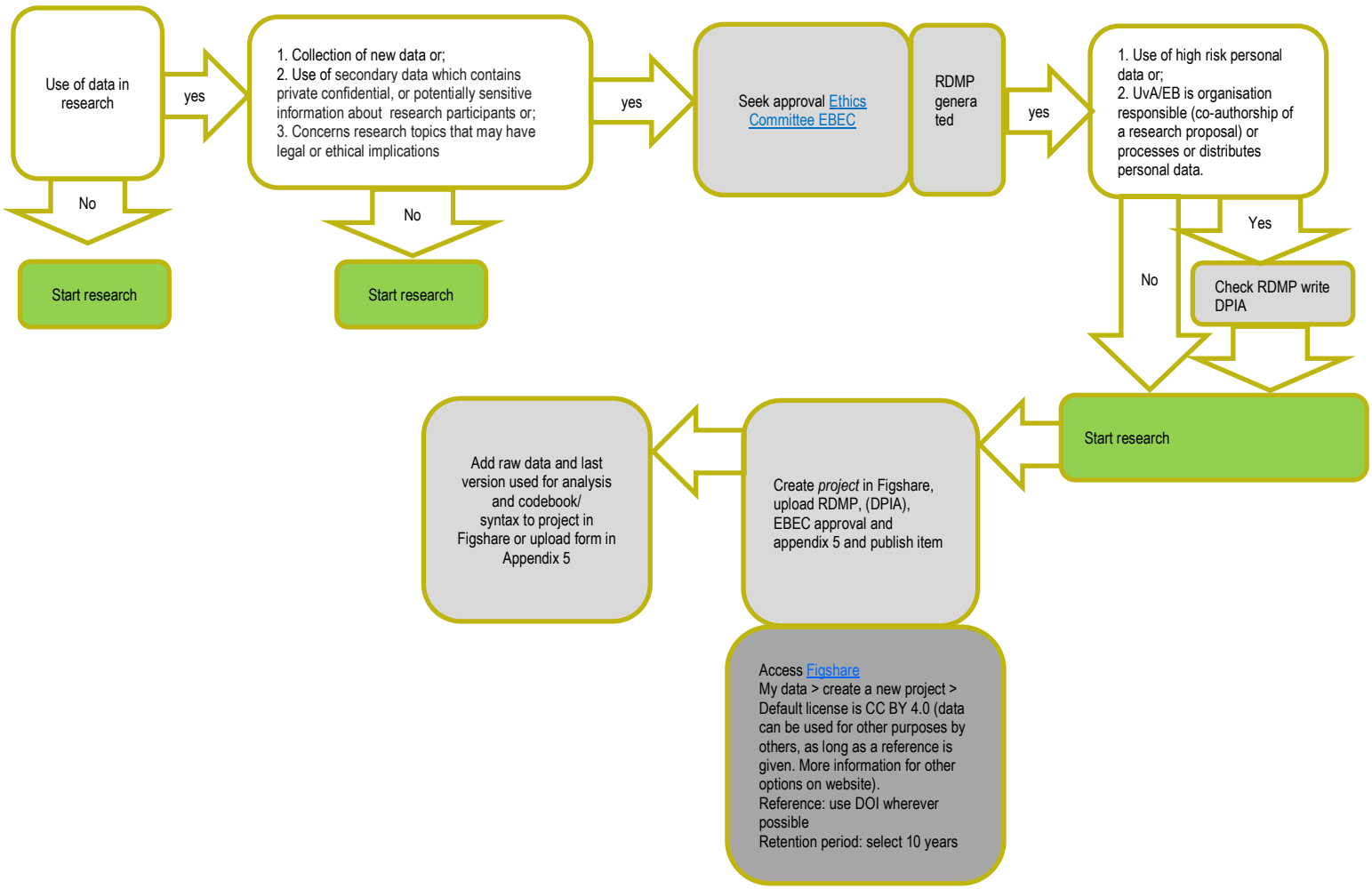
Date:
Place:

Date:
Place:

Name:
Director Research Institute

Date:
Place:

6: Process overview



7: Guidelines for safe data processing and storage

1. Anonymize data where possible and store the master-data set or the key to link names etc. to the rest of the data safely before processing further. Note that individuals can be recognizable from data other than their names. It is wise to protect the master file or key with a password.
2. Anonymization means that under no circumstances individuals can be linked to the data. Pseudonymization can help to make it more difficult to link individuals whilst keeping information important to the research intact. Names can be changed to numbers for instance, or data can be aggregated (age groups e.g.).
3. Make sure the hard drive of the computer you process the data on is encrypted.
4. When sharing data make sure to share with a trustworthy party, share only the minimum (pseudonymized data for instance) and use a safe method of transport (Figshare, Surfdrive e.g.).