

Research Data Policy University of Antwerp

1. Background

Solid management of research data leads to a better quality of research and a greater visibility and reuse of research data. In the context of developments and policies on Open Access and Open Science in general, Open Data is gaining importance within the research community. For these reasons research data management has been high on the agenda of governments, funders and publishers for some time now. The European Commission plays a leading role in this and from 2017 onwards will require inter alia that - following the mandatory open access to publications - all new Horizon 2020 projects need to provide a data management plan (see below) and submit 'open research data', be it with an opt-out for legal reasons.

Research data management (RDM) involves many, often technically complex, aspects and therefore requires a close collaboration between researchers and the various central departments of the University of Antwerp. This cooperation is based on a clear institutional vision, which is then translated into an integrated approach. In the form of a proposed "Research Data Management Policy UAntwerp" and a roadmap for implementation, this paper provides an impetus for vision and action.

2. Prior initiatives

During a policy oriented meeting of the Bureau of the Research council on July 8th 2015, it was decided that in the framework of research integrity, research data management should be a focal point from the academic year 2015-2016 onwards. This resulted in the launch of the following (interuniversity) initiatives:

- Launch structural interuniversity consultation in the form of a VLIR RDM Group. This working group meets about five times a year.
- Development of an interuniversity platform for the preparation and storage of data management plans (DMP, see below), and the development of Flemish DMP templates. This process is well under way: in the near future the platform will be accessible to researchers from the UAntwerp (see below).
- Launch consultations with the FWO on their policy on RDM, specifically with regard to DMP. This has led to a number of discussions in the VLIR Working Group Research and the final recommendation to the FWO is expected soon.
- Development of an interuniversity RDM survey, conducted at the UAntwerp during the Autumn of 2015, with a view to drawing up a well-structured advice on RDM for the Flemish Government. The preliminary results of this survey have already been discussed by the bOZR (07/06/2016) and are processed in this document (see below).

1. Introduction

“Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.”¹

Research data form the beating heart of academic research. Prudent and thoughtful research data management ensures the most hygienic handling of data, from collection to processing to archiving, with minimal data loss and maximum sharing and reuse as key assets. With this policy the UAntwerp adds a concrete dimension to the great importance it attaches to the highest scientific, ethical and integrity requirements, in particular with regard to research data management. At present, the UAntwerp is developing the necessary services and technical infrastructures to optimally support its researchers in data management throughout the data lifecycle.

2. Aim

This policy aims to formulate an ethical and practical framework for the proper handling of research data during the different stages of the full (research) data lifecycle. This includes research data generated by researchers or students from the University of Antwerp, whether or not in collaboration with others, or outside staff using the infrastructure of the University of Antwerp. This framework facilitates the alignment of the different roles and responsibilities of all parties involved, spreading best practices in research data management at the University of Antwerp, and ensuring compliance with the requirements for research data management required by funders.

3. Definitions

3.1 Data life cycle

The process of conceptualization, collection, processing, archiving, distribution, discovery and reuse of (research) data, which spans each stage of a research project. At the moment, the UAntwerp already provides specific support for a number of the steps in the data life cycle.

3.2 Data Management Plan (DMP)

A data management plan (DMP) describes the way in which the data will be collected (by conducting surveys, by trial measurements, etc.) and how they are handled during and after a project (who manages the data, who has access, how will the data be saved and stored; etc.). The DMP is a dynamic document that is prepared at the beginning of a project, but can/should be adjusted and/or completed during and after as well.

¹ Mark Wilkinson et al., ‘The FAIR Guiding Principles for scientific data management and stewardship’, *Scientific Data* 3, 2016 (zie <http://www.nature.com/articles/sdata201618>).

3.3 Data repository

Medium for the permanent storage and access to data. At present, several general, discipline-specific, sponsor-specific or commercial systems are already in place. More information can be found in the section 'Support services and infrastructure ' of this policy (see below).

3.4 DOI

DOI stands for 'Digital Object Identifier', and is the permanent reference number of a digital object, such as a publication or a dataset.

3.5 Fair Data

To allow full reuse of open data, these data should be FAIR: findable, accessible, interoperable and reusable. FAIR data minimizes the risk of data loss and maximizes the possibility of future discoveries through reuse (also see: <http://5stardata.info/en/>)

3.6 Metadata

Metadata is information on other data and describes the characteristics of the actual data, i.e. the name of the authors, year of publication, number of pages or keywords within a publication.

3.7 Researcher

All persons registered at the UAntwerp, including (doctoral) students and outside staff performing research or who make use of the infrastructure of the UAntwerp.

3.8 Research Data

Research data are all digital or physical data – regardless of the manner in which these data are collected or stored – used or analyzed to support research findings, validate research results or underlie a scientific reasoning, discussion or calculation in the study. Research data cover the entire spectrum of raw data to processed and analyzed data included or discussed in a publication. These data can be generated data, derived or composite data, as well as self-generated data and data provided by third parties.

Examples of research data are survey results, statistics, measurement results, notebooks, graphics, computer-generated data, simulations, software developed for research purposes, computational metadata, prints, video and audio tapes, organisms, gene sequences, synthetic compounds, samples of any kind, and patient records.

3.9 Open Data

Data which may be (re)used and distributed by anyone, depending on the license used with the possible obligation(s) to state the source/authors and the principle of "share alike" – i.e. the further dissemination of data under the same initial conditions. The principle of open data relates to the entire spectrum of (research) data. The guiding principle should be that the data are as open as possible and as closed as necessary.

3.10 *Supervisor*

The person carrying the full responsibility for a research project and who is marked as 'supervisor' in the Antigoon database, or the promoter of research students, including but not limited to the supervision of a Bachelor's or Master's thesis, PhD thesis, practicum or internship.

4. Responsibilities and competences

4.1

The terms on ownership, storage, data provision and storage contained in third-party agreements take precedence over the guidelines set out in this document.

4.2 *Institutional Board*

- 4.2.1 The institutional board is responsible for the elaboration, approval and - supported by the central administration - implementation of the policy on research data management in the broad sense, it includes the active promotion of the policy on research data management among researchers, offering the necessary training and administrative support and the encouragement to exchange best practices.
- 4.2.2 The competences of the Committee Academic Bibliography will be expanded to incorporate research data management. This new Committee for Academic Output will monitor the implementation and further development of this policy document.

4.3 *Central Administration*

- 4.3.1 Research data management requires bringing together the expertise of a wide range of specialists from the central administration and faculties to develop and implement guidelines and support services on practical, ethical, legal and financial matters, as well as technical infrastructure.
- 4.3.2 The central administration provides an integrated service on research data management so that researchers with any questions can contact a single point of contact. The main actors are the Library, the Department of Research Affairs and Innovation, the ICT-Department and the Legal Office (JOC).
- 4.3.3 The central administration provides a comprehensive policy and continuous substantive knowledge building on research data management in all areas.

- 4.3.4 The central administration provides training for researchers on all aspects of research data management, including but not limited to data management plans, metadata, data repositories and copyright.
- 4.3.5 The central administration gradually develops the necessary technical and/or ICT applications and infrastructure to support the data lifecycle. This will be further clarified and concretized in the section 'Support Services and Infrastructure' of this policy.

4.4 Supervisor and Researcher

- 4.4.1 All researchers involved in the data life cycle must adhere to ethical standards and legal obligations for research data management imposed by the government, the funders of the study, and those contained within this policy.
- 4.4.2 Each researcher is primarily responsible for dealing with the research data generated within their own research.
- 4.4.3 The supervisor has the final responsibility for the research data management in its entirety for the project.
- 4.4.4 The leader of the research group is ultimately responsible for the research data management of all research on behalf of his/her research group which does not fall under a regular research project.
- 4.4.5 The supervisor and the researcher need to be aware that the UAntwerp owns all generated research data and that these data should remain at the disposal of the institution in the case of a departure.
- 4.4.6 The supervisor is responsible for composing a proper data management plan for their own use, or if necessary in accordance with established templates as stipulated by an (external) funder.
- 4.4.7 The supervisor always strives to record the direct costs related to research data management in the grant application for the research project.
- 4.4.8 The supervisor is responsible for the careful - ideally FAIR - preservation of research data related to his/her project or to his/her publication for at least 10 years after the end date of the project or publication date (and up to 20 years if the collected samples / data are subject to the Protocol of Nagoya, see <https://www.cbd.int/abs/>), as well as the necessary metadata that allows a clear interpretation of the data. If research data should be deleted or destroyed, it should be in accordance with the relevant laws and regulations, the requirements imposed by the research funders, current data policies and/or applicable codes of conduct.
- 4.4.9 The supervisor is responsible for identifying which parts of the generated research data can/should be open data in the light of the obligations imposed by the funder or publisher. The supervisor is encouraged to deposit these open data in a suitable repository granting maximum access, taking into account intellectual property rights, privacy issues and other ethical and legal considerations. The guiding principle should be that the data are as open as possible and as closed as necessary.

- 4.4.10 Each researcher is encouraged to create an ORCID and use it actively for publications or when depositing datasets, which inter alia allows to automatically monitor his/her contributions to science and provide digital access in many applications.

SUPPORT SERVICES AND INFRASTRUCTURE

This section provides an overview of the current and foreseen support services and infrastructure. Given that it takes into account the different steps of the data lifecycle, it is a dynamic document that will be updated in the light of new developments. There are systematic references to the results of the survey on research data management filled out by the group leaders of the different UAntwerp research groups in the Autumn of 2015. Overall, 67 group leaders (response rate of about 50%), more or less equally divided over the three scientific disciplines, have answered to the survey. If mention is made of 'half of the surveyed group leaders' this means 'half of those who responded'.

1. Awareness raising RDM
 - 1.1 Four group leaders claim to have a formal research data management policy within their group. Nearly 40% says they have an informal policy – In the other groups an RDM policy has not been implemented.
 - 1.2 About half of the group leaders claim that the researchers within their group are informed about the research data management policies within the group, although the actual form varies from short one-on-one interviews to extensive laboratory training.
 - 1.3 More than half of the group leaders are asking for clear guidelines and specific support on RDM, one in three wants training sessions.
 - 1.4 Two-thirds of the group leaders would like to make use of better technical infrastructures, as well as receiving the necessary support (i.e. through a helpdesk).
2. Conceptualisation: the design of the research data management strategy before or at the start of a research project, preferably on the basis of a data management plan (DMP).
 - 2.1 Raising awareness: about half of the surveyed group leaders are aware that funders will require DMPs for assigned projects in the future.
 - 2.2 Composing a DMP: at the end of 2016, ADICT will develop *DMPonline* in cooperation with i.a. UGent. This tool allows to create simple data management plans, as well as share and edit them. Moreover, ADOC and ADBIB will build up the necessary know-how and develop guidelines to assist researchers.
 - 2.3 Storing DMPs: in the Spring of 2017, ADOC and ADICT will make the necessary adjustments in the Antigoon research database so that DMPs and research projects can be linked. Depending on the international developments concerning DMPs and the accessibility of *DMPonline*, the storage of the final DMP could be implemented in IRUA.
3. Collection: the data that needs to be collected as part of a project, can be collected both within the project as well as from other sources. Irrespective of this, the data collection goes hand in

hand with choices on (analysis) methods, data formats, and associated metadata, which must be carefully documented from an integrity point of view as well. How data is collected and stored has a great influence on the subsequent archiving.

3.1 Awareness raising:

3.1.1 The survey of the group leaders shows a great need for massive storage capacity: around 25% of the group leaders say they (want to) store more than 100GB of research data each year, another 25% speak of more than 1TB, one group leader even mentioned more than 1PB.

3.1.2 Two groups out of three store their research data within the research group, while one in three deposits them on a central server. A total of four group leaders makes use of a university cloud service, while one in three use an external cloud service. In total, four spokespersons keep their research data in the cloud of the university, one in three use an external cloud. One in six group leaders use specialized software such as ELN. More than 70% of the research groups store at least a substantial portion of their data on their own personal devices. More or less the same proportions apply when asked about the needs regarding storage.

3.2 Electronic Lab Notebooks: The UAntwerp supports the use of Electronic Lab Notebooks (ELN), which i.a. can map the full research data management of a laboratory or group. At the end of 2016, the contract with PerkinElmer will be renewed. The use of ELN will be further promoted among the research groups that do not use it at the moment.

3.3 Qualtrics survey: The UAntwerp supports the use of Qualtrics, which allows to compose both simple and more complex online surveys.

3.4 Storage and sharing: for storing research data, the UAntwerp provides central servers that were recently opened up via DatAnywhere (cloud-service), which also offers opportunities to share data with external partners.

4. Processing: data processing involves the validation, quality control and cleaning of the collected data, as well as additional processing after further analysis. In order to ensure integrity and later re-use, all changes to the source data should be carefully documented, i.e. in the data management plan.

4.1 StatUA: UAntwerp's core facility StatUA supports researchers in all statistical analyses or with concrete methodological advice. If additional funding can be provided, StatUA is in favour of also being actively involved in the phases of research conceptualization and the collection of research data, and/or proactively inform and advise.

4.2 Licensing: UAntwerp owns campus licenses such as Matlab, Maple, SPSS and MapInfo

5. Archiving & Open Data: when archiving research data, from source to processed data, long-term thinking is crucial. Important issues include the quality of metadata, the question of which data should be archived for future use, the cost of archiving, monitoring the integrity and context of the dataset and the sustainability of the data format. Specifically for open data, using a unique identifier or a DOI for datasets is crucial, in addition to storing the data in the best suited repository.

5.1 Awareness raising:

- 5.1.1 One in three of the surveyed group leaders indicate that 'old' research data is destroyed over time, two out of three state these are archived.
 - 5.1.2 When asked who is responsible for the archiving, one out of six group leaders state they are responsible themselves, while two out of three mentions the individual researcher. In 10 groups one academic member is responsible for the data archiving, while in another 15 groups this task is taken on by a member of the technical staff, five groups have a designated data manager, three groups consult a university data manager.
 - 5.1.3 Asked about what happens to the research data when a researcher leaves the UAntwerp, half of the group leaders state that all of their data remain within the institution. Almost half holds a copy, while the researcher retains the original data. In one in three groups the researcher takes them with them.
 - 5.1.4 Approximately 40% of the groups have experienced (unintentional) loss of research data in the past, which was mainly due to the loss of the carrier (paper, USB stick, ICT failure) or being left without a copy of the data when a researcher leaves the group.
 - 5.1.5 Around two out of three is convinced of the added value of the university storing research data for later use (mainly within their own group), although one in six is against this. Regarding accessibility for other researchers (i.e. open data), one out of four can agree to the reuse of this data by external researchers. Half of the group leaders is willing to store in open repositories.
- 5.2 Repositories: at this time, the UAntwerp does not own a repository for research data, though ADICT and ADBIB have a positive view on the development of an institutional repository. Moreover, more and more publisher/journal/discipline-specific as well as generic data repositories are available (see i.a. <http://www.re3data.org/> and <http://www.openoar.org/>).
- 5.3 Licensing: Combining multiple datasets is complicated by the different licenses that may apply to them. The use of uniform licenses ensures that Open Data can be used optimally and can lead to innovative and valuable applications. To encourage the use of Open Data as much as possible the Flemish Government has established five model licenses that can be used by all agencies.
- 5.3.1 **A Creative Commons Zero12 License**, in which the institution renounces its intellectual property rights, as much as legally possible. This allows the user to reuse the data for any purpose, without any obligation regarding attribution.
 - 5.3.2 **Free Open Data License**: under this license the institution does not relinquish their intellectual property rights, but the data can be reused for any purpose, for free and under minimal restrictions.
 - 5.3.3 **Open Data License with Fair Compensation**: Under this license, the institution still makes its data available for re-use, but they will receive fair compensation for all types of reuse.
 - 5.3.4 **Free Open Data License for Non-Commercial Reuse**: to comply with the principle of Open Data, the data must be available under minimal restrictions for both commercial and non-commercial re-use. A distinction between types of use can be made if the institution so wishes. For commercial reuse a fair compensation may be

required, while non-commercial re-use will be made free of charge. This license governs free non-commercial re-use.

- 5.3.5 **Open Data License with Equitable Remuneration for Commercial Use:** When a distinction is made on the basis of the commercial nature and a fee is asked for reuse, this license is the counterpart of the Free License for Non-Commercial Reuse.
6. Discovery: the sustainable storage of research covers only half the road to open data. In order for open data to be found and used by others, they should be findable outside of the repository in which they are stored. This requires high-quality metadata that ensure the necessary links with the author (s) of the dataset and their personal web pages (academic bibliography, FRIS), as well as with the publications and research funding linked to the datasets.
- 6.1 ORCID: UAntwerp joined the Flemish ORCID consortium in 2016. From 2017 onwards the library will integrate the Application Program Interfaces (APIs) that ORCID provides. In the initial phase, this integration will be linked to an intranet workflow that allows researchers to record a new ORCID or link their information to an existing ORCID. Later on, it will become possible to link publications from the IRUA to the ORCID profile.
- 6.2 Research Data Repository: by analogy with the repository of publications, the library will set up a repository for research data in 2017, after the completion of the ORCID-project. In this repository, metadata from research datasets will be linked to the related publications. The research datasets themselves can be located on file servers within or outside the UAntwerp.