Plan Overview

A Data Management Plan created using DMPonline

Title: Harmony project: Establishing an organizational and ICT infrastructure for data sharing to facilitate integrated analysis of COVID vaccine immune response trial data from vulnerable groups and general population

Creator: Sophie van Wingerden

Principal Investigator: Nynke Rots

Data Manager: Sophie van Wingerden, Roxane Weijenberg

Project Administrator: Marjo Knapen, Petra Jochemsen

Contributor: René Unteregger, Kevin Kort, Oostrum, Anja van, Max Nazarski, Michael Winterkorn, Mark Kroon, Elif Akbas

Affiliation: National Institute for Public Health and the Environment (RIVM)

Funder: ZonMw (Nederlands)

Template: Data management ZonMw-template 2019

ORCID iD: 0000-0002-7170-7826

Project abstract:

Een groot deel van de Nederlandse bevolking is gevaccineerd tegen COVID-19. Na vaccinatie volgt een afweerreactie in het lichaam. Zo wordt immuniteit opgebouwd. De afweerreactie van gezonde Nederlandse mensen wordt bestudeerd door onder anderen het RIVM en het Erasmus MC. Er zijn ook acht onderzoeksgroepen die de afweerreactie in patiënten met een verminderde afweer bestuderen. Die verminderde afweer kan komen door onderliggende aandoeningen zoals een aangeboren immuundeficiëntie of afweerstoornis, of door behandeling van een ziekte. Deze mensen lopen een hoger risico ernstig ziek te worden bij een COVID-19 infectie. We weten nog niet goed genoeg of zij voldoende én voldoende lang worden beschermd na vaccinatie.

Door uitkomsten van onderzoeken te vergelijken kunnen misschien meer mensen geholpen worden. Daarom heeft ZonMw verzocht om een infrastructuur en organisatiestructuur op te zetten om de resultaten van de betrokken COVID-19 vaccinstudies te delen, harmoniseren en analyses uit te kunnen voeren met de gezamenlijke database. Waardoor de resultaten uit de verschillende studies (bij verschillende patiënten) met elkaar vergeleken kunnen worden. De data is door de onderzoeksgroepen verzameld en moeten door de onderzoekers gedeeld gaan worden ten behoeve van onderzoeksgroep overstijgende analyses om project overstijgende onderzoeksvragen te kunnen beantwoorden. Er zijn wetten en regels waaraan voldaan moet worden bij het delen van data over patiënten. Daarnaast bedoelen onderzoekers soms niet precies hetzelfde met een beschrijving. Of worden verschillende woorden gebruikt terwijl wel hetzelfde bedoeld wordt. Het creëren van eenheid van taal in de verzamelde data is dus een randvoorwaarde om onderzoek overstijgende analyses uit te kunnen voeren. Daarnaast zijn er ook technische uitdagingen voor het veilig delen van data en het gezamenlijk werken in een digitale werkomgeving.

Samen met het RIVM vormen de acht onderzoeksgroepen een onderzoeksgroep; het ‘Harmony consortium’. Voor het Harmony consortium is na goedkeuring van ZonMw voor het projectvoorstel

een organisatie opgericht om de data op een goede en veilige manier te kunnen delen. Ook wordt ervoor gezorgd dat iedereen dezelfde taal gaat spreken. Voor alle hierbij optredende organisatorische en technische uitdagingen wordt een oplossing gezocht. En om te testen of de oplossingen daadwerkelijk werken, wordt met een pilot een eerste inhoudelijke onderzoeksvraag beantwoord met behulp van de gedeelde en geharmoniserde data.

ID: 120333

Start date: 01-01-2023

End date: 01-10-2024

Last modified: 12-08-2024

Grant number / URL: https://www.zonmw.nl/nl/over-zonmw/coronavirus/programmas/project-detail/covid-19-programma/inrichting-van-een-organisatie-en-infrastructuur-voor-onderzoeksgroep-overstijgende-analyses-van-co/verslagen/

Harmony project: Establishing an organizational and ICT infrastructure for data sharing to facilitate integrated analysis of COVID vaccine immune response trial data from vulnerable groups and general population

1. General features of the project and data collection

1.1 Project leader contact details

Drs. Marjo Knapen
Centrum voor Immunologie van Infectieziekten en Vaccins (IIV)
Rijksinstituut voor Volksgezondheid en Milieu
Antonie van Leeuwenhoeklaan 9 | 3721 MA Bilthoven
marjo.knapen@rivm.nl
M 06 53158335

1.2 I have composed my DMP with the assistance of a data stewardship (or management) expert. List his or her name, function, organisation/department, phone number and email address.

- The expert is connected to my department or institution (please explain his/hr expertise related to data stewardship)

Kevin Kort en Anja van Oostrum
Data Stewards
IV-organisatie | Afdeling Onderzoeks- en Datadiensten | Data en GIS
Rijksinstituut voor Volksgezondheid en Milieu
Rijksinstituut voor Volksgezondheid en Milieu
Antonie van Leeuwenhoeklaan 9 | 3721 MA Bilthoven

(no longer working at the RIVM)
Elif Akbas
Data Steward
IV-organisatie | Afdeling Onderzoeks- en Datadiensten | Data en GIS
Rijksinstituut voor Volksgezondheid en Milieu
Rijksinstituut voor Volksgezondheid en Milieu
Antonie van Leeuwenhoeklaan 9 | 3721 MA Bilthoven

1.3 In collecting data for my project, I will do the following:

- Use existing data (please specify)
- Merge different data files (please specify)
- Add new data to an existing data set (please specify)
- Generate new data

Existing data: we will use data that is already collected in various studies, namely:

1. RECOVAC IR/RV - Dr. Jan-Stephan Sanders (UMCG) - EUCTR2021-001520-18-NL
2. COVALENT - Dr. Coretta van Leer (UMCG) - NL-OMON22289
3. VACOPID - Dr. Virgil Dalm (Erasmus MC), - EUCTR2021-000515-24-NL
4. VOICE - Dr. Liesbeth de Vries (UMCG) - EUCTR2021-000872-13-NL
5. PRIDE - Dr. Louis Bont (UMCU) - EUCTR2021-002613-34-NL
6. COVIH – Dr. Kees Brinkman (OLVG) - EUCTR2021-001054-57-NL
7. TARGET2B - Dr. Filip Eftimov (AUMC) - NL-OMON49992
8. COBRA-KAI - Dr. Inger Nijhof (AUMC), Mette Hazenberg (AUMC) & Bram Goorhuis (AUMC) - EUCTR2021-001072-41-NL
9. SWITCH and SWITCH ON - Dr. Hugo van der Kuy (Erasmus MC) - EUCTR2021-000701-24-NL and NL-OMON53985
10. HCW - dr. R.D. de Vries (Erasmus MC)
11. IIV-RIVM - Dr. Nynke Rots (RIVM)
    1. VITAL (COVID part) - Dr. Josine van Beek & Dr. Nynke Rots - EUCTR2019-000836-24-NL
    2. VOCAAL (COVID part) - Dr. Annemarie Buisman - EUCTR2021-002363-22-NL
    3. IIVAC - Dr. Alienke Wijmenga - EUCTR2021-001357-31-NL

1.

All parties included in the consortium will upload their dataset in the cloud-based digital research environment (DRE) of anDREa B.V. This DRE is made available by the RIVM and agreements for usage will be signed.

Within de DRE the data can not be downloaded (unless you are an owner), but all approved researchers can perform their analyses within de DRE in their online workspace.

Within the DRE the separate datasets of the consortium partners will be retrospectively harmonized and therefore a common dataset will be generated. New (lab)data: In addition to the existing data collected by the consortium partners, Fc-mediated functionality assays will be performed on serum samples of the participants from each study in order to asses the functional antibody levels in a standardized way. NB: COVIH did not have a sufficient informed consent, therefore no data from COVIH has been included.

1.4 In my research, I will use:

- Exclusively quantitative data

Both the patient- and laboratory data are quantitative. New laboratory data will be created, which will be added at a later date. This will, however, also be quantitative.

1.5 I will be reusing or combining existing data, and I have the owner's permission for that.

- Yes, I have permission to use the data

Data sharing for overall analysis is the goal. All partners have (as agreed) asked permission of the participants in the informed consent forms to be used in their studies. To be sure the partners were asked to share these forms to do a final check (by the privacy department of the RIVM) in terms of reuse/further use of data. Unfortunately, the COVIH informed consent was insufficient, and the study was therefore not allowed to share their data with third parties.
A consortium agreement has been signed in dec 2023 to obtain and harmonize relevant data from the partners. A DPIA documenting data protection has been approved by the functionaris gegevensbescherming from the Ministry of Health.

1.6 In collecting new data, I will be collaborating with other parties.

- Yes, the new data will be (partly) provided by a project partner or supplier
- Yes, we have reached agreements on the user rights of the data used in the project

Yes, a consortium agreement has been signed by all partners in December 2023. New data will be generated in the overarching analysis which will later be added to the final dataset. An agreement on this new data has been signed.

1.7 I am a member of a consortium of 2 or more partners. Clear arrangements have been made regarding data management and intellectual property. (also consider the possible effect of changes within the consortium on issues of data management and intellectual property)

- Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement

The consortium agreement is signed where topics as ownership, terms and conditions for re-use of the date, allocation of responsibilities are agreed upon.

1.8 I can give an estimate of the size of the data collection; specifically, the number of participants or subjects (“n=”) in the collection and its size in GB/TB

- Yes (please specify)

The collection and the size will be specified. Number of participants for the harmonized dataset is 9546, however, not all participants have received vaccinations or have immunological data available. The number of participants with vaccination and/or immunological data is approximately 8000. The size of the datasets combined is 8 Mb.

1.9 The following end products I will make available for further research and verification (please elaborate briefly)

- Documentation of the research process, including documentation of all participants
- Syntaxes

- Data documentation
- (Several versions of) processed data

We will collect datasets that have already been collected in clinical trials performed by the consortium partners. These are cohort data combined with biomedical samples like serology samples.

- format: mostly csv
- pseudonimization: the data of each cohort will be pseudonimized by partners before they are uploaded to DRE
- datasharing: data will be uploaded in the digital research environment (DRE), harmonised and shared through the DRE
- quality control on pseudonimization: each cohort will check their data on free text and other traceable information in their own dataset before uploading it in the DRE.

We will

- draft an inventory of the datasets,
- perform retrospective harmonisation and
- create a harmonized common dataset.

Results of the project are

- common harmonized dataset
- data analyses on harmonized dataset including data from all consortium partners and results (publications)
- openly available syntaxes (with the rules for recoding the different variables)
- version control, so that other researchers can validate the harmonized dataset
- description and availability of metadata
- harmonization templates

1.10 During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)

- Yes, I will make use of an external provider's services for storage and backup of my data

The DRE by AnDREa will be used for storage, harmonization, and analyses.
Archiving and long term storage for harmonized dataset will be done in the DANS Data Station for Life Sciences.

2. Legislation (including privacy)

2.1 I will be doing research involving human subjects, and I am aware of and compliant with laws and regulations concerning privacy sensitive data.

- Gedragscode Goed gebruik van lichaamsmateriaal (Code of Conduct for Responsible Use of Human Tissue)
- Wet op de Geneeskundige Behandelingsovereenkomst (Medical Treatments Contracts Act)
- The Wet Medisch-Wetenschappelijk Onderzoek met Mensen (WMO, or Medical Research (Human Subjects) Act) applies to my project; I will have it reviewed by a Medical Research Ethics Committee. In addition I will comply with the Kwaliteitsborging Mensgebonden Onderzoek (Quality Assurance for Research Involving Human Subjects)
- Yes, I will involve human subjects in my research. I will comply with the Algemene Verordening Gegevensbescherming (AVG)
- GDPR - The Harmony project complies with the GDPR

The consortium partners for each study have checked which legislations apply for their study. Possible legislations are:

- WMO - Wet Medisch-Wetenschappelijk Onderzoek met Mensen
- WGBO - Wet op de geneeskundige behandelingsovereenkomst
- CTR - Clinical Trial Regulation
- Gedragscode gezondheidsonderzoek (vervanger van code goed gedrag)
- Code goed gebruik (bij sample analyse)
- Verklaring van Helsinki
- ICH-GCP
- wet BIG

2.2 I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.

- Yes, and this informed consent allows for the reuse of data (note that in the Code of Conduct for Medical Research, ‘reuse’ is also referred to as ‘further use’)

Participants have been asked to sign an informed consent form to give permission to receive, store, and use (and re-use) their data during the orginal studies.
These informed consents have been collected from the consortium partners and checked by a privacy officer of the RIVM for the reuse of data for this harmonization project. Only the data from the participants that have agreed for reuse of data has been shared with the Harmony project.

2.3 I will be doing research involving human subjects, and I will protect my data against misuse.

- Yes, the data will be pseudonymised. (please explain how this will be done, and by which organisation) and

The data will be stored and analyzed in DRE which is a GDPR compliant secure digital environment. In DRE both data and analytics are not publicly accessible but securely accessible for the researchers by 2 factor authentication. Data sharing is controlled by the data owners with restricted access rights. Consortium partners will be asked to pseudonymize the data before loading their raw data to DRE. In the pseudonimization process, personal identifiable information derived from the questionnaires will be removed before uploading the data.
Only year of birth has been be used to minimize the chance of identification. We have worked with a template in order to only receive the data that is needed to answer project overarching research questions.
Access to the harmonized dataset will only occur after approval from the consortiummembers. A projectmember will then upload the data to a DRE environment where they will provide access to the researcher(s).

2.4 I will stick to the privacy regulations of my organisation

- Yes

The RIVM and its researchers comply to the privacy guidelines of the RIVM, which are in line with the Algemene Verordening Gegevens bescherming (AVG) (**see: RIVM Privacyverklaring 1 mei 2023, Algemene privacyverklaring RIVM | RIVM)** A privacy legal professional and a DPO are being consulted. To address privacy issues we drafted a 'onderlinge regeling - mutual arrangement', in the **Appendix 6** of the Consortium Agreement. Here we may also make use of attachments/appendixes from the consortium agreement (e.g. article 15.4).

3. Making data findable

3.1 The data collection of my project will be findable for subsequent research. E.g., on a catalogue, a web portal, or through the search enginge of the repository (note: this is key item 3, which you should report to ZonMw at the end of your project).

- Yes, it can be found through an online (metadata) catalogue or web portal (please specify)

Data Station of Life Sciences by DANS will be the place where the data and metadata will be available. A DOI will be automatically generated for the data and metadata.
https://lifesciences.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/LS/8HWQUW
The metadata will be publicly available, the data will be restricted.
A form will be published were a researcher can apply to gain access to the data. This will be discussed in the stuurgroep of the consortium, and once approval is given, a RIVM projectmember will

- deploy a DRE environment with necessary software
- transfer a copy of the data from the Data station to the DRE
- give access to the researcher

Any costs for deploying a DRE environment should be provided by the researchers.

3.2 I will use a metadata scheme for the description of my data collection (note: this is key item 7, which you should report to ZonMw at the end of your project).

- Yes, I will use a generic metadata scheme (please specify)

The DANS Data Station for Life Sciences. For the separate studies links might be made to the EudraCT portal or the WHO trial portal. DANS uses their own metadata standard, which can be exported to JSON-LD and the likes.
The metadata will also be uploaded to the covid-19 data portal from Health-RI (M4M standard) and RIVMdata (ISO11915), with links to the Data Station.

3.3 I will be using a persistent identifier as a permanent link to my data collection (note: this is key item 1, which you should report to ZonMw at the end of your project).

- Yes, I will be using the DOI code

The DOI-codes of the scientific articles that will follow from this project will be shared with ZonMw.
Since the metadata and data are published in the Data Station, the DOI will automatically be generated.
Our DOI is https://doi.org/10.17026/LS/8HWQUW

4. Making data accessible

4.1 Once the project has ended, my data will be accessible for further research and verification.

- Yes, after an embargo period (please explain)

Accessibility depends on predefined terms and conditions. We will comply with the stipulation of ZonMw with respect to the embargo period. The goal is to make data accessible for further research. This project ends with the development of an infrastructure, but analysis might still be ongoing. So data will be made accessible after finishing that. The procedure as described in point 3.1 wil be followed.

4.2 Once the project has ended, my data collection will be publicly accessible, without any restrictions (open access).

- No, there will be access restrictions to my data collection (please explain)

At the end of this project, when the development of an infrastructure ends, analysis will still be ongoing on the harmonized dataset. The Harmony team will write a proposal for access rights and restrictions regarding the data repository which be discussed and decided on by the consortium partners. The proposal is described in point 3.1.

4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier; also note that this is a key item 4, which you should report to ZonMw at the conclusion of your project).

- Yes, my institution has drafted a set of terms of use with the help of a legal advisor

The terms and conditions are described in the consortium agreement. The form that can be used to request access is attached to the agreement and will be uploaded in the Data Station.

4.4 In the terms of use restricting access to my data, I have included at least the following:

- Conditions related to data security
- Collaboration in using the data set, including agreements on publication and authorship
- Whether or not the data set may be linked with another data set (for reasons of privacy)
- The manner in which the data set can be accessed
- The permitted period of use of the data set
- The reimbursement of costs, for example in obtaining the data
- A steering committee, programme committee or project leader will be charged with approving data requests
- The approval of the participants allows for further research using this data set

These terms of use restricting access to my data, collaboration in using the dataset including agreements on publications and authorship will be discussed within the consortium. A steering committee, programme committee or project leader will be charged for approving data request.

5. Making data interoperable

5.1 I will select a data format, which will allow other researchers and their computers (machine actionable) to read my data collection (note: this is key item 5, which you should report to ZonMw at the end of your project).

- Yes (please specify)

Yes, the data will be delivered in .csv format. The code book includes ontology references (LOINC and/or SNOMED) for applicable variables.

5.2 I will select a terminology for recording my data (e.g., code, classification, ontology) that allows my dataset to be linked or integrated with other datasets (note: this is key item 6, which you should report to ZonMw at the conclusion of your project).

- Yes, metadata standard (please specify)

Yes, LOINC is used for the labdata and SNOMED for clinical data, where applicable. A codebook will be published with the data in order to ensure understanding of the data and values.

5.3 I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.

- Yes, the participants have given their permission for reuse of the data, and the data have been pseudonymised

Yes, only data has been shared to the Harmony project for which the participants have explicitly given consent for sharing data with other parties. Unfortunately the informed consent of the COVIH study was insufficient, thus no data from COVIH has been included in the Harmony project.

6. Making data reusable

6.1 I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).

- I will document the research process (please explain)
- I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)
- In addition, I will take further quality assurance measures (please specify)
- I will document the software used in the course of the project (please specify)

For the reuse of data and possible replication of the research, at least the following data and documentation must be archived

- Raw data of cohorts: archiving will be done by the cohorts themselves.
- Common dataset: the data will be archived in a Data Station for Life Sciences,together with the metadata. SOP, project proposal, consortium agreements and attachments will be archived on the file platform of the RIVM.
- Metadata: Data Station for Life Sciences
- Scripts: The M queries used to create the harmonized dataset will be added to GitHub which is linked on the Data Station

6.2 I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9 and 6.1)

- Yes

The common dataset and scripts will be archived. Each study has received a compendium of the data they sent, the scripts used to transform their data, and a file with how their data ended up in the harmonized files. It is up to the researchers to properly archive their data.

6.3 Once the project has ended and the data have been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or archival.

- Yes (please specify)

Approximately 8 Mb

6.4 I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)

- Yes, and this archive meets certification criteria and intends to get certified (please explain how your data will remain accessible and reusable in the long term)
- Yes, and this archive has a data seal of approval (please specify the archive)

Archiving will take place in a DANS Data Station for Life Sciences. They are currently working on obtaining a CoreTrustSeal. Until the certification has been granted, all data are duplicated in the EASY archive which holds the CoreTrustSeal and is therefore certified.

6.5 Once the project has ended, I will ensure that all data, software codes and research materials, published or unpublished, are managed and securely stored. Please specify the period of storage.

- Yes, in accordance with other guidelines (please explain, and specify the guidelines and the number of years)

25 years, based on the CTR and archiefwet.

6.6 Data management costs during the project and preparations for archival can be included in the project budget. These costs are:

- Amount ..... (please elaborate)

Archiving in the DANS Data Station is free of charge for data under 50TB.

6.7 The costs of archiving the data set once the project has ended are covered.

- Yes (please elaborate)

Archiving in the DANS Data Station is free of charge for data under 50TB. Since our data only contains .csv files and is currently 8 MB, we expect the archiving to be free of charge.

Planned Research Outputs

Collection - "Code repository in GitHub"

Study registration - "Record in the Dutch COVID-19 portal"

Study registration - "Record in the metadata catalogue of the RIVM"

Dataset - "Record in Data Station for Life Sciences"

This is contains the metadata and data in the Data Station for Life Sciences

---

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Code repository in GitHub	https://github.com/rivm-syso/harmony-data-integrat ...	Collection	2024-08-09	Open	GitHub		European Union Public License 1.2	None specified	No	No
Record in the Dutch COVID-19 portal	https://covid19initiatives.health-ri.nl/p/Project/ ...	Study registration	2024-07-25	Open	None specified		Creative Commons Zero v1.0 Universal	None specified	No	No
Record in the metadata catalogue of the RIVM		Study registration	2024-08-06	Open	RIVMdata		Creative Commons Zero v1.0 Universal	ISO 19115	No	No
Record in Data Station for Life Sciences	10.17026/LS/8HWQUW ...	Dataset	2024-07-25	Restricted	None specified	8 MB	Creative Commons Attribution Non Commercial Share Alike 4.0 International	None specified	No	No