

# Datenmanagementplan

## „Concept Development for Collaborative Research Data Management Services“

Ein Projekt der  
Berlin University Alliance

Angela Ariza de Schellenberger (0000-0002-1005-5780), Evgeny Bobrov  
(0000-0002-2368-464X), Kerstin Helbig (0000-0002-2775-6751), Denise Jäckel  
(0000-0002-8720-6559), Monika Kuberek (0000-0002-1672-5271), Lea-Sophie Orozco Prado  
(0000-0001-8036-1479), Elisabeth Maria Schlagberger (0000-0002-2979-4997), Sibylle  
Söring (0000-0002-1698-3289), Britta Steinke (0000-0003-3320-5187)

05.12.2022

# 1 Administrative Informationen

## Änderungshistorie

Version	Datum	Änderungen
1.0	05.12.2022	Erster Entwurf

**Projektname:** Concept Development for Collaborative Research Data Management Services

**Förderkennzeichen:** 501\_CRDMS

### **Projektbeschreibung:**

Im Zuge der digitalen Transformation der Forschung hat das nachhaltige Forschungsdatenmanagement (FDM) deutlich an Relevanz gewonnen. Es umfasst den gesamten Forschungsdaten-Lebenszyklus<sup>1</sup>. von der Planung eines Forschungsvorhabens über die Datenerhebung, Aufbereitung, Analyse und Dokumentation der Daten bis zu ihrer langfristigen Sicherung, Archivierung und Veröffentlichung, um sie für Folgeprojekte und Anschlussforschung nachnutzbar zu machen.

Um Forschende der vier Verbundpartnerinnen der Berlin University Alliance (BUA) – Freie Universität Berlin, Humboldt-Universität zu Berlin, Technische Universität Berlin und Charité – Universitätsmedizin Berlin – im gesamten Forschungsdaten-Lebenszyklus zu unterstützen, hat sich das 2021 mit zweijähriger Laufzeit gestartete, im „Objective 5 - Sharing Resources“ angesiedelte Vorhaben „Concept Development for Collaborative Research Data Management Services“ (kurz: BUA-FDM) den gemeinsamen Aufbau und die Stärkung nachhaltiger FDM-Strukturen zum Ziel gesetzt. Zentraler Gegenstand ist dabei – auf der Grundlage einer standortspezifischen Bestands- und Bedarfserhebung – die Konzeption eines Netzwerks institutionenübergreifender FDM-Services. Als Teil dieses Netzwerks sollen mittels themen- und fachspezifischer Veranstaltungsreihen und Workshops FDM-bezogene Communities aufgebaut werden. Darüber hinaus wird ein Konzept für ein FDM-Büro bzw. eine zentrale Kontaktstelle entwickelt, das den Institutionen im Verbund als Koordinations- und Vernetzungsstelle im Handlungsfeld FDM dienen soll, um standortspezifische und übergreifende Services zu bündeln, zu koordinieren und regionale, nationale und internationale Anschlussmöglichkeiten, z. B. an die Nationale Forschungsdateninfrastruktur

---

<sup>1</sup>UK Data Archive: Research Data Lifecycle. <https://ukdataservice.ac.uk/learning-hub/research-data-management/> (zugegriffen: 30.05.2022).

tur (NFDI), zu sondieren. Die spezifischen FDM-Bedarfe von Exzellenzclustern, Sonderforschungsbereichen und anderen Verbundvorhaben bilden einen weiteren Schwerpunkt des zu erarbeitenden Konzepts, das perspektivisch in einem Anschlussprojekt umgesetzt werden soll.

## **Principle Investigators:**

### **Gesamtprojektleitung und Kontakt**

- Sibylle Söring, Freie Universität Berlin, Universitätsbibliothek  
E-Mail: sibylle.soering@fu-berlin.de

### **Projektleitung in den BUA-Einrichtungen**

- Sibylle Söring, Leitung Team Forschungsdatenmanagement, Universitätsbibliothek, Freie Universität Berlin  
E-Mail: sibylle.soering@fu-berlin.de
- Kerstin Helbig, Koordination Forschungsdatenmanagement, Computer- und Medienservice, Humboldt-Universität zu Berlin  
E-Mail: kerstin.helbig@cms.hu-berlin.de
- Monika Kuberek, Leitung Servicezentrum Forschungsdatenmanagement, Universitätsbibliothek, Technische Universität Berlin  
E-Mail: monika.kuberek@tu-berlin.de
- Dr. Evgeny Bobrov, Projektleitung Open Data und Forschungsdatenmanagement, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin, BIH QUEST Center for Responsible Research, Berlin  
E-Mail: evgeny.bobrov@bih-charite.de

### **Projektteam**

Freie Universität Berlin

- Lea-Sophie Orozco Prado, Projektkoordination  
E-Mail: l.orozco.prado@fu-berlin.de

Humboldt-Universität zu Berlin

- Dr. Denise Jäckel, Projektkoordination  
E-Mail: denise.jaeckel@hu-berlin.de

Technische Universität Berlin

- Elisabeth Schlagberger, Projektmitarbeit  
E-Mail: e.schlagberger@tu-berlin.de
- Dr. Britta Steinke (Eigenmittel), Koordination Forschungsdatenmanagement  
E-Mail: b.steinke@tu-berlin.de

Charité - Universitätsmedizin Berlin

- Dr. Angela Ariza de Schellenberger, Projektmitarbeit  
E-Mail: angela.ariza@charite.de

**Beteiligte Einrichtungen:**

Freie Universität Berlin

Humboldt-Universität zu Berlin

Technische Universität Berlin

Charité - Universitätsmedizin Berlin

**Forschungsförderer:**

Berlin University Alliance - Gefördert vom Bundesministerium für Bildung und Forschung (BMBF) und dem Land Berlin im Rahmen der Exzellenzstrategie von Bund und Ländern.

**Relevante Policies:**

- Deutsche Forschungsgemeinschaft. 2019. Version 1.1. Leitlinien zur Sicherung guter wissenschaftlicher Praxis (Kodex). doi:10.5281/zenodo.3923602.
- Deutsche Forschungsgemeinschaft. 2015. Leitlinien zum Umgang mit Forschungsdaten. [https://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/forschungsdaten/leitlinien\\_forschungsdaten.pdf](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf) (zugegriffen: 31.10.2022).

- Freie Universität Berlin. 2021. Forschungsdaten-Policy der Freien Universität Berlin. doi:10.17169/refubium-30560.
- Freie Universität Berlin. 2021. Open-Access-Policy der Freien Universität Berlin. doi:10.17169/refubium-30559.
- Humboldt-Universität zu Berlin. 2014. Forschungsdaten-Policy der Humboldt-Universität zu Berlin. <https://www.cms.hu-berlin.de/de/dl/dataman/hu-fdt-policy/view>.
- Humboldt-Universität zu Berlin. 2006. Open-Access-Policy der Humboldt-Universität zu Berlin. <https://www.ub.hu-berlin.de/de/forschen-publizieren/open-access/open-access-erklaerung>.
- Technische Universität Berlin. 2019. Forschungsdaten-Policy der Technischen Universität Berlin. <https://www.tu.berlin/go35000/>.
- Technische Universität Berlin. 2017. Open-Access-Policy der Technischen Universität Berlin. <https://www.tu.berlin/go4430/>.
- Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis. 2018 [https://www.charite.de/fileadmin/user\\_upload/portal/charite/presse/publikationen/amtl-mitteilungsblatt/2018/AMB180329-208.pdf](https://www.charite.de/fileadmin/user_upload/portal/charite/presse/publikationen/amtl-mitteilungsblatt/2018/AMB180329-208.pdf).
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities der Max-Planck-Gesellschaft. 2003 <https://openaccess.mpg.de/Berliner-Erklärung>.
- Sorbonne Declaration on research data rights. 2020 <https://www.leru.org/files/Sorbonne-declaration.pdf>.

## 2 Datenbeschreibung

Für das Vorhaben sind keine aktuellen bzw. geeigneten Forschungsdaten zur Nachnutzung verfügbar. Die im Projekt erzeugten Daten wurden daher im Zuge der Bestands-

und Bedarfserhebung zum Forschungsdatenmanagement<sup>2</sup> an den beteiligten Universitäten erhoben. Im Projekt werden verschiedene Datensätze erzeugt, die weiter unten in diesem Kapitel näher beschrieben werden.

Die Ergebnisse für die im Projekt durchgeführte Selbstevaluation mit RISE-DE (Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement)<sup>3</sup> wurden durch die Befragung verschiedener Stakeholder an den einzelnen Institutionen erzeugt. Die Bewertung des Ist- bzw. Sollzustands wurde anhand einer vierstufigen Bewertungsskala vorgenommen. Die Auswertung der Daten erfolgte mithilfe des RISE-DE-Evaluationstools<sup>4</sup>. RISE-DE basiert auf dem lose an Reifegradmodellen orientierten Research Infrastructure Self-Evaluation Framework (RISE v1.1) des Digital Curation Centre (DCC)<sup>5</sup>.

Die Ergebnisse der Bestands- und Bedarfserhebung zum Forschungsdatenmanagement wurden mithilfe einer quantitativen Online-Befragung erzeugt. Hierzu wurde die vom Computer- und Medienservice der Humboldt-Universität zu Berlin bereitgestellte Software LimeSurvey<sup>6</sup> von der Freien Universität Berlin und der Humboldt-Universität zu Berlin genutzt. An der Technischen Universität Berlin fand eine UB-eigene LimeSurvey-Instanz Verwendung. An der Charité - Universitätsmedizin Berlin wurden die Daten mit dem Befragungstool „Sociolutions QUAMP-Softwaremodul Surveyor“<sup>7</sup> erhoben, das unter einer Institutionslizenz verfügbar ist.

Im Rahmen der Bestands- und Bedarfserhebung zum Forschungsdatenmanagement entstanden verschiedene Forschungsdaten:

---

<sup>2</sup>Ariza de Schellenberger, A., Bobrov, E., Helbig, K., Jäckel, D., Kuberek, M., Orozco Prado, L.-S., Schlagberger, E. M., Söring, S., & Steinke, B.. (2022). Bestands-und Bedarfserhebung zum Forschungsdatenmanagement an den BUA-Einrichtungen. DOI: 10.5281/zenodo.7060446

<sup>3</sup>Hartmann, N. K.; Jacob, B.; Weiß, N. (2019). RISE-DE: Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement. Version 1.0. DOI: 10.5281/zenodo.2549343; Universität Potsdam (2020).

<sup>4</sup>Das Evaluationstool wurde von Hakan Pinarak (ORCID: 0000-0003-4858-9355) von der Hochschule für Angewandte Wissenschaften Hamburg entwickelt, in Zusammenarbeit mit den Autor\*innen des RISE-DE Referenzmodells Niklas K. Hartmann, Boris Jacob und Nadin Weiß.

<sup>5</sup><https://www.dcc.ac.uk/guidance/how-guides/RISE> (zugegriffen 30.05.2022).

<sup>6</sup>LimeSurvey. <https://www.cms.hu-berlin.de/de/dl/netze/vpn/software/LimeSurvey/standardseite> (zugegriffen: 30.05.2022).

<sup>7</sup>Sociolutions QUAMP <https://www.sociolutions.de/article/quamp-software.html> (zugegriffen: 23.05.2022).

- ein Datensatz pro Institution, der die Rohdaten der Ergebnisse enthält
- ein anonymisierter und formatierter Datensatz pro Institution, der die Umfrageergebnisse enthält
- ein gemeinsamer, formatierter und anonymisierter Datensatz mit den Umfrageergebnissen aller vier BUA-Einrichtungen für die generischen Fragen
- ein Datensatz, der die Verteilung der Freitextantworten auf Kategorien abbildet
- die Fragebogen-Datei, welche die generischen Fragen enthält

Hauptsächlich fielen textuelle und tabellarische Daten an. Diese werden nach Möglichkeit in offenen Formaten gespeichert:

- Textuelle Daten: txt, rtf, oder pdf
- Tabellarische Daten: csv, xlsx
- Skripte für Analysen u. Auswertungen: R oder Rmd
- Archivdateien aus der Auswertung mit OpenRefine: tar.gz

Während der Projektlaufzeit wurden die Analysen und Auswertungen mit der frei verfügbaren Programmiersprache R Statistics (Version 4.0.3)<sup>8</sup> sowie deren etablierten Open-Source-Bibliotheken (u. a. tidyverse<sup>9</sup>) angefertigt.

Für die Auswertung der qualitativen Daten der Erhebung (Freitextantworten) wurde das Open-Source-Tool OpenRefine<sup>10</sup> verwendet, das auf dem Rechner lokal installiert wird, um den Schutz von personenbezogenen Daten, falls vorhanden, zu gewährleisten.

Für die Fachliteratur wurde im Open-Source-Literaturverwaltungssystem Zotero<sup>11</sup> ein Gruppenzugang angelegt, in dem Forschungsliteratur zum Thema „Bedarfserhebungen im institutionellen Forschungsdatenmanagement“ gesammelt wird. Zugriff auf das System erhalten sämtliche Projektmitglieder.

---

<sup>8</sup>The R Project for Statistical Computing <https://www.r-project.org/> (zugegriffen: 31.10.2022).

<sup>9</sup>R packages for data science. <https://www.tidyverse.org/> (zugegriffen: 31.10.2022).

<sup>10</sup>OpenRefine. <https://openrefine.org/> (zugegriffen: 31.10.2022).

<sup>11</sup>Zotero. <https://www.zotero.org/> (zugegriffen: 31.10.2022).

Die Nachnutzung von Software anderer Urheber\*innen wird im Sinne der guten wissenschaftlichen Praxis gemäß der Software Citation Principles<sup>12</sup> zitiert.

### 3 Dokumentation und Datenqualität

Die anonymisierten Forschungsdaten der Bedarfserhebung und die Skripte der Software R werden auf dem fachübergreifenden Repositorium Zenodo<sup>13</sup> veröffentlicht. Von einer Veröffentlichung der Ergebnisse von RISE-DE wurde abgesehen (vgl. Abschnitt 6).

Im Sinne der FAIR-Prinzipien<sup>14</sup> werden die Daten in Zenodo durch Metadaten beschrieben, die sich am DataCite-Metadaten-Schema<sup>15</sup> orientieren (u.a. Abstract, freie Schlagwörter und DCC-Klassifikation). Darüber hinaus wird den Daten eine Dokumentation in Form einer README-Datei hinzugefügt, welche die durchgeführten Arbeitsschritte zur bestmöglichen Nachnutz- und Reproduzierbarkeit umfasst, und ein Codebook, das die Variablen und ihre Werte, die als Spaltenbezeichnungen im Datensatz genutzt wurden, erklärt.

Dem Datensatz wird durch das Repositorium Zenodo ein persistenter Identifier (DOI) automatisch hinzugefügt, der den Datensatz eindeutig referenzierbar macht und die Wiederauffindbarkeit gewährleistet. Zenodo ermöglicht zudem die Weitergabe der Metadaten an das frei verfügbare europäische Forschungsinformationssystem OpenAire<sup>16</sup> und an die DOI-Vergabeagentur CrossRef<sup>17</sup>.

Die Projektdaten werden getrennt nach Typ und Format in verschiedenen Verzeichnissen gespeichert (z. B. CSV-Dateien in einem „data/tabular“-Verzeichnis, alle R-Skripte gesammelt in einem „src“- oder „scripts“-Verzeichnis).

---

<sup>12</sup>Smith, Arfon M., Daniel S. Katz, Kyle E. Niemeyer, und FORCE11 Software Citation Working Group. 2016. Software citation principles. PeerJ Computer Science 2: e86. doi:10.7717/peerj-cs.86.

<sup>13</sup>Zenodo. <https://zenodo.org/> (zugegriffen: 31.10.2022).

<sup>14</sup>Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, u. a. (2016). The FAIR Guiding Principles for scientific data management and stewardship . Scientific Data, 3 (1), 160018. doi.org/10.1038/sdata.2016.18).

<sup>15</sup>DataCite Metadaten-Schema. <https://schema.datacite.org/> (zugegriffen: 18.10.2022).

<sup>16</sup>vgl. OpenAire Explore. <https://explore.openaire.eu/> (zugegriffen: 31.10.2022).

<sup>17</sup>vgl. die „Zenodo principles. “. <https://about.zenodo.org/principles/> (zugegriffen: 31.10.2022).



Die Benennung von Dateien und Verzeichnissen erfolgt nach einem einheitlichen Schema, z.B. werden Datumsangaben nach ISO 2014 formatiert: [JJJJ]-[MM]-[TT]<sup>18</sup>. Die Benennung erfolgte nach folgender Konvention:

BUA-FDM\_\_[Art der Daten]\_\_[ggf. Einrichtung]\_\_[ggf. Jahr der Erhebung/Entstehung]\_\_[Ressourcentyp plus ggf. Nummerierung]\_\_[JJJJ]-[MM]-[TT]\_\_[ggf. Versionierung].[Dateiformat]

- Beispiel 1: BUA-FDM\_\_Umfragedaten\_\_HU\_\_2021\_\_R-Skript\_\_2022-06-03.src
- Beispiel 2: BUA-FDM\_\_RISE-DE-Evaluation\_\_2021\_\_Abbildung-1\_\_2022-06-01.tiff
- Beispiel 3: BUA-FDM\_\_Projekt-DMP\_\_2022-06-03\_\_v01.pdf

Die R-Skripte werden dem tidyverse-Styleguide entsprechend formatiert<sup>19</sup>. Zur Sicherung der Nachnutzbarkeit wird der Code dokumentiert und kommentiert. Die (Nach-)Nutzung der Daten und Skripte ist durch Open-Source-Standardtools möglich. Kosten für spezialisierte Software zum Lesen, Bearbeiten oder Ausführen der Dateien fallen nicht an.

Eine fortlaufende Dokumentation und die Aufbereitung der Daten und Skripte findet bereits während der Projektlaufzeit im internen Wiki der Freien Universität Berlin statt; ihre Finalisierung erfolgt in den letzten drei Monaten der Projektlaufzeit.<sup>20</sup>

## 4 Speicherung und technische Sicherung während des Projektverlaufs

### Freie Universität Berlin:

Das erwartete Speichervolumen von maximal 5 GB wird durch einen wissenschaftlichen Speicherbereich<sup>21</sup> (im Folgenden als „Netzlaufwerk“ bezeichnet) des institutio-

---

<sup>18</sup>vgl. File naming and folder structure | CESSDA TRAINING. <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure> (zugegriffen: 18.10.2022).

<sup>19</sup>Müller K, Walthert L (2022). styler: Non-Invasive Pretty Printing of R Code. <https://github.com/r-lib/styler>, <https://styler.r-lib.org>.

<sup>20</sup>vgl. UB Wiki der Freien Universität Berlin. <https://wikis.fu-berlin.de/> (zugegriffen: 30.10.2022).

<sup>21</sup>Speicherbereiche für wissenschaftliche Daten. <https://www.fu-berlin.de/sites/scientific-data-storage> (zugegriffen: 18.10.2022).

nellen Rechenzentrums ZEDAT bereitgestellt. Dieses Netzlaufwerk unterliegt einer automatisierten, regelmäßigen, dateibasierten Backup-Routine durch das Rechenzentrum. Die Sicherungen werden vom zentralen Backup-Service der ZEDAT<sup>22</sup> auf Magnetbänder kopiert und in einem Datentresor vorgehalten. Werden Daten und Skripte lokal auf den Arbeitsrechnern der Projektgruppe erzeugt, erfolgt eine tägliche Synchronisierung durch die Mitarbeitenden mit dem Netzlaufwerk, um Datenverlust vorzubeugen. Zugriff auf die Projektdaten haben Sibylle Söring (Gesamt-Projektleitung) und Lea-Sophie Orozco Prado (Projektkoordination). Ein Zugriff für weitere Personen ist nicht erforderlich, kann aber ggf. auf Anfrage erteilt werden.

### **Humboldt-Universität zu Berlin:**

Bei der Humboldt-Universität zu Berlin erfolgt die Speicherung der Daten in der HU-Box, welche eine Kapazität von maximal 2 TB (Soft Quota) über den Computer- und Medienservice erlaubt. Derzeit haben ausschließlich Kerstin Helbig (Projektleitung) und Denise Jäckel (Projektkoordination) Zugriff auf die Projektdaten. Ein Zugriff für weitere Personen ist nicht erforderlich, kann aber ggf. auf Anfrage erteilt werden. Die tägliche automatische Speicherung und das Backup der Daten werden durch die redundante Speicherung an mehreren Standorten in gesicherten sowie gekühlten Räumen durch den Computer- und Medienservice gewährleistet.

### **Technische Universität Berlin:**

Die im Projekt erzeugten Daten werden auf Netzlaufwerken auf virtuellen Servern der Universitätsbibliothek gespeichert. Diese sind in die virtualisierte Serverumgebung der Zentraleinrichtung Campusmanagement (ZECM) integriert; die UB sichert ihre Daten regelmäßig und automatisiert bei der ZECM über das dort betriebene Backup-System. Die ZECM betreibt ein Storage-Area- Network (SAN) in gesicherten und gekühlten Räumen, das sich logisch auf zwei Standorte verteilt und in sich vollständig redundant ist.<sup>23</sup> Die von der Technischen Universität Berlin kostenfrei zur Verfügung gestellte Speicherkapazität reicht für die im Projekt anfallenden Forschungsdaten und Dokumente aus. Zugriff auf die Projektdaten haben ausschließlich Monika Kuberek (Projektleitung), Elisabeth Schlagberger und Britta Steinke (beide Projektmitarbeit).

---

<sup>22</sup>Backup-Service für Server im FU-Netz. <https://www.zedat.fu-berlin.de/Backup> (zugegriffen: 18.10.2022).

<sup>23</sup>vgl. Backup-Sicherung des ZECM. [https://www.campusmanagement.tu-berlin.de/menue/dienst\\_e/daten\\_server/speicherplatzsan/](https://www.campusmanagement.tu-berlin.de/menue/dienst_e/daten_server/speicherplatzsan/) (zugegriffen: 31.10.2022).

Ein Zugriff für weitere Personen ist nicht erforderlich, kann aber ggf. auf Anfrage erteilt werden.

Die an der Technischen Universität Berlin erhobenen Daten und Ergebnisse werden in anonymisierter Form auf einem Laufwerk der Universitätsbibliothek in der virtualisierten Serverumgebung der Technischen Universität Berlin für zehn Jahre gespeichert.

### **Charité - Universitätsmedizin Berlin:**

An der Charité - Universitätsmedizin Berlin dient der cloudbasierte Dienst „Microsoft Office 365 OneDrive for Business“ als institutionelle Infrastruktur zur Versionierung, Speicherung und Sicherung der täglichen Arbeit. Gesamtdaten werden in mit dem institutionellen Intranet verbundenen Arbeitsrechnern generiert und kontinuierlich im Rechenzentrum gespeichert. Das gleiche cloud-basierte System wird auch für die interne Zusammenarbeit über SharePoint implementiert. Für die Aufrechterhaltung dieses Workflows sind Angela Ariza de Schellenberger (Projektmitarbeit) und Evgeny Bobrov (Projektleitung) verantwortlich. Darüber hinaus sind mit diesen Daten arbeitende Mitarbeitende derselben Abteilung für die Aufrechterhaltung des Workflows verantwortlich. Zusätzlichen Zugriff auf die Daten erhielt Jan Taubitz, der in derselben Gruppe arbeitete. Die Sicherung und Speicherung von Daten mit hohem Datenschutzniveau z. B. Umfragerohdaten einschließlich Freitextantworten erfolgt auf dem Server-S. Verantwortlich für die Datensicherung sind Angela Ariza (Projektmitarbeit) und Jan Taubitz.

**Institutionsübergreifende Datenhaltung:** Darüber hinaus werden Tabellen und Skripte in Box.FU, dem Cloud-System der Freien Universität Berlin<sup>24</sup> in einem Projektordner gespeichert, um kollaboratives Arbeiten zu ermöglichen. Die Box.FU erfüllt die Anforderungen an Datenschutz-Konformität.

Daten und Skripte, die im Rahmen der Auswertung der Bestands- und Bedarfserhebung entstehen, werden mit dem Versionskontrollsystem git<sup>25</sup> versioniert. Eine zentrale Organisation des Projekt-Repositorys ist über die GitLab-Instanz der Humboldt-Universität Berlin<sup>26</sup> gegeben. Daten und Skripte in R werden zusätzlich auch auf der

---

<sup>24</sup>Box.FU- Sync & Share. <https://www.zedat.fu-berlin.de/BoxFU> (zugegriffen: 09.02.2022).

<sup>25</sup>Git. <https://git-scm.com> (zugegriffen: 18.10.2022).

<sup>26</sup>GitLab der Humboldt-Universität zu Berlin. <https://scm.cms.hu-berlin.de/> (zugegriffen: 09.02.2022).

GitLab-Instanz der Technischen Universität Berlin<sup>27</sup> gesichert.

Die Analyse und Speicherung der Daten innerhalb der Projektlaufzeit und die Datenpublikation bei Zenodo sind kostenfrei. Nach Ende der Projektlaufzeit werden sämtliche zur Veröffentlichung vorgesehenen Daten am angegebenen Ort publiziert. Über die Laufzeit hinaus findet keine weitere lokale Kuratierung der Daten statt. Die Dokumentation im Wiki der Freien Universität Berlin bleibt für die Nutzung in einem Folgeprojekt erhalten.

## **5 Rechtliche Verpflichtungen und Rahmenbedingungen**

### **Datenschutz**

Die Teilnahme an den Befragungen erfolgte an allen vier Einrichtungen anonym. Bei den Bedarfserhebungen der Freien Universität Berlin, der Humboldt-Universität zu Berlin und der Technischen Universität Berlin mit LimeSurvey konnten die IP-Adressen der Teilnehmenden nicht zurückverfolgt werden. An der Charité - Universitätsmedizin Berlin wurden die Links zur Umfrage personalisiert verschickt, die direkt personen-bezogenen Daten (Name und Mailadresse) jedoch in der Umfragesoftware QUAMP durch technische und organisatorische Maßnahmen von den Umfragedaten getrennt gehalten und nur zusammengeführt, wenn die Teilnehmenden Rechte nach DS-GVO wahrnahmen, die dies erfordern.

Für die Teilnahme mussten die Befragten an jeder Einrichtung eine Einwilligung in die Verarbeitung ihrer personenbezogener Daten geben. Die entsprechenden Texte unterschieden sich und waren mit den jeweiligen Datenschutzbeauftragten abgestimmt worden. Falls die Einwilligung nicht gegeben wurde, war keine Teilnahme an der Befragung möglich. An der Charité - Universitätsmedizin Berlin wurde die Möglichkeit gegeben, durch Hinweis an eine Funktionsadresse nachträglich der Nutzung zu widersprechen. Dies galt allerdings nur für die Dauer der Befragung selbst, da anschließend die Verbindung zwischen Umfragedaten und direkt identifizierenden Daten unwiderruflich gelöscht wurde.

---

<sup>27</sup>GitLab der Technischen Universität Berlin. <https://git.tu-berlin.de> (zugegriffen: 31.10.2022).

Die Umfragedaten enthalten quantitative und qualitative Antworten (Freitextantworten). Freitextantworten sind wiederum zu unterscheiden in Antworten auf reine Freitextfragen („Freitexte“) und in Angaben unter „Sonstiges“ oder „Andere“ („Erläuterungen“). Die quantitativen Antworten enthalten ausschließlich indirekt identifizierende demografische Daten zur Statusgruppe und Förderung sowie im zusammengeführten Gesamtdatensatz die jeweilige Einrichtung. Die Freitextantworten konnten dagegen sowohl direkt identifizierende Informationen enthalten (z.B. eigener Name oder Namen anderer, Dienstnummer, Projekttitel) als auch indirekt identifizierende Informationen (z. B. Abteilung, Arbeitsgruppe, Gerätenutzung, Art der Forschung). Bei der Freien Universität Berlin, der Humboldt-Universität zu Berlin und der Charité - Universitätsmedizin Berlin sieht die Einwilligung eine entsprechende Bearbeitung der Freitextantworten vor; in den Umfragen der Technischen Universität Berlin mussten die Teilnehmenden explizit zustimmen, in den Freitextantworten keine personenbezogenen Daten, auch nicht von Dritten, anzugeben.

Die Freitextantworten wurden zwecks Anonymisierung kategorisiert und lediglich die Ergebnisse der Kategorisierung der Antwort offen geteilt. Die Erläuterungen wurden grundsätzlich beibehalten, jedoch dort, wo sie Informationen enthielten, die voraussichtlich auf eine Person oder wenige Personen schließen lassen, vor der Veröffentlichung anonymisiert.

Die Verarbeitung der Daten erfolgte im Einklang mit der DSGVO<sup>28</sup> sowie dem Berliner Datenschutzgesetz (BDSG)<sup>29</sup>.

### **Informationssicherheit**

Neben personenbezogenen Informationen wurden auch aus Institutionssicht vertrauliche Informationen im Zuge der Zusammenführung zu einem Gesamtdatensatz getilgt. In Bezug auf die Daten ist zu beachten, dass sowohl die Datensätze von RISE-DE als auch die Datensätze aus der Bedarfserhebung sensible, nicht-personenbezogene Daten enthalten, die vertraulich behandelt und entsprechend geschützt werden müssen.

---

<sup>28</sup>EU-DSGVO. <https://www.datenschutz-grundverordnung.eu/> (zugegriffen:31.05.2022).

<sup>29</sup>[https://www.datenschutz-berlin.de/fileadmin/user\\_upload/pdf/publikationen/informationmaterialien/2018-BlnBDI\\_BlnDSG.pdf/](https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/publikationen/informationmaterialien/2018-BlnBDI_BlnDSG.pdf/) (zugegriffen: 31.05.2022).

## 6 Datenaustausch und dauerhafte Zugänglichkeit der Daten

Die anonymisierten Daten und Skripte für die Bestands- und Bedarfserhebung zum Forschungsdatenmanagement bieten sich für die Nachnutzung durch Dritte an. Daher werden Daten sowie Skripte im öffentlichen Repositorium Zenodo<sup>30</sup> publiziert. Der Veröffentlichung wird eine Dokumentation beigelegt (vgl. Abschnitt 2). Zwischenergebnisse von Verarbeitungs- und Analyseschritten, die sich sämtlich aus den bereitgestellten Daten und Skripten erzeugen lassen, sind nicht Teil der Publikation.

Die Veröffentlichung folgt den jeweiligen Empfehlungen der Open-Access-Policies und Forschungsdaten-Policies der Einrichtungen. Die Ergebnisse werden unter einer offenen Lizenz (voraussichtlich Creative Commons CC0<sup>31</sup> oder CCBY<sup>32</sup>) lizenziert.

Für das Begleitmaterial in Form von Skripten wird eine offene Software-Lizenz, (GNU-Lizenz)<sup>33</sup> vergeben.

Im Sinne der Leitlinien der DFG zur Sicherung guter wissenschaftlicher Praxis<sup>34</sup> werden die Daten für mindestens zehn Jahre durch das Repositorium öffentlich, d. h. ohne Zugangsbeschränkung, bereitgestellt. Eine Sperrfrist ist nicht erforderlich. Die Veröffentlichung findet so schnell wie möglich, spätestens jedoch innerhalb der letzten drei Monate der Projektlaufzeit statt.

Die im Kontext der Evaluation mit RISE-DE entstandenen Ergebnistabellen werden nicht publiziert, da sie sensible, institutionsspezifische Daten enthalten.

---

<sup>30</sup>vgl. Zenodo. <https://zenodo.org/> (zugegriffen: 25.05.2022).

<sup>31</sup>Creative Commons - CC0 1.0 Universell. <https://creativecommons.org/publicdomain/zero/1.0/deed.de> (zugegriffen: 25.05.2022).

<sup>32</sup>Creative Commons - Attribution 4.0 International - CC BY 4.0. <https://creativecommons.org/licenses/by/4.0/> (zugegriffen: 25.05.2022).

<sup>33</sup>vgl. GNU <https://www.gnu.org/gnu/> (zugegriffen: 27.05.2022).

<sup>34</sup>Deutsche Forschungsgemeinschaft. 2019. Leitlinien zur Sicherung guter wissenschaftlicher Praxis (Kodex). doi:10.5281/zenodo.3923602.

## **7 Verantwortlichkeiten und Ressourcen**

Hauptverantwortlich für den Umgang mit den im Projekt erzielten Forschungsdaten sind Sibylle Söring (Gesamtleitung), Kerstin Helbig, Monika Kuberek und Evgeny Bobrov (Projektleitungen) des Vorhabens gemeinsam mit ihren Institutionen. Die Einhaltung und Aktualisierung des Datenmanagementplans (DMP) wird durch die Projektmitarbeitenden sichergestellt.