Model

Guidelines for the management of Research Data and Records

Amsterdam University op Applied Sciences
School of Economics and Management

March 17, 2016

Dr. G.J. van Bussel

A. Introduction

1. The preparation and retention of appropriate research data is an essential component of all research.

2. The University, its faculty and its students have a shared responsibility to assure that research is appropriately recorded, archived for the required length of time, and made available for review under appropriate and legal conditions.

3. At a minimum, original research records are essential to protect intellectual property rights, to answer ongoing questions regarding management of a research program, and to address possible questions that may arise regarding the propriety of research conduct.

4. A research dataset typically has a longer lifespan than the research project that creates it. A research project begins and ends with its funding. The reality is that researchers continue to work on the datasets generated/collected well after funding has ceased, and follow-up projects that continue to analyse or add to the datasets may subsequently be funded.

5. These Guidelines should be read in conjunction with the current University Policy on Research Data Management that outlines its regulatory and policy framework at the Amsterdam University of Applied Sciences. These procedures outline the critical steps for meeting the requirements of the Policy and provide practical advice for achieving compliance.

6. The execution of these Guidelines is the responsibility of the University researchers, and will be audited by University Data Steward(s).

B. Legal Issues

1. <u>Ownership.</u>

2.1. Clarification of ownership and rights associated with research should be determined early in the project planning.

2.2. Documentation of these details should be stored with the research data to ensure appropriate management and access to the research data and records associated with the project.

2.3. In particular where research is undertaken in accordance with a contractual agreement or under (commercial) sponsorship the ownership of the research data and records and responsibilities should be determined prior to commencement of the research contract. It should be specified in the research contract.

3. <u>Copyright</u>

3.1. Under Amsterdam University of Applied Sciences policy, *scholarly works* are owned by the University, where 'scholarly works' means any article, book, musical composition, creative writing or like publication or any digital or electronic version of these works that contains material written by any member of academic staff, an honorary appointee, a visitor or a student based on that person's scholarship, learning or research.

4. <u>Privacy</u>

4.1. Researchers who are collecting information from or about individuals for their research should be aware of the requirements and implications of privacy legislation, and any privacy policy of relevant organizations and how this may affect the data collection, storage, use and disclosure of the information they wish to collect.

4.2. Researchers are required to implement the *Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek* (VSNU) for all research projects that are collecting information from or about individuals. (http://www.vsnu.nl/files/documenten/Domeinen/Accountability/Codes/Gedragscode%20persoonsgegevens.pdf)

4.3. Especially researchers should:
  - Include clear documentation within data management about:
      o The nature of any private, sensitive or confidential information that may be collected;
      o Non-disclosure agreements and any restrictions on use of the data;
      o Consequences/penalties for breaches of confidentiality, and,
      o Steps to be taken to safeguard privacy and confidentiality.
  - Ensure that consent to participate is both informed and freely given by the participants of their research
  - Ensure that participants have available to them a full description of the project in language they can understand, the nature of their participation and the implications in terms of risks and

benefits of participating in the research, including information about what will happen to their information, how it will be used, stored and when it will be disposed of.

- Give participants a plain language information sheet with information about the project and a consent form that outlines what the participants will do if they agree to take part and researchers agree.
- Capture the signed consent form and the information sheet together within the research archives of the Amsterdam University of Applied Sciences as proof of the process of informed consent and as evidence that the consent to participate was informed and freely given, during or after completion of the project, and will ensure its (secured) accessibility.
- Gather directly identifiable personal data within 'communication data' file, separated from the research data, but administratively linked.
- Destroy this file of 'communication data' when it is not needed anymore for reaching the aims of the project.
- Implement art. 3.10 of the *Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek* (VSNU) when archiving research data.

5. <u>Confidentiality</u>

5.1. Research data generated or compiled in research projects may be confidential or secret. Examples of confidential material include, but are not limited to:
- Research data which link the individual human participant with the study like signed consent forms, master lists of names and addresses or matching codes for a current study or similar listings, which may be held for a period of time for a follow up study. Personal information is protected under Privacy legislation.
- Research data which is sensitive, for example: identified highly personal data; data which may be incriminating either to the provider of the data or to a third party; personal data which although not identified by name is in such a form (such as a case study or life history) that it may be able to identify the subject, and data, which even if not sensitive, may identify people (for example photographs, videotape, audiotape).
- Research data that may cause harm to a third party should it be released.
- Research data protected by a contract of secrecy or non-disclosure. The research may be considered 'commercially valuable' or 'trade secret'.

5.2. Research material of a sensitive or confidential nature, which has possible patent, trademarks or Intellectual Property implications is considered confidential for commercial purposes. Information about relevant documents and contracts relating to these agreements should be stored with the data.

5.3. Confidential research data should be stored securely, with controlled access that complies with any agreements in place for the research activity. Such precautions include password access and 'locking' data files. The signed consent forms for a particular project should be stored separately from the collected research data for that project.

5.4. Breach of confidentiality agreements and requirements must be reported to the Data Steward as soon as the knowledge of the breach occurs. A breach is considered to have occurred where:

- Disclosure of research information with imposed confidentiality restrictions has occurred.
- A formal confidentiality or non-disclosure agreement has been broken
- Confidential data has entered the public domain, for example, published on the Internet.

6. Patents

6.1. Where a patent has been granted all research data must be retained for the life of the patent.

6.2. In the cases of commercially exploitable research, and research data that concern a patent application filed by the University, it is necessary for original research data to be retained at the University.

6.3. All correspondence, deeds and contracts associated with the commercial exploitation of the patent must be retained within the repository of the University.

6.4. Researchers are required to disclose inventions to the University. This will provide a means of assessing the potential value of the intellectual property.

7. Sponsored research

7.1. Funding bodies may have specific requirements for retention of research data. Researchers should be aware of the conditions of any awards or contracts supporting their research.

8. Discipline specific practices or codes

8.1. Researchers should be aware of, and adopt, the relevant practices or codes within their research discipline that establish norms or best practice for the retention of research data and records researchers.

C. Documenting and Archiving Research Data and Records

1. <u>Data Documentation</u>

1.1. Research data and records need to be documented with metadata when creating, organizing and managing research data. Metadata are a subset of core standardized and structured data documentation that explains the origin, purpose, time reference, geographic location, creator, access conditions and terms of use of a data collection.

1.2. Researchers are expected to provide provenance and contextual information for the data so that it can be understood in the future. Documentation includes metadata on:
   - The context of data collection: project history, aim, objectives and hypotheses
   - Data collection methods: sampling, data collection process, instruments used, hardware and software used, scale and resolution, temporal and geographic coverage and secondary data sources used
   - Dataset structure of data files, study cases, relationships between files
   - Data validation, checking, proofing, cleaning and quality assurance procedures carried out
   - Changes made to data over time since their original creation and identification of different versions of data files
   - The legal implications of the research
   - The retention (periods) and disposal of (parts of) the dataset
   - Information on access and use conditions or data confidentiality

   At the data-level:

   - Names, labels and descriptions for variables, records and their values
   - Explanation or definition of codes and classification schemes used
   - Definitions of specialist terminology or acronyms used
   - Codes of, and reasons for, missing values
   - Derived data created after collection, with code, algorithm or command file
   - Weighting and grossing variables created
   - Data listing of annotations for cases, individuals or items

1.3. Data-level descriptions may be embedded within a data file itself. Many data analysis software packages have facilities for data annotation and description, as variable attributes (labels, codes, data type, missing values), data type definitions, table relationships, etc. Other documentation may be contained in publications, final reports, working papers and lab books or created as a data collection user guide.

1.4. Different disciplines develop and adopt various metadata standards and/or practices for the management of their research data and materials. The extent to which any single discipline might adopt these standards is variable. It is expected that researchers use the metadata standards that are prevalent within their discipline.

1.5. Researchers are expected (when there is not a discipline specific metadata standard available) to use

existing international metadata standards or schemes such as Dublin Core, ISO 19115 for geographic information, Data Documentation Initiative (DDI), Metadata Encoding and Transmission Standard (METS) and Encoded Archival Description (EAD).

2. <u>Data Archiving</u>

2.1. Researchers are expected to keep clear and accurate records of the dataset itself (and its different versions, if used in stages of research), the research methods used, the data sources, the used data analytics, developed database queries and algorithms, and used metadata schedules, including any approvals granted, during and after the research process. This enables the documentation of how data collections relate to results, which in turn support reproduction and validation of the results, and potentially, data re-use.

2.2. Researchers are expected to use the University repository for long-term preservation of research data or the repository that is used as commodity for long-term preservation within their specific discipline. At all times, they need to register the deposit in the University index of research data sets.

2.3. Researchers are expected to use file formats for long-term preservation of research data as specified by DANS in its overview of Preferred and Acceptable Formats (Appendix 1).

2.4. Researchers are expected to:
   - Decide how many versions of a file to keep, which versions to keep, for how long and how to organize versions
   - Identify milestone and master versions to keep
   - Uniquely identify files using a systematic naming and categorization convention
   - Record version and status of a file, e.g. draft, interim, final, internal
   - Record what changes are made to a file when a new version is created
   - Record relationships between items where needed, e.g. relationship between code and the data file it is run against; between data file and related documentation or metadata; or between multiple files
   - Track the location of files if they are stored in a variety of locations
   - Regularly synchronize files in different locations
   - Maintain single master files in a suitable file format to avoid version control problems associated with multiple working versions of files being developed in parallel
   - Assign responsibility for master files to a single project team member
   - Regulate write access to master versions of data files
   - Archive copies of master files at regular intervals
   - Check data integrity of stored data files at regular intervals
   - Develop a formal procedure for the destruction of master files

2.5. Researchers are expected to regularly making backups to protect against accidental or malicious data loss, regardless of the University back-up policy.

2.6. Researchers are expected to keep research data secure. This includes:

- Not storing confidential data such as those containing personal information on servers or computers connected to an external network, particularly servers that host internet services
- Locking computer systems with a password and installing a firewall system
- Protecting servers by power surge protection systems through line-interactive uninterruptible power supply (UPS) systems
- Implementing password protection of, and controlled access to, data files, e.g. no access, read only, read and write or administrator-only permission
- Controlling access to restricted materials with encryption and password-protection
- Imposing non-disclosure agreements for managers or users of confidential data
- Not sending personal or confidential data via email or through File Transfer Protocol (FTP), but rather transmit as encrypted data
- Destroying data in a consistent manner when needed

2.7. Data that contain personal information should be treated with higher levels of security than data that do not. Security can be made easier by:
- Anonymizing or aggregating data
- Separating data content according to security needs
- Removing personal information, such as names and addresses, from data files and storing them separately
- Encrypting data containing personal information before they are stored - encryption is certainly needed before transmission of such data.

2.8 Researchers are expected to adhere to the retention period for research data sets of at least ten years, after their deposit in a repository for long-term preservation. The University adheres in this to the *Nederlandse Gedragscode wetenschapsbeoefening.*
(http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2014).pdf).

D. Data Management Planning

1. Researchers are expected (before starting a research project) to develop a Data Management Plan.

2. A data management plan should contain information about:
    - Which data will be generated during research
    - Metadata, standards and quality assurance measures
    - Plans for sharing data
    - Ethical and legal issues or restrictions on data sharing
    - Copyright and intellectual property rights of data
    - Data storage and back-up measures
    - Data management roles and responsibilities
    - Costing or resources needed

3. Researchers are expected to use the model of the Data Management Plan as is specified in Appendix 2.

APPENDIX 1

DANS Preferred and Acceptable Formats

	Preferred format(s)	**Acceptable format(s)**
Text documents	• PDF/A (.pdf)	• OpenDocument Text (.odt) • MS Word (.doc, .docx) • Rich Text File (.rtf) • PDF (.pdf)
Platte text	• Unicode TXT (.txt, …)	• Non-Unicode TXT (.txt, …)
Marked-up language[(*)]		• XML (.xml) • HTML (.html)
Spreadsheets	• OpenDocument Spreadsheet (.ods) • Comma Separated Values (.csv)	• MS Excel (.xls, .xlsx) • PDF/A (.pdf) • OOXML (.docx, .docm)
Databases[(*)]	• ANSI SQL (.sql, …) • Comma Separated Values (.csv)	• MS Access (.mdb, .accdb) • dBase III or IV (.dbf)
Statistical data[(*)]	• R • SPSS Portable (.por) • SAS transport (.sas) • STATA (.dta)	
Images (raster)	• JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png)	• JPEG 2000 (.jp2)
Images (vector)	• Scalable Vector Graphics (.svg)	• Adobe Illustrator (.ai) • PostScript (.eps)
Video[(*)]	• MPEG-2 (.mpg, .mpeg, …) • MPEG-4 H264 (.mp4) • Lossless AVI (.avi) • QuickTime (.mov)	
Audio[(*)]	• WAVE (.wav)	• MP3 AAC (.mp3)
Computer Aided Design (CAD)	• AutoCAD DXF versie R12 (.dxf)	• AutoCAD other versions (.dwg, .dxf)
Geographic Information (GIS)	• Geographic Markup Language (.gml) • MapInfo Interchange Fomat (.mif/.mid)	• ESRI Shapefiles (.shp and associated files) • MapInfo (.tab and associated files) • Keyhole Markup Language (.kml)
Images (geo-referenced)	• GeoTIFF (.tif, .tiff)	• TIFF World File (.tfw en .tif)
Raster GIS	• ASCII GRID (.asc, .txt)	• ESRI GRID (.grd and associated files)
3D	• WaveFront Object (.obj) • X3D (.x3d)	• COLLADA (.dae) • Autodesk FBX (.fbx)
RDF	• W3C standards	

[(*)] In research

Appendix 2.

Model of a Data Management Plan

Name of student/researcher(s)	
Name of group/project	
Description of your research	*Briefly summarise the type of your research to help others understand the purposes for which the data are being collected or created.*
Funding body(ies)	
Grant number	*A grant number provides unique identification for the grant.*
Partner organisations	
Project duration	Start: **MM-DD-YYYY** End: **MM-DD-YYYY**
Date written	**MM-DD-YYYY**
Date last update	
Version	*A new version of the DMP should be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.*
Name of researcher(s) with roles/responsibilities for data management	*Naming anyone with specific roles and responsibilities for data management is especially important for collaborative projects that involve many researchers, institutions, and/or groups.*

A. Data Collection

Description of the data to be collected within the research project

Checklist:

1. How will data be collected?
2. What type of data will be collected? (measurements, observations, models, software....)

3. In what file formats?

4. Will it be reproducible? What would happen if it gets lost or becomes unusable later?

5. What is the estimated size of the data, and what growth rate?

6. How do you handle version control to maintain all changes that are made to the data?

7. Which tools or software are needed to create/process/visualize the data?

8. Will you also use pre-existing data? From where?

B. Data Storage and Back-up

Description of the way(s) research data during research are stored securely and backed up or copied regularly.

Checklist:

1. How will the raw data be stored and backed up during the research?

2. How will the processed data be stored and backed up during the research?

3. Which storage medium will you use for your storage and backup strategy? Network storage; personal storage media (CDs, DVDs, USBs, portable hard drives); cloud storage

4. Are backups made with sufficient frequency so that you can restore in the event of data loss?

5. Are the data backed up at different locations?

C. Data Documentation

Description of the way data will be documented to help future users to understand and reuse it.

Checklist:

1. What standards will be used for documentation and metadata? If there is not a standard already available for your data, outline how and what metadata will be created.

2. How will your data be documented during your research and for long-term storage?

3. What directory and file naming convention will be used to enable the titling of your folders, documents and records in a consistent and logical way?

4. What project and/or data identifiers will be assigned? (e.g. DOI or Digital Object Identifier)

D. Data Access

Description of access and security management.

Checklist:

1. How will you manage copyright and Intellectual Property Rights issues? E.g. Who owns the data? How will the data be licensed for reuse?

2. Are there any limitations on the access of your data?

3. What are the access criteria for the data (open/restricted access, embargo period, etc.)?

4. Who controls data access (e.g. PI Principal Investigator, student, lab, University, funder) ?

E. Data Sharing and Reuse

Description of how research data is shared.

Checklist:

1. If you allow others to reuse your data, how will the data be shared? In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
2. Any sharing requirements (e.g., funder data sharing policy)?
3. Audience for reuse? Who will use it now? Who will use it later?
4. When will you publish it and where? Is your data underlying a scientific publication?
5. Which tools/software are needed to view/visualize/analyse the data?

F. Data Preservation and Archiving

Description of which and how data should be preserved, and/or shared?

Checklist:

1. Which criteria will you use to decide which data has to be archived for preservation and long-term access.
2. Which data has to be destroyed?
3. How long should it be preserved (10-20 years, permanently)?
4. What file formats? Are they long-lived?
5. Which data repository is appropriate for archiving your data?
6. What costs (if any) will your selected repository charge?