A Data Management Plan created using DMPonline.be

Creators: Vera Hoorens, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Vera Hoorens, n.n. n.n.

Data Manager: Vera Hoorens, n.n. n.n.

Project Administrator: Vera Hoorens

Grant number / URL: G053623N

ID: 197155

Start date: 01-01-2023

End date: 31-12-2028

Project abstract:

Verbal messages transmit beliefs about social groups (stereotypes) not just through their content, but also their wording. An important aspect of that wording is the messages' comparative format – the extent to, and the manner in which they emphasize an intergroup comparison. This project examines how the valence and the distribution of properties in and across social groups shape communicators' selection of comparative formats. It also examines how comparative formats shape observers' translation of verbal claims into distributions of the involved properties. Finally, it examines how both processes jointly contribute to the transmission of stereotypes. Through seven social psychological experiments, the project tests three competing hypotheses, derived from theories and findings from social and cognitive psychology, linguistics, and philosophy. The project will clarify processes that are involved in intuitive truth perception and show why claims that are in many respects comparable provoke such different truth judgments. It will also raise awareness among policy makers and journalists of the role of formal aspects of communication in the transmission, perpetuation, and sharpening of stereotypes, and show to what extent and how the strategic use of comparative formats may be an instrument in efforts to mitigate problematic stereotypes.

Last modified: 05-03-2023

Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects ...) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

We added the information below to the DMP Online platform after FWO had granted our application (as DMP Online was launched after the submission date). Rather than literally copying the information from our original application, we here report an updated version on the basis of the actual DMP. This information deviates from the information in the grant application due to, e.g., progressive insights in recommended data types.

All data will be newly generated.

The project will generate, per study:

- Experimental materials (Informed consent, protocol, debriefing)
- Ethical/GDPR clearance application file (& approval file)
- Preregistration file
- · Experimental data, plus a codebook to accompany the file
- Syntax file
- Output file
- [Depending on participant recruitment: possibly file with contact information]

The project will also generate (per combination of studies):

- Publication manuscripts
- Conference presentations
- Outreach (dissemination) activities

The experimental data file [& possibly the file with contact information] will include personal data

File types:

- Original files: .docx (texts), .sav (experimental/longitudinal data), .sps (statistical syntax), .spv (output of statistical analyses), .xlsx (contact information), .ppx (presentations)
- Portable types: .pdf (texts, presentations), .txt (texts), .dat (experimental/longitudinal data), .csv (experimental/longitudinal data), .rtf (syntax)

Size: < 500GB

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

The information below is an updated version of the information in the grant application. The only change is that it is more detailed

- 1. Designation of responsible person (If already designated, please fill in his/her name. Vera Hoorens Head of LESP & PI
- Storage capacity/repository as per KU Leuven policy (access between brackets):
 - 1. During
 - 1. KULeuven OneDrive for Business (research team)
 - 2. Shielded space on J:-drive KU Leuven (research team)
 - Personal device of PIs (involved PI)
 - 2. During & After, duration:
 - Preregistrations: OSF Frankfurt (public), in principle indefinitely
 - 2. Ethical/privacy clearance documents (research team, SMEC & DPO): PRET-platform KULeuven, at least 10 years after the project
 - 3. After (public), duration after comma:
 - 1. Publications: Lirias Repository, in principle indefinitely
 - 2. Anonymized datasets, codebooks, syntax, output: KULeuven Research Data Repository (RDR), at least 10 years after the project

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

The information below is an updated version of the information in the approved grant application. The only change is that it is more detailed.

We do not intent to deviate from the principle. In fact, we will preserve all but one type of datasets for at least 10 years. The exception involves the _contact files. We will delete these as soon as possible.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Below, we literally copy information from the approved grant application.

We will gain participants' consent for storing data at least 10 years and sharing anonymous data. Identification information, if present, will be separated from the data immediately after data collection, and deleted after participants are rewarded. When data are collection through data collection platforms or participant panels, we will enter a written agreement that their managers will not disclose identities to us and that they cannot access the data. Students involved in the project will sign confidentiality agreements and only have access to variables relevant to, & during their tasks. Researchers' workfiles will be on encrypted drives. Datafiles on RDR will not include personal data.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Below, we literally copy information from the approved grant application.

We will submit an ethical application to the Social and Societal Ethics Committee KU Leuven (SMEC) and via that register the project/studies in KU Leuven's Register of Data Processing for Research and Public Service Purposes. For transparency, we will pre-register the studies on OSF. The data stored on RDR will have a persistent identifier and rich metadata, and come with a create commons license. Besides storing files as described above, we will submit the data underlying the results reported in specific papers as supplemental materials with those papers. Our readiness to share materials has already been shown (see http://www.psychfiledrawer.org/files/8291448480580more_less_replication).

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

• Not applicable

GDPR

Have you registered personal data processing activities for this project?

• Not applicable

FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

The project will include 7 studies. Studies 1-3 each have two parts (a & b). The materials will be construed, and ethical/juridical clearance sought per Study. Parts a and b of Studies 1-3 address different questions and will be conducted separately. Thus, they will yield separate files from the preregistration on.

Because the datasets will be identically structured across studies, we here describe the datasets once per type of dataset / study.

The names of the datasets are construed as follows:

ComSter_[study]_[type]

where

ComSter = acronym for the project

[study] = the number of the study within the project

[type] = the type of data

[type] = the type of data	1			1		,	
				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
ComSter_1_mat ComSter_2_mat ComSter_3_mat ComSter_4_mat ComSter_5_mat ComSter_6_mat ComSter_7_mat	Research materials (1 file/study): Information for informed consent Experimental protocol, incl. manipulations, instructions, stimuli, self-developed measures, references to published measures, software (incl. version), coding scheme for open answers Debriefing information	New	Digital	Textual and graphical Construed in MSWord by the research team (will for actual use in study entered in Qualtrics, Inquisit, or PsychoPy).	.docx, .pdf (original format + open format)	<100MB	
ComSter_1_eth_appli ComSter_2_eth_appli ComSter_3_eth_appli ComSter_4_eth_appli ComSter_5_eth_appli ComSter_6_eth_appli ComSter_7_eth_appli	Ethical/juridical clearance application (1/study), incl. informed consent form, information on participant recruitment & inclusion criteria	New	Digital	Textual Construed online on the submission portal of SMEC and downloaded by research team.	.pdf	<100MB	
ComSter_1_eth_appr ComSter_2_eth_appr ComSter_3_eth_appr ComSter_4_eth_appr ComSter_5_eth_appr ComSter_6_eth_appr ComSter_7_eth_appr	Confirmation ethical/juridical clearance (1/study)	New	Digital	Textual Prints of mails from Social and Societal Ethical Committee	.pdf	<100MB	
Note. Studies 1-3 include parts a & b, to be preregistered separately. ComSter_1a_pre ComSter_1b_pre ComSter_2a_pre ComSter_2b_pre ComSter_3a_pre ComSter_3b_pre ComSter_4_pre ComSter_5_pre ComSter_6_pre ComSter_7_pre	Preregistration of study on Open Science Framework (1/study)	New	Digital	Textual Construed in MSWord by research team	.docx, .pdf (original format + open format)	<100MB	
ComSter_1a_data ComSter_1b_data ComSter_2a_data ComSter_2b_data ComSter_3b_data ComSter_3b_data ComSter_4_data ComSter_5_data ComSter_4_data ComSter_5_data ComSter_6_data ComSter_5_data	Participants' responses to experimental tasks, process measures, suspicion probes, demographics, and individual differences measures if any (1/study)	New	Digital	Textual and numerical Responses of participants	.sav, .dat .csv (original format + open format)	<100MB	
ComSter_1a_code ComSter_1b_code ComSter_2a_code ComSter_2b_code ComSter_3a_code ComSter_3b_code ComSter_4_code ComSter_5_code ComSter_6_code ComSter_6_code	Codebook (1/[part of] study), incl. information on data collection procedures and events, explanation of variables in _data files	New	Digital	Textual, numerical Construed in MSWord by research team	.docx, .pdf (original format + open format)	<100MB	
ComSter_1a_syntax ComSter_1b_syntax ComSter_2a_syntax ComSter_2b_syntax ComSter_3a_syntax ComSter_3b_syntax ComSter_4_syntax ComSter_5_syntax ComSter_6_syntax ComSter_7_syntax ComSter_7_syntax	Syntax of data analysis, annotated (1/[part of] study)	New	Digital	Software	.sps, .txt (original format + open format)	<100MB	

ComSter_1a_output ComSter_1b_output ComSter_2a_output ComSter_2b_output ComSter_3a_output ComSter_3b_output ComSter_4_output ComSter_5_output ComSter_6_output ComSter_6_output	Output of statistical analysis (1/[part of] study)	New		Textual, numerical, graphical Output produced by applying syntax to data	.spv, .rtf .pdf (original format + open format)	<100MB	
Comster_1a_contact Comster_1b_contact Comster_2a_contact Comster_2b_contact Comster_3a_contact Comster_3b_contact Comster_4_contact Comster_5_contact Comster_6_contact Comster_6_contact Comster_7_contact	Contact information of participants, cf. question 2 (1/[part of] study)	New	Digital	Textual Observational	.sav, .xlsx, .csv (original formats + open format)	<100MB	
Comster_DRM	Data Management Plan	New	Digital	Textual Written in DMP Online & converted into text formats by research team	.docx, .pdf (original format + open format)	<100MB	
Comster_paper_1 Comster_paper_2 Comster_paper_3 Comster_paper_4 Comster_paper_5	Manuscript based on research (n=5)	New	Digital	Textual Written by research team	.docx, .pdf (original+open)	<100MB	
Comster_pres_[n] [n] = number	Conference presentation (exact number still unknown)	New	Digital	Textual / graphical Prepared by research team	.ppx, .pdf	<100MB	
Comster_outreach	Overview of dissemination activities (with links to media appearances or appendices)	New	Digital	Textual	.docx, .pdf	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Not applicable

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

Yes, human subject data

Yes, as we will collect personal data (in all studies: gender, age, educational level; more personal data in some studies, see below). We will apply for ethical and GDPR clearance from the Social and Societal Ethical Committee (SMEC) during the preparation stage of each study. We will appropriately update both the DMP and the Compliance Monitoring Tool form (E-2023-3782) with the relevant SMEC-numbers.

The datasets with the suffix _data will contain data obtained from experimental research on adult healthy volunteers, who will have given informed consent (opt-in approach). Studies 2-3 and 5-7, on distributions of features in, or claims about features of real-life groups, the features will include desirable and undesirable features. In Studies 3, 6, and 7, these features will be either stereotypical or counter-stereotypical. Although we will not present the information as scientifically valid (such that there will not be any deception), a careful debriefing will be in place to make clear that the items were construed for the purpose of the study and do not reflect actual distributions of features in groups. Adequate information about aftercare will need to be provided. We are not yet certain if we will need to create the datasets with the suffix _contact. If our participants will be from the iVox panel or from a platform like Prolific, we will not need to create them. If the circumstances at the time of the study force us to recruit participants directly (e.g., through a course where they participate for course credit), we will need to get in touch with participants directly to reward them for their participation. In that case, we will collect e-mail addresses, names, and/or participant codes (e.g., EMS numbers) and will create the _contact files by separating contact information from the original datasets. We will delete the _contact datafiles as soon as the reward is given.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

Yes

The datasets with the suffix _data will include participants' age, gender, and educational level. In the case of Studies 2-3 and 5-7, they will include membership of participants in stimulus groups (e.g., if claims are about smokers, we will ask if they smoke). Depending on the specific stimulus groups, these personal data may or may not fall under special categories of personal data (highly sensitive personal data). Because the stimulus materials are under construction, we cannot give more specific information about the nature of the to-be-processed sensitive personal data yet. This information will be included in the ethical/GDPR clearance applications and will be added to the DMP.

The datasets with the suffix _contact (if created) will also include personal data (e-mail addresses, names), possibly in an pseudonymized form (EMS-numbers, which only one member of the faculty, who is not a member of the research team, can identify).

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

No

Depending on the results of the research, the obtained data may have potential for valorization (e.g., in the form of trainings and consultancy for professional communicators). Leuven Research & Development (LRD) will be contacted for guidance and support with valorization.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

Yes

We will enter a collaboration agreement between KU Leuven and Friedrich-Alexander-University Erlangen-Nürnberg. The intellectual property rights of the datasets will per this agreement be shared by those members of the research team who contribute to them. As per FWO regulations, this agreement will be formalized in the first 9 months of the project.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

No

Not applicable

2 Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded)

We will organize the files in an overarching DRM folder called ComSter. It will include

- 1. A copy of this Data Management Plan (ComSter_DMP)
 2. A DRM folder per study (ComSter_[number of study]), to hold all files of the involved study (i.e., 7 folders with 1 Study each, folders of Studies 1-3 holding part a and b).
- 3. A publications folder (ComSter_Publication) holding _paper, _pres, _outreach files

Information at the level of each study

The DRM folder of each study will include a README file in .txt format, named Comster_[number of study]_README. To create the README files we will fill out the template on README files-Research Data Management (kuleuven.be). They will in any case include the project name, keywords, name of involved researchers and their ORCID ID, name of funding code, start- and end date of the project, DOIs of shared datasets, creative common license, the link to the preregistration on OSF, approval/registration code of ethical committee, links to publications. We will also add the version of the software used to analyze the experimental data.

For version control procedures of defining milestones (e.g., draft, final) and subversions (e.g., V1, V2,...) will be implemented.

Information at the level of the variables in datasets

Each dataset with a name ending on '_data' (= datasets with experimental data) will go accompanied by a codebook (dataset ending on '_code') that will be stored in the same folder (see Question 1). Because the '_data' and '_code' files will have identical names except for that part of the name and because they will be stored together, it will be clear which ones will belong together.

Information at the level of papers and conference presentations

For ease of cross-reference between publications and studies, the publications folder will include, per _paper and _pres file, a simple README file in .txt format, named ComSter_paper_[n]_ or ComSter_pres_[n], respectively. This file will include links to the studies whose data were used for that publication. If applicable, the file will detail which part of the data from a study were used for that particular publication.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reus

Because we will upload the datasets in KU Leuven's RDR, we will use the metadata standard used there: full-metadata-model (kuleuven.be).

We will for each dataset listed in the table under Question 1 fill out all required, recommended and optional fields except the following:

- alternative title field (there will not be an alternative title)
- alternative ID (there will be a single ID per dataset)
- geospatial fields (filling these fields out would require to obtain the coordinates from where people participate in our research, which we cannot do to preserve anonymity).

3. Data storage & back-up during the research project

Where will the data be stored?

All Datasets, insofar as they do not contain highly confidential information:

- In a shared folder on KULeuven OneDrive for Business.
- Storage facilities of the research unit at KU Leuven: In a shared folder on a J: drive.
- 3. Personal devices of PIs and PhD student working on the project.

The files above will only be accessible to the PIs and PhD student working on the project. The offline copies on personal devices will be on encrypted media only. Datasets which include highly confidential information (e.g., _contact files) will be stored on a shielded J: drive, which can only be accessed by the data manager of the research unit. Preregistrations (& other datasets during editorial phase of manuscripts, except _contact, _eth_appli, and _eth_appr): on the OSF server in Frankfurt. Ethical/juridical clearance files: on KU Leuven's Digital Privacy & Ethics Platform (PRET).

How will the data be backed up?

The automatic version management in OneDrive provides a backup of up to 100 versions. There will be 2 additional backups:

- 1. On a shielded network drive of KU Leuven (J: drive, with a daily automatic backup)
- 2. On an encrypted external hard disk owned by PI Vera Hoorens.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

Yes

The available capacity on OneDrive for Business, the J-drive and personal devices of the PI Vera Hoorens is more than sufficient for the storage of all datasets pertaining to it.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

OneDrive for Business is a Microsoft cloud solution to securely store documents and files. The files will only be accessible by the involved researchers. Only PI Vera Hoorens can provide team members access to the OneDrive folder of the project.

Access to the OneDrive for Business files and the files on the J:-drive will be protected by multifactor authentication with the KU Leuven Authenticator app.

All personal devices used for backups or on which the members of the research team will temporarily store data to conduct analyses offline will be Bitlocker-protected.

The links to OSF files will be read-only

The data will be pseudonymized by removing highly sensitive data and storing these data separately from the research data on the shielded J: drive at KU Leuven. Access to these files will be controlled by the data manager, with PI Vera Hoorens as a backup. All other researchers who are involved in the project will have access to the pseudonymized data only.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

No costs are expected.

All KU Leuven personnel has access to 2 TB of data storage for free on OneDrive. As the estimated sizes of the datasets <100 GB, sufficient storage and backup capacity is available.

PI Vera Hoorens possesses several hard disks with storage capacities that are large enough to hold the datasets of all her projects. In case additional hard disks are necessary or costs do arise, these will be covered with the obtained project funding.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All datafiles, except those with the suffix '_contact', will be retained for at least 10 years, in compliance with KU Leuven RDM policy. The files with the suffix '_contact' will be deleted as soon as participants have been rewarded for their participation (or, if the reward is financial, when the project is finished and due financial reporting to FWO has taken place).

Where will these data be archived (stored and curated for the long-term)?

Besides the KU Leuven OneDrive for business, all datasets will be stored and curated for the long-term on the KU Leuven institutional Research Data Repository (RDR)

Research Data Repository (kuleuven.be).

Research Data Repository (kuleuven.be).

The datasets that will be uploaded on OSF will in principle be archived indefinitely. However, the main archive for most files will be RDR. The main archive for papers and conference presentations will be LIRIAS.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs are expected because of the limited size of the datasets. If costs arise nonetheless, they will covered with the obtained funding for the project and/or the research group of PI Vera Hoorens.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

• Yes, in an Open Access repository

We will make the datasets for the relevant studies and of the following types publicly available: _pre, _mat, _data, _code, _syntax, _output.

Publications will be made available through LIRIAS (and possibly also through other repositories, such as ResearchGate). Funding permitting, we will maximally publish Open Access such that publications are also openly available through the involved outlets.

If access is restricted, please specify who will be able to access the data and under what conditions.

The datasets that are listed under the previous question (i.e., datasets of the following types: _pre, _mat, _data, _code, _syntax, _output) will become publicly available after the publication of the involved research. These datasets will during the editorial process already be available to editors/reviewers of the publication outlets to which we submit manuscripts, under the condition that they should be kept confidential and only serve to inform an editorial decision.

There are no further restrictions to share the pseudonymized data as the data do not contain highly sensitive information and participants will have given informed consent to share these data with other researchers.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

Yes, Privacy aspects

The datasets ending on '_contact' cannot be shared because of privacy issues.

The datasets ending on '_eth_appli' and '_eth_appr' will not be shared because the specific description of the intended samples may also create privacy issues.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

KU Leuven institutional Research Data Repository (RDR):

Research Data Repository (RDR) - RDR - Research Data Repository (kuleuven.be).

Publications will be made available through LIRIAS (and possibly also through other repositories, such as ResearchGate).

When will the data be made available?

The datasets (of the following types: _pre, _mat, _data, _code, _syntax, _output) will be made available after the involved studies have been accepted for publication. They will during and after the project thus gradually become available. At the end of the project, the data of unpublished studies will be made available after an embargo period of 5 years to give the research team enough time to consider publication after the end of the project.

These datasets will during the editorial process already be available to editors/reviewers of the publication outlets to which we submit manuscripts, under the condition that they should be kept confidential and only serve to inform an editorial decision.

Publications will become available on LIRIAS (or other repositories) as soon as the copyright agreement with the involved outlet allows. However, we will (funding permitting) maximally publish Open Access such that publications immediately become available.

Which data usage licenses are you going to provide? If none, please explain why.

CC Attribution 4.0 International

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

Yes

DOIs will be available through RDR, but are not yet available.

What are the expected costs for data sharing? How will these costs be covered?

RDR and LIRIAS are free for KU Leuven personnel, hence, no costs are expected for data sharing.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

The PhD researcher will manage the data, supervised by Pls Vera Hoorens and Susanne Bruckmüller

Who will manage data storage and backup during the research project?

The PhD researcher will manage data storage and backup, supervised by Pls Vera Hoorens and Susanne Bruckmüller

Who will manage data preservation and sharing?

During the project, the PhD researcher will manage data preservation and sharing. After the project, Pls Vera Hoorens and Susanne Bruckmüller will do so.

Who will update and implement this DMP?

The PhD researcher will update this DMP, but PIs Vera Hoorens and Susanne Bruckmüller bear the final responsibility