Plan Overview

A Data Management Plan created using DMPonline

Title: Research Transparency Check

Creator: Rene Bekkers

Principal Investigator: Daniel Lakens, Rene Bekkers

Data Manager: Rene Bekkers

Project Administrator: Rene Bekkers

Affiliation: Vrije Universiteit Amsterdam

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: 1 - VU DMP template 2021 (NWO & ZonMW certified) v1.4

ORCID iD: 0000-0002-0247-239X

ORCID iD: 0000-0002-4403-7222

Project abstract:

Project to develop software that provides an automatic assessment and suggestions for the improvement of the transparency of data and methods in research reports before they are published.

ID: 179914

Start date: 01-06-2025

End date: 31-05-2027

Last modified: 15-06-2025

Grant number / URL: ICT.001.TDCC.018

Research Transparency Check

0. General information

0.1 Document version & date

Version 1.0
Date: 10.06.2025

0.2 Project title

Research Transparency Check

0.3 Project summary

We propose to develop Research Transparency Check, including software that provides an automatic assessment and suggestions for the improvement of the transparency of data and methods in research reports before they are published. Research Transparency Check serves people at universities and research institutes, academic journals, and funding agencies. We engage seven data communities for the development of transparency criteria, and checklists for documentation of these criteria specific to the most common sources of data used in the Social Sciences and Humanities.

0.4 At which VU Faculty is this project situated?

- Faculty of Social Sciences (FSW)

0.5 Your contact details

Full name: Rene Bekkers
Contributor Roles: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing
Telephone number: 020-5986493
Email: r.bekkers@vu.nl
ORCID: 0000-0002-4403-7222
University: VU Amsterdam
Faculty/Institute: Faculty of Social Sciences
Department: Sociology

0.6 List other people involved, including those at partner organisations in the project (if applicable)

Co-PI: Daniel Lakens, d.lakens@tue.nl
ORCID: 0000-0002-0247-239X
Technical University Eindhoven

Programmer/analyst: Cristian Mesquida Caldenty, c.mesquida.caldentey@tue.nl
ORCID: 0000-0002-1542-8355
Technical University Eindhoven

Programmer/analyst: Lisa DeBruine, lisa.debruine@glasgow.ac.uk
ORCID: 0000-0002-7523-5539
University of Glasgow

0.7 Funding organisation & grant number (if applicable)

Funding organisation: NWO
Grant number: ICT.001.TDCC.018

0.8 Project code (if applicable)

NA

0.9 Consulted data management expert(s)

Name: Emily Barabas
Email: e.k.barabas@vu.nl

1. Data description

1.1 Will you collect and/or process personal data in this project?

- Yes

1.2 Will you use existing data? If yes, what is their source?

Name source: public OSF projects
Conditions for reuse: none
Reference to dataset: NA
URL of the dataset: https://osf.io

Name source: public OSF preprints
Conditions for reuse: none
Reference to dataset: NA
URL of the dataset: https://osf.io/preprints

Name source: public GitHub repositories
Conditions for reuse: none
Reference to dataset: NA
URL of the dataset: https://github.com/

Name source: OpenAlex
Conditions for reuse: none
Reference to dataset: NA
URL of the dataset: https://openalex.org/

1.3 Will you collect or produce new data? If yes, please describe how.

We organize a series of deliberations within communities of researchers using seven different data sources: synthetic data, registers, observations, surveys, news and social media, interviews and focus groups, and participant observation. We do not record audio or video of the meetings.

With each of these data communities, we create a rank ordered list of preferences for data transparency indicators through structured group discussions and online surveys.

Group discussions: we create a shared document to collect a longlist of candidate transparency indicators and arguments why they are important for what purpose.

Online surveys: we create anonymous surveys inviting participants to rank order the indicators emerging from the group discussions.

Research quality indicators: we produce seven FAIR datasets of research quality indicators, one for each data community, including descriptions of transparency criteria as well as examples of good documentation of these criteria provided by data community members.

1.4 Describe the population/participants/subjects that will be studied

Participants in the group discussions and online surveys are researchers, librarians, and data support staff in the social sciences and humanities interested in transparency.

1.5 Do you process any of the following (personal) data?

- Contact details
- Name

Names and email addresses may be used to recruit participants.

1.6 Do you process the personal data based on informed consent?

- Yes, using digital consent

In online surveys and in focus groups

1.7 On what legal ground will the data processing take place if it is not based on informed consent?

- Not applicable, I use informed consent

NA

1.8 Does the data collection include any of the following types of personal data?

We do not plan to collect these types of personal data.

1.9 If your research involves special categories of personal data (previous question) and you will not use explicit informed consent, what is the legal ground for the exemption?

NA

1.10 What kinds of outputs will you produce in this project? Please describe these data assets.

Raw data

Data asset: candidate transparency indicators from group discussions

Description: a machine readable text file will contain information about preferences for transparency indicators from group discussions

Format: .txt

Data asset: opinions on transparency indicators in online survey

Description: machine readable comma separated datafiles will contain information about preferences for transparency indicators from online surveys.

Format: .csv

Data asset: ground truth for transparency indicators

Description: machine readable comma separated datafiles will contain human coded information on transparency indicators to be used for validation purposes

Format: .csv

Data asset: tagged research reports

Description: machine readable XML datafiles will contain tagged sections of preprints and other publicly available research reports

Format: .xml

Processed data

Data asset: opinions on transparency indicators in online survey

Description: machine readable comma separated datafiles will contain information about the data practices and opinions of participants in the online surveys.

Format: .csv

Data asset: transparency indicators

Description: machine readable comma separated datafiles will contain information about research quality indicators.

Format: .csv

Data asset: knowledge graph

Description: persistent identifiers and data transparency assessments

Format: .json and .rdf

Other materials

Data asset: codebooks

Description: we will produce codebooks files for all data files.

Format: .pdf

Data asset: read-me files

Description: we will produce read-me files for all data files.

Format: .txt

Data asset: analysis syntax

Description: we will make analysis syntax producing the results.

Format: .r and .do

1.11 How much digital data storage will your project require?

- 0 - 50 GB

We will not collect or store audiovisual information.

1.12 Will you collect physical data? If yes, please describe these.

No

1.13 Will you take measures to ensure data quality? Please describe these, if applicable.

By sharing the documents produced in the deliberations with the data communities, all participants can check them.

2. Legal and ethical requirements, codes of conduct

2.2 What legislation applies to your research project? Please tick the relevant boxes for your project.

- General Data Protection Regulation (GDPR)/Algemene verordening Gegevensbescherming (AVG)

2.3 Do you require approval of an ethical committee for this project? If yes, please indicate which ethical committee and whether you have obtained approval for this project.

- No

Data collection and analysis at the School of Social Sciences is subject to an ethics check, which can result in the requirement to obtain advice from the Research Ethics Review Committee if ethics issues arise. I conducted the check on 10 June 2025 (#2025-6-10-420), based on the plans in the proposal posted at https://osf.io/cpv4d. The result of the check is that full ethics review is not necessary.

2.4 Will you work with data for which intellectual property and/ or confidentiality are an issue? If yes, please describe.

- No

Our commitment to open infrastructure sets Research Transparency Check apart from commercial initiatives. Research Transparency Check works exclusively with open materials. We will not use confidential data. The data communities we engage with are open, the dataset that we use and produce are open, and the software we build is open.

2.5 Do you plan on generating a marketable product from your research project? if yes, please describe

- No

3. Storage and back-up during the research process

3.1 What measures will you take to secure and protect data during the research process? Please describe, for each separate data asset you described for question 1.10, how you will ensure data security, where the data assets are stored & backed up, and who has authorization to access the asset.

Existing data

Will not be stored or archived, and instead be queried with APIs.

Raw data

All data assets:
Storage: Google Drive (GD)
Backup: Automated daily backup (GD)
Access: Authorizations (GD)

Processed data

Data asset: opinions on transparency indicators in online survey
Storage: Open Science Framework (OSF)
Backup: Automated daily backup (OSF)

Data asset: transparency indicators
Storage: Open Science Framework (OSF)
Backup: Automated daily backup (OSF)

Data asset: knowledge graph
Storage: Open Science Framework (OSF)
Backup: Automated daily backup (OSF)

Other materials

Data asset: codebooks
Storage: Open Science Framework (OSF)
Backup: Automated daily backup (OSF)

Data asset: read-me files
Storage: Open Science Framework (OSF)
Backup: Automated daily backup (OSF)

Data asset: analysis syntax
Storage: Open Science Framework (OSF)
Backup: Automated daily backup (OSF)

3.3 Which tools are used in the collection, processing or storage of data during research?

- Other (please specify below)
- Microsoft Teams
- R (software) *

- Python (software) *
- Open Science Framework (OSF)

Google Drive (VU Amsterdam account)

3.4 What other tools or software do you intend to use during your research?

Name: GitHub
Role: Software repository
Country: United States of America
URL: https://github.com/

3.5 Is it necessary to transfer the (physical or digital) data assets to other locations or research partners? If yes, please describe how you secure the file transfer.

- No

3.7 Do you transfer personal data outside of the European Economic Area (EEA)? If Yes, please provide additional information

- No

4. Data archiving and publishing

4.1 Which data assets will be archived and which will be published?

Raw data
All raw data will be archived.
Data asset: data practices and opinions in online survey

Processed data
All processed data will be published.
Data asset: data practices and opinions in online survey

4.2 Where will you archive your data assets?

Data asset: data practices and opinions in online survey
Archived: Yoda

4.3 What other archive(s) do you intend to use to archive data assets?

Name: Open Science Framework (OSF)
Role: Data repository
Country: Germany
URL: https://osf.io/pbkun/

4.4 For how long will the data be available in the archive?

Indefinitely

4.6 Where will you publish your data assets?

Open Science Framework (OSF), https://osf.io/z3tr9

4.8 How will you ensure your dataset gets a persistent identifier (e.g. a DOI-code)?

A DOI will be assigned automatically by the OSF for each datafile deposited.

4.9 Will you register your datasets in an online registry other than PURE? If yes, where?

OSF

4.10 Are there restrictions to data publishing? If yes, please specify the reasons and list the data assets you do not wish to share publicly.

No

4.12 When will you share the data? If not immediately after completion of the project, please specify the reasons.

As soon as a datafile is available

4.13 Please indicate the license and/ or terms of use under which you share your data.

CC0 for public data; a restricted license for data archived in Yoda.

5. Documentation

5.1 What documentation will accompany the data?

Each processed data file will be provided in a data package, including

1. a read-me file documenting the contents of the package and containing instructions for users;
2. the processed data file;
3. a codebook documenting the variables (names, labels and measurement properties).

5.2 What metadata will accompany the data?

Datacite

5.3 What methods, software or hardware are needed to access and use your data?

As much as possible, we provide data assets in open formats (.csv, .txt, .pdf) that do not require specific and proprietary software.

6. Data management responsibilities and resources

6.1 Who will be responsible for management of the data assets during the project? Please specify their name, position, role in the project, and faculty/ institution/ group.

The PI, René Bekkers, professor of philanthropy at the Department of Sociology at VU Amsterdam, will be responsible for data management during the project. Email: r.bekkers@vu.nl. ORCID: 0000-0002-4403-7222.

6.2 Who will be responsible for management of the data assets after completion of the project (e.g. the project lead/ dedicated data manager/ department head)? Please specify their name, position, role in the project, and faculty/ institution/ group.

The PI, René Bekkers, professor of philanthropy at the Department of Sociology at VU Amsterdam, will be responsible for data management after the project. Email: r.bekkers@vu.nl. ORCID: 0000-0002-

4403-7222.

6.3 For data that are only available upon request, what methods will be used to handle requests for access and how will data be made available to those requesting access?

NA

6.4 What resources (for example financial and time) will be dedicated to research data management? Please estimate their cost.

Data management tasks are included in the PI research time on the project.
We publish data and preprints on the OSF, for which VU Amsterdam has an institutional license that does not incur costs on this project.