Ellen Giebels 26-10-2016

Faculty Data Management Policy – Behavioural, Management and Social Sciences (BMS)

Background

In July 2015, the University of Twente adopted a new research data policy. This document sets out minimum criteria for the management and storage of data. Since the way in which research data is handled varies from one academic discipline to another, it is up to each faculty to supplement the university's policies by setting up their own faculty data policy. These policies specify how scientists within the faculty are required to handle research data.

Process

In early 2016, the Dean of the BMS faculty appointed a committee to deal with this matter. This committee, which consists of stakeholders from the faculty, the Institute for Innovation and Governance Studies (IGS) research institute, and Library, IT Services & Archive (LISA), was tasked with surveying all current initiatives and facilities (infrastructure) in the area of BMS/IGS data policy (1). The committee was also asked to issue recommendations, where necessary, concerning the steps that need to be taken to make the best possible use of available research data in the upcoming years, and to guarantee adequate storage, management and security during this period (2).

Based on the above, the committee has set itself the target of developing a comprehensive data policy. The goal is to make it clear to new and existing members of staff exactly what steps and conditions are involved – from beginning to end – in the handling of research data. In addition to developing a policy, an action plan to 'hammer the message home' among the members of staff is needed (4).

Details of the data management process are depicted in Figure 1.

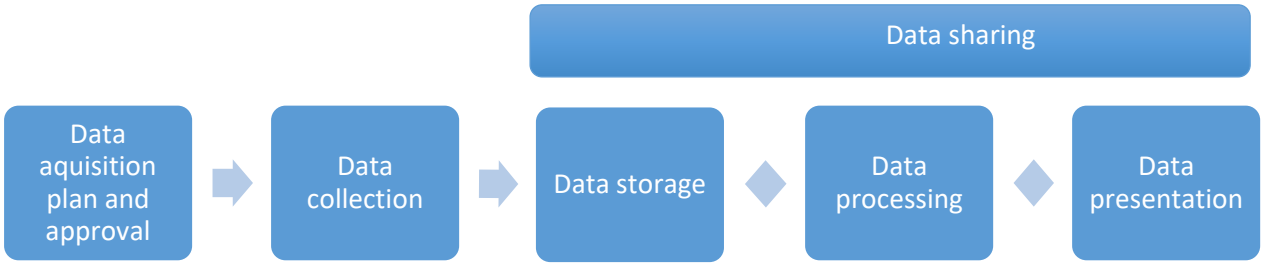

Figure 1. The data management process

It should be noted that these steps may sometimes take place in a different sequence, or that some steps may be skipped completely (in the case of externally sourced files, for example). In general, however, the process steps shown above are followed.

The faculty has recently invested mainly in tools for data collection (Tech4People lab), in new servers for the storage and processing of data, and in the requisite personnel support. Work has also begun on the software for data processing and data visualization. The key issue here (in this document too) involves data infrastructure and an overarching policy.

1. *Principles of the University of Twente's Data Policy*

Based on the university's central policy, the data policy being drawn up by each individual faculty must consist of the following components, at the very least:

Provisions concerning responsibility for the data policy of the group/faculty/institute in question

Details proposed for BMS:

At faculty level, it is the Dean who is responsible for **establishing** and **implementing** (directly or indirectly) **data policy**, for **communicating** this policy to the departments and staff, and for establishing accurate **basic facilities**. The Dean must ensure that there is optimum communication about these tasks, and that they are effectively delegated to the various department chairs involved. The Dean is also responsible for furnishing the basic facilities needed to implement data policy within the faculty. The department chairs, PIs, thesis supervisors, and supervisors involved are asked to actively and consciously adopt a responsible attitude. Their role is to further communicate details to the individual researchers, and to draw their attention to the proper implementation of the data policy. It is also important for them to be fully aware of their function as role models, and notify the Faculty Dean's office concerning any unethical or improper conduct (known or suspected).

It is the individual researcher/first author of the project who always bears ultimate responsibility for the **proper handling** of data. Before the start of a project, researchers must always ensure that they have the appropriate expertise and capacity available to carry out the research in line with the VSNU's guidelines and code of conduct. This is one aspect of a professional attitude in science. Furthermore, in the case of junior researchers and PhD students, the first supervisor of the project (with a PhD) bears **joint responsibility** for proper data management, including verification of the data packages involved. The faculty will monitor events at a distance, and will perform occasional data package audits.

Provisions concerning the long-term storage of data, in accordance with legal regulations, the contractual requirements of third parties, funding bodies, etc.;

Details proposed for BMS:

- Every study carried out within the faculty (whether by students or members of staff) that necessitates the direct or indirect involvement of human subjects must be submitted to – and approved by – the Ethics Committee.
- Where necessary, the data will be anonymized (by removing links to identifying personal information) immediately after being collected. This applies both to data that researchers have collected personally, and to data obtained from external sources. 'Anonymized' means that the data cannot be traced back to a specific individual. In addition, the linking files (which establish a link between an individual and their data, for the use of researchers) are encrypted and stored offline.
- The environment in which the encryption key/linking files are stored must be properly secured.

Provisions concerning both the short-term and long-term storage of data, and the way in which such data is defined;

Criteria relating to the selection of research data for long-term storage; this may vary from one study to another. The general criteria in this context are: reproducibility, verifiability and reusability

Details proposed for BMS:

- Data should never be stored on personal and/or local media only.
- Upon receipt of data (including raw data), a copy should be stored on BMS's central and secure server, at the very least. The file (or files) should be stored in the same directory as the file containing the Ethics Committee's approval, to identify the research project to which it relates.
- At the end of the study, the data is either stored in a trusted repository (DANS), or it is permanently stored on one of the faculty's secure servers. This concerns the raw dataset, at the very least.
- The guideline states that raw research data should be stored for at least 10 years[1].
- The data is archived in such a way that it remains accessible, even over the longer term. This requires good metadata, involving the use of a persistent identifier (Digital Object Identifier or DOI), for example, so that the data is not lost and is easy to retrieve. This also concerns the documentation needed to ensure that the data remains understandable, as well as facilities for converting files to a new format, before the old format becomes obsolete and potentially unreadable.
- For the purposes of archiving, there is the option of storing data externally, when it is no longer in active use; some archives or repositories have a data preservation mission and a conversion facility, as mentioned in the preceding point (these are certified as trusted digital repositories).

[1] Provided that this does not conflict with Art. 10 of the Personal Data Protection Act, which prohibits the storage of data for longer than is strictly necessary.

Conditions under which research data is made available to third parties

Details proposed for BMS:

- The general rule here is 'open if possible, closed if necessary'. The preconditions involved correspond to legal requirements (anonymization) or contractual requirements, e.g. funding bodies or clients.
- Data can only be shared if prior authorization has been given, by means of the informed consent procedure.
- Wherever possible, the data will be shared for scientific purposes (with scientific partners).
- It is also possible to share data with end users/partners in the professional field for the purposes of commercial knowledge transfer (social utility). Ideally, the data is not transferred to these partners. A provision is made to allow these parties to inspect/re-analyse this data within the University of Twente.
- Researchers should be extremely cautious about sharing data for commercial purposes.

Some additional frameworks for the proper handling of data are:

- The VSNU Codes of Conduct. In principle, all universities and their scientific staff should make every effort to familiarize themselves with this code. They must also see to it that the code of conduct is discussed within the academic community. This triggers an awareness of what constitutes good scientific teaching and research. The code requires more of scientists than compliance alone. It also involves a responsibility to promote compliance throughout the academic environment and to identify and report any abuses. The university's governing bodies are obliged to promote and enforce compliance with the code. The university has public and binding rules governing the independent handling of complaints about breaches of academic integrity. There is also a Scientific Integrity Committee.
- The advisory report formulated by the DSW[2] Committee concerning academic integrity, data storage and reproducibility (dated 06/01/2016). This advisory report is based on an appraisal of the data protocols used by the Social Science faculties of nine Dutch universities. There is consensus on a number of principles. The DSW Data Committee recommends that these be used as minimum standards. It is important for BMS to adopt these standards, to ensure that there is a common standard within this academic discipline as a whole. It also facilitates cooperation between researchers from different universities and the transfer of researchers from one university to another. Details of these minimum standards have been incorporated into this advisory report.

[2] Meeting of representatives of the academic discipline of Social Sciences, involving the deans of various faculties of social sciences.

- In addition, under the Data Leaks Reporting Obligation (which came into force on 1 January 2016), it is mandatory for University of Twente staff to report all data leaks (known or suspected). This reporting requirement is the result of an amendment to the Personal Data Protection Act (see PDPA/GDPR[3]).
- With effect from 1 October 2016, before starting any research on human subjects, all BMS researchers and students are required to submit an application to the faculty's Ethics Committee.
- Finally, in 2015, an information security policy for the entire University of Twente was adopted. The four basic rules are:

1. As a public-law organization, the University of Twente abides by the law. Many consider this to be self evident. Even though the University of Twente is an entrepreneurial university, it does not subscribe to the view that an organization's decision on whether or not to abide by the law should be based on a cost-benefit analysis. Having said that, of course, it should be pointed out that the university is not a policeman either.

2. Information about students and staff is handled as carefully as possible. Prospective and current students must be able to rely on the university to handle their information as carefully as possible. Much of this information is related to students' courses of study. Exercising due caution in matters of privacy is one of the challenges we face, as a university.

3. All University of Twente staff are expected to adopt a proactive attitude, particularly with regard to information security in all processes and activities. There are many aspects to information security, which touches on virtually all processes and activities. Taking risks is part of the University of Twente's entrepreneurial attitude. Part of this involves exploring potential effects in advance and taking steps to mitigate any unacceptable risks.

4. The information security policy in no way interferes with the University of Twente's entrepreneurial and creative nature. All necessary security precautions must be taken, of course, even though some individuals may not be particularly happy about it, but only after the matter has been carefully considered. Proportionality is called for here. No radical or restrictive measures that are disproportionate to the actual risk reduction involved will be taken.

In addition, the document distinguishes between three types of data: basic, sensitive and critical.

2. *Existing infrastructure for research data (collection, storage and support)*

What aspects relating to research data (at faculty level, throughout the entire university, and nationally) are already undergoing reorganization? What type of support has been offered in this connection?

There are a range of facilities within BMS/IGS:

[3] PDPA = Personal Data Protection Act (now current), GDPR = General Data Protection Regulation (will come into effect on 25 May 2018).

- **IGS DataLab**: data infrastructure that enables researchers to create, store and maintain data in compliance with the conditions governing professional and responsible data handling. The DataLab can also provide researchers with methodological recommendations regarding the use of data. From the very beginning, DataLab has used an open source tool (LimeSurvey) to create online surveys. IGS provided support for the administrative aspects, as well as methodological advice. In the course of 2016, it became evident that it was no longer possible to provide appropriate technical management. Moreover, the software was no longer downward compatible, which resulted in problems with updates. In addition, LISA identified a security issue relating to the external storage of data and, potentially, to applicable US legislation. Based on a list of wishes and requirements, two alternatives are currently being assessed. One is a tool from Statistics Netherlands (CBS), and the other is an application developed by Easyonsurvey – a new local startup. An added benefit of the latter option is that this company's software includes a sequence of steps that can be used to anonymize data, and there are many applications for extras in line with Tech4People projects (apps etc.). Ultimately, these two new features will be compared to the tools that are currently in use. This is because these providers also appear to modify their policies and options in line with developments in data management.
- **Tech4People lab**: this lab offers various data collection options, involving a wide range of new technological tools. Various developments have recently taken place within the Tech4People lab, including the installation of a file server (containing general information, manuals, support information, and project directories). This server also includes a software section containing all the programs needed to operate the new tech tools for research, and to analyse this research. In addition to the file server, an FTP server has been installed for safe and secure file transfers in the case of joint projects with third parties (such as the MST). Next autumn, all of the BMS faculty's researchers will be provided with secure SSD drives (Samsung TD). The researchers will be asked to make no further use of other, non-secure solid state drives. Our objective is to promote greater awareness of secure storage and transport options for University of Twente research data, and to get more people to make use of these options.

  A comprehensive list of what Tech4People has to offer can be found on the lab's website https://bmslab.utwente.nl/. A great deal of consideration has also been given to the issue of data storage. To this end, two new servers have been purchased, which will be managed by LISA (the former ICT and B&A).
- **CBS Remote Access Terminal**: gives researchers access to CBS microdata, anonymized data at the level of individuals or businesses.

The following facilities/services are available at the University of Twente:

- **LISA** (throughout the entire university) offers researchers assistance with the long-term archiving and storage of research data. In addition, LISA offers support with the transition to the national university facilities offered by the 4TU Centre for Research Data and DANS.

- **Health DataLab** (initiative by the MIRA research institute/Health Sciences). There is also a similar initiative within CTIT (Centre for Telematics and Information Technology).

The following facilities are available in university settings at national level:

- **4TU Datacenter** (long-term archive for scientific research data, with permanent access to research data together with tools for reusing such data). The 4TU Datacenter advises and supports researchers with regard to data management. (DOI added, meets the requirements of institutions and research funding bodies.)
- **DANS** (institute for the long-term archiving of datasets, where previously collected data can be reused). DANS also advises on data management and on the certification of digital archives.
- **Dataverse.nl (DVN)** (a joint venture between DANS, 4TU Datacenter, Dutch universities and universities of applied sciences, for the short term archiving and sharing of non-static research data between project teams or research departments) offers researchers facilities for storing and sharing data, and for recording it online, during their research and for the following ten years. This facility is jointly offered by participating institutions.
- **SURF** (collaborative IT organization for teaching and research in the Netherlands) offers researchers data storage, the option of sharing data, and support with data handling.

Appendix 4[4] Data Management Plan - Guidelines, Policies, Paragraphs

There are many different Data Management Plan guidelines and templates. Exact data management requirements and conditions differ per funding agency, journal or institution. An overview of the most important guidelines is given below.

Funder requirements[5]:

DANS

Extensive guidelines for writing a Data Management Plan for scientific research, including a DMP-checklist, are provided by DANS. DANS (Data Archiving and Networked Services) is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO).

4TU.Datacentrum

The 4TU.Datacentrum focusses on technical studies and offers a template to help setup a Data Management Plan. The template can be found here. Please also check the University of Twente template which can be opened or downloaded here at the right side of the screen.

Horizon 2020

Horizon 2020 is a European Commission programme that finances and stimulates European research and innovation projects. Applicants and recipients of subsidies can find the Data Management Policy in the document Guidelines on Data Management in Horizon 2020. A handy tool to write a Data Management Plan that meets the Horizon 2020 policy, you can use the tool DMP Online.

ZonMW

ZonMW, the Netherlands Organisation for Health Research and Development, promotes health research and innovation. ZonMW stimulates re-use of existing datasets and accessibility of new data sets. Checklists for writing a Data Management Plan or a Data Paragraph (in Dutch) can be found in the report "Data: Digitale Diamanten" (2013).

KNAW

The policy of KNAW, the Royal Netherlands Academy of Arts and Sciences, focuses on open access and digital preservation of research data. Research proposals should include a "data paragraph" containing information on data management.

NWO

NWO, the Netherlands Organisation for Scientific Research, funds research programmes and manages the national knowledge infrastructure. NWO stimulates permanent archiving and re-use of research data. To receive funding, researchers are required to sign a data contract with DANS. More information on NWO regulation on granting can be found here and information on the datamanagement protocol NWO here

Publisher requirements[6]:

More and more publishers have guidelines on research data management. The Radboud University provides an extensive overview on the different requirements on their website on Research Information Services. In addition to the overview of publisher requirements – adopted below – they have published an Excel sheet with an extensive list of journal requirements. This Excel file can be found here.

Elsevier

Elsevier supports the principle that research data are made freely available to all researchers, but does not oblige authors to make these data available. Authors are, however, strongly encouraged to deposit their datasets in relevant

[4] We want to thank Sytske Wiegersma for drawing up this overview.
[5] Adopted from https://www.utwente.nl/igs/datalab/datamanagement/guidelinesdmp/
[6] Adopted from http://www.ru.nl/research-information-services/institutional-policy/research-managers/appendix-7/

data repositories or to make them available through other channels. If a suitable data repository cannot be found, Elsevier enables its authors to store additional information relevant to the article as a supplementary file. Authors retain their copyright with respect to this additional information.

PLOS

All data and relevant metadata have to be deposited in a relevant and publicly accessible data repository, unless the data already form part of the submitted article. The *data availability* statement should specify the names of the repositories, DOI's and/or access numbers for the relevant data sets. Smaller data sets may be uploaded as supporting-information files. These should be submitted in a format that allows the data to be extracted easily. For instance, where data in tables are concerned, a spreadsheet is preferable to a PDF file.

Whenever ethical or legal considerations prevent the author from handing over the data to a repository or submitting them together with an article, he or she may indicate that the data will be made available upon request to all interested researchers. Particular considerations in a given case, such as possible implications for patent proceedings or possible future research, are not looked upon as valid exemptions to the rule. PLOS requires that authors meet the standards specific to their discipline when preparing and registering their data. Repositories should preferably adhere to accepted standards, such as the criteria formulated by the *Centre for Research Libraries* or the *Data Seal of Approval*. Licenses should be no more restrictive than CC-BY. Information obtained through research on human trial subjects should be handled in such a way that the privacy of the individuals is fully protected.

Nature Publishing Group

When publishing in Nature journals, authors are upon request obliged to put all their material, data, code, and all related protocols immediately at the disposal of their readers. Possible restrictions on availability must be reported to the editors and indicated in the article itself.

As of May 2013, papers submitted in any field within life sciences must be accompanied by detailed information regarding the experimental and analytical design of the research. This is done by filling out a checklist. Supporting information must be available to editors and peer reviewers. Data should be deposited in a publicly accessible repository. *Scientific Data*, an open access sister journal to *Nature*, maintains a list of approved and recommended repositories, sorted by discipline.

A less desirable alternative would be to make data sets available as supplementary-information files. These become freely accessible upon publication through *www.nature.com*. If it appears technically impossible to provide a dataset, authors must provide a URL or another unique identifier.

Authors are encouraged to publish a data descriptor in *Scientific Data*. Data on cell lines should be deposited at a repository that has a certificate of authenticity. Consult www.nature.com for links to additional sources to check the identification of cell lines. Computer code should be made available upon request to editors and reviewers. It is recommended that experimental protocols are shared through Protocol Exchange. Separate guidelines exist for clinical trials.

Science

All of the information necessary to understand, evaluate and elaborate upon the conclusions presented in the manuscript should be available to the readers of *Science*. Datasets must be deposited in approved repositories and accession numbers must be mentioned in the article. Adherence to the MIBBI guidelines (*Minimum Information for Biological and Biomedical Investigations*) is encouraged. The same holds true for all computer codes that have been used to create or analyse the data.

Upon publication, all reasonable requests to provide information and materials must be met, and any exception or limitation must be reported to the editors. Fossils or other rare specimens have to be deposited at a public museum or repository and remain available for inspection.

Large datasets for which no approved repository can be found may be deposited as supplementary materials at *Science* or, if that is not possible, on an institutional website provided a copy of the data is deposited at *Science* in order to guarantee their availability to readers.