# Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** National Survey on Research Integrity

**Creator:**Gowri Gopalakrishna

**Principal Investigator:** Gowri Gopalakrishna

**Data Manager:** Gowri Gopalakrishna

**Project Administrator:** Sara Behrad

**Affiliation:** Vrije Universiteit Amsterdam

**Funder:** ZonMw (Nederlands)

**Template:** Datamanagement ZonMw

**Project abstract:**

Introduction To merit society's trust and investment in science, the scientific enterprise has the responsibility to ensure that the science it produces is trustworthy. One major threat to the trustworthiness of science arises when scientists engage in 'detrimental research practices' (QRPs): behaviors that range from subtle trespasses of ethical and methodological principles to outright scientific fraud. Although the importance of preventing these behaviors is widely recognized, the scope of engagement in QRPs is currently unknown. Estimates of occurrence cannot be validly and reliably based on previous studies because these studies employed methods that did not account for bias due to the sensitivity of admitting to QRPs, employed small samples, or focused on specific scientific disciplines. It is also unclear which strategies should be employed to decrease the occurrence of QRPs. Objectives The National Survey on Research Integrity is a globally unique endeavor in which the Dutch scientific community will take the lead in acquiring the necessary solid empirical basis for building strategies to reduce QRPs and thereby foster Responsible Research Practices (RRPs) in the Netherlands. With the National Survey, we seek to obtain a) valid and reliable domain-specific estimates of the occurrence of QRPs in the Netherlands, b) thorough domain-specific comprehension of the associations between explanatory variables and engagement in QRPs, and c) domain-overarching and domain-specific shared understanding of the roles that each of the stakeholders in science can play in reducing QRPs and fostering RRPs. Methods We will pursue the first two of these objectives by surveying all scientific personnel employed for research at all universities and university medical centers in the Netherlands through a web-based questionnaire. To elicit more honest responses to the questions about engagement in QRPs and to guarantee anonymity, we will apply a highly advanced Randomized Response (RR) technique. We will pursue the third objective through extensive stakeholder consultation in which the obtained knowledge will be translated into concrete stakeholder-tailored action plans on how to reduce QRPs and thereby foster RRPs across and within each of the four science domains: a) the Life Sciences and Medical Sciences; b) the Natural Sciences and Engineering Sciences; c) the Humanities, Language, Information, Communication, Law, and Arts; and d) the Social Sciences and Behavioral Sciences. Moreover, the National Survey takes an approach that acknowledges the complexity of the science system. It therefore addresses the roles of all of the five major stakeholder groups in the science system: a) researchers, b) academic research institutions, c) the umbrella organizations of these institutions, d) funders of academic research, and e) publishers of academic research. The National Survey will be conducted over a period of three years. The first year will be devoted to thorough preparation and extensive piloting of the different aspects of the survey and the RR techniques. The second year will be assigned to data collection, analysis, and reporting. The third year will be dedicated to translation of the survey results into action plans and their dissemination. Conclusion Employing extensive stakeholder consultation, the obtained knowledge will then be translated into concrete stakeholder tailored action plans on how to reduce QRPs and thereby foster responsible research practices.

**ID:** 29309

**Last modified:** 08-04-2021

**Grant number / URL:** 445004002

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

# National Survey on Research Integrity - Data management ZonMw (English version)

## 1. General features of the project and data collection

### 1.1 Project leader contact details

Prof. dr. L M Bouter
VU Medical Center Amsterdam, Department of Epidemiology and Biostatistics, Postbus 7057, 1007 MB Amsterdam, Netherlands
lm.bouter@vu.nl

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

- The expert is connected to my department or institution

Klaudia Onnasch
Datamanagement Consultant
Locatie VUmc | MF - H132 / Postbus 7057, 1007 MB Amsterdam
T: 020-44 49827 | k.onnasch@vumc.nl

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

- Generate new data

The National Survey on Research Integrity will gather data from a web based survey of the academic population of the Netherlands i.e. all academic researchers employed as professors, associate professors, assistant professors, postdoctoral researchers, or PhD students at a Dutch university or university medical center (UMC).

In addition to this we will generate and analyze data collected from focus groups conducted among scientists from the four major science domains with the aim of discussing and understanding the most problematic areas in research in their domain. In the first year of this project we will conduct a thorough pilot of the survey instrument using a sample selected from a representative research population to the Netherlands. This data as data from the main study is considered as new data.

### 1.4 In my research, I will use:

- A combination of quantitative and qualitative data

A web based survey with a combination of open and closed ended questions, as well as data from focus groups will be generated.

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

- No, I will not be reusing or combining existing data

This study will generate new data.

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

- Yes, I will collect the new data in conjunction with other researchers or research groups

Yes, as we will employ a trusted and respected data management organization who will execute and collect the survey data of the study in year 2 for us so as to ensure the raw data collected will not be directly available to the research team at any point in time of this study.

**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)**

- No, I am not working with 2 or more partners

However it is worthwhile to note that the data collection of the survey will be outsourced to a trusted experienced and respected research organization in Netherlands which will adhere to information security norms ISO 27001, 27002, and NEN 7510, and fall under European and Dutch data protection laws. An attorney from the Innovation Exchange Amsterdam (IXA) (see http://www.ixa.nl/en/home) will compose a processing agreement with the selected organization following a standard format. See also 2.3 for more information

**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)

We are expecting to be able to survey about 30-40,000 researchers who will be required to fill in a web based survey that does not exceed 15-20 minutes in order to get an optimal response rate. These questions are expected to be a combination of open and closed ended types. For the focus groups we have yet to determine the actual size but based on typical sizes of a focus group we expect to have about 10 participants and 5 focus groups. the sample size of the pilot study is yet to be determined but is likely to be smaller than the actual survey in year 2.

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

- Several versions of processed data
- Data documentation
- Syntaxes

In order to protect the confidentiality of all respondents in this study we will only be able to provide processed data which has been annonymized. We plan to also include syntax (the syntax is in R) from the analyses) for reproducibility.

**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 my institution's standard facilities for storage and backup of my data

## 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.**

- The Wet Medisch-Wetenschappelijk Onderzoek met Mensen (WMO, or Medical Research (Human Subjects) Act) does not apply to my project.

All invited researchers will be provided with access to our privacy policy. This privacy policy describes how we comply with all requirements of the General Data Protection Regulation (GDPR) Act, by which Ethics Committees the study protocols will have been reviewed and a statement declaring that the study does not fall under the scope of the Medical Research Involving Human Subjects Act (WMO). The privacy policy will further explain that data set will not contain any personal data other than in which of the four major science domain the respondent works, their gender, and whether they are PhD student, non-tenured scientist, or tenured scientist. Individuals are thus neither directly nor indirectly identifiable.

In addition, it will specify how and under which restricted circumstances the data may be shared for verification or further analyses with third parties, how we will prevent data breaches, and how we ensure that only authorized persons can access the anonymous

data set. Finally, it will specify how and where we will disseminate the results. All invited researchers will also have access to the full study protocol and the data analysis plan which will be preregistered on the Open Science Framework. During the pilot stage of the survey in the first year of the study, we will ensure that a similar privacy policy is applied and the pilot study protocol will also be preregistered on the Open science Framework.

**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')

A trusted, professional data collection organization will execute the survey. This organization will send a message to all e-mail addresses containing an invitation to participate in the survey, a brief explanation of the goal of the study, the privacy policy, and the link to the survey itself. When respondents click on this link, they will first be directed to the informed consent form. We will also ensure the same consent requirements will be followed during the pilot phase of this study in year 1.

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

- Yes, the data will be anonymised. I realise that this will limit the options for re-use of my data. (explain)

The privacy policy of the study including the pilot which will follow the GDPR requirements will explain that the data collected from this study will not contain any person identifiable data other than which of the four science domains the respondent works in, gender and occupational level i.e. PhD student, non-tenured or tenured scientist and so on. Individual data will neither directly or indirectly be identifiable. The data collection of the actual survey in Yr 2 will be outsourced to a trusted experienced and respected research organization in Netherlands which will adhere to information security norms ISO 27001, 27002, and NEN 7510, and fall under European and Dutch data protection laws. An attorney from the Innovation Exchange Amsterdam (IXA) (see http://www.ixa.nl/en/home) will compose a processing agreement with the selected organization following a standard format.
Further to this, we will also ensure the following for the survey and its piloting:
- Collecting and storing the data on a protected server in the Netherlands.
- Automatic immediate removal from the recorded data of the personal URLs used by respondents to access the survey.
- None of the researchers will at any time have access to data files in which respondents personal data and their answers have not yet been delinked and anonymized. At no point in the study will any of the researchers be able to repair the delinked data.

**2.4 I will stick to the privacy regulations of my organisation**

- Yes

## 3. Making data findable

**3.1 The data collection of my project will be findable for subsequent research (note: this is a key item, which you should report to ZonMw at the end of your project).**

- Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)
- Yes, it can be found through an online (metadata) catalogue or web portal (please specify)

We archived all the protocols and research information on the Open Science Framework (https://osf.io/dp6zf/). OSF is an international online public repository focused on
storing scientific/scholar documents.
At the same time we created a website with the information related to the project and
advancements. Please see www.nsri2020.nl

**3.2 I will use a metadata scheme for the description of my data collection.**

- Yes, I will select a metadata scheme from the list published by Datacite (please specify)

We will use a metadata scheme from Datacite as the DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes.
We will also use Digital Academic Repository of VU (VU-DARE) data scheme in the 2nd and 3rd year once we have written and approved scientific papers, as it is our University main repository.
VU-DARE is part of the national NARCIS archive, an initiative of the Dutch universities to publish their research results in open access as much as possible.
Once the results from the National survey on research Integrity (NSRI) are gathered for anonymization purposes there will be a coding file. The coding file with answers won't be accessible to secondary users. for privacy purposes the data will be destroyed after analysis have been made.
We will also use The Data Documentation Initiative (DDI) as it is designed to document and manage data across the entire life cycle, from conceptualization to data publication and analysis and beyond.

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

- Yes, I will be using the DOI code

Once we have the results of the National Survey on Research Integrity (NSRI) we will use TU Delft Library as it has established DataCite Netherlands to serve as a central research data-set registration service that helps Dutch research organizations register research data-sets and assigns Digital Object Identifiers (DOIs) to them.

## 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)

The data collected from this study has a detailed implementation plan in order to meet the project's objectives which are:
1. To report on the occurrence of Questionable Research Practices (QRPs) as well as their associations with explanatory variables across the four specific science domains.
2. To translate this knowledge into specific action plans which will be disseminated to the participating institutions. All data generated as a result of fulfilling these objectives including data from the pilot phase will be presented at international and national conferences and published in open access renowned journals. All data which has been anonymized will be made available upon request. Because of the size and gravity of the project, we have a Steering Committee that will decide when interested researchers can access the data-set. We will make the information available about how to contact the Steering Committee together with the data-set in the Open Science Framework (OSF) and DANS repository.
The embargo period will be no longer that 3 months before it is publicized.

**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)

Data resulting from the study and pilot preceding the full study will be made available upon request to the project leader. Even though the original data-set will be anonymized as described in section 3, given the sensitivity of the research topic and importance of ensuring respondents feel fully confident that their responses are well protected, we prefer to only provide the data-set upon request.

**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, which you should report to ZonMw at the conclusion of your project).**

- Not yet, my institution will draft a set of terms of use with the help of a legal advisor

**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
- The sharing of data for commercial purposes, taking into account the provisions of state aid law
- Whether or not the data set may be linked with another data set (for reasons of privacy)
- 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

## 5. Making data interoperable

### 5.1 I will select a machine actionable data format, which will allow other researchers and their computers to read my data collection.

- Yes (please specify)

The data will be collected by a professional data management organisation for the actual survey in yr 2 and made available to the project team for analysis in cvs or SPSS.sav format which is the format we will likely make available for further use. The data will be collected for the pilot in-house and likely made available in a similar format.

### 5.2 I will select a metadata standard to allow my data collection to be linked to other collections (note: this is a key item, which you should report to ZonMw at the conclusion of your project).

- Yes, I will select a metadata standard from the list published by Biosharing (please specify)

We will use the metadata standard from Biosharing/Fairsharing (specifically the Data Documentation Initiative (DDI) [DOI:10.25504/FAIRsharing.1t5ws6] as this fits our project better.

### 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.

- No (please explain)

The survey data will be fully anonymized hence will not allow for combination with other data sets.

## 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)

The research process has been documented in the project proposal and will be further detailed in each of the resulting publications in peer reviewed journals of this study. In addition to this, a respected and well known data collection organization will manage the collection of the data for the survey aspect of this study. All processes pertaining to this will be documented and reviewed by the project leader and/or steering committee as required to ensure the process of data collection, quality assurances, and checks are in place. In the first year of the project the survey will be extensively piloted including aspects relating to how quality assurance will be maintained and checks on the data quality conducted. All software and methodology (e.g. code manual) used will be documented in the preregistered study & pilot protocols as well as in the methods sections of all resulting publications.

### 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)

- Some or all of the data must be destroyed once the project has ended, because of a contract or law

To help us determine which research data constitute valuable source material for further research, we will consult the DANS checklist that sets out 'general guidelines for the selection of research data for storage' as a guidance. We will also ask our data manager at Amsterdam UMC.

**6.3 Once the project has ended and the data has 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)

The web based survey is expected to recruit 30-40,000 respondents in all universities and university medical centers across The Netherlands across 4 science domains. In addition to this there will be a smaller pilot in the first year and 5 focus groups of about 10 participants each in the second year. To maximize response rate we intend for the web based survey to be no longer than 15-20 minutes with a likely maximum of 20-25 open and/or closed ended questions. During the first year of this study, we will select the QRP's and explanatory variables to be included in the survey. At the end of this first year we expect to have piloted the survey and finalized the selection of desirable research practices and explanatory variables as well as final number of questions that will be included.
We consider that the size of the data collection (in GB/TB) is not an important element for our project development, but we might be able to specify the size once we have created the Survey, however the estimated number of respondents will determine the size of the data files.

**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 has a data seal of approval (please specify the archive)
- Not yet

We consider DANS as our data repository since it has a data seal of approval and it would be appropriate for our project anonymizing personal data of National Survey on Research Integrity (NSRI) respondents.

**6.5 Once the project has ended, I will uphold the recommended data preservation period of at least 10 years.**

- Yes, in accordance with VNSU guidelines (please specify the number of years)

we will comply with at least 10 years

**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)

Based on preliminary quotations from a number of highly professional organizations the costs of this service will not exceed € 25.000 excluding tax. To be able to choose the organization that will deliver the highest quality and highest levels of confidentiality and to account for 21% tax we reserve € 30,250 of the research budget for the survey data collection. As for the data archiving costs, these have not yet been estimated and will depend on the costs for archiving at DANS.

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

- Yes (please elaborate)

see 6.6 above.

# National Survey on Research Integrity - Datamanagement ZonMw

### 1. Kenmerken van het project en de dataverzameling

#### 1.1 Contactgegevens projectleider

Question not answered.

#### 1.2 Ik heb mijn DMP opgesteld in samenwerking met een expert op het gebied van datamanagement. Noem naam, functie, organisatie/afdeling, telefoonnummer, e-mailadres.

Question not answered.

#### 1.3 Bij het verzamelen van de data voor mijn project ga ik als volgt te werk:

Question not answered.

#### 1.4 Ik maak binnen mijn onderzoek gebruik van:

Question not answered.

#### 1.5 Ik ga bestaande data hergebruiken en/of koppelen en ik heb toestemming van de data-eigenaar/eigenaren voor het gebruik/ koppelen van zijn/hun data.

Question not answered.

#### 1.6 Bij het verzamelen van nieuwe data werk ik samen met andere partijen.

Question not answered.

#### 1.7 Het project voer ik uit in een consortium met twee of meer partners. Binnen het consortium zijn duidelijke afspraken gemaakt over datamanagement en intellectueel eigendom. (Denk hierbij ook aan wat het effect is van veranderingen binnen het consortium voor datamanagement en intellectueel eigendom)

Question not answered.

#### 1.8 Ik kan een inschatting maken van de omvang van de dataverzameling, namelijk het aantal deelnemers of subjects ("n=") van de dataverzameling en de grootte in giga-/terabytes.

Question not answered.

#### 1.9 De volgende eindproducten van het project stel ik beschikbaar voor vervolgonderzoek en verificatie. (licht kort

**toe)**

Question not answered.

**1.10 Gedurende het project heb ik voldoende opslaglocaties en -capaciteit en heb ik een back-up van de data beschikbaar. (Geef een korte toelichting.)**

Question not answered.

## 2. Wet- en regelgeving (incl privacy)

**2.1 Ik ga mensgebonden onderzoek doen en ik ben op de hoogte van en mij houd aan de wet- en regelgeving betreffende privacygevoelige data**

Question not answered.

**2.2 Ik ga mensgebonden onderzoek doen en ik heb geregeld dat ik de data verkrijg met (een vorm van) toestemming van de deelnemers.**

Question not answered.

**2.3 Ik ga mensgebonden onderzoek doen en ik ga privacygevoelige data beschermen tegen datalekken en/of misbruik.**

Question not answered.

**2.4 Ik houd me aan het privacyreglement van de organisatie waaraan ik verbonden ben.**

Question not answered.

## 3. Data vindbaar maken

**3.1 De dataverzameling uit mijn project is vindbaar voor vervolgonderzoek. (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven.)**

Question not answered.

**3.2 Voor de beschrijving van de dataverzameling gebruik ik een metadataschema.**

Question not answered.

**3.3 Ik zal gebruik maken van een Persistent Identifier (PI) om duurzaam naar het databestand te verwijzen. (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven.)**

Question not answered.

## 4. Data toegankelijk maken

**4.1 Na afloop van het project zullen de data toegankelijk zijn voor verificatie en vervolgonderzoek.**

Question not answered.

**4.2 Na afloop van het project wordt het databestand openbaar toegankelijk, zonder aanvullende voorwaarden (open access).**

Question not answered.

**4.3 Ik heb gebruiksvoorwaarden beschikbaar waarmee ik de voorwaarden voor toegang tot mijn databestand uitleg na afloop van het project. Geef een link of Persistent Identifier (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven.)**

Question not answered.

**4.4 In de voorwaarden die ik stel aan het gebruik van mijn data (restricted access), heb ik in ieder geval de hieronder aangekruiste punten opgenomen.**

Question not answered.

## 5. Data interoperabel maken (uitwisselbaar, koppelbaar)

**5.1 Ik kies een dataformat zodat mijn data leesbaar zijn voor andere onderzoekers en hun computers ('machine actionable').**

Question not answered.

**5.2 Ik kies een metadatastandaard zodat mijn data gekoppeld kunnen worden aan andere data. (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven).**

Question not answered.

**5.3 Ik ga mensgebonden onderzoek doen en heb bij de privacybescherming rekening gehouden met hergebruik van de data en eventuele koppeling met andere databestanden.**

Question not answered.

## 6. Data herbruikbaar maken en duurzaam opslaan

**6.1 Ik zorg voor een goede kwaliteit en documentatie van de data opdat andere onderzoekers ze kunnen interpreteren en gebruiken ('replication package').**

Question not answered.

**6.2 Ik heb selectiecriteria om aan het eind van het project te bepalen welk deel van de data moet worden bewaard. (zie uw antwoord op vraag 1.9)**

Question not answered.

**6.3 Aan het einde het project kan ik, na de selectie van de data, een inschatting maken van de omvang van het databestand (in Gb/Tb) dat ik voor lange termijn ga opslaan/ archiveren.**

Question not answered.

**6.4 Ik zorg dat ik aan het eind van het project een keuze heb gemaakt voor een archief of repository voor duurzame lange termijnarchivering (gecertificeerd) van mijn databestand. (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven.)**

Question not answered.

**6.5 Na afloop van het project, zal ik voor mijn data de aanbevolen bewaartermijn van minimaal 10 jaar hanteren.**

Question not answered.

**6.6 De kosten voor datamanagement gedurende het project en de voorbereiding van archivering mogen op de projectbegroting. De kosten zijn:**

Question not answered.

**6.7 De kosten van lange termijn archivering van de dataset, na afloop van het project, zijn gedekt.**

Question not answered.