IJDC | *Conference Pre-print*

Towards a Risk Catalogue for Data Management Plans

Franziska Weng
Kiel University

Stella Thoben
Kiel University

Abstract

Although data management and its careful planning are no new topics, there is only little literature on risk mitigation in data management plans (DMPs). We consider it a problem that DMPs do not include a structured approach for the identification or mitigation of risks, because it would instil confidence and trust in the data and its stewards, and foster the successful conduction of data-generating projects, which often are funded research projects. In this paper, we present a lightweight approach for identifying general risks in DMPs. We introduce an initial version of a generic risk catalogue for funded research and similar projects. By analysing a selection of 13 DMPs for projects from multiple disciplines published by the Research Ideas and Outcomes (RIO) journal, we demonstrate that our approach is applicable to DMPs and transferable to multiple institutional constellations. As a result, the effort for integrating risk management in data management planning can be reduced.

Received 15 December 2019 ~ *Accepted* 19 February 2020

Correspondence should be addressed to Franziska Weng, Kiel University, currently changing to GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstr. 1-3, 24148 Kiel, Germany, Email: franziskaweng@web.de, and Stella Thoben, Kiel University, Christian-Albrechts-Platz 4, 24118 Kiel, Germany, Email: thoben@rz.uni-kiel.de.

This paper was presented at International Digital Curation Conference IDCC20, Dublin, 17-19 February 2020.

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/

http://dx.doi.org/10.2218/ijdc.v15i1.697
DOI: 10.2218/ijdc.v15i1.697

Introduction

"A data management plan (DMP) is a document that describes how you will treat your data during a project and what happens with the data after the project ends" (Michener, 2015, p. 1). DMPs "serve to mitigate risks and help instil confidence and trust in the data and its stewards" (Donnelly, 2012, p. 83). "Planning for the effective creation, management and sharing of your data enables you to get the most out of your research" (Jones, 2011, p. 2). Therefore, the creation of a DMP should not only happen for obtaining a grant but also for successfully conducting the proposed project.

According to ISO 31000 (International Organization for Standardization, 2009, p. 1) a risk is "an effect of uncertainty on objectives". Data management plans should help to decrease effects of uncertainty on project objectives. We consider it a problem that neither DMPs nor funders' DMP evaluation schemes include a structured approach for the identification or mitigation of risks, since this would foster the successful conduction of data-generating projects, which often are funded research projects. We believe our approach will help funders evaluate risks of proposed projects and hence the risks of their investment options.

Data management maturity models like the Data Management Maturity (DMM)^{SM} Model (Capability Maturity Model Integration) (CMMI Institute, 2019) or the Enterprise Information Management (EIM) maturity model (Newman and Loga, 2008) are primarily designed for enterprises and may not be feasible for higher education institutions (HEIs). A rigid model for HEIs to coordinate support of data management and sharing across a diverse range of actors and processes to deliver the necessary technological and human infrastructures "cannot be prescribed since individual organisations and cultures occupy a spectrum of differences" (Jones, Pryor and Whyte, 2013, p. 4). Also, there is a potential conflict between organisational demands and scientific freedom. The Charter of Fundamental Rights of the EU contains scientific freedom as a constitutional right and researchers may view the imposition of specific data management processes as a restriction of their scientific freedom. On an even more international level, the UNESCO recommends that "Each Member State should institute procedures adapted to its needs for ensuring that, in the performance of research and development, scientific researchers respect public accountability while at the same time enjoying the degree of autonomy appropriate to their task and to the advancement of science and technology" (UNESCO, 2018, p. 119).

We consider it important, that researchers commit themselves to data management practices like e.g., ISO 31000. However, ISO 31000 (International Organization for Standardization, 2009, p. 14) defines the risk management process as a feedback loop to be conducted in organisations. Projects tend to have a much more limited scope with regard to funding and duration than organisations. Therefore, we regard the ISO 31000 risk management process as too time-consuming and of limited suitability for funded research and similar projects.

In this paper, we propose a lightweight approach for the identification of general risks in DMPs. We introduce an initial version of a generic risk catalogue for funded research and similar projects. By analysing a selection of 13 DMPs for projects from multiple disciplines[1] published by the Research Ideas and Outcomes (RIO) journal, we

[1] Anderson and Fey, 2016; Canhos, 2017; Fisher and Nading, 2016; Gatto, 2017; McWhorter, Thomas and Wright, 2016; Neylon, 2017; Stolze and Nichols, 2016; Pannell, 2016; Traynor, 2017; Wael, 2017; White, 2016; Woolfrey, 2017; Xu, Ishida and Wang, 2016

demonstrate that our approach is applicable and transferable to multiple institutional constellations. As a result, the effort for integrating risk management in data management planning can be reduced.

Related Work

Jones, Pryor and Whyte (Jones et al., 2013, p. 2) developed a guide for HEIs “to help institutions understand the key aims and issues associated with planning and implementing research data management (RDM) services”. In this guide, the authors mention data management risks for HEIs. While the upfront costs for cheap storage of active data “may be only a fraction of those quoted by central services, the risks of data loss and security breaches are significantly higher, potentially leading to far greater costs in the long term” (Jones et al., 2013, p. 13). There are “potential legal risks from using third-party services” (Jones et al., 2013, p. 14). Data selection counters the risks of “reputational damage from exposing dirty, confidential or undocumented data that has been retained long after the researchers who created it have left” (Jones et al., 2013, p. 15).

The OSCRP (Open Science Cyber Risk Profile) working group developed the OSCRP, which “is designed to help Principal Investigators (PI) and their supporting Information Technology (IT) professionals assess cybersecurity risks related to Open Science projects” (Peisert et al., 2017, p. 2). The OSCRP working group proposes that principal investigators examine risks, consequences and avenues of attack for each mission critical science asset on an inventory list, whereas assets include devices, systems, data, personnel, workflows, and other kinds of resources (Peisert et al., 2017). We regard this as a very detailed alternative to our approach, but FAIR guiding principles (Wilkinson et al., 2016, p. 8) and long-term preservation need to be added.

In 2014, Ferreira et al. (Ferreira et al., 2014, p. 41) “propose an analysis process for eScience projects using a Data Management Plan and ISO 31000 in order to create a Risk Management Plan that can complement the Data Management Plan”. The authors describe an analytical process for creating a risk management plan and “present the previous process’ validation, based on the MetaGen-FRAME project” (Ferreira et al., 2014, p. 42). Within this validation Ferreira et al. (Ferreira et al., 2014, p. 50) identify project task specific risks like “R6: Loss of metadata, denying the representation of the output information to the user via Taverna”. This risk is tailored to the use of Taverna and hence may not be relevant for the majority of funded research and similar projects. There may be projects, for which analysing specific risks for all resources may be crucial. However, a detailed risk analysis may require a considerable amount of work.

Methods

We propose a lightweight approach that can serve as a starting point to include risk management in research data management planning. It doesn’t preclude detailed approaches like OSCRP (Peisert et al., 2017) or ISO 31000 (International Organization for Standardization, 2009).

Table 1. General risk catalogue

Risk category	Risk	Possible risk source
LEGAL	Penalty for conducting unreported notifiable practices [RLEGU]	Physical sample collection
	Penalty for unpermitted usage of external data [RLEGE]	Processing external data
	Penalty for unpermitted usage of personal data [RLEGP]	Processing personal data
	Penalty for conducting inadequate data protection practices [RLEGD]	Using an external service provider for processing personal data
PRIVACY	Loss of confidentiality through sending data to an unintended recipient [RPRIR]	Correspondence
	Loss of confidentiality through interception or eavesdropping of information [RPRII]	Online data transmission
	Loss of confidentiality through loss or theft of portable storage media or devices [RPRIS]	Portable storage media or devices
	Loss of confidentiality through careless data handling by an external party [RPRIE]	Sharing data with an external party without publication purposes
TECH-NICAL	Unavailability through data corruption [RTECC]	Data processing
	Unavailability through data loss [RTECL]	Data storage
SCIENCE	Poor knowledge discovery or reusability for stakeholders cannot find the data [RSCIF]	Searchable information not planned
	Poor knowledge discovery or reusability for stakeholders cannot access the data [RSCIA]	Sharing location not planned
	Poor knowledge discovery or reusability for stakeholders cannot integrate the data [RSCII]	File format not planned
	Poor knowledge discovery or reusability for stakeholders cannot reuse the data [RSCIR]	Licensing and context information not planned
PRESER-VATION	Unsustainability in the long-term through unavailability or discontinuity of financial support [RPREU]	Preservation location not planned

Instead, we propose an approach which tries to reduce and maybe avoid the burden of a full risk management process like e.g. ISO 31000. Our approach is based on a pre-tailored and extensible general risk catalogue (Table 1) to lessen the effort required for risk management. We derived part of this risk catalogue from 29 interviews with researchers from multiple disciplines[2], which we conducted as part of project *SynFo – Synergy Creation on the operational Level of Research Data Management*. One goal of project

[2] Geo sciences (12), biology (5), humanities (5), social and behavioural sciences (4), computer science, systems engineering and electrical engineering (2) and medicine (1)

SynFo was the development of a transferable approach to improve research data management in multiple organisational constellations. In generalized content from the interviews, we identified risks entailed by interfaces of information, e.g. between researchers and data subjects or between researchers and external service providers. For the development of our approach, we also consulted the catalogues for threats and measures from the supplement of the “IT-Grundschutz” catalogues (Federal Office for Information Security (BSI), 2016) by the German Federal Office for Information Security (BSI), the FAIR guiding principles (Wilkinson et al., 2016) as well as the report and action plan from the European Commission expert group on FAIR data (Collins et al., 2018).

Our risk identification includes risks, their possible risk sources, mitigation approaches, and consequences. By analysing occurrences and mitigations of risks from our catalogue within a selection of 13 DMPs from multiple disciplines[3], published by the RIO journal, we demonstrate that our lightweight approach is applicable to DMPs and transferable to multiple institutional constellations. We evaluate the occurrences of the 15 risks in our catalogue by identifying possible risk sources in each of the selected DMPs and analyse the risk mitigations in accordance to what the authors wrote.

Risks

Legal Risks

A breach of a regulation like the GDPR or the Nagoya Protocol can result in high fines. At worst, compliance breaches can lead to reputational damages, legal disputes and enormous cost.

Penalty for conducting unreported notifiable practices

Research may include reportable research practices like the collection of physical samples regulated by the *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization*, which was transposed into EU law by Regulation (EU) No 511/2014. Under this regulation, there is a reporting obligation if the research on genetic resources is financially supported (Regulation (EU) No 511/2014, Art. 7, Sec. 1) and if the final stage of development of a product that is based on the utilisation of genetic resources (Regulation (EU) No 511/2014, Art. 7, Sec. 2). Article 11 says that “Member States shall lay down the rules on penalties applicable to infringements of Articles 4 and 7 and shall take all the measures necessary to ensure that they are applied” (Regulation (EU) No 511/2014). The Nagoya Protocol “and EU documents themselves give no guidance on penalties, each country has the liberty to determine these” (van Vegchel, 2018). Consequences may be fines of up to EUR 810,000 or even imprisonment (van Vegchel, 2018). To avoid penalties, the parties should comply strictly with the rules. The Convention on Biological Diversity publishes a detailed list of parties to the Nagoya Protocol[4].

[3] Biology (4), geo sciences (4), social and behavioural sciences (3), computer science, systems engineering and electrical engineering (1) and humanities (1)

[4] Parties to the Nagoya Protocol: https://www.cbd.int/abs/nagoya-protocol/signatories/

Penalty for unpermitted usage of external data

In many countries, data by themselves do not have inherent legal protection. Licence contracts can reach various agreements concerning terms of use. Free licences make (data) objects available for utilisation to everyone, but usage can be restricted or conditioned. Creative Commons (CC) licences and the GNU General Public License (GPL), which is specialised for free software, are widely used. Nonetheless, using CC licences can lead to conflicting rights of third parties. Publicity, personality, and privacy rights "not held by the licensor are not affected and may still affect your desired use of a licensed work" (Creative Commons, 2019). "If there are any third parties who may have publicity, privacy, or personality rights that apply, those rights are not affected by your application of a CC licence, and a reuser must seek permission for relevant uses" (Creative Commons, 2019). This e.g. holds for pictures of persons. Also, the GNU GPL licence imposes transitive obligations, e.g. "derivative programmes must also be subject to the same initial GPL conditions of ability to copy, modify, or redistribute" (Lipinski, 2012, p. 312). To mitigate the risk of unpermitted usage of external data, it is recommended to abide by the licence terms. In general, an overview about the data and the related licences can be developed in the DMP or within the framework of a data policy.

Penalty for unpermitted usage of personal data

In the EU, the General Data Protection Regulation (GDPR) governs the processing of personal data. Articles 6 and 7 of the GDPR regulate the lawfulness of processing and the conditions of consent. On an international level, the European Commission can conduct an assessment to "ensure that the level of data protection in a third country or international organization is essentially equivalent to the one established by the EU legislation" (Article 29 Data Protection Working Party, 2018, p. 5). Canada (commercial organisations), Israel, Switzerland, Japan and the USA (limited to the Privacy Shield Framework) offer an adequate level of data protection (European Commission, 2019). To avoid penalties, it is recommendable to receive written consents from data subjects including information about purpose and procedures of data processing.

Penalty for conducting inadequate data protection practices

Article 5 of the GDPR enumerates principles related to processing of personal data: the principle of lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality as well as accountability. According to Article 45 of the GDPR, "A transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection. Such a transfer shall not require any specific authorisation" (Council of the European Union and European Parliament, 2016, p. 61). Countries without adequacy, which are not classified as safe third countries, can guarantee protection in other ways, for example by appropriate safeguards (Art. 6, GDPR) or binding corporate rules (Art. 7, GDPR). To avoid penalties, it is recommendable to abide by the applicable laws. In case of doubt, researchers can contact the (data protection) authorities.

Privacy Risks

A loss of confidentiality can have adverse effects on an organisation like financial effects (Federal Office for Information Security (BSI), 2016, p. 396). These effects may also apply to a researcher who additionally may want to keep research data confidential before scientific output is published, so that research data will not be subject to theft of work.

Loss of confidentiality through sending data to an unintended recipient

Correspondence has the intrinsic potential that a researcher transmits data to an unintended recipient. This may happen accidentally or as the result of a fraudulent attack like social engineering and leads to loss of confidentiality. "Social engineering is a method used to gain unauthorised access to information or IT systems by social action" (Federal Office for Information Security (BSI), 2016, p. 419). Researchers should take extra care when sending confidential information and be aware of fraudulent attacks.

Loss of confidentiality through interception or eavesdropping of information

In the supplement of the IT-Grundschutz catalogues, the BSI specifies the threats of interception or eavesdropping of information, which entail the risk of loss of confidentiality (Federal Office for Information Security (BSI), 2016, p. 396). "Since data is sent using unforeseeable routes and nodes on the internet, the sent data should only be transmitted in an encrypted form, as far as possible" (Federal Office for Information Security (BSI), 2016, p. 3105).

Loss of confidentiality through loss or theft of portable storage media or devices

"Portable terminal devices and mobile data media in particular can be lost easily" (Federal Office for Information Security (BSI), 2016, p. 394) or even be stolen. "Whenever possible, mobile data media such as USB sticks and laptops should always be encrypted completely even if they are only occasionally used for confidential information" (Federal Office for Information Security (BSI), 2016, p. 3877).

Loss of confidentiality through careless data handling by an external party

We regard the event that researchers share data with an external party without the purpose of publication as entailing the risk of loss of confidentiality. The external party may handle confidential data carelessly. "It can frequently be observed that there are a number of organisational or technical security procedures available in organisations, but they are then undermined through careless handling of the specifications and the technology" (Federal Office for Information Security (BSI), 2016, p. 767). We recommend that researchers who share their research data to always grant specific usage rights in written form to the external party or to check if appropriate security measures are applied by the external party.

Technical Risks

Data can lose their integrity or be lost (Federal Office for Information Security (BSI), 2016, pp. 422–423) leading to the major risk of unavailability of data. Unavailability of the correct data through silent corruption can lead to usage of incorrect data and hence to the production of incorrect results. If data are unavailable, either the project may fail or researchers need to repeat their data collection and the project will be behind schedule.

Unavailability through data corruption

"The integrity of information may be impaired due to different causes, e.g. manipulations, errors caused by people, incorrect use of applications, malfunctions of software or transmission errors" (Federal Office for Information Security (BSI), 2016, p. 423). "If only accidental changes need to be detected, then checksum procedures (e.g. cyclic redundancy checks) or error-correcting codes can be used" (Federal Office for Information Security (BSI), 2016, p. 2991). Nonetheless, there may be other scenarios where these verification techniques are insufficient.

Unavailability through data loss

Data may "be lost when devices or data media are damaged, lost or stolen" (Federal Office for Information Security (BSI), 2016, p. 422), hence become unavailable. Approaches to mitigate irretrievable losses of data are for example regular backups (Federal Office for Information Security (BSI), 2016, p. 4432) or keeping copies in multiple storage locations (Reich and Rosenthal, 2000).

Science Risks

Consequences of poor discoverability and reusability of data are that researchers may unnecessarily repeat work and that scientific outputs derived from it may fail to be comprehensible, reproducible, or traceable. Problems with reproducibility and replication "can cause permanent damage to the credibility of science" (Peng, 2015, p. 32). For this reason, we named this category "Science risks".

Poor knowledge discovery or reusability for stakeholders cannot find, access, integrate or reuse the data

Making data findable, accessible, interoperable and reusable to human and computational stakeholders is a best practice approach described in the *The FAIR Guiding Principles for scientific data management and stewardship* (Wilkinson et al., 2016). Therefore, we include the risks that stakeholders cannot find, access, process or reuse data in our risk catalogue. Authors of DMPs can mitigate these risks as described by Wilkinson et al. (Wilkinson et al., 2016). We abbreviated the risk names under this risk category using the term 'poor knowledge discovery or reusability' but refer to all FAIR principles by Wilkinson et al. (Wilkinson et al., 2016).

Preservation Risk

If data are not suitably preserved, scientific outputs derived from them may fail to be comprehensible, reproducible, or traceable in the long run. Data should be stored in a trusted and sustainable digital repository (Collins et al., 2018, p. 22).

Unsustainability in the long-term through unavailability or discontinuity of financial support

A digital preservation location has the intrinsic technical risk that data become unavailable through data loss or corruption. However, preservation locations also entail the risk of becoming unavailable when their funding ends. For example, Canhos states that discontinuity of financial support is a threat to Brazil's Virtual Herbarium and its data sources (Canhos, 2017, p. 5). Authors of DMPs should consider these risks when selecting a preservation location. They can mitigate the risk that data are not preserved long-term by reviewing the external preservation location's longevity, certificates, and funding. We also suggest that attention is paid to possible migration and exit strategies like exporting and handing over data to a national data archive. This may particularly be important when the preservation location is not external.

Table 2. Risk occurrences (+) and risk occurrences with at least one mitigation (-) in the sample of 13 DMPs

DMP	RLEGU	RLEGE	RLEGP	RLEGD	RPRIR	RPRII	RPRIS	RPRIE	RTECC	RTECL	RSCIF	RSCIA	RSCII	RSCIR	RPREU
Anderson and Fey, 2016		+					+		+	-	-	-	-	-	-
Canhos, 2017		-							+	-	-	-	+	-	+
Fisher and Nading, 2016		+	+	+		+		+	+	-	-	-	-	+	-
Gatto, 2017		+							+	-	-	-	+	-	-
McWhorter, Thomas and Wright, 2016									-	-	+	-	-	-	-
Neylon, 2017			+	+	+	+		+	+	-	-	-	-	-	-
Pannell, 2016	+	+				+		-	+	-	-	-	-	-	-
Stolze and Nichols, 2016									-	-	-	-	-	-	-
Traynor, 2017			-	+	+	+	+	+	+	-	+	-	-	-	-
Wael, 2017		+	+	+	+	+	+	-	+	-	+	-	+	-	+
White, 2016		-							+	-	+	-	+	-	+
Woolfrey, 2017		-							+	-	+	-	-	-	-
Xu, Ishida and Wang, 2016	+								+	-	-	-	+	-	-

Evaluation

When applying the risk catalogue (Table 1) to the sample of 13 DMPs, we distinguish between risk occurrences themselves and risk occurrences with at least one mitigation as show in Table 2.

Table 3. Summary of risk evaluation results

Risk from catalogue	% of risk occurrences mitigated	Most often used mitigation strategy (in No. of DMPs)
Unavailability through data loss [RTECL]	100.00	Backup (8)
Poor knowledge discovery or reusability for stakeholders cannot access the data [RSCIA]	100.00	Specific repository (7)
Poor knowledge discovery or reusability for stakeholders cannot reuse the data [RSCIR]	92.31	Specific licence (9)
Unsustainability in the long-term through unavailability or discontinuity of financial support [RPREU]	76.92	Specific file formats (4); Specific data archive (4)
Poor knowledge discovery or reusability for stakeholders cannot integrate the data [RSCII]	69.23	Specific file formats (8)
Poor knowledge discovery or reusability for stakeholders cannot find the data [RSCIF]	61.54	Metadata (2)
Loss of confidentiality through careless data handling by an external party [RPRIE]	40.00	Agreement for IP rights (1); Secure external infrastructure (1)
Penalty for unpermitted usage of external data [RLEGE]	37.50	Respect usage permissions of external data (2)
Penalty for unpermitted usage of personal data [RLEGP]	25.00	Signed consent forms (1)
Unavailability through data corruption [RTECC]	15.38	Compare data from before and after transmission (1); Data quality control (1)
Loss of confidentiality through interception or eavesdropping of information [RPRII]	.00	
Penalty for conducting inadequate data protection practices [RLEGD]	.00	
Loss of confidentiality through sending data to an unintended recipient [RPRIR]	.00	
Loss of confidentiality through loss or theft of portable storage media or devices [RPRIS]	.00	
Penalty for conducting unreported notifiable practices [RLEGU]	.00	

Because risk sources and mitigations were not always explicitly mentioned in the 13 sample DMPs, we needed to make interpretations. Appendix A shows our interpretation notes. According to these interpretations, we found the mitigations shown in Appendix B.

Evaluation Results

Each of the 15 risks of our catalogue occurred in at least two of the selected 13 DMPs. Table 3 summarises our evaluation results.

Within the small sample of 13 DMPs, we found 34 distinct strategies to mitigate ten of the 15 risks of our proposed catalogue. Hence, we also found that for five of the 15 risks from our catalogue the authors did not describe any mitigation in the corresponding DMP. These risks are legal and privacy risks and they do have possible consequences like loss of reputation or project failure through theft of work. The authors of the selected DMPs overall attach highest importance to mitigating data unavailability through data loss, making data findable, accessible, interoperable and reusable as well as their long-term digital preservation. We found that two risks from our catalogue the authors mitigated in all of the selected DMPs. These risks are unavailability through data loss (RTECL) and poor knowledge discoverability or reusability for stakeholders cannot access the data (RSCIA).

Conclusion

Since we identified each risk of our catalogue in at least two of the selected DMPs, we conclude, that our risk catalogue is applicable to DMPs from multiple areas of research. In the selected DMPs, we overall find 53 of 125 (42.4%) risk occurrences not mitigated and hence see the necessity of DMP quality improvement through risk identification and mitigation planning in the data management planning phase.

We consider our approach useful to identify general risks in DMPs. We propose that after filling out a funder’s DMP template, authors of DMPs refer to a risk catalogue to identify possible risk sources and hence risks. Next, the authors should add mitigations to their DMP in the corresponding paragraph, if their DMP does not already contain one. For example, in a DMP’s paragraph in which authors write about the usage of external hard disks they should add a sentence indicating that these external hard disks will be encrypted to mitigate the risk of loss of confidentiality through loss or theft of storage media, if their DMP does not yet contain any measures mitigating this risk.

The risk catalogue may also be useful to funders, since it makes it possible for them to evaluate basic investment risks of proposed projects.

Many of the legal assertions in this article hold within the EU. Applicability to non-EU countries may vary.

We think further research on suitable risk management approaches concerning the data management of funded research and similar projects needs to be conducted.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) under the funding ID 16FDM024.

We thank Peter Wullinger and Klaus Stein for the constructive feedback.

References

Anderson, S. & Fey, J. (2016). Boulder Creek Critical Zone Observatory Data Management Plan. *Research Ideas and Outcomes*, *2*, e9419. doi:10.3897/rio.2.e9419

Article 29 Data Protection Working Party. (2018). *Adequacy Referential [WP 254 rev.01]*. Working document. European Commission. Brussels. Retrieved December 1, 2019, from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=614108

Canhos, D. A. L. (2017). Data Management Plan: Brazil's Virtual Herbarium. *Research Ideas and Outcomes*, *3*, e14675. doi:10.3897/rio.3.e14675

CMMI Institute. (2019). *Data Management Maturity (DMM) ^{SM} Model: At-a-Glance.* CMMI Institute. Retrieved November 22, 2019, from https://cmmiinstitute.com/data-management-maturity

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., ... Wittenburg, P. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data.* European Commission, Directorate-General for Research and Innovation. doi:10.2777/1524

Council of the European Union & European Parliament. (2014). *Regulation (EU) No 511/2014 of the European Parliament and of the Council of 16 April 2014 on compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization in the Union.* Retrieved December 1, 2019, from http://data.europa.eu/eli/reg/2014/511/oj

Council of the European Union & European Parliament. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).* Retrieved November 30, 2019, from http://data.europa.eu/eli/reg/2016/679/2016-05-04

Creative Commons. (2019). *Frequently Asked Questions.* Creative Commons. Retrieved December 8, 2019, from https://creativecommons.org/faq/

Donnelly, M. (2012). Data management plans and planning. In G. Pryor (Ed.), *Managing Research Data* (Chap. 5, pp. 83–104). Facet Publishing.

European Commission. (2019). *Adequacy decisions: How the EU determines if a non-EU country has an adequate level of data protection.* Retrieved December 1, 2019, from https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

European Union. (2012, October 26). *Charter of Fundamental Rights of the European Union [2012/c 326/02].* Retrieved November 30, 2019, from http://data.europa.eu/eli/treaty/char_2012/oj

Federal Office for Information Security (BSI). (2016). *IT-Grundschutz Catalogues: 15. EL. Supplement.* Version Draft. Federal Office for Information Security (BSI). Germany. Retrieved November 19, 2019, from https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/International/GSK_15_EL_EN_Draft.html

Ferreira, F., Coimbra, M. E., Bairrão, R., Viera, R., Freitas, A. T., Russo, L. M. S. & Borbinha, J. (2014). Data Management in Metagenomics: A Risk Management Approach. *International Journal of Digital Curation, 9*(1), 41–56. doi:10.2218/ijdc.v9i1.299

Fisher, J. & Nading, A. (2016). A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy. *Research Ideas and Outcomes, 2*, e8720–1. doi:10.3897/rio.2.e8720

Gatto, L. (2017). Data Management Plan for a Biotechnology and Biological Sciences Research Council (BBSRC) Tools and Resources Development Fund (TRDF) Grant. *Research Ideas and Outcomes, 3*, e11624. doi:10.3897/rio.3.e11624

International Organization for Standardization. (2009). *ISO 31000: 2009: Risk Management: Principles and Guidelines.*

Jiménez, R. C., Kuzak, M., Alhamdoosh, M., Barker, M., Batut, B., Borg, M., ... Crouch, S. (2017). Four simple recommendations to encourage best practices in research software. *F1000Research, 6*, 876. doi:10.12688/f1000research.11407.1

Jones, S. (2011). *How to Develop a Data Management and Sharing Plan.* Digital Curaton Centre. Edinburgh. Retrieved November 19, 2019, from http://www.dcc.ac.uk/resources/how-guides

Jones, S., Pryor, G. & Whyte, A. (2013). *How to Develop Research Data Management Services - a guide for HEIs.* Digital Curation Centre. Edinburgh. Retrieved November 19, 2019, from http://www.dcc.ac.uk/resources/how-guides

Lipinski, T. A. (2012). *Librarian's legal companion for licensing information resources and legal services.* Neal-Schuman Publishers, Inc.

McWhorter, J., Thomas, J. & Wright, D. (2016). Coastal Data Information Program (CDIP). *Research Ideas and Outcomes, 2*, e8827. doi:10.3897/rio.2.e8827

Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. *PLOS Computational Biology, 11*(10), e1004525. doi:10.1371/journal.pcbi.1004525

Newman, D. & Loga, D. (2008, December 5). Gartner Introduces the EIM Maturity Model. *Gartner Research Publication, ID*, (G00160425).

Neylon, C. (2017). Data Management Plan: IDRC Data Sharing Pilot Project. *Research Ideas and Outcomes, 3*, e14672. doi:10.3897/rio.3.e14672

Pannell, J. (2016). Data Management Plan for PhD Thesis "Climatic Limitation of Alien Weeds in New Zealand: Enhancing Species Distribution Models with Field Data". *Research Ideas and Outcomes, 2*, e10600. doi:10.3897/rio.2.e10600

Peisert, S., Welch, V., Adams, A., Bevier, R., Dopheide, M., LeDuc, R., ... Stocks, K. (2017). *Open Science Cyber Risk Profile (OSCRP). Version1.2*. March 2017. Retrieved November 19, 2019, from hdl:2022/21259

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance, 12*(3), 30–32. doi:10.1111/j.1740-9713.2015.00827.x

Reich, V. & Rosenthal, D. S. H. (2000). LOCKSS (lots of copies keep stuff safe). New *Review of Academic Librarianship, 6*(1), 155–161. doi:10.1080/13614530009516806

Secretariat of the Convention on Biological Diversity. (2011). *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity.* United Nations Environmental Programme. Montreal. Retrieved November 30, 2019, from hdl:20.500.11822/27555

Stolze, S. & Nichols, H. (2016). Migration of legacy data to new media formats for long-time storage and maximum visibility: Modern pollen data from the Canadian Arctic (1972/1973). *Research Ideas and Outcomes, 2*, e10269. doi:10.3897/rio.2.e10269

Traynor, C. (2017). Data Management Plan: Empowering Indigenous Peoples and Knowledge Systems Related to Climate Change and Intellectual Property Rights. *Research Ideas and Outcomes, 3*, e15111. doi:10.3897/rio.3.e15111

UNESCO. (2018). *Records of the General Conference, 39th session, Paris, 30 October-14 November 2017, v. 1: Resolutions*: 39 c/resolutions. Retrieved November 30, 2019, from https://unesdoc.unesco.org/ark:/48223/pf0000260889.page=116

van Vegchel, M. (2018, June 18). *Implementation of Nagoya Protocol: A comparison between The Netherlands, Belgium and Germany*. V.O. Patents & Trademarks. Retrieved November 30, 2019, from https://publications.vo.eu/implementation-of-nagoya-protocol

Wael, R. (2017). Data Management Plan: HarassMap. *Research Ideas and Outcomes, 3*, e15133. doi:10.3897/rio.3.e15133

White, E. (2016). Data Management Plan for Moore Investigator in Data Driven Discovery Grant. *Research Ideas and Outcomes, 2*, e10708. doi:10.3897/rio.2.e10708

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*(1). doi:10.1038/sdata.2016.18

Woolfrey, L. (2017). Data Management Plan: Opening access to economic data to prevent tobacco related diseases in Africa. *Research Ideas and Outcomes, 3*, e14837. doi:10.3897/rio.3.e14837

Xu, H., Ishida, M. & Wang, M. (2016). A Data Management Plan for Effects of particle size on physical and chemical properties of mine wastes. *Research Ideas and Outcomes, 2*, e11065. doi:10.3897/rio.2.e11065

Appendix

Appendix A

Interpretation notes

In Anderson and Fey's DMP (Anderson and Fey, 2016), we interpret the choice of txt- and csv-formats as open file formats for interoperability and we interpret the use of metadata standards as to mitigate the risk that data are not findable.

Canhos states that discontinuity of financial support is a threat to Brazil's Virtual Herbarium and its data sources (Canhos, 2017, p. 5). We interpret this as the risk that data are not preserved long-term.

In Fisher and Nading's DMP (Fisher and Nading, 2016), we find only geospatial metadata mentioned. Neither documentation nor licence information are mentioned in Fisher and Nading's DMP (Fisher and Nading, 2016).

We assume, data are enriched or combined so that a licence of the resulting data set should be derived from the source data licences in Gatto's DMP (Gatto, 2017). We evaluate Gatto's DMP (Gatto, 2017) according to the FAIR-principles for research software as proposed by Jiménez et al. (Jiménez et al., 2017).

Concerning McWhorter, Wright and Thomas' DMP (McWhorter et al., 2016), we assume that data do not have protection requirements and no risk of interception because data are made freely available for public use.

Neylon's DMP (Neylon, 2017) does not include collecting signed forms of consent from interviewees. Neylon decides not to make all data anonymous and accessible (Neylon, 2017, p. 5). Neylon's DMP (Neylon, 2017) does not include using file formats that are interoperable or allow re-use.

Concerning Pannell's DMP (Pannell, 2016), we think that it would be adequate to inform the responsible authority about the planned research project. Pannell (Pannell, 2016) does not address rights of use of external data. We interpret the used term "filterable" in the context of metadata documentation (Pannell, 2016, p. 6) as "findable".

Stolze and Nichols' DMP (Stolze and Nichols, 2016) describes migration of data from old storage media to xlsx-format and their publication.

In Traynor's DMP (Traynor, 2017), it is not clear if personal data are anonymised before they are uploaded in the infrastructure of an external service provider. Traynor's DMP (Traynor, 2017) contains no decisions for metadata capture or a specific long-term preservation location.

Wael plans to hire a consultant to do technical planning and system set up (Wael, 2017, p. 4). Wael's DMP (Wael, 2017) does not include making data interoperable. We think that utilizing academic contacts to make the research community know that the data exist as stated by Wael (Wael, 2017, p. 5) is not the same as making data findable.

According to White's DMP, the project members will develop data and software "in the open" (White, 2016, p. 3) which we interpret as making data accessible. In his DMP, White does not mention long-term preservation (White, 2016). We regard the metadata capture and user-focused documentation stated in White's DMP (White, 2016, p. 2) as making data re-usable.

In Woolfrey's DMP (Woolfrey, 2017, p. 4), metadata are captured for re-usability. Woolfrey's DMP (Woolfrey, 2017) does not include making data findable.

Xu, Ishida and Wang (Xu, Ishida and Wang, 2016) do not explicitly state in which states or countries they plan to collect physical samples. We make the interpretation, that physical samples are registered with a persistent identifier as described by Xu, Ishida and Wang (Xu, Ishida and Wang, 2016, p. 2) to make their metadata findable. Xu, Ishida and Wang (Xu, Ishida and Wang, 2016) write that for re-use and distribution, IEDA (Interdisciplinary Earth Data Alliance) would have a persistent identifier assigned to the data sets (Xu, Ishida and Wang, 2016, p. 3).

Appendix B

Table 4. Risk mitigations

DMP	Risk mitigations
Anderson and Fey, 2016	Backup (RTECL); Provide rights of use (RSCIR); Specific data archive (RPREU, RSCIA); Metadata (RSCIF); Metadata standard (RSCIF); Publicly accessible server (RSCIA); Specific file formats (RSCII)
Canhos, 2017	Aligned licensing of all data (RLEGE); Backup (RTECL); Maintain blog and social media account (RSCIF); Publicly accessible server (RSCIA); Specific file formats (RSCIR); Specific licence (RSCIR); Metadata or citation of external data (RSCIR); Standard data model (RSCIR)
Fisher and Nading, 2016	Backup (RTECL); Specific data archive (RPREU, RSCIA); Listing in national discipline specific Wiki (RSCIF); Listing on funders website (RSCIF); Specific file formats (RSCII)
Gatto, 2017	Multiple storage locations (RTECL); Specific licence (RSCIR); Collaborative software development (RPREU, RSCIF); Specific repository (RSCIA); Documentation (RSCIR); Metadata or citation of external data (RSCIR)
McWhorter, Thomas and Wright, 2016	Data are freely available for public use (RSCIR); Data quality control (RTECC); Backup (RTECL); Multiple preservation locations (RTECL, RPREU); Specific data archive (RPREU); Publicly accessible server (RSCIA); Specific file formats (RSCII); Metadata (RSCIR)
Neylon, 2017	Multiple storage locations (RTECL); Specific licence (RSCIR); Multiple preservation locations (RPREU); Specific file formats (RPREU, RSCII); Persistent identifier (RSCIF); Specific repository (RSCIA); Documentation (RSCIR)
Pannell, 2016	Secure external service infrastructure (RPRIE); Replicas in external service infrastructure (RTECL); Specific licence (RSCIR); Preservation at institution's library (RPREU); Specific file formats (RPREU, RSCII); Specific repository (RSCIA); Metadata (RSCIF); Persistent identifier (RSCIR)
Stolze and Nichols, 2016	Compare data from before and after transmission (RTECC); Backup (RTECL); Multiple storage locations (RTECL); Specific licence (RSCIR); Repository guarantees long-term availability (RPREU); Publish data descriptor in open access journal (RSCIF, RSCIR); Specific repository (RSCIA); Specific file formats (RSCII); Documentation (RSCIR); Metadata (RSCIR); Persistent identifier (RSCIR)
Traynor, 2017	Signed consent forms (RLEGP); Multiple storage locations (RTECL); Backup (RTECL); Specific licence (RSCIR); Specific file formats

DMP	Risk mitigations
	(RPREU); Specific repository (RSCIA); Specific file formats (RSCII); Documentation (RSCIR)
Wael, 2017	Multiple storage locations (RTECL); Agreement for IP rights (RPRIE); Specific licence (RSCIR); Publicly accessible server (RSCIA); Anonymization (RSCIR); Documentation (RSCIR); Specific file formats (RSCIR)
White, 2016	Respect usage permissions of external data (RLEGE); Backup (RTECL); Specific licence (RSCIR); Specific repository (RSCIA); Metadata (RSCIR); Documentation (RSCIR)
Woolfrey, 2017	Respect usage permissions of external data (RLEGE); Backup (RTECL); Data are freely available for public use (RSCIR); Specific data archive (RPREU, RSCIA); Specific file formats (RSCII); Documentation (RSCIR); Metadata (RSCIR); Metadata standard (RSCIR)
Xu, Ishida and Wang, 2016	Multiple storage locations (RTECL); Specific licence (RSCIR); Specific file formats (RPREU); Specific repository (RPREU, RSCIA); Persistent identifier for physical samples (RSCIF); Documentation (RSCIR); Metadata (RSCIR); Persistent identifier (RSCIR); Standardised vocabulary (RSCIR)