

Implementing a Research Data Policy at Leiden University

Fieke Schoots
Centre for Digital Scholarship
Leiden University

Laurents Sesink
Centre for Digital Scholarship
Leiden University

Peter Verhaar
Centre for Digital Scholarship
Leiden University

Floor Frederiks
Senior Policy Officer, Academic Affairs
Leiden University

Abstract

In this paper, we discuss the various stages of the institution-wide project that lead to the adoption of the data management policy at Leiden University in 2016. We illustrate this process by highlighting how we have involved all stakeholders. Each organisational unit was represented in the project teams. Results were discussed in a sounding board with both academic and support staff. Senior researchers acted as pioneers and raised awareness and commitment among their peers. By way of example, we present pilot projects from two faculties. We then describe the comprehensive implementation programme that will create facilities and services that must allow implementing the policy as well as monitoring and evaluating it. Finally, we will present lessons learnt and steps ahead. The engagement of all stakeholders, as well as explicit commitment from the Executive Board, has been an important key factor for the success of the project and will continue to be an important condition for the steps ahead.

Received 20 October 2016 ~ Revision received 31 January 2017 ~ Accepted 31 January 2017

Correspondence should be addressed to Fieke Schoots, University Libraries Leiden, PO Box 9501, 2300 RA Leiden, The Netherlands. Email: f.schoots@library.leidenuniv.nl

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introducing Leiden's Data Management Policy

In April 2016, Leiden University adopted its Regulation for Data Management.¹ The regulation was one of the outcomes of an institution-wide programme, coordinated centrally at the department of Academic Affairs. The primary reason for the formulation of an institutional policy, was the need felt to help Leiden researchers comply with data management requirements from funders and other external parties. Furthermore, it was recognised that data management is essential to enhance the transparency and integrity of research. Other outcomes of the university-wide project were related to training and support and infrastructure.²

The data management project was directed centrally by the Academic Affairs department, the office which supports the University's Executive Board. The senior policy officer for research coordinated the central project and managed the Policy and the Training and Support projects. Being close to the University Board, she could integrate data management into the University agenda and directly engage Faculty Boards. Early in the process, the University's Rector Magnificus chaired a meeting for all deans and scientific directors that contributed greatly to the awareness of the importance of data management and helped to identify common goals. The data management programme was directed by a steering committee consisting of the University's Rector Magnificus as the business executive, the deputy chairman of the Executive Board and the head or directors of the organisational units involved, Academic Affairs, Information Management, University Libraries and central ICT centre (ISSC).

The project members were representatives from all organisational units mentioned above, as well as of the Grant Office. They all brought specific expertise. The University Library had already set up services for data management. Subject librarians trained as data librarians created an information portal, gave lectures on demand and through a front office – back office construction, provided access to long term storage and expertise from the DSA certified national data archives DANS and 4TU.Research Data. The pilot projects at the faculties gave the data librarians a chance to work closely with data management pioneers, learn about specific needs and develop tailor-made services.

Representatives from the Grant Office, familiar with the planning phase of research projects as well as the valorisation or impact phase, brought knowledge on funder requirements for data management and the relation to ethical issues and data protection. Information and security managers were closely involved in the faculty pilots and familiar with the needs for storage facilities of sensitive and large volumes of data. Furthermore, the ICT department was represented by an architect and a service manager. The infrastructure project, which aimed in the first place to create a catalogue of services, was led by the head of the library's innovation department.

Most importantly, directly at the start of the university-wide programme in 2014, academic staff were actively invited to participate. Within each faculty, one or two

¹ Research Data Management Regulations Leiden University: <http://regulations.leiden.edu/research/research-data-management-regulations-leiden-university.html>

² The infrastructure project provided a catalogue of services and facilities to help researchers make a reasoned choice according to the Leiden policy. The future developments of the catalogue are described in our paper for *Liber Quarterly* (Verhaar et al., 2017).

senior researchers were asked to volunteer as data management pioneers.³ While the central teams worked on the policy, training and support and infrastructure projects, the pioneers organised pilot projects within their faculties or institutes to raise awareness of data management and to identify obstacles and concerns. The pioneers, together with support staff from the faculties, formed a sounding board, which gathered every two or three months to discuss project results, progress and issues encountered, both in the projects and at the faculties. Thus, Leiden academic staff played a crucial role in the formulation of the regulation, as well as in the definition of a subsequent three-year implementation programme. Representatives of the Leiden University Medical Center (LUMC) joined the sounding board to share experiences and exchange ideas. The LUMC had already set up a comprehensive Research ICT strategy programme, partly in national collaboration with other university medical centres and was ahead of the university programme.

Leiden's data management policy applies to three distinct stages: before, during and after the research project. It states – in line with most funder's requirements – that a data management plan (DMP) must be drawn up before data collection for a research project begins. The DMP elaborates upon the data management policy of the faculty or institute for the specific research project in question. During the research project, research data must be securely preserved, meaning that the integrity, availability and – if required – confidentiality of the data, must be guaranteed. After the project, the data must be managed in such a way that they are findable, accessible, comprehensible and reusable in the long term, from the last publication on the data or the formal completion of the research project onwards.⁴ This means that data must be stored together with the metadata, documentation and possibly the software required for its reuse. To guarantee the secure and sustainable access to the data, they should preferably be preserved in a Trusted Digital Repository. The minimum retention term for research data is ten years. The regulation also states that faculty boards have to further elaborate the regulation, either at the level of the faculty or its institutes, by working out details such as the assignment of responsibilities or the type of data that needs to be retained. An implementation period of three years, starting 2017, was agreed upon.

Faculty Pilots

In order to comment on the preliminary regulation and to identify local, discipline-specific issues regarding the management of data at their institutes, each of the six faculties was asked to do pilot projects and to report on them in the sounding board of the institutional programme. As the current state of data management practices and facilities varied from one discipline to another, some institutes, like the Faculty of Archaeology, already having a clear protocol, while others were just starting the conversation, faculties were free to choose the type of pilot project to conduct, as long as it would give them insight into the conditions for responsible data management at their institutes. All of them organised discussion sessions or interviews with fellow researchers, some did a general survey and two institutes piloted tailor-made data management training for their PhDs.

³ Leiden is a broad university with faculties for Archeology, Governance and Global Affairs, Law, Humanities, Social and Behavioural Sciences and Science. The Leiden Academic Medical Centre conducted a parallel project in national collaboration.

⁴ In line by and large with the – at the moment of writing the Regulation less prevalent – FAIR guidelines.

Social Sciences

At the Faculty of Social and Behavioural Sciences, all steps that lead from a research proposal to its final results were identified through a prospective and retrospective analysis of current research projects. Workshops and interviews quite clearly showed that there are differences in the methods of collection, the nature of the data and the obstacles to data management between the four institutes⁵. Political scientists and anthropologists, for instance, focused on the willingness of informants to participate in studies if data has to be stored in a central facility. Psychologists and educationalists, on the other hand, insisted on the long-term availability of data in view of research integrity⁶. Therefore, each institute worked out its own protocol. The Institute of Psychology, for instance, has developed a provisional storage protocol which states that all data and supplementary materials from published studies are to be stored in a central archive immediately upon completion. The protocol must ensure that all data underlying publications will be available for verification and replication. Access will probably be restricted, in most cases, to a data manager, responsible for the correct archiving, and the scientific director of the institute. The ICT department started a pilot project to explore the potential of iRODS for this purpose.⁷ Furthermore, on the basis of the university template, an institutional data management plan was developed, in which preferred options were already filled out according to the data storage protocol. The plan is now used in the mandatory data management training for PhD that is provided twice a year by the data librarians and a senior researcher from the autumn of 2016 onwards.⁸

Science

The Science Faculty adopted a central approach that focused on four themes: quality of research, education, external requirements and integrity. To enhance the quality of research and research reporting, the introduction of an electronic lab notebook was explored and finally opted for at one of the institutes. To improve data management practice, data management is to be integrated in bachelor courses. A template for a data management plan was designed in close collaboration with the data librarians for use in the training of PhD.⁹ The Grant Office provided a template to help write data management paragraphs according to the requirements of the Dutch funding agency NWO. Last but not least, the scientific directors performed a case study to assess the availability of research data. They retrieved data from a number of recent theses. As there were no standard procedures in place yet, this proved to be a rather time-consuming task. As a result of this exercise, a protocol was introduced for the Science Faculty stating that PhDs need to hand in a document, a so called ‘data form’, no later than three weeks before defending their theses. In this document, they need to indicate

⁵ Cultural Anthropology and Development Sociology, Education and Child Studies, Political Science, Psychology.

⁶ In 2012 Professor Dr. Naomi Ellemers from the Institute of Psychology co-authored with Dutch colleagues an internal report on the ‘Stapel Case’. One of the recommended steps to help reduce the risk of fraud or other forms of scientific misconduct is to ‘provide central facilities for the storage of data’ (Van Lange et al., 2012).

⁷ The iRODS consortium: <https://irods.org>.

⁸ See the Institute of Psychology scientific committee’s website: <http://www.socialsciences.leiden.edu/psychology/organisation/committees/dep-com.html>

⁹ This template, based on examples from Bath and Delft among others is, with slight modifications, in use as the university’s template. See Leiden Research Data Blog: <http://blogs.library.leiden.edu/researchdata/about-this-blog/data-management-plan>.

where the data underlying their thesis and, if applicable, the relevant software to reuse the data, is to be found. If the data is not publicly available, they transfer the data sets themselves. The document has to be signed by the supervisor and the Scientific Director.

One particular institute at the Science Faculty, the Leiden Academic Centre for Drug Research (LACDR) opted for a comprehensive approach to training, policy and infrastructure.¹⁰ As part of ‘Good Academic Research Practice’ some groups of the LACDR already shared best practices for data management, such as a common system for file naming and folder structures.¹¹ The need was felt, however, to put a provision in place to ensure that all data would be available and comprehensible after a project ended. The institute chose to integrate mandatory data management training into their PhD Education and Supervision programme, provided by data librarians and a senior researcher from the institute.¹² During such training, participants write a data management plan according to the university’s template. The plan, and its updates, is explicitly taken into account in the evaluation meetings. As all PhDs discuss their plan within their group, we see an increase, not only in the awareness of data management principles for the individual student, but also in the use of standards within groups, after six training sessions for about 70 PhDs since 2014. Thanks to this bottom-up approach, research groups started to include data management principles in their ‘Standard Operation Procedures’. One group even started to organise regular ‘Data Organisation Days’ during which all members get instructions and take time to organise their files and folders. The imminent introduction of an electronic lab notebook in the LACDR, is expected to stimulate the standardisation of workflows even more.

As the LACDR is taking the lead in data management measures in the Science Faculty, it is to be assumed that the quality of the data management as well as the sustainable availability to research data will be more advanced at this institute than at others. A recent evaluation of one hundred ‘data forms’ filled out by PhD students from all Science institutes show that only a very small minority, namely 11 projects, have made their data publicly available. Even less PhD mention having a persistent identifier. In a large number of projects there are good reasons for not sharing the data publicly, perhaps because of contractual or commercial reasons. It is more likely, however, that the majority of these datasets are not in sustainable archives simply because of the lack of appropriate facilities, or an unfamiliarity with their existence. One way of assessing the impact of the LACDR Education and Supervision programme for PhD is by reviewing the findability, the availability and the reusability of thesis data of all science institutes from 2018 onwards, when the first participants in the LACDR programme are likely to defend their thesis.

Feedback on the Regulation

While the pilots took place, data management committees were formed at the level of faculties or institutes by pioneers, scientific directors and researchers with expertise in data management. These committees commented on a previous draft of the central regulation and started to define the discipline-specific details for institutional protocols.

¹⁰ Maritta de Vries MSc, information management trainee, coordinated the pilots at the LACDR.

¹¹ See our interview with Dr. Liesbeth de Lange on the Leiden Research Data Blog:

<http://blogs.library.leiden.edu/researchdata/2016/06/03/best-practices-file-names-and-folder-structures/>

¹² PhD Portal Leiden Academic Centre for Drug Research:

<https://www.universiteitleiden.nl/en/science/drug-research/education/phd>

The policy officer from Academic Affairs, support staff from the library and the ICT department paid a visit to all committees to assess support and infrastructure needs. Among the main issues discussed across all disciplines were the protection of sensitive data and facilities to easily and safely share (large volumes of) data among colleagues, both from within and outside of the university.

Throughout 2015, all faculties reported on the results of their pilots and on the regulation, following a template in which they were actively invited to mention barriers and indicate issues to be discussed with funding agencies. In general, the reactions to the Regulation were positive. None of the faculties had encountered prohibitive objections. All welcomed the implementation period of three years, acknowledging that the actual practice is far from being the ideal situation. As to the text of the Regulation, some issues needed extra clarification, for instance the definition of research data, the target group of the regulation, or the (formal) closure of a research project. Also, responsibilities needed to be designed more precisely and better distributed among faculties and institutes. As a result, it was decided that the Faculty Board should be responsible for the further elaboration of the regulations and that the academic director of the institute should be responsible for the correct implementation of the policy. All responsibilities should be recorded in the data management plans. Finally, an extra article was added on the yearly review of the Regulation during the implementation phase, because of the pace with which developments take place. Remarks referring to the implementation of the policy, rather than the policy itself, have been integrated into the implementation programme. With some minor textual changes, the regulation has been officially adopted by the Executive Board on April 19, 2016.

The pilot projects, the meetings with the sounding board and the reports from the data management committees taught us that, overall, the regulation was welcomed as a means to improve the quality of research and research reporting. We also found, interestingly, that besides obvious differences in nature, volume and method of data collection, faculties have many concerns in common, especially with regard to large volumes or sensitive data. Subsequent training events show that researchers at Leiden University are more and more aware of the importance of sound data management. Consequently, the need for adequate facilities increases as well. Thus, we drew a list of recommendations from the pilots and projects, and translated these into a comprehensive implementation programme.

Implementation and Evaluation of the Policy

It was acknowledged that the proposed policy would have such an impact on the organisation, that it would not be realistic to make it compulsory within the short term. Therefore, an implementation term of three years was entered in the Regulation, with the exception of the immediate appointment of a Data Protection Officer. The extended implementation period gives faculties and institutes the opportunity to apply their own phasing-in plan. It also allows the organisational units responsible for providing the necessary facilities and services to set these up. The implementation programme was written in the course of 2016 and officially began in 2017.

The central aim of the programme is to implement the Regulation. There are a number of more specific objectives:

- Coordinate activities for the purpose of efficiency;

- Offer services for the entire data life cycle to support faculties and institutes during the implementation of data management;
- Develop and apply a domain architecture to make sure that all data management facilities will be manageable in the future;
- Contribute to the awareness of the benefits and necessity of data management;
- Cooperate nationally and internationally to work efficiently and join in new developments;
- Make training and support specific when needed, and generic when possible;
- Make all scholarly output visible and measurable;
- Provide insights in the costs of data management and develop costing models;
- Monitor, evaluate and review the Regulation for data management.

The programme contains a list of activities that must fulfil these objectives. While some activities can be taken up by the current organisation, a new project organisation needs to be installed for some of the other activities. The first option will be the case, for example, for the coordination of the programme, updates of the services catalogue, the exploration of costs, communication and the monitoring of the policy. The exploration and creation of storage and archiving facilities will mostly require a project organisation. The infrastructure project and the reports by the faculties showed evident gaps in the infrastructure that need to be filled as a priority. Researchers lack adequate facilities for storage during the project to protect sensitive data, freeze data, do version control, or distribute access rights. Some research projects also lack adequate provisions to archive their data or their code after the project. The pilot project with iRODS at the Institute of Psychology will be carried on to further explore its usefulness as a data vault for this institute, other disciplines, or the entire organisation. To ensure the coherent and sustainable integration of all systems involved in the research life cycle, a project aims at developing a research data domain architecture.¹³ Another project must balance pros and cons for sourcing decisions regarding data management solutions. The university must provide guidelines on the use of external facilities. A costing model for the infrastructure must be developed. Decisions may need to be made to identify which costs will be covered centrally, and which expenses will be charged to the faculties? How can indirect funding (through funding agencies) or funding from third parties contribute?

Other projects aim at implementing tools and instruments to quantify and make data management activities visible. First, a plan will be designed to monitor and review the Regulation. To start with, the development of protocols, at the level of the faculties and institutes, will be monitored on a yearly basis. At a later stage, once the protocols are in place and all the required facilities are available, the university would want to check if the protocols are put into practice according to the Regulation. Several projects in the implementation programme can help to get the management information needed for this purpose.

The first project aims at registering DMPs centrally, preferably in the Current Research Information System (CRIS). The Regulation states that each new project should have a DMP, based either on the university's, an institute's or a funder's

¹³ In a national working group under the aegis of the National Coordination Point RDM (LCRDM) the services catalogue developed in the Leiden infrastructure project will be linked to a research domain architecture. See LCRDM: <https://www.surf.nl/en/lcrdm/issues/facilities-and-data-infrastructure>

template. These plans must be kept for a period of 20 years, ten years longer than the datasets themselves, because they can account for responsibilities and decisions regarding the data, like deletion or embargos etc. Plans need to be approved by the scientific directors and information in it should be available for administrative or service objectives. The number of plans and the information in them gives valuable information to evaluate Leiden's data management activities. In a later phase, a project will be set up for the registration of data sets in the CRIS.

To produce data management plans, a digital template is to be implemented. Using an online tool will simplify the writing, registering and safe keeping of the plans, as existing tools like DMPonline show. It would be interesting to implement a link from the tool to the repository and the CRIS in such a way that all plans are automatically archived and registered.

Another project aims at making research output accessible through persistent identifiers (PI). Eventually, all data sets produced at Leiden University should be registered centrally with a persistent identifier, whether the data is in the public domain or in a local, protected facility. Parallel to the data management activities, Leiden University participates in a national project to implement the ORCID identifier for all researchers. All university processes and systems should be made ready for the optimal use of PIs. Thus, the complete scholarly output of a researcher can be presented in a comprehensive view.

In order to make sure that all researchers are aware of the implications of the new Regulation, a strategy for communication and awareness is developed in the first year and executed in the second year of the implementation programme. The emphasis will not be exclusively on the requirements, but will advocate the benefits of proper data management and the availability of tools that will make reporting easier, increase impact and allow for the reuse of data.

A one-stop-shop is planned which researchers can turn to for all questions related to data management. The idea is to publish as much generic information as possible on the intra- or internet to filter out those questions that require one-to-one contacts. Questions and answers will be registered to improve the general information. The 'shop' will be located at the newly founded Centre for Digital Scholarship at the University Libraries but represents a network of specialists from the ICT department, the Grant Office, the legal department and the copyright office, the Data Protection Officer and the data librarians. The data librarians will continue to provide tailor-made training for specific institutes, in addition to workshops on how to write data management plans. DMP services will be expanded to help with proposals according to the funding calendar.

Lessons Learnt and Steps Ahead

The success and the impact of the centralised programme that led to the adoption of the Data Management Regulation depends, for a large part, on the engagement of all relevant stakeholders. The programme manager, the senior policy officer for research, played a crucial role by actively encouraging the participation of all faculties, organising sessions to collect feedback and doing justice to legitimate concerns. The fact that this important role was taken up by the department of Academic Affairs, instead of one of the other service providing units, formed proof of the commitment of the Executive Board.

Another important factor for success was the combination of a top-down and bottom up approach, that accounted for differences in the disciplines and allowed each institute to work out details of the regulation in their own protocols. At the moment of writing, all faculties have the outlines for their protocols. Most protocols will be officially adopted once the implementation programme has provided the facilities needed.

It is important to note that the implementation programme and the steps to take are not officially laid down as yet. The programme manager only started in January 2017. The programme still needs to be discussed with all stakeholders and project plans still have to be written. This may all lead to different priorities or even different activities. Eventually, this can even lead to changes in the Regulation itself and the requirements as explicitly mentioned in the Regulation: “If there is due cause, these regulations will be revised by the Execution Board during the implementation phase”.¹⁴ We can, however, give a rough sketch of the plans.

The first step will be to collect input on the programme from all stakeholders. Members of the sounding board with representatives from all faculties are asked to collect input from their faculties as to the omissions, improvements or priorities with regard to the programme. The input – and the current state of affairs for each faculty – will be discussed in the meetings of the sounding board. A coordination team, consisting of representatives from Academic Affairs, Information management, the ICT department and the Centre for Digital Scholarship, will assist the programme manager in working out details and defining priorities. An updated plan must be evaluated and approved by the steering committee, which consists of the vice president, the Rector Magnificus and the directors of all organisational units. After that, project plans will be written, teams will be formed etc. It is important to keep up the momentum and follow up on the results of the preliminary pilot projects.

A major challenge in the implementation phase will be the simultaneous development, and introduction, of a large number of new facilities. It is crucial to assign responsibilities to all parties involved in the implementation for the ownership of systems, technical or functional management, user support, development, etc. Depending on the requirements, some systems or services will be organised centrally, others will be organised at the level of faculties. For each facility, the University must choose whether to build or to buy a solution.

Also, during the implementation phase, we will continue to collaborate with local, national and international partners. All implementation activities will be geared to the activities from the Leiden University Medical Center to ensure uniformity in services and regulations. In the existing Dutch working groups we will explore national solutions and services. We will actively follow and participate in international alliances.

The implementation programme must allow Leiden University to evaluate in three years’ time if Leiden researchers are indeed able to perform responsible data management according to the Regulation. The programme will be crucial to fulfil the ambitions of Leiden University. Sound data management contributes to the quality of research, promotes cooperation among research institutions and fits in the Open Science agenda of policy makers and funding agencies.

¹⁴ Article 20 of the Research Data Management Regulations Leiden University:
<http://media.leidenuniv.nl/legacy/research-data-management-regulations-leiden-university-def.pdf>

Acknowledgements

We would like to thank all those that contributed to the University programme for sharing their experiences, especially Prof. dr. Sander Nieuwenhuis (Institute of Psychology), and Prof. dr. Meindert Danhof (LACDR). Thanks to dr. Jacko Koster, programme manager for the implementation programme, for his input on this paper.

References

- Leiden University. (2016). Research data management regulations. Retrieved from <https://www.universiteitleiden.nl/en/research/quality-and-integrity/academic-integrity>
- Verhaar, P., Schoots, F., Sesink, L., Frederiks, F. (2017). Fostering effective data management practices at Leiden University. *LIBER Quarterly*, 27(1), 1–22. doi:10.18352/lq.10185
- Van Lange P.A.M., Buunk A.P., Ellemers N., Wigboldus D.H.J. (2012). Sharpening scientific policy after Stapel. Retrieved from <http://www.sociale-psychologie.nl/wp-content/uploads/2012/10/Van-Lange-et-al.-Sharpening-scientific-policy.pdf>