HEALTHYCLOUD
Health Research & Innovation Cloud

D1.3 Data Management Plan

VERSION 1.2

Document Information

Contract Number	965345
Project Website	http://www.healthycloud.eu/
Contractual Deadline	M6, August 2021
Dissemination Level	PU
Nature	R
Author(s)	Juan González-García (IACS)
Contributor(s)	Salvador Capella (BSC), Nadia Tonello (BSC)
Reviewer(s)	Adrian Thorogood (UniLu), Michaela Th. Mayrhofer (BBMRI-ERIC)
Keywords	Management, data management, data processing, GDPR

***Notice**: The HealthyCloud project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement Nº965345*

Table of contents

Executive Summary

This document provides the HealthyCloud Data Management Plan (DMP) which according to the Description of Action is to be 1st delivered in Month 6 and then updated periodically. The DMP outlines what kind of data will be collected or generated and how it will be handled, processed and shared. It describes the standards that will be incorporated and the related methodology for data collection, processing and data sharing. This deliverable is based on the template and the guidelines provided by the European Commission (see Section 7 on issues to cover in your Horizon 2020 DMP).

HealthyCloud will deliver a Strategic Agenda including a Ready-to-implement Roadmap for the European Health Research and Innovation Cloud, a major cornerstone piece in the future European Health Data Space. The Strategic Agenda will incorporate the consolidated feedback of a broad range of stakeholders: the European Commission, the Member States and regional, national, European and international relevant initiatives.

1 Data summary

1.1 Purpose of the data collection/generation and the relation to the objectives of the project

HealthyCloud has the following objectives:

- Promote a dialogue with a broad range of stakeholders for identifying needs, challenges and opportunities so they can be modelled and prioritized in the HealthyCloud's Strategic Agenda.
- Develop strategies to overcome legal and administrative barriers of a borderless sharing of human data in Europe and establishing a robust and trustable governance that combines efficient support for research with full endorsement of citizens' rights and interests.
- Provide recommendations, guidelines and best practices to enable accessing, using and reusing health data for better research outcomes across Europe within an ethically sound and legally compliant framework.
- Drive the adoption of mechanisms for the sustainable use of European capabilities on computational systems including both Cloud and High-Performance Computing.

To fulfil these objectives, HealthyCloud is expected to generate or collect data in the following activities:

(1) qualitative interviews with Data Protection Authority representatives and key stakeholders (WP2), to validate proposed governance models;
(2) health data metadata catalogues description and FAIR maturity levels (WP3), to understand the particularities of existing health data catalogues and collections;
(3) a glossary on health data and health data management and curation terms (WP3, WP4), to unify the terminology among the project and related projects;
(4) a catalogue of characteristics of health data curation and governance on data hubs (WP4), to understand the particularities of data hubs operation;
(5) user expectations surveys (WP6), to capture the desires and expectations of the future HRIC users;
(6) the different documents produced as intermediate or final achievements of HealthyCloud;
(7) scientific manuscripts produced as a consequence of the project;
(8) any relevant link to repositories and related initiatives;
(9) (optionally) synthetic data to conduct the use cases theoretical exercises (WP7).

1.2 Types and formats of data generated/collected

The expected types and formats are the following

(1) Qualitative interviews with Data Protection Authority representatives and key stakeholders
    a. Types: opinions on governance models proposed, opinions legal frameworks proposed;
    b. Formats: Video recordings - H264 or similar; Typescripts – Plain text;
(2) Health data metadata catalogues description and FAIR maturity

a. Types: identifiers of data catalogues, description of data catalogues, characteristics of data catalogues, FAIR level of data catalogues;
b. Formats: Structured tables – CSV/Relational

(3) Glossary on health data and health data management and curation terms
a. Types: name of the term, common description of the term
b. Formats: Plain text – Structured tables (Key/Value)

(4) Catalogue of characteristics of health data curation and governance on data hubs
a. Types: identifier of the data hub, description of the data hub, location of the data hub, operational characteristics of the data hub, governance model of the data hub
b. Formats: Structured tables – CSV/Relational

(5) Profile-based user expectations surveys;
a. Types: Self-reported user type, opinions on HRIC available services, opinions on HRIC interaction mechanisms, opinions on HRIC architecture
b. Formats: Structured tables – CSV/Relational

(6) Documents produced as intermediate or final achievements of HealthyCloud;
a. Types: reports, brochures
b. Formats: Plain text – Word files – PDF Files

(7) Scientific manuscripts produced as a consequence of the project;
a. Types: reports, manuscripts
b. Formats: Plain text – Word files – PDF Files

(8) Any relevant link to repositories and related initiatives;
a. Types: links
b. Formats: HTML links – DOIs

(9) (optionally) Synthetic data to conduct the use cases theoretical exercises
a. Types: identifiers of the data sets, identifiers of the synthetic patients, synthetic clinical variables of interest
b. Formats: Structured tables – CSV – Relational

1.3 Re-use of existing data

Existing data may be re-used for the following purposes:

- Existing health data metadata catalogues (WP3)
- Glossary on health data and health data management and curation terms (WP3; WP4)
- Existing links to repositories or initiatives

1.4 Data origins

(1) Qualitative interviews with Data Protection Authority representatives and key stakeholders
a. Produced during the project development: interviews and surveys.

(2) Health data metadata catalogues description and FAIR maturity
a. Produced during the project development: catalogues scrutiny
b. Re-use of existing catalogues

(3) Glossary on health data and health data management and curation terms
a. Produced during the project development: focus groups, experts' consensus
b. Re-use of existing glossaries

(4) Catalogue of characteristics of health data curation and governance on data hubs

a. Produced during the project development: landscape analysis

(5) Profile-based user expectations surveys (WP6);
   a. Produced during the project development: surveys.
(6) Documents produced as intermediate or final achievements of HealthyCloud;
   a. Produced during the project development: experts writing
(7) Scientific manuscripts produced as a consequence of the project;
   a. Produced during the project development: experts writing
(8) Any relevant link to repositories and related initiatives;
   a. Re-use of existing links: landscape analysis
(9) (optionally) Synthetic data to conduct the use cases theoretical exercises
   a. Produced during the project development: synthetic data generator (software)

1.5 Expected size of the data

(1) Qualitative interviews with Data Protection Authority representatives and key stakeholders
   a. Videos: few GBs
   b. Typescripts: few MBs
(2) Health data metadata catalogues description and FAIR maturity
   a. 100s MBs
(3) Glossary on health data and health data management and curation terms
   a. Few MBs
(4) Catalogue of characteristics of health data curation and governance on data hubs
   a. Few MBs
(5) Profile-based user expectations surveys (WP6);
   a. Few MBs
(6) Documents produced as intermediate or final achievements of HealthyCloud;
   a. Few MBs
(7) Scientific manuscripts produced as a consequence of the project;
   a. Few MBs
(8) Any relevant link to repositories and related initiatives;
   a. Few MBs
(9) (optionally) Synthetic data to conduct the use cases theoretical exercises
   a. Few GBs / 10s of GBs

1.6 Data utility: to whom it may be useful

(1) Collected data of the qualitative interviews with Data Protection Authority representatives and key stakeholders
   a. ELSI guidelines evaluation (WP2)
(2) The health data metadata catalogues description and FAIR maturity
   a. To create the data collections typology and FAIRness evaluation (WP3)
   b. External stakeholders, for example researchers whose aim is to characterise the data availability in other contexts or in another setups (other continents or markets).
(3) Glossary on health data and health data management and curation terms
   a. Overall project
   b. External stakeholders from other projects or health data analysis communities.
(4) Catalogue of characteristics of health data curation and governance on data hubs

a. Data Hubs characterization (WP4) and external stakeholders

(5) Profile-based user expectations surveys;
   a. FAIR data portal analysis and design (WP6)
   b. External stakeholders with the aim of designing equivalent cloud environments for other communities;

(6) Documents produced as intermediate or final achievements of HealthyCloud;
   a. Overall project
   b. External stakeholders to capture the knowledge produced in the project (for those publicly released documents)

(7) Scientific manuscripts produced as a consequence of the project;
   a. Overall project
   b. External stakeholders to capture the knowledge produced in the project

(8) Any relevant link to repositories and related initiatives;
   a. Overall project
   b. External stakeholders aiming to reuse the knowledge gathered in the project

(9) (optionally) Synthetic data to conduct the use cases theoretical exercises
   a. Use cases development (WP7, WP5, WP6)
   b. External stakeholders with the aim of providing alternative analysis to the uses cases proposed in the project

2 FAIR Data

Beyond the final Strategic Agenda, the generation of data is not intended to be a central aim of the Healthycloud project. Regarding to data FAIRness, it will focus mainly on the following data:

(1) Health data metadata catalogues description and FAIR maturity
   a. Data collections typology and FAIRness evaluation (WP3) and external stakeholders

(2) Glossary on health data and health data management and curation terms
   a. Overall project and external stakeholders

Both elements are expected to be produced in an electronically processable manner, for example the metadata catalogues will be created in structured tables, later stored in CSV files or relational databases, same applicable for the glossary of terms. The latter will be also stored in plain text and in a text processor format. Standard codifications will be explored to increase the interoperability. This data will be released under open licenses such as CC-BY 4.0.

The archive of the data will be done in the Zenodo repository and the persistent ID provided by this repository will be used to link the data in the HealthyCloud website, to increase its visibility, and this, its findability and re-usability.

3 Allocation of resources and responsibilities

During HealthyCloud, the costs of curation and preservation are related to the storage of data and results produced in the different those WPs that do not use personal data. In the case of the interviews and surveys of WP2 and WP6 respectively, where personal data may be collected

(to be detailed in Sections 6.2 and 6.3 in future revisions of this document), will also incur the cost of secure storage.

These costs of maintenance of the project website and the private document sharing system will be part of the Barcelona Supercomputing Center (BSC) budget. BBMRI-ERIC, co-leader of the WP2 and WP6, will manage the specific personal data gathered in the interviews and surveys on each WP, respectively, including its resource allocation and curation responsibility.

Long-term preservation of publicly available data will make use of publicly available systems like Zenodo or EUDAT B2Share repositories. Any internal/private data to be stored after project duration will be evaluated specifically and detailed in further versions of the DMP.

4 Data Security

Due to the responsibility of the data storage, the BSC and the BBMRI-ERIC will be in charge of the data security.

In the case of BSC, it will follow the security measures included in their *internal security policy and control mechanisms (administrative, technical and physical)*, available upon request. BSC policies states the prevention of execution of unauthorized software at their systems; a governance framework for privacy, confidentiality and security; the requirement of special authorization when accessing restricted data aligned a Central BSC Password Policy as well as restricted access to physical infrastructures.

In case of BBMRI, it will follow the BBMRI's "*Security and Privacy Architecture*"[1]. In this document, among specific information of biobank sampling securities elements, it is included 27 security measures to guarantee the proper operation, which includes elements: User Management for Operations (Me-1 User accounts are untransferable), Physical Security (Me-8 Server infrastructure must be physically accessible only to the designated IT personnel); System Protection, System Separation, and Network Protection (Me-12 Default passwords must not be used by any system connecting to the network or by systems they deliver network functionality); Software Development & Deployment (Me-18 Any software installed must come from trusted installation sources (original media, signed software packages, etc); Security Incident Handling (Me-24 BBMRI-ERIC IT and Data Protection Manager must be informed about any security incidents concerning BBMRI-ERIC IT infrastructure. Any affected third parties shall be notified, too.); and, User Training (Me-27 Data protection and privacy aspects shall be included in all the relevant training curricula produced by or supported by BBMRI-ERIC).

[1] https://doi.org/10.5281/zenodo.159444

5 Ethical Aspects

Interviews and surveys will be anonymous when possible. Otherwise, a consent will be provided to respondents (see informed consent template for WP2 in Section 6.2.1). Data controllers will be the responsible of the respective WPs (BBMRI for WP2, BSC for WP6)

Informed consents as well as opinions and recommendations of specific ethical committees regarding the contents of the interviews and surveys will be included in Sections 6.2 and 6.3 of the present documents, as well in Deliverables 9.1 and 9.2.

In addition, ethical aspects are discussed in section 5 of the DoA. All Consortium Partners are committed to comply with relevant international and EU level fundamental ethical, privacy and security legislation and regulations:

• Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

• The Declaration of Helsinki, "Ethical Principles for Medical Research Involving Human Subjects", that sets out the ethical standards.

• The Oviedo Convention on Human Rights and Biomedicine.

• The Belmont Report "Ethical Principles and Guidelines for the Protection of Human Subjects of Research".

• The Charter of Fundamental Rights of the European Union (7 December 2000) that sets out civil, political, economic and social rights of European citizens and all persons resident in the EU.

• The Recommendation of the Committee of Ministers No. R (90) 3 concerning medical research on human beings.

6 Other

6.1 DPOs contact list

Data Protection Officers (see table 5.1 below) of the coordination team and the BBMRI-RERIC, in charge of interviews and surveying activities will be involved to guarantee that, among others, the data minimisation principle of the personal data acquisition is respected.

Table 1 DPOs contact list

Organisation	DPO Name	DPO e-mail
IACS (Coordinator)	Mercedes Gómez Escribano	protecciondatos.iacs@aragon.es
BBMRI-ERIC (WP2/WP6 Lead)	Ayodeji Adeniran	dpo@bbmri-eric.eu

6.2 WP2 Interviews detailed information

6.2.1 Informed consent

Informed consents for the validation interviews will be based in the templates proposed in the document attached.

Project Logo

Participant Information/Privacy Notice

About the project xy?

xxx

About this study xy?

The main objective of this study within project XY is to xxx.

How will this be achieved?

The study uses a variety of research methods to engage with a wide range of stakeholders, including members of the xy consortium and the research field more broadly as well as ethical and legal experts: (1) Interviews will be conducted to explore the experiences and perspectives of various stakeholders; (2) Observations will involve the study team participating in and observing meetings and workshops; (3) The ECOUTER stakeholder engagement method will be carried out to capture the process of how challenges and solutions are identified and addressed (Murtagh, M. J. et al. 2017). The data collected will take the form of fieldnotes, photographs of flipcharts and mind maps, as well as audio and/or video recordings.

What will participation involve?

(1) Interviews will be conducted based on your agreement either face-to-face, by phone or online at your convenience. Interviews will last approximately 30 minutes and you may be asked for a subsequent interview. (2) Observations will be conducted during xy-related meetings and workshops, where members of the study team will watch, listen, take notes and, if possible, make audio and/or video recordings and take pictures. You will be asked for consent if you attend a meeting that includes observations. (3) If you agree to participate in an ECOUTER exercise, you will be asked to contribute to a mind-map based on questions posed by the study team. You are welcome to contribute as much or little as you like, and your contributions are anonymous.

Who is conducting the research?

The research is conducted by a study team, lead by: xy

Do I have to participate? Can I change my mind?

Participation in the study is absolutely voluntary and based on your consent. If you do not want to be interviewed, observed at a workshop or participate in an interview and/or ECOUTER session, just let us know. We will not take notes of what you said during the workshops or ECOUTER sessions or will remove your contributions from fieldnotes or transcripts of audio and/or video recordings. If other participants respond to what you said, their comments will be included in fieldnotes, transcripts and recording, as long as the information does not identify you.

If you change your mind after agreeing to participate, you can withdraw your consent until data is aggregated in analysis or after the end of the study (please find contact details at the end of this document). You can also change your mind at any time during an interview, workshop or ECOUTER session with no questions asked. Please be aware that we cannot remove your contributions after the analysis has been completed or if they were anonymous in the first place (in particular ECOUTER

contributions), and if the results of the study have been published in presentations, reports and papers.

How will data be kept secure and confidential?

All data collected will be kept secure at all times. Fieldnotes written by researchers during participant observations will be typed up digitally as necessary following each meeting; hand-written originals will be scanned into electronic form before the paper version is stored in a locked location and ultimately destroyed. Interviews – and observations where appropriate and possible – will be audio recorded digitally, transcribed under terms of a confidentiality agreement, and checked for accuracy. Electronic copies of all data will be password protected and stored securely in BBMRI-ERIC servers or its cloud service. Paper documents – if any – containing data will be stored in a secure locked location at BBMRI-ERIC. Copies of ECOUTER mind maps (e.g. electronic downloads or digital photographs) will also be password protected and stored securely on BBMRI-ERIC servers or its cloud service. While every effort will be made to keep all data collected confidential to prevent identification of individual participants (coded/pseudonymized), this study by its nature will likely include a number of highly specialized professionals and specialists working in a narrow field of study. Consequently, researchers will strip data from direct identifiers wherever possible and keep sensitive comments strictly confidential in order not to disclose the identity of the interviewee. Audio recordings of the interviews will be permanently deleted one year after analysis is complete. Other data will be kept for ten years to facilitate scientific reporting (e.g. papers, conferences), and then deleted.

How long will the study take? Will I see the results?

Data will be collected throughout the duration of the xy project (expected end date after xmonths on x x 202x). Researchers will undertake analysis as soon as data collection begins, and we expect this process to continue past the formal end date of the project. Results will made available where possible through reports, webinars and scientific journals, and presented at conferences and meetings.

What are the possible benefits and risks of my participation?

With your participation you will help increase the knowledge regarding the research goals of the study and xxx at large, which we expect to be of use for the broader research field, as well. We do not expect any sensitive issues to arise for you or other participants resulting from your participation. Appropriate safeguards have been put in place to minimize potential risks for you in case of unauthorized access or disclosure. The processing of the data collected is in accordance with Austrian national law (Datenschutz-Anpassungsgesetz 2018) and the EU General Data Protection Regulation (GDPR).

Who is funding the research?

The research is funded as part of the H2020 project xxx (Grant agreement no.: xxx).

Has the study received ethical approval?

The study was approved by the Ethics Committee [Ref.]. This approval was received before the empirical study started.

Contact information:

If you have questions or want to get in touch with the study team, please contact:

Investigator: xxx	DPO: xxx
Email: xxx Phone: xxx	Email: xxx

Project Logo

Informed Consent for Interviews

I hereby provide voluntary consent to a research interview conducted as part of the xy project (GA xx).

I confirm that I have read and understood the information leaflet provided, that I was informed about the nature, content and extent of the study and that I had the opportunity to ask questions. I was assured that my participation in the interview is completely voluntary and I can stop the conversation at any time without needing to give a reason. I understand that I am free to withdraw my data during or after the interview, but that I will not be able to withdraw my data once analysis is complete or from reports or publications.

I agree that my interview is being recorded, transcribed and used for further analysis. I agree that anonymized statements from my interview can flow directly into a publication.

I was assured that neither the audio nor the textual form of the conversation would be disclosed in any way, that the records would be confidential and that the research outputs would not include personally identifiable information. I was informed that audio recordings of the interviews will be permanently deleted one year after analysis is complete. Other data will be kept for ten years after this date to facilitate scientific reporting (e.g. papers, conferences), and then deleted. I understand that I am free to ask that particular comments, phrases or language be excluded from data collected and transcribed.
Processing of this data will be in accordance with Austrian national law (Datenschutz-Anpassungsgesetz) and the GDPR.

[Include only in consents sent by email.] By typing my name below, I am electronically signing this consent form. I understand that doing so constitutes a legal signature.

Date: _____

Name: _____

Signature: _____

Contacts:

xx
Email: xx Phone: xx

Page 2 of 4

Informed Consent for Participant Observations

I hereby provide voluntary consent to being involved in participant observations at events, meetings or workshops as part of the xx project (GA xx).

I confirm that I have read and understood the information leaflet provided, that I was informed about the nature, content and extent of the study and that I had the opportunity to ask questions. I was assured that my participation in the study is completely voluntary. I understand that I am free to withdraw my data during or after any event at which I am observed without needing to give a reason, but that I will not be able to withdraw my data once analysis is complete or from reports or publications.

I agree that my contributions at events, meetings or workshops conducted as part of xy are being audio or video recorded, photographed, transcribed and used for further analysis. I agree that anonymized statements can flow directly into a publication.

I was assured that neither the audio or video nor the textual form of the observations would be disclosed in any way, that the records would be confidential and that the research outputs would not include personally identifiable information. I was informed that audio recordings of the interviews will be permanently deleted one year after analysis is complete. Other data will be kept for ten years after this date to facilitate scientific reporting (e.g. papers, conferences), and then deleted. I understand that I am free to ask that particular comments, phrases or language be excluded from data collected and transcribed. Processing of this data will be in accordance with Austrian national law (Datenschutz-Anpassungsgesetz) and the GDPR.

[Include only in consents sent by email.] By typing my name below, I am electronically signing this consent form. I understand that doing so constitutes a legal signature.

Date: _____

Name: _____

Signature: _____

Contacts:

xx
Email:xx, Phone.: xx

Page 3 of 4

6.2.2 Interviews contents

TBC in further revisions of the DMP

6.3 WP6 Surveys detailed information

6.3.1 Informed consent

Informed consents for the user's expectation surveys will be based in the templates proposed in the following documents, among others is yet to be define and may follow the template detailed in the Section 6.2.1.

6.3.2 Surveys contents

TBC in further revisions of the DMP