Hogeschool van Amsterdam

URBAN VITALITY 'ARCHIVING PACKAGE'

Opmerkingen

1. Werk in uitvoering
2. Betreft archivering van computercode, data en bijbehorende beschrijvingen om resultaten te kunnen **verifiëren** of **repliceren**, en om data (eventueel) te kunnen **hergebruiken**
3. Betreft dus niet archivering van alle essentiële projectdocumentatie (i.e. het gehele onderzoeksdossier)
4. Gebaseerd op:
   a) HvA RDM richtlijnen (2018), artikel 8 t/m 12
   b) Nederlandse gedragscode wetenschappelijke integriteit (2018), artikel 3.3.25, 3.4.25, 3.4.45, 4.4
   c) Guidelines for the archiving of academic research for faculties of behavioural and social sciences in the Netherlands (2018)
   d) FAIR-principes (2016), zie bijlage
5. Hoewel voor WMO-plichtig onderzoek dezelfde principes gelden, kan het zijn dat de WMO of het richtsnoer GCP strengere eisen stelt. Dit is nog niet volledig onderzocht
6. In eerste instantie gericht op onderzoeksdata die ten grondslag liggen aan een wetenschappelijke publicatie of andere (wetenschappelijke) rapportage. Andere situaties waar archivering van belang kan zijn:
   a) Einde van een project, ongeacht of de data worden gepubliceerd in een artikel (met name als de data van belang kunnen zijn voor hergebruik in een nieuwe studie)
   b) Als de dataverzamelingsfase is beëindigd en je de originele 'bevroren' ruwe data veilig wilt bewaren om verlies of modificatie van de ruwe data te voorkomen
   c) 'Snapshots' van datasets van langlopende cohort studies of 'registries' op basis waarvan je analyses wilt gaan doen

Hogeschool van Amsterdam

UITGANGSPUNTEN/OVERWEGINGEN

- Het gaat om onderzoek(sresultaten) waarvoor de HvA verantwoordelijk is
  - Bij promotieonderzoek zijn door de (co)promoteren afspraken gemaakt tussen de HvA en de instelling waar gepromoveerd wordt over (onder andere) archivering en eventueel hergebruik van onderzoeksdata
  - Bij WMO-plichtig onderzoek waarvan de HvA niet verrichter is, ligt de verantwoordelijkheid voor databeheer (en dus archivering) bij de partij die verrichter is, tenzij daar andere afspraken over zijn gemaakt
- De te gebruiken data repository voor het archiveren van onderzoeksdata is UvA/HvA Figshare tenzij de lector een reden heeft om de onderzoeksdata ergens anders te archiveren
- De (ruwe) datasets zijn zo opgebouwd dat ze geen direct identificerende (contact) gegevens bevatten die niet nodig zijn voor de analyse en uiteindelijke resultaten. Direct identificerence (contact) gegevens staan in een apart opgeslagen sleutelbestand, gescheiden van de onderzoekgegevens en worden, indien nodig, gescheiden gearchiveerd
- De datasets worden vóór publicatie van een artikel gearchiveerd zodat in het artikel naar de persistent identifier (DOI) van de (meta)data kan worden verwezen
- Het archiving package wordt in het Engels gedocumenteerd / gearchiveerd
- Onderzoeksdata worden zoveel mogelijk gearchiveerd in open, preferred formats. Van data in deze formats kunnen data repositories garanderen dat ze langdurig bewaard kunnen worden en ook in de toekomst leesbaar en toegankelijk blijven. Zie hier voor de preferred formats volgens DANS

Hogeschool van Amsterdam

ARCHIVING PACKAGE 1/2

BEHORENDE BIJ GEPUBLICEERDE WETENSCHAPPELIJKE RAPPORTAGE

Archiving component	Remarks
A brief description of the problem definition, research design, conceptual framework, data collection (sampling, selection and representativeness of informants) and methods used	- Not required as long as the manuscript or study protocol provide this information in detail - If previously archived, documented in a published preregistration or described in detail in a published manuscript, cite the persistent identifier (location) of this information
The instructions, procedures, the design of the experiment and stimulus materials that can reasonably be deemed necessary in order to replicate the research	- For example, topic list, interview guide, questionnaires, standard operating procedures - The materials must be available in the language in which the research was conducted - If previously archived, documented in a published preregistration or described in detail in a published manuscript, cite the persistent identifier (location) of this information
Raw (pseudonymized) data files	- For example, a locked Castor database, export file of an online survey, recordings or transcripts of interviews, descriptions of observations, a snapshot of databases that constantly change (e.g. registries or long-term cohorts) - If the raw data is already archived elsewhere, provide a reference to where the archived raw data can be found - The raw data should not contain directly identifying personal information - The variables and values of the raw data file should be labeled and/or described in a codebook/data dictionary
Analyzed data files	- The processed data files that were eventually analyzed when preparing the article (e.g. an SPSS data file after transforming variables, after applying selections, etc.) - Not required if the raw data file was directly analyzed and the computer code is provided (see below) - It is important that the correct version of the file is submitted - Any intermediate files created during the process of raw data into analyzed data do not need to be archived, as long as the code showing the processing steps are provided (see below)
Computer code and syntax	- For example, Atlas.ti, SPSS syntax file, R code, Python analysis script - Computer code raw data file > analyzed data file. Describing the steps taken to process the raw data into analysis data, including brief explanations of the steps in English, for example a brief description of the steps taken in the qualitative analysis of research data, i.e. themes, domains, taxonomies, components - Computer code of the (statistical analyses) - Computer code analyzed data file > manuscript results (e.g., figures, tables, etc). Describing the steps taken to process the analysis data into results in the manuscript, including brief explanations of the steps in English - Version control parameters should be provided and the correct version of the code must be submitted with the archiving package - The computer code should 'run' (work) on the provided raw and analyzed data files

Hogeschool van Amsterdam

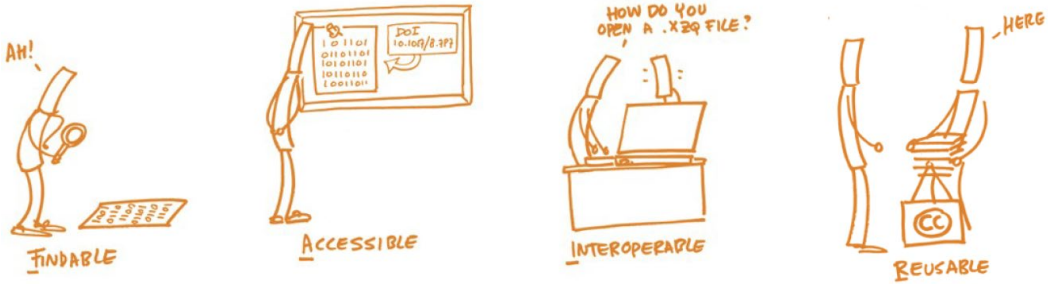
ARCHIVING PACKAGE 2/2

BEHORENDE BIJ GEPUBLICEERDE WETENSCHAPPELIJKE RAPPORTAGE

Archiving component	Remarks
Additional documentation and metadata necessary to replicate the research or reuse the data:	
Codebook / data dictionary	- A data dictionary is critical to making your research more reproducible because it allows others to understand your data. The purpose of a data dictionary is to explain what all the variable names and values in your spreadsheet really mean. See here and here. - For example, units used, definition and labels of (categorical) variables, codes for missing values, description of how derived variables were created, allowed values
A readme file describing which documents and files can be found where and how they should be interpreted	The readme file must contain (part of) the following information: - Name of the person who stored the documents and files - Date, period, location of data collection - Names of people who collected data - How the various files are related to each other and where related documents, data, preregistrations, protocols can be found - Descriptions of measurement instruments used (device, brand, version, calibration procedures) - Hardware and software used (including version) - Description of quality assurance (checking for errors, validation methods) The readme file should be saved in a non-proprietary format, such as .txt or .xml
Statistical Analysis Plan and Data Management Plan	- If previously archived, documented in a published preregistration or described in detail in a published manuscript, cite the persistent identifier (location) of this information
Documents relating to the ethical approval	- The empty informed consent form (information sheet + consent form) should be archived because it shows what participants consented for (e.g. data reuse or not) - The signed informed consent form should be archived separately from the archiving package. In case of paper informed consent, see non-digital archiving.
Logbooks or lab journals	

Hogeschool van Amsterdam

BIJLAGE F.A.I.R. DATA



- Door onderzoeksdata in een data repository zoals UvA/HvA Figshare te zetten, zet je een grote stap in het F.A.I.R. maken van je data
- F.A.I.R. data ≠ OPEN data
- How FAIR are your data? (Jones & Grootveld, 2017). https://doi.org/10.5281/zenodo.3405141

FINDABLE

It should be possible for others to discover your data.
Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- A persistent identifier is assigned to your data
- There are rich metadata, describing your data
- The metadata are online in a searchable resource e.g. a catalogue or data repository
- The metadata record specifies the persistent identifier

INTEROPERABLE

Data, metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- Data is provided in commonly understood and preferably open formats
- The metadata provided follows relevant standards
- Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- Qualified references and links are provided to other related data

ACCESSIBLE

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open!

- Following the persistent ID will take you to the data or associated metadata
- The protocol by which data can be retrieved follows recognised standards e.g. http
- The access procedure includes authentication and authorisation steps, if necessary
- Metadata are accessible, wherever possible, even if the data aren't

REUSABLE

Documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- The data are accurate and are well described with many relevant attributes
- The data have a clear and accessible data usage license
- It is clear how, why, when and by whom the data have been created and processed
- The data and metadata meet relevant domain standards