DMP Questions Quantitative data

This DMP is based on a fictional research project with basis in reality. For each topic of the DMP there are example questions and answers where applicable. The examples are not country specific. Some of the information are generic. Project is still running.

Administrative information

Version

DMP version 3, 2017-02-05 (Current DMP)	Funding for the project is decided and the project can start. DMP is completed with more details about data management during the project.
DMP version 2, 2016-10-11	After questions from funding agency the plan has been completed with more specific information about data sharing.
DMP version 1, 2016-09-23	First version, sent to funder agency.

Project name

State the name on the project.

Media, Opinions and Society 2017 (MOS 2017)

Project Description

Short description of the project. For example, the nature of the project, the research questions that are addressed and the purpose for which are data being collected or generated.

This survey is conducted in order to identify how the evolution of society affects attitudes and behaviour. MOS addresses three areas – mass media, opinion and society. The survey include a large number of questions related to media, society, politics and social background. In addition to the survey data, information of sex, age, income and education are completed to the dataset from national registries. The population is all individuals aged 18-35 years, resident within private households in the studied country (XX).

PI / Researcher (person, institution or organization)

Person, institution or organization that is responsible for the material and the intellectual content of the project? Name, telephone number, email contact details as well as organization. State researcher ID if possible, e.g. ORCID (http://orcid.org).

PI (Name, organisation, tel number, email, ORCID.)

Participating researcher and/or organizations

If there are other researchers and/or organizations involved in the project, state their name, telephone number and email contact details. If possible, state their roles in the project

The research group consists of:
PI (Name, organisation, tel number, email, ORCID.)
Project Coordinator (Name, organization, tel number, email)
Researcher X (Name, organisation, tel number, email, ORCID)
Researcher Y(Name, organisation, tel number, email, ORCID)
Statistician (Name, organization, tel number, email)

External partners:
The Survey Institute of YY are conducting the survey as well as collecting data from national registries.

Project Data Contact

Name, telephone number and email contact details (to contact under and after the project if needed).

During the project:Questions about the data or project, under the research project, is answered by the Project Coordinator (Name, organization, tel number, email).

After project: When data is delivered to the data repository, the PI is responsible for answering questions (Name, organization, tel number, email).

Owner of the material

Who owns the data and are responsible for the material?

University of...

Producer

Which organization has the administrative responsibility for the project? (This can be an institution at a university.)

Department of Social science, University of...

Related Policies

Are there any relevant policies/guidelines at the university/department, preferably with information about where those documents can be found, as well as the version used? Does your funder or university have a Research Data Management policy?

The National Research Trust requires that the data and its results are OA within one year after project ends. DMP is to be deposited together with data at the National Digital Repository for dissemination and preservation.

The university has a policy for IT security, that all research project at the university are required to follow. (More information, see Policy for IT version 3 2014-03-15, available at the university website)

Funder

State research funder if relevant. Later, also state the reference number of funding that has been granted.

The National Research Trust is funding the survey together with the Department of Social Science, Univ of...

Roles

Who is responsible for what within the research project? This includes naming the individuals responsible for data management in the research project as well as writing and updating the data management plan. If there are external partners, write down what they are going to do and their responsibilities.

The overall responsible for the DMP is assigned to the project leader (PI).

Data collection: The Survey Institute is responsible for conducting the survey. Details about the data collection will be described in a technical report.

Data management: Since several researchers are going to work in the project (analyzing data) there are principles for naming files, structuring maps, versioning datasets etc. The Research Project Coordinator is responsible to ensure that data

handles according to good practice, introduce researchers and answer questions related to the dataset, and also to prepare data for availability in the end of the project.

Ethics and Legal Compliance

Ethical review

Is there a need for an ethical review?

The research project is approved by the Ethical Review Board.

Privacy officer

If the data of the research project is going to have data that includes confidential information or information that requires informed consent, is there a requirement to notify a privacy officer?

The privacy officer at the University is informed about the project. Though, the data will be anonymized before the datafile is received from the Survey Institute.

Informed consent

Is informed consent required? Does it include aspects of how data is planned to be shared?

Informed consent will be sent to all participants together with the questionnaire. A copy of the informed consent will be available in the technical report and also in the project map at the University. The informed consent includes information about the plan for data sharing.

Protection of the identity of participants

How will the identity of the participants be protected?

The dataset delivered from the Survey Institute will be anonymized. That means that the dataset will not include any personal data. The dataset will not include any direct identifiers, ID-numbers och other types of identification codes. Controls for re-identification will be done by the Survey Institute, to ensure that there will not be any possibility to identify individuals in the dataset. If necessary, variables will be coded (e.g detailed geographic information can be coded to broader levels of geography, or income to income ranges or categories).

Access to un-anonymized data at the Survey Institute will be strictly limited. Only anonymized data will be transferred to the PI at the University.

Confidential information

Is there any confidential information within the material that requires special treatment and/or limits the access to it during/after the project? Describe the technical and procedural protections for information and how permissions and restrictions will be enforced.

The data collected from questionnaires will include personal data until the registry data has been linked to the dataset. The Survey Institute has a strict policy according to current legislation to ensure that respondents are protected. This includes technical, administrative and physical security. Before data are transferred to the University, all data are anonymized.

Intellectual property rights/copyright

Are there any copyright and/or intellectual property right issues to consider? Will permission be needed for the collection of the material that is going to be used? Name persons or entities who will hold these rights and note any constraints. Will these rights be transferred to another organization for data archiving and distribution?

Not applicable for this project.

Agreements with other organizations

Are written agreements with other organizations needed?

Agreement with the Survey Institute. A copy is available in the project map at the University (orig. archived at the University)

Restrictions

Will there be any access restrictions? Describe the access procedures for dissemination. Will access be open or granted only to specific user groups?

There will not be any restrictions regarding dissemination of the data. Dataset will be retrieved from the Repository after ordering dataset via the website.

Embargoes

Are there any embargo periods for all of, or parts of, the material?

Embargo period: The funder has accept that the research group can analyse data during the first 2 years (starting from the date when dataset is received to the university).

Data collection

Type of data

What kind of data are used in the project? Already existing data (e.g. administrative data, demographic data, public communication data) and/or will data be collected (e.g. survey data)?

The project will include both existing data from national registries (with information about sex, age, income, education) and data collected via questionnaires.
The questionnaire, that consists 55 questions, will be sent to 10,000 individuals. All data will be structured in one datafile (using SPSS) and the estimated size of the datafile will be 5-8 MB.

Existing data

Are there any relevant existing data material that can be reused in the project? If so, describe the scope, quantity and format of the material. Will any of that material be reused and if so how will it be integrated? If new data are collected even though there are data that can be reused, explain why it is important to collect new data.

In addition to the survey data, national registries with data on sex, age, income, education will be used. Registry data (as well as data collected from questionnaires) are gathered by the Survey Institute.

Data collection

Who is responsible for collecting the data? How will the questionnaire be sent out? (e.g. by post or through a web form?) and under which period? Will a reminder be sent out? What procedure will be used to combine the registry data with the survey data?

The Survey Institutet are responsible for collecting data, questionnaires as well as registry data. They are also responsible for linking the registry data with the collected data from questionnaires.

The sampling procedure is based on a simple random sample, from the register of the population. Data will be collected between Feb - Apr 2018. The questionnaires will be send out by post, with a reminder after 3 weeks. Registry data on age, sex,

education and income will be received during March 2018, and refers to data registered per 2018-01-01.

The questionnaires are marked with ID numbers. Each ID number can be linked to a personal number by using a key code, which enable to collect and add registry data for each individual in the dataset. The key code will only remain until registry data are collected. The technical report will include a full description of the data collection.

Method of data collection/creation

How will data be collected/created?

Data will be collected by self-administered questionnaires, and also from national registries.

Documentation and metadata

Documentation

A lot of information is produced during a research project describing methods used, decisions made as well as procedures during the data collection. What documentation will be created during the different phases of collection and analysis (e.g. logbooks, variable lists, protocols)? How will the documentation be structured? Consider what information will be needed to interpret and read the data in the future.

Survey Institute: The Survey Institute will create a technical report describing the data collection.
Research group: For documenting discussions and decisions during the project a logbook is used, and administered by the Research Coordinator. Principles for naming files, versioning datasets, structuring folders etc. are stated in a document called 'DM Principles' and are available in the project map. Curation of data in SPSS will be saved in a syntax file. For each new dataset, a variable list (including variable names, codes and content) will be created.

Metadata

Some of the documentation that are created during a project is structured information with purpose to explain, describe and localize data. This type of documentation can be described as metadata. Describe the metadata that will be provided with the collected/generated/reused data. What metadata will be produced on project-level (e.g. title, purpose, population, sampling procedure, time method etc.)? What metadata will be produced on data-level? Will it be possible to include metadata of each variable in the dataset, as a label? Are there any limitations in number of characters that can be used as a

To prepare data for dissemination the data management program, Nesstar Publisher, will be used. This is a recommendation from the data repository that are going to preserve and disseminate the data. By using the program, detailed information about each variable in the dataset can be added and a codebook including both metadata on project-level and data-level can be created (which are useful for a secondary user).

The research group are using SPSS, and the sav. file from SPSS can be imported to Nesstar. From Nesstar it´s possible to convert the dataset and documentation to several formats.

Metadata standard

What metadata standard(s) will you use and why? Data Documentation Initiative (DDI) is a metadata standard, commonly used for describing research data within social science. Will DDI, or any other metadata standard, be used?

Metadata will be provided by using an online deposit form at the data repository website. The metadata is provided in DDI. A description of the study and data will be published by the data repository, and available at the website. All metadata will be provided in English.

Terminologies, ontologies etc.

It is useful to use terminologies/ontologies or similar to describe and document the material. Will established terminologies/ontologies be used in the project? If not, will the created ones be mapped towards any established ones? For social science data, CESSDA topical classification can be used to categories the data. Will CESSDA topical classification or any other terminologies be used?

Keywords and topics generated from the data repository will be used. The keywords are based on the ELSST thesaurus, and topics on CESSDA topical classification.

Data management during the project

Folder structure

How will you structure and name your folders?

The folder structure is described in a document called 'DM Principles', available in the project map

Organizing your data

How will you organize data that are collected/re-used? Will the data be organized in simple files or more complex databases? Describe procedures for ensuring data quality during the project. How will all the answers from the respondents be entered into a dataset? Will the procedure be automatically or made manually?

The responses on the questionnaires will be automatically scanned into a program used by the Survey Institute. From the program a data cleaning section will be produced, maintaining a list of which cases that have been scanned, interpreted and if there are any issues. All issues and how those are handling are documented in the technical report. When registry data are linked with the data collected from questionnaire all cases will be anonymized. Data from the Survey Institute will be sent to the PI in a SPSS-file.

Research Data Management policy

Relate to the guidelines and policies regarding data protection or data security that exist at the university and define what it implies for the project. If there are different security levels defined, which one applies for the data of the project and what security measures must be taken to protect the data?

How will access to the data during the project be controlled and how will collaborators be granted access to the data in a secure way?

According to the IT-security principles, all information (in example data) will be classified. The information is classed by using the criterias; confidentiality, integrity, and availability. The PI and Research Coordinator are responsible for classifying the information and will be supported by the IT-staff if needed. (This section in the DMP will be updated).

File naming

How will you structure and name your files? Set naming conventions to be followed so that it will be easy to understand which file is which one.

All files that are created during the research project will be named in a consistent way, and include following information:
- Short description of content
- Version number (one-level structure e.g. V1, V2, V3)

- Date of creation, by using ISO 8601 (YYYY-MM-DD)

Example:

"Questionnaire V3 2017-06-10"

"Dataset from survey institute V1 2017-08-01"

"Technical report V1 2017-09-30"

File format

In what formats will the data be generated, maintained and in the end archived/made available? Is it possible to choose appropriate formats from the beginning of the project? If a statistical program will be used for the data, which one? Other files that will be produced under the project, in what formats will they be?

Data from the Survey Institute will be received as a .sav file, used for SPSS (Statistical Package for the Social Sciences), and also as a xml-file (Extensible Markup Language. Xml is a flexible format, which enables import of data and metadata into different information systems. Data will be preserved at the data repository according to best practices. Data, disseminated from the data repository, can be delivered in differents formats (SPSS, Stata, SAS, Excel).

Versioning

What conventions for versioning datafiles should be followed during the project? Who will be responsible for securing that a "masterfile" will be maintained, documented and versioned according to the project guidelines? How to separate different versions of a datafile?

Versioning principles are stated in a separate document, available in the project map. The versioning principles differentiates between substantive and non-substantive changes. To reflect this difference, version numbering system is based on a structure with two-levels (examples: 1.0, 1.1, 1.2, 2.0, etc.). To ensure that all researchers are using the latest version of data, all version changes are handled by the Research Project Coordinator.

Storage and backup

What physical resources and facilities will be used for the preservation and storage of the research data? Describe how and where data is stored as well as the backup procedures for the data. How will the data be recovered in the event of an incident?

Data will be stored in a secure environment at the university, with access restricted to the project members. All data are backed up twice a day, on the university server.

Budget

Staff

Estimate what resources that will be needed to document the data material during the project? Included here are cost for staff to manage, process and document the material during the project but also for the preparation of the material for long term preservation and possibly dissemination.

Costs for manage data during the project are included in the Research Project Coordinator activities and are included in the budget.

Hardware and software

Budget for possible costs to obtaining necessary hardware and software (for example systems for backup, security and software for documentation)

No costs related to hardware or software.

Storage

Budget possible costs which may occur to make storage possible.

Repository does not charge any costs for storing/disseminating the data.

Long-term preservation and archiving

Archiving and long-time preservation

Which data should be retained and/or preserved? Are there any data that must be retained/destroyed for legal, regulatory or contractual purposes? How long will the data be retained and preserved? Who is going to be contact person towards the repository?

A preservation plan will be created, with support from the archivist at the university. Data deposited to the data repository will be preserved for as long as possible (meaning as long as the data repository will exist or until any other changes in the policy or agreement changes).

The Research Project Coordinator will have handle the contact with the data repository.

Limitations (hardware & software)

Will a specific software or tool be required to use the data material? Is it possible to deposit the software and/or tool together with the data material and what extra documentation is required if that is the case? If it is not possible to deposit the software/tool, what is required to be able to use the data material?

A secondary user can order data from the data repository in different formats (SPSS, Stata, SAS, Excel).

File format for long-time preservation

What file format will be used for preservation and dissemination? Suitable formats should be commonly used, be well documented, open technical specification and not being proprietary.

Data will be preserved at the data repository according to best practices. Data, disseminated from the data repository, can be delivered in differents formats (SPSS, Stata, SAS, Excel).

Contact a data repository

If applicable: Contact a data repository to find out what can be done with the material to make a deposition easier at a later stage. Also, make clear who's responsible for the contact with the repository

A contact with the data repository has been established. The Research Project Coordinator are responsible for maintaining the contact during the project.

Data sharing

Making data available

Which data are of long-term value and should be shared? How will data be shared? When? Will the data be accessible via a repository, domain specific database, an institutional repository at the university or will the research group have a dedicated website that the group will create and maintain?

Data will be deposited at a data repository for long time preservation and dissemination.

Limitations

Will all data or only parts of it be published/made available? Are there any legal/ethical restrictions that prevents the publication of all the material? Will these restrictions mean that action must be taken before the material can be made available?

There will not be any need of restrictions. All data will be anonymized and can be available after the embargo time.

Delays

Is there a risk of delayed publication/making data available (all or parts of]? Explain why and what might be needed to do to avoid this?

No risks are identified.

Citation

Are there any specific requests of how data should be cited when reused?

The data repository have a suggested citation that will be used.

Persistent identifier (PID)

If possible, state whether or not the data material will receive a persistent identifier (PID)? If data is made available via a data repository a PID, for example a Digital Object Identifier (DOI), will in most cases be given the data material

The dataset, deposited at the data repository, will get a Digital Object Identifier (DOI).