

---

# National Survey on Research Integrity

*A Data Management Plan created using DMPonline*

**Creators:** Gowri Gopalakrishna, Sara Behrad, PLEASE UPDATE YOUR DETAILS, PLEASE UPDATE YOUR DETAILS, PLEASE UPDATE YOUR DETAILS

**Affiliation:** Vrije Universiteit Amsterdam

**Funder:** ZonMw (Netherlands)

**Template:** Datamanagement ZonMw

**Grant number:** 445004002

## **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' (DRPs): 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 DRPs 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 DRPs, employed small samples, or focused on specific scientific disciplines. It is also unclear which strategies should be employed to decrease the occurrence of DRPs. 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 DRPs 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 DRPs in the Netherlands, b) thorough domain-specific comprehension of the associations between explanatory variables and engagement in DRPs, and c) domain-overarching and domain-specific shared understanding of the roles that each of the stakeholders in science can play in reducing DRPs and fostering RRP. 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 DRPs 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 DRPs and thereby foster RRP 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 DRPs and thereby foster responsible research practices.

**Last modified:** 02-09-2019

**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

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

- 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](mailto:k.onnasch@vumc.nl)

- 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.

- 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.

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

This study will generate new data.

- 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.

- 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.ix.nl/en/home.html>) will compose a processing agreement with the selected organization following a standard format. See also 2.3 for more information

- 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 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.

- Other products (please specify)
- 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 anonymised. We plan to also include SPSS syntax (from the analyses) for reproducibility. Legal information, code book, validation procedures.

- Yes, I will make use of my institution's standard facilities for storage and backup of my data

## 2. Legislation (including privacy)

- 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.

- 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.

### Participant 's consent form:

- I have read the information letter. Also, I could ask questions. My questions have been adequately answered. I had enough time to decide whether I wanted to participate. I will treat the information from other participants in this focus group as confidential.
- I am aware that participation is voluntary. I also know that I decide to no longer participate in this focus group at any moment without having to provide a reason.
- I consent to the collection, use and reuse if necessary of my data for the purposes stated in the information letter. I give permission to keep my data for 2 years after this investigation.
- I want to participate in this study.
- I give permission to approach me again after this investigation for a follow-up investigation

- 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.ix.nl/en/home.html>) 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

- Yes

## 3. Making data findable

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

We archived all the protocols and research information in the DANS (using their metadata scheme), Dataverse.nl, as well as the Open Science Framework (OSF.io). This last one, 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](http://www.nsri2020.nl)

- 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.

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.

- Yes, I will be using the DOI code

Once we have the results (data collection) of the National Survey on Research Integrity (NSRI), we will use the DANS and DataCite Netherlands to serve as a central research datasets registration services that helps Dutch research organizations register research datasets and assigns Digital Object Identifiers (DOIs) to them.

#### 4. Making data accessible

- 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 DRPs as well as their associations with explanatory variables across the four specific science domains, and 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 dataset. We will make available the information about how to contact the Steering Committee together with the dataset in the Open Science Framework and DANS repository. The embargo period will be no longer than 3 months before it is publicized.

- 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 project leader. Even though the original data set will be anonymised as described in section 3, given the sensitivity of the research topic and importance of ensuring respondents feel fully confident their responses are well protected, we prefer to only provide the dataset upon request.

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

- 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.

- No (please explain)

Given that the type of data that we will collect is not yet under any metadata standards, we will deliver a Data model with explanations about the variables used in this research. We will use a code book and make it available in the Open Science Framework for any one that needs to use it.

- No (please explain)

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

## 6. Making data reusable

- 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 pre registered study & pilot protocols as well as in the methods sections of all resulting publications.

- 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.

We will only keep the anonymized data, and given the response technique used: Randomize-response technique, plus our privacy policy, we will destroy any personal information, such as name, email address, IP address, etc.

For the raw data we will create a read only version for the future reuse of data.

We will comply with the GDPR and AVG regulations.

- 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 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 DRP 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 which desirable research practices and explanatory variables as well as final number of questions that will be included.

We anticipate a response rate of about 20% for the main web-based study, but this is only an approximation that could change due to response rate. In the near future we might be able to specify the data size, once we have created the Survey.

Based on this size the size of the data collection won't be bigger than 1 TB.

- Yes, and this archive has a data seal of approval (please specify the archive)

We consider DANS as our data repository since it has a data seal of approval and it would be appropriate for our project.

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

we will comply with at least 10 years

- 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 at DANS, this together with the expenses for the additional management costs are already covered by the project's budget.

- Yes (please elaborate)

see 6.6 above.