LIBER QUARTERLY
The Journal of the Association of European Research Libraries
Vol. 27, no. 1 (2017) 153–170 | e-ISSN: 2213-056X

The Development of a Research Data Policy at Wageningen University & Research: Best Practices as a Framework

Hilde van Zeeland

Wageningen University and Research
hilde.vanzeeland@wur.nl, orcid.org/0000-0003-0354-3230

Jacquelijn Ringersma

Wageningen University and Research
jacquelijn.ringersma@wur.nl, orcid.org/0000-0001-7465-9172

Abstract

The current case study describes the development of a Research Data Management policy at Wageningen University & Research, the Netherlands. To develop this policy, an analysis was carried out of existing frameworks and principles on data management (such as the FAIR principles), as well as of the data management practices in the organisation. These practices were defined through interviews with research groups. Using criteria drawn from the existing frameworks and principles, certain research groups were identified as 'best-practices': cases where data management was meeting the most important data management criteria. These best-practices were then used to inform the RDM policy. This approach shows how engagement with researchers can not only provide insight into their data management practices and needs, but directly inform new policy guidelines.

Key Words: research data management; policy; research data; data storage; data archiving

Uopen Journals | http://liberquarterly.eu/ | DOI: 10.18352/lq.10215

1. Introduction

Over the last years, there has been a growing recognition that research data are of significant value. The importance of transparency, reuse, and verifiability of research data is often emphasised. Many funders encourage or demand researchers to critically plan their Research Data Management (RDM) at the start of a project, and to retain and possibly publish datasets once the research has been completed (e.g., European Commission, 2016). Publishers, too, are showing awareness of the value of data, asking or sometimes requiring authors to share the datasets underlying their publications (e.g., PLOS, 2017). Researchers recognise the value of RDM too, since it brings incentives of various kinds, such as efficiency of research, more societal impact, and increased chances of getting funding (Hoetink, Broekhoven, & van den Hoogen, 2016). Overall, the scholarly landscape reflects a growing concern with making data FAIR: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016).

To ensure that they meet the above-discussed requirements, and that their research data assets are safely guarded and fully exploited, more and more universities are implementing research data management policies. While data policies have been most prolific in the UK and Australia (Shearer, 2015), they have also been found to exist at 44% and 41% of North American and European universities, respectively (Briney, Goben, & Zilinski, 2015; Tenopir et al., 2017).[1] Where policies are not yet in place, they are often planned or in the process of being established. Overviews of implemented data policies are provided and maintained by the Digital Curation Centre (DCC) for the UK (Horton & DCC, 2016), the Australian National Data Service (ANDS) for Australia (ANDS, 2017a), and the National Coordination Point Research Data Management (LCRDM) for the Netherlands (LCRDM, 2017).

Data policy development has often been supported or even led by university libraries. Tenopir et al. (2017) report that two-thirds of the European libraries they surveyed is involved in the development or planning of policy related to research data. This key role of the library is not surprising, as libraries tend to already provide services on data-areas such as metadata and archiving. Moreover, libraries often have a central role in the organisation, collaborating with other departments such as IT Services, as well as being close to researchers (Erway, 2013).

Although there are some common themes that most data policies seem to cover (Berchum & Grootveld, 2016; Shearer, 2015), there is great variation between the policies of different universities, and universities have taken different approaches to defining their policies (Jones, 2011). To provide policy makers with a starting point, several templates have been developed (ANDS, 2010; Budroni, Sánchez Solís, & Traub, 2017; Hall, Corey, Mann, & Wilson, n.d.). These model policy guidelines cover a set of important research data themes, such as data retention, ownership, and the responsibilities of stakeholders. Institutions can select and adapt these themes to suit their own context.

While such formats can certainly provide a useful basis, a final policy needs to meet the data practices and requirements of an organisation. Context-specific information such as available storage and archiving services need to be taken into account, together with the actual data management needs of researchers. There are tools available to help institutes with such assessments, such as the RDM Readiness Survey (LEARN, 2017), the RDM Capability Maturity Guide (ANDS, 2017b), and the Data Asset Framework developed by the University of Glasgow and the DCC (http://www.data-audit.eu/). These resources help to identify which services are in place and which ones might need further development. While these tools are not developed to inform data management policies per se, they can provide information that is key to RDM policy development, such as the data assets present and the storage/archival practices used.

To get such information first-hand, universities often approach researchers directly. One example is the Monash University in Australia, which engaged their researchers in the framing of an overall RDM strategy, including a policy. This engagement gave the policy makers valuable information about data practices, and simultaneously allowed them to communicate the benefits of good research data management, minimising the sense of risk and compliance that often comes with RDM policies (Jones, 2013). The University of Leiden in The Netherlands also actively involved research staff in the creation of their policy (Verhaar, Schoots, Sesink, & Frederiks, 2017). They did so not only to ensure that researchers could comment on the policy, but also to create a sense of support amongst them. In the UK, the Universities of Southampton and Surrey have used surveys and interviews with researchers to set up their RDM strategies. As argued by Rans and Jones (2013), such

engagement offers a complete and accurate view of the data practices and needs in an organisation:

> *"It is essential that policies and strategies are developed with reference to the particular institutional context. In order to build up an accurate picture of this context it is necessary to engage in a period of requirements-gathering and analysis. You should be aware of the scale and nature of the data to be managed. By engaging with researchers you can understand the key RDM issues they face and identify any gaps in infrastructure and desired support."* (Rans & Jones, 2013, p. 4)

As will be discussed, the RDM policy at Wageningen University & Research was also set up with the help of involvement with researchers. This approach was chosen not only to gain understanding of the existing data management practices, but mainly to identify certain 'best-practices' in the organisation: use cases that can be considered exemplary in how they manage their research data. Such best-practices are highly informative in the definition of a RDM policy, for two reasons: (1) they outline how data should be managed, thus providing a basis for policy guidelines, and (2) they relate to the data management workflows of researchers, thereby providing concrete examples of how the established policy guidelines might work in practice. As will be discussed, these best-practices were identified through the assessment with certain criteria, which were set up with the help of existing regulations and frameworks on data management.

2. The Development of a New RDM Policy

2.1. Motivation for the New Policy

Wageningen University & Research is an institution that constitutes a University and various National Research Institutes in the domains of natural sciences and life sciences. Wageningen University & Research was the first university in the Netherlands to introduce a data policy in 2014. This policy, initially only meant for the University part of the organisation, states that all chair groups and all PhD students should have a Data Management Plan (DMP). The DMP must outline what data researchers collect and how they deal with issues such as storage, version control, archiving, and sharing. The

philosophy behind this rather lean policy was the belief that once researchers would be thinking about how to manage their data, they would automatically put this data management in practice. In 2016, Academic Affairs evaluated the policy. They found that although most PhD students and chair groups indeed made a DMP, the actual implementation fell short. It was unclear whether researchers followed the practices they described in their DMPs, and what data management practices they used at all. Moreover, only a limited number of researchers archived their datasets. Academic Affairs realised that more insight and a more binding policy was needed. This was needed in particular with regards to where researchers stored their data during research and if, where and for how long they archived their datasets.

For this reason, Academic Affairs asked the existing Data Management Support unit[2] to advise them on guidelines for research data management. They requested an overview of which criteria should be met in terms of safety, accessibility, and findability. These criteria could then be translated into guidelines for an RDM policy. Data Management Support carried out this project between October and December 2016.

In its request for RDM criteria, Academic Affairs emphasised two things:

1. The criteria should follow existing laws, guidelines and frameworks that apply to research data from Wageningen University & Research.
2. The criteria should reflect the diversity of Wageningen University & Research as an organisation, in terms of the characteristics of its research data.

Regarding the first, the sources identified as relevant were the national Archives Act ('Archiefwet'), the Netherlands Code of Conduct for Scientific Practice (VSNU, 2014), and the FAIR principles of data management (Wilkinson et al., 2016). To achieve the second, several research groups from the organisation were selected (hereafter referred to as 'use cases'). Of these use cases, individual researchers or support staff were interviewed about their research data characteristics and management.

It is important to mention that the current project involved only a subset of data management aspects (data storing, archiving and registration). The final RDM policy will cover more aspects, such as research data ownership, which is a key concern in most institutional data policies (Briney, Goben, & Zilinski, 2017).

However, as these policy guidelines were set up in a separate project, the current article focuses only on the areas of data storage, archiving and registration.

2.2. Defining the Criteria for Data Management

The entire project was framed around the two phases *during research* and *after research*, with data storage taking place *during*, and data archiving and registration *after research* (Table 1).

We specifically included data registration in the process. Data registration has various benefits for researchers: it allows them to produce evidence of published datasets, to indicate the links between their datasets and published articles or other output, and to make their datasets more findable. In addition, data registration enables chair groups to provide a full overview of their research output, and gives research institutes a means of showing that they take RDM seriously. Data registration is aimed at a future situation in which research data has become an integral part of the total research output.

To define criteria of data storage, it was analysed what criteria were followed by the IT storage systems at WUR and by the data storage system used at the University of Utrecht (received through personal communication). Together, this provided a list of criteria that data storage solutions might or might not meet. Table 2 gives examples of three criteria. Appendix 1 provides the entire list.

Table 1: The phases and practices distinguished in the project, used as a basis to define the criteria and structure the interviews.

Phase	Practice	Definition
During research	Data storage	The storage of data while the research is being conducted, i.e. while data are being produced or collected (for reuse), analysed and/or processed, and prepared for publication.
After research	Data archiving	The long-term saving of data after the research has been completed, i.e. the archiving of data in a durable and searchable environment, possibly with access rights.
	Data registration	The registration of an archived dataset in a Research Information System, and the establishment of links between datasets and the publication(s) they underlie. This allows researchers to showcase their work and facilitates the creation of publication reports.

Table 2: Examples of criteria defined for the storage, archiving and registration of research data.

Practice	Criteria
Data storage	The storage solution allows the encryption of sensitive data files. The stored data is protected against physical access and disasters (e.g. fire), and has an emergency power system. The storage solution enables the sharing of data files.
Data archiving	The archive provides datasets with persistent identifiers. The archive provides metadata fields so that datasets can be described and found. The archive has a back-up and recovery system in place.
Data registration	The registration system allows the establishing of links between articles and datasets. It is possible to set up links between the datasets and other organizational records. It is possible to provide URLs to the location of the datasets.

All these were categorised as Must-have criteria (M). Appendix 1 provides the full list of criteria.

For criteria of data archiving, the above-mentioned laws and frameworks were used. The national Archive Act, the National Code of Scientific Conduct, and the FAIR principles of data management all apply to the retention and findability of datasets, providing a set of criteria regarding the archiving of data. In addition, criteria were extracted of the Data Seal of Approval (DSA, 2017), a certification for trusted digital repositories. Together, these sources provided an extensive list of criteria that applied to archiving data after research. Table 2 again gives three criteria, and Appendix 1 provides the complete list.

Criteria also needed to be defined for the registration of datasets. This is a rather new area without established frameworks. To set up criteria, the possibilities offered by the output registration systems at Wageningen University & Research and at the University of Amsterdam (personal communication) were used. Based on these, criteria of the registration of datasets were defined (see Table 2 and Appendix 1).

Once all criteria were established, the MOSCOW-method was used. This is a technique in which requirements are categorised as Must-have, Should-have, Could-have, or Won't-have. All criteria were categorised as one of these four. The outcome of this was an overview of which criteria were considered crucial to follow in data storage, archiving and registration (Must-haves), and which were not. Criteria were mostly marked as Must-haves when they

followed the law or the frameworks used (e.g. a criteria key to making data Findable according to the FAIR principles), although in some cases features were marked as Must-have based on the experience of Data Management Support staff. The Appendix provides an overview of the various criteria categorised using the MOSCOW-method.

This MOSCOW-step was an important one in the forming of the final criteria, because it clarified not only the criteria that might be met, but also their relative importance in this institutional context. The Must-haves and Should-haves in particular could be translated into guidelines for the RDM policy. Moreover, they provided a basis to evaluate the use cases' practices, i.e. to explore to what extent these already aligned to the criteria set up (most importantly, the Must-have criteria).

2.3. Comparing the Criteria to use Cases' Data Management Practices

2.3.1. Selecting use Cases

To ensure that the criteria and guidelines take into account the diverse data practices and needs in the organisation, staff members from various research groups were interviewed. Eight use cases were contacted that varied in terms of research domains, and in data characteristics such as expected dataset sizes and confidentiality levels. Table 3 shows an overview of the domains and foci of the use cases. The use cases came from both the university and the research institutes, with the majority (six) from the latter. One interview was conducted with each use case.

Table 3: Research domain and type of the eight use cases interviewed.

Research domain	Type of research
Food	Microbiology in horticulture
Plant	Molecular biology, genetics
Food	Consumer Science
Soil	Digital Soil Mapping
Animal	Genomics
Environmental	Risk assessment
Environmental	Software code development
Animal Sciences	Monitoring movements (high-speed cameras)

To assess if the pool of use cases indeed reflected this diversity, a separate analysis was carried out after completion of the interviews. A Research Data Management classification model was used for this, under development by the LCRDM (personal communication). This model distinguishes nine types of datasets of increasing complexity, based on five dimensions (the data's legal complexity, confidentiality level, lifetime span, value, and the number of disciplines it covers). The interview results indicated that the eight use cases covered eight of the nine categories. This confirms that the use cases work with datasets that vary widely, and that the interview data reflects the diversity of the organisation.

2.3.2. Preparing and Conducting the Interviews

Twenty-six interview questions were set up, covering aspects of data storage, archiving and registration practices (see Appendix 2). The interview questions were based on the defined criteria, but did not go into technical details. For example, the criteria of whether a storage solution makes back-ups and allows access control, could be covered by the general question: 'Where do you store your data during the research?' The researcher's answer to this question would suffice to asses if his/her data storage practices met the relevant criteria (e.g. if data was stored on the university's shared network drives, this would meet the criteria of back-ups and access control). This structured approach allowed us to mark whether each use case met the MOSCOW-criteria of data storage, archiving and registration.

The interviews lasted approximately one hour each. Most of the individuals interviewed were researchers, and in some cases, they also managed the data of their research group.

2.3.3. Interview Results

The interviews resulted in an overview of how the various MOSCOW-categorised criteria of storage, archiving and registration were (not) met by the various use cases. This article does not provide the results of individual use cases, as interviewees were not asked permission for this. Instead, this section presents the main findings.

In terms of storage, most of the use cases stored their research data on the institution's shared network drive or on servers at their own research unit (in some cases maintained by IT Services). These storage solutions meet almost all Must-have criteria. Some research groups used external storage solutions (e.g. USBs and hard drives), but most were conscious of the vulnerability of such storage, and therefore in the process of moving their data to other solutions. Some also used cloud storage, mostly to collaborate on data files.

Only two of the eight use cases had an established archiving protocol in place. They archived their datasets in domain-specific repositories that ensured the sustainability and findability of datasets. These repositories thereby followed the FAIR-principles. In all other use cases, researchers archived datasets only where this was explicitly requested by a journal or a financing body. Occasionally, researchers used the institutional network drives for long-term archiving. While this is a safe and durable storage environment, it does not provide data curation, nor does it allow other researchers to find and access the data files.

Only one of the eight use cases registered their datasets in the Research Output System. The researchers in this group realised the importance of making their datasets, and the links between their datasets and publications, visible. The lack of registration activities among the other use cases was often due to them not being familiar with the possibility of data registration. In other cases, researchers were aware of this, but hardly saw the additional value of registering their datasets because data citations are not rewarded like article citations are.

Finally, the interviews indicated that there was a great variation in the types of datasets that researchers worked with, in terms of both size and confidentiality level. While this was not a surprising find, it certainly emphasised that the policy guidelines should be applicable to a wide range of data types and needs.

2.4. Combining the Criteria and Interview Results to Inform Policy

As the aim was for the RDM policy guidelines to relate closely to data management practices in the organisation, the interview results were used as a framework wherever possible.[3]

The guidelines for storage were built around the storage solutions used by the use cases. These were the organisation's network drives, external storage devices, servers managed within the research groups, and cloud storage services. Of these solutions, some (largely) met the Must-have and Should-have criteria, and could be marked as 'mandatory' or 'allowed' in the guidelines. This way, the best-practices directly informed the policy guidelines. Use cases also reported solutions that could not be marked as 'allowed' or 'not allowed,' for it was impossible to predict if individual services would meet the criteria. Cloud services, for example, differ in their safety and functionality. In these cases, the guidelines approve such solutions in case the individual service used meets a set of (Must-have) criteria. This way, the guidelines cover all general storage types mentioned in the interviews, thereby closely relating to researchers' data management practices.

As two use cases reported a clear archiving protocol, the guidelines for data archiving were based on these two best-practices and on the pre-defined Must-have criteria for archiving (as listed in Appendix 1). Following the criteria, the policy lists several possible archiving options that ensure that datasets are safely retained, findable and citable. As the two use cases archived their data at domain-specific repositories that met the FAIR-principles, the guidelines stipulate that such freely chosen archiving options are allowed if certain (Must-have) criteria are met.

With Wageningen University & Research having a Research Information System in place, the only registration guideline to establish was the requirement to register archived datasets' underlying publications. This should increase the findability of datasets (the F of the FAIR principles), and hopefully in the long term also enhance researchers' perceived value of registering datasets as research output. The use case that already registered its datasets was a true best-practice, for they registered their datasets consistently, as an integrated part of their research workflow.

As explained, the interviews also uncovered a wide variety of dataset types, both in confidentiality level and in size (from large DNA sequencing files to considerably smaller Excel files with survey results). To address this variety, the policy provides guidelines for eight 'types' of data files. These eight types are based on the four information confidentiality levels used by Wageningen University & Research,[4] as well as on the distinction between large (>2TB) versus small (<2TB) datasets. This enables researchers to select storage and

archiving solutions that are most appropriate considering the characteristics of their data.

3. Discussion and Conclusion

While the above section has focused mainly on the factual information extracted during the interviews, the engagement with researchers has proven fruitful in other ways. For one, it has provided some insight into researchers' familiarity with the data support services at the institution. This has indicated what aspects of our services are known (e.g. data storage), and which might need a little push in terms of visibility (e.g. data registration). When informing researchers about the policy in the future, the less familiar services might need highlighting. Another valuable experience has been to learn more about the general data management practices in different research groups. Besides the preset questions, various issues were discussed during the interviews, such as file synchronization, version control, and the use of laboratory notebooks. Engaging with researchers about these issues is key to staying up to date on their daily data management practices and, in turn, to provide them with the necessary support.

What truly helped in setting up the guidelines was to consider the diversity of the organisation from the start. Use cases differed in what data they kept in terms of format, size and confidentiality level, and in where they stored and archived their data. This indicated that the guidelines should be sufficiently general. The final guidelines therefore outline which solutions are allowed, while also approving other (group- or domain-specific) solutions if these meet the M-criteria.

Another decision made was to take the institutionally available storage and archiving solutions as a starting point. The use cases indicated that the available solutions largely met their needs: researchers either already managed their data using the available solutions, or they could if they were to transfer their data. The guidelines could therefore be set up quickly, as the infrastructure was in place. Simultaneously, the interviews indicated for which data types the existing solutions were *not* appropriate (secret and very large datasets). A follow-up project explores what solutions might be appropriate for these data types in particular. Once identified and implemented, these solutions will be added to the guidelines.

This article has discussed how use cases have helped us define a policy, but their contribution is certainly also valuable in the follow-up steps. At the moment, the use cases are interviewed again to assess if the established guidelines are achievable. The policy guidelines are compared to their current data management practices, and any gaps encountered are bridged (e.g. datasets are registered to meet the guidelines). This is a highly useful follow-up step. It allows researchers to prepare themselves for the policy at an early stage, while it helps Research Data Management Support to focus and possibly expand our support services where needed. In doing so, the library's central role in the organisation is used to its advantage (Erway, 2013), working together with other departments such as IT and Legal services to ensure that the necessary support is established. Moreover, continuing the conversation is important for creating support among researchers and for ensuring that the guidelines are credible (Rans & Jones, 2013; Verhaar et al., 2017). Credibility of the guidelines is likely enhanced due to them being built on best-practices in the organisation. Use cases that store their data on safe solutions, archive their data in trusted repositories following the FAIR principles, and/or consistently register their datasets, show other research groups how the guidelines can be followed in practice. In fact, if these best-practice use cases do not object, their data management practices will be used in communicating the new policy, serving as exemplars to other researchers.

Needless to say, the establishment of RDM policies takes time. It is a lengthy process of which defining the guidelines is only an initial step. Once in place, policies need to be integrated into researchers' behaviour, accepted as part of their culture, and continuously reviewed and audited (ANDS, 2017b). However, as found at various universities (Rans & Jones, 2013; Verhaar et al., 2017) engagement with researchers is helpful in the further roll-out of a policy. Thanks to having involved use cases in this study, Data Management Support has an established network to engage with, which should be beneficial in the next steps of policy communication, implementation, and support.

References

ANDS (2010). *Outline of a research data management policy for Australian Universities/Institutions*. Retrieved from http://www.ands.org.au/__data/assets/pdf_file/0004/382072/datamanagementpolicyoutline.pdf.

ANDS (2017a). *ANDS project registry.* Retrieved from https://projects.ands.org.au/policy.php.

ANDS (2017b). *Creating a data management framework*. Retrieved from www.ands.org.au/guides/creating-a-data-management-framework.

Berchum, M. van, & Grootveld, M.J. (2016). Het beheren van onderzoeksdata. In *Handboek Informatiewetenschap*. [IV B 475]. Doetinchem: Vakmedianet.

Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. *Journal of Librarianship and Scholarly Communication*, *3*(2), 1–25. https://doi.org/10.7710/2162-3309.1232.

Briney, K., Goben, A., & Zilinski, L. (2017). Institutional, funder, and journal data policies. In L.R. Johnston (Ed.), *Curating research data: Practical strategies for your digital repository* (pp. 61–78). Chicago: ACLR.

Budroni, P., Sánchez Solís, B., & Traub, I.D. (2017). Development of a model policy for RDM at Austrian research institutions. In *LEARN toolkit of best practice for research data management* (pp. 14–18). https://doi.org/10.14324/000.learn.03.

DSA (2017). *Data seal of approval.* Retrieved from https://www.datasealofapproval.org/en/.

Erway, R. (2013). *Starting the conversation: university-wide research data management policy*. Retrieved from http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf.

European Commission (2016). *H2020 Programme: Guidelines on FAIR data management in Horizon 2020.* Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

Hall, N., Corey, B., Mann, W., & Wilson, T. (n.d.). *Model language for research data management policies.* Atlanta: ASERL & SURA. Retrieved from https://www.fosteropenscience.eu/content/model-language-research-data-management-policies.

Hoetink, P., Broekhoven, M., & van den Hoogen, H. (2016). *Working towards incentives.* [Powerpoint slides]. Retrieved from https://surfdrive.surf.nl/files/index.php/s/StQkBXK2cbO2RFd/download?path=%2F&files=Sessie3-1-HenkvdH.pptx.

Horton, L., & DCC (2016). *Overview of UK institution RDM policies*. Retrieved from http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies.

Jones, S. (2011). *Research data policy briefing.* Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/resource/policy/DCC_policy_briefing_2011.pdf.

Jones, S. (2013). *Bringing it all together: a case study on the improvement of research data management at Monash University.* Retrieved from http://www.dcc.ac.uk/resources/developing-rdm-services/improving-rdm-monash.

LCRDM (2017). *RDM bij universiteiten.* Retrieved from https://www.edugroepen.nl/sites/RDM_platform/RDMbijinstellingenn/RDMbijuniversiteiten.aspx.

LEARN (2017). *SURVEY: Is your institution ready for managing research data?* Retrieved from http://learn-rdm.eu/wp-content/uploads/LEARNSurvey.pdf.

PLOS (2017). *Data availability*. Retrieved from http://journals.plos.org/plosone/s/data-availability.

Rans, J., & Jones, S. (2013). *RDM strategy: Moving from plans to action*. Retrieved from http://www.dcc.ac.uk/resources/developing-rdm-services/rdm-strategy-moving-plans-action.

Shearer, K. (2015). *Comprehensive brief on research data management policies*. Retrieved from https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., … Allard, S. (2017). Research data services in European academic research libraries. *LIBER Quarterly, 27*(1), 23–44. https://doi.org/10.18352/lq.10180.

Verhaar, P., Schoots, F., Sesink, L., & Frederiks, F. (2017). Fostering effective data management practices at Leiden university. *LIBER Quarterly*, *27*(1), 1–22. https://doi.org/10.18352/lq.10185.

VSNU (2014). *The Netherlands code of conduct for scientific practice*. Retrieved from http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code_of_Conduct_for_Scientific_Practice_2012.pdf.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18.

Notes

[1] Of the 44% found by Briney et al. (2015) two-thirds concerned stand-alone data policies, and one third intellectual property policies mentioning data. Note also that this analysis covered 206 universities from the 2014 Carnegie list of universities with research activity classified as 'high' or 'very high'. Tenopir et al. (2017) surveyed 109 European university libraries (LIBER members) from 22 countries.

[2] Data Management Support is a collaborative effort between the Library, the IT department, Document Management & Logistics, and Corporate Governance & Legal Services at Wageningen University & Research.

[3] At the time of writing, the policy has not been officially approved yet. For this reason, the guidelines provided here are general. Once approved and established, the policy will appear on the website of Wageningen University & Research: http://www.wur.nl/en.

[4] These are: open (may be accessed by everybody), internal (may be accessed by students and staff from Wageningen University & Research), confidential (may be accessed by a closed group), and secret (may be accessed by a select number of individuals, often data that is sensitive or has commercial value).

Appendix

Appendix 1: Criteria for data storage, archiving and registration and whether these criteria were considered as Must-have, Should-have, Could-have, or Won't-have (MOSCOW).

Criteria	MOSCOW
Data storage	
No unauthorised access is possible.	M
A string password policy is in place.	M
The storage solution allows the encryption of sensitive data files.	M
The storage solution has an access rights system, with user accounts, roles, and authentication.	M
The stored data is protected against physical access and disasters (e.g. fire), and has an emergency power system.	M
Data can be found back for one year if it is removed by accident (by a file bin, previous versioning, daily back-up that is kept for a longer period, etc.).	M
Data can be recovered in case of a disaster (fire, flood, hardware crash, etc.).	M
The storage solution allows virus scanning.	S
The storage solution offers an availability guarantee of >99.5%.	M
The storage solution can safely keep data of all four information confidentiality levels: open, internal, confidential, and secret.	M
The storage solution allows the storing of file sizes up to TBs.	M
The system can group data files visibly into collections (datasets, collections of datasets, etc.), and it must be possible to store a file in more than one collection.	C
The storage solution enables the sharing of data files.	M
It is possible for an administrator to give other individuals certain roles with rights.	M
Admins can link files to other users.	M
The storage solution provides a version control system.	S
The data are accessible from any operating system (Windows, Linux, Mac OS).	M
The data are accessible from any smart device (Smartphone, Tablet).	C
The data can be accessed at any time.	M
The data can be accessed from anywhere.	M
It is possible to do a full-text search of the content.	C
It is possible to work together on data files simultaneously.	W

Appendix 1: (continued)

Criteria	MOSCOW
Data archiving	
Preferred formats are: .pdf .txt .sgm(l) .xml .jpg .tif .wav .shp (and otehr ESRI) .csx .tab .nc (and .cdf).	M
File formats are migrated in case of obsoleteness.	S
A Linked Data format can be used.	C
Md5 type checksums are carried out on datafile and on metadatafile level, and replacements are made in case of degradation.	M
The archive has a back-up and recovery system in place.	M
The archive provides datasets with persistent identifiers.	M
It is possible to establish links between datasets and publications.	M
Preservation time is at least 10 years.	M
The archive provides metadata fields so that datasets can be described and found.	M
Data documentation can be added to the dataset (a description of methods and techniques used to collect and analyse the data).	M
Published data are indexed in Google (and in other search engines).	S
The archive provides metadata fields so that datasets can be described and found.	M
Restricted access to datasets with end user licence is possible.	M
The archive offers an access rights system, with user accounts, roles, and authentication.	M
OAI harvesting of metadata is possible.	S
The archive is accurate, complete, authentic and reliable.	M
The archive provides clear guidelines for data citation.	S
The archive shows the number of downloads of datasets.	C
The archive shows who downloaded which datasets.	C
Quality of data entry, data storage, and data processing is controled for.	M
Data registration	
The registration system allows the establishing of links between articles and datasets.	M
It is possible to set up links between the datasets and other organizational records.	M
It is possible to provide URLs to the location of the datasets.	M
Bibliographical metadata can be added to datasets.	M

Appendix 2: The interview questions used.

Category	Question
General questions	What kind of research do you do? What kind of data do you work with (raw and analysed)? What file formats do you use for these data? Do you work with sensitive data? If so, how do you deal with this (e.g. anonymisation)? Do you use data of external sources? If so, have you made certain agreements as to how you can use this data? What is the confidentiality level of your data (open, internal, confidential, or secret)?
Storage of data	Where do you store your data? Is there a difference in the storage solution used for raw data, analysed data, and data of external sources? Do you store all the data that you generate/use? How much capacity do you need for storing your data, both during and after the research? How often do you make back-ups during the research? How and where do you do this? How long to you store data during the research ('active data')? How long do you keep the data after the research has finished? Do you ever destroy data after the research has finished? If so, how? Do you also take the value of data into account when you decide on data storage, e.g. more back-ups or safer storage solutions for data that are more difficult to reproduce? If so, how? Is it possible for other people to access and change the data after the research has finished? If so, how? Do you ever check the data during or after the research, to ensure that files are still in good condition? If so, how often and how do you do this? What kind of documentation, if any, do you add to your data?
Accessibility of data	How many people work on one dataset during the research? Do you also share data externally, or only internally? How does this sharing work (e.g. cloud storage, e-mail)? Do you use metadata to describe datasets? If so, what kind (discipline-specific or general standards)? Once the research has finished, do you save the data or metadata somewhere where other people can find them? If so, where? And are the data and/or metadata searchable there? Do you want your data and/or metadata to be findable for everybody, or only researchers in your own discipline? Are you ever asked to share your data with external parties? If so, how do you do this? Do you use licences or other agreements when you share your data with others? Who has ownership rights over the data? If you collaborate with other (commercial) organisations and/or countries, how does this influence ownership of the data?
General RDM support questions	Are there any data management practices that are going well, or not so well? Are you aware of the available data support provided by Wageningen University & Research, i.e. in terms of data storage, archiving and registration? If so, how (if at all) do you use these services?