



# Generations and Gender Programme Preparatory Phase Project (GGP-5D) Data Management Plan

Work package 7: **Management and D-E-C coordination**

Grant Agreement Number: 101079357

Project acronym: **GGP-5D**

Project full title: **The Generations and Gender Programme Preparatory Phase Project**

Due delivery date: 31<sup>st</sup> March 2023

Actual delivery date: 30<sup>th</sup> March 2023

Organization name of lead participant for this deliverable: GGP Central Hub, Netherlands Interdisciplinary Demographic Institute (NIDI), Royal Netherlands Academy of Arts and Sciences (KNAW)

Dissemination level: Public



**Funded by  
the European Union**

## Document Control Sheet

|                          |  |                           |
|--------------------------|--|---------------------------|
| Deliverable number:      | 7.2  |                           |
| Deliverable responsible: | Vytenis Juozas Deimantas, Olga Grünwald                          |                           |
| Work package:            | 7  |                           |
| Editor:                  | Anne Gauthier, Martin Bujard, Laurent Toulemon, Arianna Caporali |                           |
| <b>Author (s)</b>        |  |                           |
| <b><i>Name</i></b>       | <b><i>Organization</i></b>                                       | <b><i>E-mail</i></b>      |
| Vytenis Juozas Deimantas | GGP central hub, NIDI, KNAW                                      | deimantas@nidi.nl         |
| Olga Grünwald            | GGP central hub, NIDI, KNAW                                      | grunwald@nidi.nl          |
| Anne Gauthier            | GGP central hub, NIDI, KNAW                                      | gauthier@nidi.nl          |
| Martin Bujard            | BiB  | Martin.Bujard@bib.bund.de |
| Laurent Toulemon         | INED   | toulemon@ined.fr          |
| Arianna Caporali         | INED   | arianna.caporali@ined.fr  |

| <b>Document Revision History</b> |                    |  |                           |
|----------------------------------|--------------------|--|---------------------------|
| <b><i>Version</i></b>            | <b><i>Date</i></b> | <b><i>Modifications Introduced</i></b> |                           |
|                                  |                    | <b><i>Modification reason</i></b>      | <b><i>Modified by</i></b> |
| V1                               | 30.03.2023         | Deliverable D7.2                       |                           |
| V2                               |                    |  |                           |
| V3                               |                    |  |                           |

## Executive summary

This data management plan identifies the data sets that will be produced as a result of the Generations and Gender Programme Preparatory Phase Project (GGP-5D, duration October, 2022 to September, 2026) and defines the key elements that will facilitate the potential reuse of the data collected and processed within the project. The data management plan will ensure that this data will be findable, accessible, interoperable and reusable (FAIR).

It also outlines other potential research outputs, allocation of resources to ensure the FAIRness of the project data, its security, ethical considerations and other issues.

## Table of Contents

|                                       |   |
|---------------------------------------|---|
| <i>Definitions and acronyms</i> ..... | 5 |
| <b>1. Introduction</b> .....          | 6 |

|              |   |           |
|--------------|---|-----------|
| <b>1.1.</b>  | <b><i>Project background</i></b> .....                                | <b>6</b>  |
| <b>1.2.</b>  | <b><i>Data management plan objectives</i></b> .....                   | <b>7</b>  |
| <b>2.</b>    | <b><i>Project data</i></b> .....                                      | <b>8</b>  |
| <b>2.1.</b>  | <b><i>Project datasets</i></b> .....                                  | <b>9</b>  |
| <b>2.2.</b>  | <b><i>Data summary</i></b> .....                                      | <b>16</b> |
| <b>2.2.1</b> | <b><i>Personal data</i></b> .....                                     | <b>16</b> |
| <b>2.2.2</b> | <b><i>Statistical data</i></b> .....                                  | <b>17</b> |
| <b>2.2.3</b> | <b><i>Textual data</i></b> .....                                      | <b>17</b> |
| <b>2.2.4</b> | <b><i>Other data types</i></b> .....                                  | <b>17</b> |
| <b>2.2.5</b> | <b><i>Data formats</i></b> .....                                      | <b>18</b> |
| <b>3.</b>    | <b><i>FAIR data</i></b> .....   | <b>19</b> |
| <b>3.1.</b>  | <b><i>Making data findable, including provisions for metadata</i></b> | <b>19</b> |
| <b>3.2.</b>  | <b><i>Making data accessible</i></b> .....                            | <b>20</b> |
| <b>3.3.</b>  | <b><i>Making data interoperable</i></b> .....                         | <b>21</b> |
| <b>3.4.</b>  | <b><i>Increase data re-use</i></b> .....                              | <b>21</b> |
| <b>4.</b>    | <b><i>Other research outputs</i></b> .....                            | <b>22</b> |
| <b>5.</b>    | <b><i>Allocation of resources</i></b> .....                           | <b>22</b> |
| <b>6.</b>    | <b><i>Data security</i></b> .....                                     | <b>23</b> |
| <b>7.</b>    | <b><i>Ethics</i></b> .....  | <b>23</b> |
| <b>8.</b>    | <b><i>Other issues</i></b> .....                                      | <b>23</b> |

## Definitions and acronyms

**CC** Creative Commons

**CESSDA** Consortium of European Social Science Data Archives

**Data set** is a group of data

**DDI** Data Documentation Initiative

**GDPR** General Data Protection Regulation

**GGP** Generations and Gender Programme

**GGP CCT** GGP Central Coordination Team

**GGP-5D** The Generations and Gender Programme Preparatory Phase  
Project: five dimensions of readiness

**GGS** Generations and Gender Survey, the main data product of the GGP

**DMP** Data management plan

**FAIR** Findable, Accessible, Interoperable and Reusable

**KNAW** Koninklijke Nederlandse Akademie van Wetenschappen (Royal  
Netherlands Academy of Arts and Sciences)

**Primary data** refers to the first hand data gathered by GGP-5D partners.

**Secondary data** means data collected earlier than GGP-5D has started.

**MoU** Memorandum of Understanding

**NIDI** Netherlands Interdisciplinary Demographic Institute

**RI** Research infrastructure

**WP** Work package

## 1. Introduction

The scope of the first version of the the Generations and Gender Programme Preparatory Phase Project (GGP-5D, duration from October, 2022 to September, 2026) Data Management Plan (DMP) is to identify the data sets that will be produced as a result of the project and define the key elements that will facilitate the potential reuse of the data collected and processed within the project. The DMP will ensure that this data will be findable, accessible, interoperable and reusable (FAIR).

### 1.1. Project background

The Generations and Gender Programme (GGP) is an interdisciplinary research infrastructure (RI) on population and family dynamics. It collects, processes and disseminates cross-nationally comparable longitudinal data on young adults, families, generational exchanges, and the life courses of women and men. In turn, these data help address key scientific and societal questions related to demographic changes. According to the 2021 EU Strategic Foresight Report, demographic shifts are one of the mega-trends affecting the future and sustainability of societies.

The core of the GGP is the Generations and Gender Survey (GGS). The GGS collects individual level data on key demographic events such as leaving parental home, union formation and dissolution, childbirths and deaths that structure individual life courses and family/kinship networks. The survey focuses on how they interrelate to each other as well as to migration and employment through family life. The current GGS archive includes individual-level data from more than 31 countries worldwide with more than 300,000 respondents (to present).

The GGP therefore responds to large needs for solid data on the life trajectories of individuals, on families, gender roles, and relationships between generations.

The aim of the Generations and Gender Programme Preparatory Phase Project (GGP-5D) is to enhance the long-term sustainability and the overall 'investment-readiness' of the GGP. Specifically, this means bringing the GGP to a stage at which governmental stakeholders will be willing to

commit and invest into. To do so, the GGP-5D project will pursue five main objectives, each of them tapping into one of the five dimensions of readiness: technical, impact, financial, legal, and strategic.

## 1.2. Data management plan objectives

This document is the initial DMP for the GGP-5D project. The DMP identifies the datasets that will be generated, used and re-used throughout the duration of the project. It also defines the datasets' purpose and life cycle.

The objectives of the DMP are as follows:

- To identify the datasets that GGP-5D will produce.
- To define which of these datasets can be made FAIR and how to do it.
- To define the allocation of resources (costs and responsibility) for data management during and after the project.
- To outline procedures for data security and storage during and after the project.

Deliverable 7.2, the Data management plan for the GGP-5D project, follows the Horizon Europe, OpeanAIRE and DMPonline guidelines. The DMP is to be a living document which is meant to be continuously updated and revised in connection to the development and progress of the GGP-5D project.

The DMP provides an outline of the types of data that is to be collected, processed and/or generated (e.g. using previously existing data). It also indicates how the data is going to be shared among the partners and whether the data is going to be made open access to the use of the broader public. Legal matters in relation to collecting and/or processing of personal data are acknowledged and addressed both in terms of data processing and data protection.

In connection to personal data, the consortium of the GGP-5D shall ensure protection of the aforementioned data by following the requirements stipulated in national and European regulations. In particular, it shall abide to the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679).

If there is more data collected, processed or in other ways generated throughout the duration of the GGP-5D, that is not referred to in this DMP; or if there are changes on the project consortium policies or in the consortium composition; or if external factors undermine the validity of the current DMP, an updated DMP shall be prepared and released reflecting the changes faced. In addition, It is the responsibility of each partner to notify the coordinator of changes in the data they are to collect, process or generate during the project.

## 2. Project data

In this section, all GGP-5D datasets are listed. A large share of these datasets is to be generated during the GGP-5D project and will be treated as a primary data source. Other data sets (e.g. teaching data sets) are going to be generated using secondary data from the GGP as a RI. GGP infrastructure has its own data management guidelines that are approved by the Royal Netherlands Academy of Arts and Sciences (KNAW). The management of national GGS datasets are under the responsibility of national teams, while the standard GGS data are under the responsibility of the GGP Central Coordination Team (GGP CCT).

### 2.1. Project datasets

Table 1 below provides an outline of the project datasets and lists additional information on these datasets. This refers to data set annotations that are meant to provide an insight into the contents of each dataset, data type, one of four reasons why data is going to be processed (i. *conceptualising, collecting, curating and enhancing the data*; ii. *improving the FAIR architecture of secondary data*; iii. *promoting the use of secondary data*; iv. *building networks of partners, stakeholders and users*), a partner institution leading the data collection and/or processing and the work package a dataset is assigned to. The table also indicates a dataset risk level (very low, low, medium, high, very high) and respective needs for data protection following the KNAW guidelines. See section 8 'Other issues' for more information.



**Table 1. Datasets in GGP-5D**

| Dataset name                          | Dataset annotation  | Data type                      | Main data processing purpose                                 | Risk level | Leading partner | WP | Task |
|---------------------------------------|---|--------------------------------|--|------------|-----------------|----|------|
| DS_GGP5D_1_Project_Partners           | GGP-5D Project partners   | Personal data                  | Building networks of partners, stakeholders and users        | Low        | NIDI            | 7  |      |
| DS_GGP5D_2_Permanent_Indentifiers_GGS | GGS data sets and their permanent identifiers   | Statistical data, textual data | Improving the FAIR architecture of secondary data            | Very low   | INED            | 1  | 1.1  |
| DS_GGP5D_3_Metadata_Files_GGS         | GGS Metadata files compliant with the CESSDA model  | Textual data                   | Improving the FAIR architecture of secondary data            | Very low   | INED            | 1  | 1.1  |
| DS_GGP5D_4_Mixed_Mode_GGS_Guidelines  | Data from completed mixed mode (CAPI, CATI, CAWI) GGS, on response rates, incompleteness, data representativeness, etc. | Statistical data, textual data | Conceptualising, collecting, curating and enhancing the data | Very low   | INED            | 1  | 1.2  |

## GGP-5D Data Management Plan

|                             |  |                                |  |          |      |   |     |
|-----------------------------|--|--------------------------------|--|----------|------|---|-----|
| DS_GGP5D_5_Technical_Design | Data on GGP's technical capacities e.g. measures on sampling management procedures, capacity of the Central Hub to handle the extra work, the interest of the scientific community for more frequent data collection, GGP community interest in the possibility of fielding minimodules between waves. | Statistical data, textual data | Conceptualising, collecting, curating and enhancing the data                               | Very low | INED | 1 | 1.3 |
| DS_GGP5D_6_Other_Data_Forms | Data on the use of administrative and register data to enhance the GGS survey data. It will cover the sources and availability of data, but also their legal and institutional requirements for use and linkages to GGP survey data.   | Statistical data, textual data | Conceptualising, collecting, curating and enhancing the data                               | Very low | INED | 1 | 1.4 |
| DS_GGP5D_7_KPIS             | Revising and collecting data for the GGP KPIS to better align them to the needs of the GGP and the latest developments in this field (i.e. reports of ESFRI and the European Commission)   | Statistical data               | Promoting the use of secondary data; Building networks of partners, stakeholders and users | Very low | SGH  | 2 | 2.1 |

## GGP-5D Data Management Plan

|   |   |                                |  |          |     |   |     |
|---|---|--------------------------------|--|----------|-----|---|-----|
| DS_GGP5D_8_Impact_Showcasing                | Data necessary to produce eight impact case studies on the use of GGP in policy debates and eight briefs to showcase how the GGP contributes to tackling societal challenges. | Statistical data, textual data | Promoting the use of secondary data; Building networks of partners, stakeholders and users | Very low | SGH | 2 | 2.1 |
| DS_GGP5D_9_User_Module_Committee            | Data on user module assessment committee  | Personal data                  | Building networks of partners, stakeholders and users                                      | Low      | SGH | 2 | 2.2 |
| DS_GGP5D_10_User_Module_Evaluation_Criteria | Data necessary to evaluate the user modules submitted   | Statistical data, textual data | Building networks of partners, stakeholders and users                                      | Very low | SGH | 2 | 2.2 |
| DS_GGP5D_11_User_Module_Applications        | Data on user module applications (name and email address of applicant etc.)   | Personal data                  | Building networks of partners, stakeholders and users                                      | Medium   | SGH | 2 | 2.2 |
| DS_GGP5D_12_Stakeholders_Dialogue_Forum     | Data on Stakeholders Dialogue Forum members (name, institution, email, stakeholder field of expertise ( scientific, governmental and societal)).                              | Personal data                  | Building networks of partners, stakeholders and users                                      | Low      | SGH | 2 | 2.3 |

## GGP-5D Data Management Plan

|   |  |                                |   |          |      |   |     |
|---|--|--------------------------------|---|----------|------|---|-----|
| DS_GGP5D_13_Funding_Opportunities         | Data set on funding opportunities and respective funding institutions  | Statistical data, textual data | Building networks of partners, stakeholders and users | Very low | BiB  | 3 | 3-3 |
| DS_GGP5D_14_National_Roadmap_Inclusion    | Data on GGP partners, national contact points, national roadmaps   | Personal data, textual data    | Building networks of partners, stakeholders and users | Low      | HDRI | 5 | 5-3 |
| DS_GGP5D_15_Future_GGP_Enlargement        | Regional (Southern Europe, Eastern Europe, South-Eastern Europe) workshop participant, stakeholder data, data on how to fund GGP in respective counties. | Personal data, textual data    | Building networks of partners, stakeholders and users | Medium   | HDRI | 5 | 5-4 |
| DS_GGP5D_16_Board_of_Govt_Representatives | Government representatives to GGP, their names, institutional affiliation, emails.   | Personal data                  | Building networks of partners, stakeholders and users | Low      | UB   | 6 | 6.1 |
| DS_GGP5D_17_Board_ERIC_Readiness          | GGP ERIC readiness assessment scores as given by the Board of Government Representatives   | Statistical data               | Building networks of partners, stakeholders and users | Very low | UB   | 6 |     |
| DS_GGP5D_18_CB_Meetings                   | Attendance sheets of GGP-5D Consortium Board meetings  | Personal data, textual data    | Building networks of partners,                        | Medium   | NIDI | 7 | 7.2 |

## GGP-5D Data Management Plan

|                                    |  |                             |   |          |      |   |     |
|------------------------------------|--|-----------------------------|---|----------|------|---|-----|
|                                    |  |                             | stakeholders and users                                |          |      |   |     |
| DS_GGP5D_19_AB_Meetings            | Attendance sheets of GGP Advisory Board meetings   | Personal data, textual data | Building networks of partners, stakeholders and users | Medium   | NIDI | 7 | 7.2 |
| DS_GGP5D_20_EB_Meetings            | Attendance sheets of GGP Ethics Board meetings   | Personal data, textual data | Building networks of partners, stakeholders and users | Medium   | NIDI | 7 | 7.2 |
| DS_GGP5D_21_Teaching_Data          | Anonymous teaching datasets generated from existing Generations and Gender Surveys made to be easily used for educations | Statistical data            | Promoting the use of secondary data                   | Very low | NIDI | 7 | 7.3 |
| DS_GGP5D_22_Metadata_Teaching_Data | Metadata files for the anonymous teaching datasets compliant with the CESSDA model                                       | Textual data                | Promoting the use of secondary data                   | Very low | NIDI | 7 | 7.3 |

## 2.2. Data summary

The purpose of data collection, processing and/or generation in the scope of the GGP-5D project is differentiated into four groups that have been introduced in the previous subsection.

The four categories are:

- i. Conceptualising, collecting, curating and enhancing the data.
- ii. Improving the FAIR architecture of secondary data.
- iii. Promoting the use of secondary data.
- iv. Building networks of partners, stakeholders and users.

The existing data collected during GGS fieldwork in numerous countries, and existing metadata on this data are going to be the primary sources of data re-use. For the purpose of GGP-5D, data re-use will also refer to secondary data use. Consequently, data is going to be re-used in connection to categories i, ii and iii.

### 2.2.1 Personal data

The following types of data are going to be collected and/or generated in GGP-5D. Personal data (for members of the Consortium, Advisory Board, Ethics Board, User Module Assessment Committee, User Module Applicants, Stakeholders Dialogue Forum members, contact people for national roadmap inclusion, participants and stakeholders in regional workshops, Board of Government Representatives) in connection to WP 2, 5, 6 and 7. The personal data to be collected may include the categories below, however in some cases some of these categories may be omitted:

- Name
- Affiliation
- Country
- Mailing address
- Professional background
- Job title

- Gender

## 2.2.2 Statistical data

Statistical data will refer to the following data:

- KPIs and impact measures (WP2).
- GGP ERIC readiness assessment scores as given by the Board of Government Representatives.
- Teaching datasets generated from existing Generations and Gender Surveys made to be easily used for education. These datasets will be fully anonymized before diffusion.

## 2.2.3 Textual data

At this stage, purely textual datasets are expected to be generated in relation to metadata files. However, datasets that will also include textual data will be processed. For more details consult subsection 2.2.4 below.

## 2.2.4 Other data types

Other data types will refer to data with characteristics of more than one data type, such as:

- Relevant metadata and fieldwork performance data (modes, response rates, cut-off rates) from previously conducted GGS (WP1).
- Data on GGP Central Hub's capacity for data collection (WP1).
- Data on administrative data and its potential linkage to GGS (WP1).
- User Module evaluation criteria (WP2).
- Data set on funding opportunities, grant sizes, eligibility criteria and respective funding institutions (WP3).

## 2.2.5 Data formats

The formats of the data will be as follows:

## GGP-5D Data Management Plan

- Data and metadata will be requested, stored and transferred in comma-separated values CSV format, individual file size reaching up to 100mb.
- To facilitate the data exchange, MS Excel compatible files including comma separated and .xls(x) format will be also accepted.
- For statistical purposes, other formats include .dta (Stata), .RData (R), .SAV (SPSS).
- Where applicable data formats may be migrated when new technologies become available and are proved robust enough to ensure digital continuity and continued availability of data.

The datasets generated in GGP-5D will directly contribute to the five goals of the project. They will help to:

1. Enhance the technical excellence of the GGP.
2. Enhance the scientific and socio-economic impact of the GGP and the engagement of stakeholders
3. Enhance the financial sustainability of the GGP.
4. Develop the legal frameworks for the ERIC.
5. Consolidate the strategic position of the GGP in the landscape of Ris.

Acknowledging that many of the datasets will serve in ensuring continuous and improved functioning of the GGP, the GGP data users outside of the GGP-5D project will also benefit from GGS data permanent identification and data use for educational purposes.

### 3. FAIR data

This section is based on the Horizon Europe Data Management Plan Template v1.0<sup>1</sup>.

#### 3.1. Making data findable, including provisions for metadata

---

<sup>1</sup> <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=HORIZON>



A part of effort in WP<sub>1</sub> will be devoted to provide persistent identifiers for already existing GGS data. Persistent identifiers shall be provided for teaching datasets that will be created as a part of WP<sub>7</sub>. Other datasets that concern personal or GGP performance or capacity data etc. shall not be attributed with a persistent identifier.

Data Documentation Initiative (DDI) metadata standards shall be applied in connection to aforementioned teaching datasets. Already existing GGS metadata following DDI metadata standards shall inform the process of assigning permanent identifiers to GGS (WP<sub>1</sub>). DDI files will be made compliant with the CESSDA model.

Other data produced in GGP-5D will have an associated metadata document (stored as a .txt file) which describes key aspects of the data.

Search keywords for optimising data re-use of the GGP-5D project shall also be provided.

### 3.2. Making data accessible

For the duration of the project, personal data will be stored on the local secured server of the partner responsible for taking care of the data following national and EU-level data and privacy regulations. Data that is to be made available publicly (e.g. teaching data sets) will be stored on the GGP website<sup>2</sup> through which it will be accessible following the one-stop-shop principle free of charge pursuing standard access protocols for anonymised data. Individuals accessing the data will have to agree to the terms and conditions of ethical GGP data use which excluded commercial use. The functionality to express agreement shall be embedded on the website as a part of the standard access protocol. The teaching data sets that will be made publicly accessible as part of the GGP-5D are going to be completely anonymous, no additional measures (e.g. data access committee) for the project are envisioned.

Part of WP<sub>1</sub> is dedicated not only to provide datasets with persistent identifiers, but also to analyse, find and apply data licenses (e.g. Creative Commons) for existing and future GGP data products. Although non-

---

<sup>2</sup> <https://www.ggp-i.org/>

personal data that also does not contain sensitive information on the administrative, budgetary and performance aspects of the GGP is generally agreed to be made publicly accessible, there currently is no decision under what license it shall be done. The decision shall be reported in the updated version of the GGP-5D DMP.

Data generated in GGP-5D which will be made publicly accessible will remain available and findable after the project is over. This will be ensured by the GGP as an RI. Metadata pertaining to these datasets will continue to be available from GGP's Colectica portal<sup>3</sup>. The .txt format metadata on datasets that cannot be made public will remain stored on the GGP website.

Statistical data generated in the GGP-5D project will be stored in a machine readable format accessible by most open-source or licensed pieces of software. Other data will be stored in CSV or .xls(x) formats.

### 3.3. Making data interoperable

GGP has adopted DDI metadata standards. Part of work in WP<sub>1</sub> will be focused on making the documentation of the GGP data compliant with the CESSDA controlled vocabularies and the requirements of the Metadata Model implementing the OAI-PMH (Open archives initiative protocol for metadata harvesting) to allow the harvesting of our data by other catalogues (e.g. CESSDA).

Some data sets that will be produced in GGP-5D (e.g. teaching datasets) will include qualified references to existing datasets in the GGP.

### 3.4. Increase data re-use

Syntax and read-me files will be provided as part of the package with project data that is to be accessible and re-used. Once selected, CC licenses will be used for data to be preserved. During the project and after the end of GGP-5D, documented non-personal data (e.g. teaching datasets) will be made available for third party users as well. All the data generated in the project will undergo a quality assurance process that will involve a leading

---

<sup>3</sup> <https://ggp.colectica.org/>

institution for the dataset (see Table 1) and other consortium partners who will be selected for the task.

## 4. Other research outputs

Datasets generated in connection to WPs 1,2,3,4,5 and 6 will generally feed into reports that will be presented as project deliverables. All reports/deliverables except for the GGP Costbook (WP3), report on funding opportunities (WP3), report on the strategies for national roadmap inclusion (WP5) and final document on the Readiness reports and signed Memorandum of Understanding (MoU) (WP6) will be made publicly available on the CORDIS website. Among other things, these reports will include such organisational protocols as co-operation with other RIs in social sciences, stakeholders or the FAIR status of GGP data.

## 5. Allocation of resources

At the project level, the cost of making data fair is estimated at €49.246. These costs are directly related to the GGP website that is used as a tool to ensure data FAIRness and the Colectica license. These costs are included in the GGP-5D Grant Agreement.

Consortium partners will use their own budgets to archive personal data produced during the GGP-5D project (see Table 1 for more details) in their own repositories.

The overall coordination and data management of the project shall fall with NIDI. However, each partner leading a WP is responsible for preparing the datasets to make FAIR the data collected within its own activities. Furthermore, consortium partners have the responsibility to make sure their activities are in line with all applicable local, government and international laws, regulations and guidelines.

Long term data preservation will be decided upon by the project consortium.

## 6. Data security

The data security of GGP-5D project datasets will be ensured by following GGP Ethics Board recommendations and KNAW privacy impact assessment. In particular, any sensitive (personal, non-public) data will be stored on password protected partner institutional hardware and is not to be shared outside the project consortium and stored for longer than the end of the project in September 2026. Data transfers between partners are to be done either using SURFdrive or WeTransfer both recommended by the KNAW.

## 7. Ethics

In general, GGP produced data is sensitive and therefore subject to higher ethical and data subject right protection standards which is overseen by the GGP Ethics Board and the Data Protection Officer of the host institution. In practice, the GGP Ethics Board and KNAW-NIDI Data Protection Officer will oversee the implementation of GGP-5D as well.

## 8. Other issues

For the data management of the GGP-5D project, KNAW privacy impact assessment risk matrix is used (see Table 1). There are four five risk levels: very low, low, medium, high and very high. The risk levels in the matrix are a product of the probability of an event happening and its impact on data (subject) security. Table 2 below illustrates the principle, where green cells refer to low risk, yellow refer to medium risk and red indicate high risk.

**Table 2. Risk to data assessment matrix**

|               | Low probability | Medium probability | High probability |
|---------------|-----------------|--------------------|------------------|
| Low impact    | Very low risk   | Low risk           | Medium risk      |
| Medium impact | Low risk        | Medium risk        | High risk        |

## GGP-5D Data Management Plan

|             |             |           |                |
|-------------|-------------|-----------|----------------|
| High impact | Medium risk | High risk | Very high risk |
|-------------|-------------|-----------|----------------|