SFB 1463
Offshore–
Megastrukturen

Projekt:	SFB 1463 Integrierte Entwurfs- und Betriebsmethodik für Offshore-Megastrukturen
Titel:	Richtlinie zum Umgang mit Forschungsdaten
Ersteller:	AG Datenmanagement
Letzte Änderung am:	17.01.2022
Verabschiedet am:	
Version:	2

Inhaltsverzeichnis

Präambel 3

§1 Zusammenarbeit 4

§2 Planung des Umgangs mit Forschungsdaten und der erforderlichen Ressourcen 4

§3 Nachnutzung existierender Daten 5

§4 Beachtung bestehender Standards und Dokumentation 5

§5 Metadaten 5

§6 Benennung und Ablage von Dateien 6

§7 Schutz vor Datenverlust 7

§8 Schutz vor Datenmissbrauch 8

§9 Auswahl aufhebenswerter Daten 10

§10 Langfristige Archivierung und Publikation von Forschungsdaten nach den FAIR-Prinzipien 11

Literaturverzeichnis 14

Präambel

Wissenschaft ist Teamarbeit. Dennoch haben alle Forschenden ihre eigene Arbeitsweise. Um eine effektive Zusammenarbeit zu gewährleisten, finden Sie in diesem Dokument Hinweise zu einem projektübergreifenden Umgang mit Forschungsdaten im Sonderforschungsbereich „Integrierte Entwurfs- und Betriebsmethodik für Offshore-Megastrukturen" (SFB 1463). Diese stellen sicher, dass Teilergebnisse von allen Beteiligten gleichermaßen gut verstanden und weiterverarbeitet werden können. Diese Richtlinie legt Mindeststandards fest und soll den Projektbeteiligten Sicherheit und Orientierung im Umgang mit ihren Daten geben.

Diese Richtlinie gilt ab dem Beginn des Projekts für alle projektbeteiligten Personen, deren Arbeit einen wissenschaftlichen Umgang mit Forschungsdaten erfordert. Forschungsdaten sind alle Daten, die im Verlauf des wissenschaftlichen Arbeitsprozesses erhoben und verarbeitet werden. Sie bilden die Grundlage von wissenschaftlichen Forschungsergebnissen. Darunter fallen je nach Fachgebiet unterschiedliche Arten von Daten. Der SFB 1463 schließt sich damit dem Verständnis von Forschungsdaten der Leibniz Universität Hannover (LUH) und der Deutschen Forschungsgemeinschaft (DFG) an, die hierzu u. a. Datentypen wie Messdaten, Laborwerte, audiovisuelle Informationen, Texte, methodische Testverfahren wie Fragebögen und Simulationen zählen [1, 2].

Der SFB 1463 bekennt sich zur Bedeutung von Forschungsdaten im wissenschaftlichen Erkenntnisprozess.

Zur Nachvollziehbarkeit der Forschung sowie zur Verbreitung von wissenschaftlichen Erkenntnissen ist ein verlässlicher und verantwortungsvoller Umgang mit Forschungsdaten unabdingbar. Das Forschungsdatenmanagement sichert den Zugang, die Nachnutzung, die Reproduzierbarkeit und die Qualität von Forschungsergebnissen. Daher sollen die Erstellung, Bearbeitung, Dokumentation, Sicherung sowie Aufbewahrung und nachhaltige Bereitstellung von Forschungsdaten nach anerkannten fachspezifischen Standards erfolgen und sich an den *FAIR Data Principles* orientieren [3]. So wird sichergestellt, dass Daten und Metadaten auffindbar (**f**indable), zugänglich (**a**ccessible), kompatibel zu anderen Systemen (**i**nteroperable) und nachnutzbar (**r**e-usable) sind.

Neben dieser Richtlinie gelten übergeordnete Regelungen. Zum einen sind gesetzliche Vorgaben wie die Datenschutzgrundverordnung und das Urheberrecht einzuhalten. Darüber hinaus gelten die Leitlinien zur Sicherung guter wissenschaftlicher Praxis der DFG und LUH [4, 5]. Spezielle Leit- und Richtlinien zum Umgang mit Forschungsdaten und zu Open Access sind ebenfalls einzuhalten. Dazu gehören die an der LUH geltenden und von der DFG erläuterten Leitlinien zum Umgang mit Forschungsdaten [1, 2] und die Open Access Resolution der LUH [6].

Bei Fragen zur Umsetzung dieser Grundsätze ist die Arbeitsgruppe Datenmanagement des SFB 1463 (AG Datenmanagement) zu kontaktieren.

§1 Zusammenarbeit

Mit dem Forschungsdatenmanagement verfolgt der SFB 1463 das gemeinsame Ziel, die neu gewonnenen Forschungsergebnisse nachvollziehbar und nutzbar zu machen. Die Nachvollziehbarkeit ist bei interdisziplinär vernetzten Forschungsprojekten von großer Bedeutung. Vor allem die Struktur und Zugänglichkeit von großen Datenmengen, die durch interdisziplinäre Zusammenarbeit im Forschungsbereich anfallen, sollen mithilfe dieser Richtlinie gesichert werden. Dabei sind eine gemeinsame und einheitliche Verarbeitung, Speicherung und Austausch der Forschungsdaten unabdingbar. Besonders bei der Ablage von Daten soll darauf geachtet werden, Daten gut zu dokumentieren und zu strukturieren,

sodass auch KollegInnen, die nicht an der Erstellung beteiligt waren, einen Mehrwert aus den Forschungsergebnissen ziehen können. Weitere Details zu der Speicherung und Benennung von Dateien sind unter §6 zu finden. In diesem Zusammenhang ist einer der wichtigsten Aspekte die **gemeinsame** Verfolgung der Ziele des Datenmanagements. Durch die Umsetzung der Ziele dieser Richtlinie für die interdisziplinäre Zusammenarbeit kann eine erfolgreiche Nutzung der Forschungsdaten über die Projektlaufzeit sowie über die Dauer möglicher weiterer Förderperioden und darüber hinaus gewährleistet werden. In den folgenden Paragraphen sollen die gemeinsam angestrebten Ziele des SFB 1463 zum Datenmanagement erläutert und festgehalten werden.

§2 Planung des Umgangs mit Forschungsdaten und der erforderlichen Ressourcen

Die Planung und der Umgang mit Forschungsdaten ist ein Hauptaspekt dieser Richtlinie. Wichtig für eine gemeinsame Arbeit ist, dass gemeinsame Regelungen und Festlegungen getroffen werden. Dabei müssen sowohl die technischen als auch die personellen Ressourcen auf das Forschungsprojekt abgestimmt sein.

Ganz allgemein werden unter Forschungsdaten solche Daten verstanden, die im Laufe der Bearbeitung des Projektes entstehen und nachgenutzt werden (siehe §3). Darunter fallen bspw. Messdaten, Rohdaten, Programmiercodes und Berechnungsergebnisse. Zu den klassischen Forschungsdaten gehören immer auch Metadaten, welche zusätzliche Informationen zu den vorliegenden Forschungsdaten beinhalten und in §5 näher beschrieben werden.

Für eine strukturierte Ressourcenplanung im Umgang mit Forschungsdaten soll von allen Teilprojekten des SFB 1463 ein Datenmanagementplan (DMP) erstellt werden. Der DMP beinhaltet bspw. Datenformat-Standards sowie Lieferbeziehungen und –zeitpunkte, die teilprojektspezifisch zu definieren sind. Eine allgemeingültige und zu verwendende Vorlage wird von der AG Datenmanagement erarbeitet und allen Beteiligten zur Verfügung gestellt. Die Anpassung und Erstellung der DMP liegt in der Zuständigkeit der einzelnen Teilprojekte und ist auf diese spezifisch zuzuschneiden. Mit einem gut geplanten Datenmanagement, welches diese Richtlinie und die DMP berücksichtigt, werden die wissenschaftlichen Ergebnisse nachvollziehbar und messbar.

§3 Nachnutzung existierender Daten

Die beteiligten Teilprojekte des SFB 1463 sind angehalten zu überprüfen, ob eine Nachnutzung existierender Daten sinnvoll ist. Eine Recherche zu relevanten existierenden Daten, eine sog Sekundäranalyse, sollte daher durchgeführt werden. In jeglicher Hinsicht ist bei der Nachnutzung von existierenden Daten auf die Regeln zum richtigen Zitieren zu achten. Dabei ist bei der Wiedergabe von anderen Forschungsergebnissen, Artikeln, Büchern etc. darauf zu achten, dass die korrekten Angaben zu Autor, Verlag und Jahr angegeben werden. Ggf. kann bei der Nachnutzung großer Datenmengen die Nutzung einer Literaturverwaltungssoftware (z. B. Citavi) sinnvoll sein. In jedem Fall ist eine korrekte und einheitliche Zitierweise bei der Veröffentlichung von Daten einzuhalten.

Die Nachnutzung der im SFB generierten Daten ist ebenfalls einer strukturierten Aufarbeitung zu unterziehen. Nachvollziehbarkeit ist ein essenzieller Aspekt bei der Aufbereitung von Forschungsdaten, die bspw. durch zusätzliche Metadaten gesichert wird. Das Vorgehen zur Auswahl aufhebenswerter und zu publizierender Daten wird in §9 beschrieben.

§4 Beachtung bestehender Standards und Dokumentation

Übergeordnete Regelwerke und Richtlinien sind neben dieser Richtlinie uneingeschränkt gültig und zwingend zu beachten. Dazu zählen insbesondere gesetzliche Grundlagen wie die europäische Datenschutzgrundverordnung [7] oder das Urheberrecht [8]. Neben den gesetzlichen Grundlagen sollen weitere Leit- und Richtlinien eingehalten werden. Dazu zählen die Leitlinien zur Sicherung guter wissenschaftlicher Praxis der DFG [2] und der LUH [1]. Ein Hauptziel ist es, die Forschungsdaten uneingeschränkt verfügbar zu machen. Dafür soll die Open Access Resolution der LUH [6] und die von der DFG erläuterten Leitlinien zum Umgang mit Forschungsdaten [2]eingehalten werden.

Bei der Veröffentlichung der Forschungsdaten sind Geheimhaltungsvereinbarungen u. ä. besonders beim Umgang mit sensiblen Daten zu beachten. Getroffene Vereinbarungen sind einzuhalten und bei der Veröffentlichung von Forschungsergebnissen zu berücksichtigen, sodass unter Umständen nur Teilergebnisse veröffentlicht werden können.

§5 Metadaten

Metadaten beinhalten Beschreibungen und Details zu vorliegenden Forschungsdaten. Dabei sollen diese das Verständnis sowie die Weiter- und Nachnutzung der eigentlichen Forschungsdaten ermöglichen. Unter Metadaten werden z. B. Angaben zu Ort, Inhalt, Methodik, Erzeugungsprozess, Technologie, Dokumentation benötigter Software, Datum/Zeit, Quellen, Identifikationen (z. B. DOI) etc. verstanden. Die Angaben sind vollständig für die jeweiligen Daten zu erfassen und bei Änderungen anzupassen. Dabei sind die Angaben in den Metadaten datentypabhängig. Bei der Erfassung von Messdaten sind bspw. die Messrate und die Art der angewendeten Messtechniken in den Metadaten zu berücksichtigen, wohingegen bei einer Softwarenutzung umfangreiche Metadaten bezüglich Betriebssystems, Version etc. relevant sind.

Das Beispiel der projektinternen Nutzung der Software DeSiO soll die Vorteile bei Erstellung von Metadaten verdeutlichen. DeSiO ist ein Finite-Elemente-Programm, welches vom Teilprojekt Z01 weiterentwickelt und für die Modellierung des digitalen Zwillings herangezogen wird. In diesem Fall sollten die Metadaten mindestens folgende Informationen beinhalten:

- Angaben zum Betriebssystem, auf dem DeSiO angewendet werden kann
- weitere Voraussetzungen (z. B. Installation zusätzlicher Programme etc.)
- Version und ggf. Änderungen gegenüber vorherigen Versionen
- Eingabeformate (z. B. windIO)

Metadaten sollten genau solange erfasst und gespeichert werden wie die generierten Forschungsdaten [9]. Dabei ist darauf zu achten, dass die Metadaten einfach zugänglich sind. Es empfiehlt sich für die Erstellung von Metadaten die Verwendung des *.TXT- oder *.PDF-Formats. Ein Katalog für fachspezifische Standards von Metadaten kann in [10] gefunden werden.

§6 Benennung und Ablage von Dateien

Forschungsdaten sind bevorzugt auf den institutseigenen Servern der Teilprojekte zu speichern. Zum Ablegen von projektübergreifenden Daten wie bspw. Protokollen kann das Seafile des SFB 1463 genutzt werden [11]. Die Speicherkapazität ist jedoch auf 40 GB beschränkt und ist daher nicht als Alternative zum Institutsserver zu sehen. Die doppelte Sicherung der Daten ist zu gewährleisten, z. B. durch Backups der Server, vgl. §7.

Für den Austausch von Dateien ist im Seafile ein Austauschordner eingerichtet, welcher regelmäßig gelöscht wird und daher nicht zur dauerhaften Ablage dient. Der Pfad lautet:

Bibliotheken/sfb-megastrukturen/02_Projektphase/07_Austauschordner

Sind größere Datenmengen zu tauschen oder zu verschicken, wird die Nutzung des Download Ticket Service des LUIS Hannover empfohlen [12]. Dort können alle Projektbeteiligten einen Link zum Herunterladen der gewünschten Daten erstellen, welcher dann an die gewünschte Person (sowohl LUH-interne als auch externe Personen) bspw. per E-Mail versendet werden kann.

Für den Austausch und die Ablage von Programmiercode soll GitLab verwendet werden [13]. Als Bestandteil der LUH-Projektablage erfolgt eine Anmeldung mit den gleichen Zugangsdaten wie zur Projektverwaltung über Seafile.

Eine Konvention für die Benennung der Seafile-Ordner wurde bereits eingeführt. Bei der Ergänzung von Projektordnern in der ersten Ebene (bspw. *07_Austauschordner*) erfolgt die Benennung nach Folgendem Prinzip: *00_Ordnername*. Dabei ist *00* durch eine noch nicht vergebene fortlaufende Zahl zu ersetzen und dem Ordnernamen ein aussagekräftiger Titel zu geben. Die Struktur der untergeordneten Ordnerebenen wird nicht durch diese Richtline geregelt.

Die Benennung von Dateien erfolgt nach einem ähnlichen Prinzip wie die der Ordner. Es wurde entschieden, die folgende Konvention bei der Dateibenennung einzuhalten:

JJMMTT_KürzelDerBearbeiterMitDreiBuchstaben_NameDerDatei_v01

Dabei stehen die ersten sechs Zahlen für das Erstellungsdatum nach dem Format JJMMTT. Durch einen Unterstrich getrennt wird das aus drei Buchstaben bestehende Kürzel für das Teilprojekt, die BearbeiterInnen oder die Arbeitsgruppe/Cluster angegeben. Eine entsprechende Tabelle mit Kürzeln für Arbeitsgruppen und Cluster ist in Tabelle 1 zu finden. Ein weiterer Unterstrich leitet den Namen der Datei ein. Die einzelnen Wörter werden dabei durch Groß- und Kleinschreibung kenntlich gemacht. Jedes neue Wort wird mit einem Großbuchstaben ohne Leerzeichen angefangen. Der letzte Teil des Dateinamens besteht aus einer Angabe zur Version der Datei. Es soll hier mit der Abkürzung „v“ für Version und zwei Ziffern für die Nummerierung gearbeitet werden. Auf Umlaute und Sonderzeichen ist zu verzichten.

Tabelle 1: Abkürzungen der Arbeitsgruppen

Name der Arbeitsgruppe oder Clusters (Deutsch / Englisch)	Kürzel für Dateinamen
AG Datenmanagement / Research Data Management	RDM
AG Nachwuchsförderung / Early Career Support	ECS
AG PhD Workshop / PhD Workshop	PHD
AG Entwurf und Lastfalldefiniton / Design and Load Case Definition	DLC
Cluster Einwirkungen aus Wind und Wellen / Cluster Wind and Waves	CWW
Cluster Einbindung Meeresboden / Seabed integration	CSI
Cluster Entwurf Tragstruktur / Support structure design	CSD
Cluster Entwurf Rotor / Rotor design	CRD
Cluster Regelung und Überwachung / Control and Monitoring	CCM
Cluster gekoppeltes Gesamtsystem / coupled system	CCS

Die Benennung dieser Richtlinie lautet daher:

210714_RDM_RichtlinieZumUmgangMitForschungsdaten_v01

Die Datei hat einen Bearbeitungsstand vom 14.07.2021, wurde von der AG Datenmanagement (engl. *Research Data Management,* RDM) erstellt und befindet sich in der ersten Version.

§7 Schutz vor Datenverlust

Schon vor der eigentlichen Datenerhebung ist sicherzustellen, dass eine ausreichend geeignete Infrastruktur zur Verfügung steht. Hierfür sind insbesondere folgende Fragen wichtig:

- Ist ausreichend Speicherkapazität vorhanden?
- Ist ein automatisiertes tägliches Backup gewährleistet?
- Ist die Bandbreite ausreichend?
- Sind die Speichersysteme ausreichend sicher?

Für das tägliche Backup ist darauf zu achten, dass die Backup-Kopie nicht im selben Gebäude wie das Original liegt, ansonsten können Daten infolge unvorhergesehener Ereignisse (wie bspw. Brand) verloren gehen. Die Bandbreite ist wiederum wichtig für die regelmäßige Übertragung großer Datenmengen über ein Intra- oder Internet (bspw. für ein ortsverteiltes Backup oder zum Austausch mit externen Partnern). Die Sicherheit der Daten ist vor allem für sensible Informationen wichtig. Die Aspekte zum Schutz vor Datenmissbrauch werden in §8 ausführlich beschrieben.

Darüber hinaus stellt die LUH LUIS-Dienste zum Speichern und Sichern von Daten zur Verfügung. Für projektübergreifende Inhalte und zum Austausch kleinerer Datenmengen steht das Projekt-Seafile als Projektablage zur Verfügung [11]. Für die langfristige Datenaufbewahrung sollte das LUH-Datenarchiv genutzt werden [14]. Dieses Datenarchiv ist für Daten bestimmt, die aus rechtlichen und ethischen Gründen nicht veröffentlicht, aber aufgrund der guten wissenschaftlichen Praxis aufgehoben werden müssen. Für zu

veröffentlichende Daten, für die kein geeignetes Fachrepositorium gefunden werden kann, ist das LUH-Datenrepositorium zu nutzen [15]. Die Daten des laufenden Projekts sind täglich auf Institutsservern zu speichern, indem entweder direkt auf den Servern gearbeitet oder auf Software zur automatisierten Synchronisierung von Ordnern zurückgegriffen wird (z. B. PureSync)..

§8 Schutz vor Datenmissbrauch

Forschungsdaten vor unbefugtem Zugriff zu schützen ist ein wichtiger Bestandteil des Forschungsdatenmanagements. In diesem Paragraphen wird erläutert, welche Methoden und Hilfsmittel zur Verfügung stehen. Es wird

- der Umgang mit Passwörtern,
- das Rechtemanagement,
- die Verschlüsselung von Dateien und Datenträgern und
- weitere organisatorische Maßnahmen

in den nachstehenden Absätzen erläutert.

Umgang mit Passwörtern

Für den Umgang mit Passwörtern gibt es einige zu beachtende Aspekte. Zunächst ist festzuhalten, welche Aspekte beim Umgang mit Passwörtern unbedingt zu vermeiden sind:

- Passwörter nicht an den Bildschirm kleben
- Passwörter niemals im Klartext per E-Mail versenden
- nicht dasselbe Passwort für unterschiedliche Accounts verwenden (zumindest bei sensiblen Daten und Infrastrukturen)

Mit diesen grundlegenden Empfehlungen wird verhindert, dass Unbefugte Zugriff auf die Passwörter haben, sensible Daten durch achtloses Weiterleiten von E-Mails weitergegeben werden und der Zugriff auf sämtliche Dateien bei einem einmaligen „Knacken“ des Passworts ermöglicht wird.

Für die Erstellung sicherer Passwörter sind folgende Tipps zu beachten:

- mindestens 20 Zeichen
- schwer zu erraten (keine Geburtstage etc. verwenden)
- Passphrase (Zusammenstellung mehrerer Wörter ohne erkennbaren Sinn, also keine Sprichwörter etc.) statt ein einzelnes Wort
- Schreibweise verfremden (z. B. „Tr!ck“ statt „Trick“)
- mehrere Sprachen kombinieren

Weitere Tipps können bspw. den Empfehlungen des Bundesamtes für Sicherheit in der Informationstechnik (BSI) entnommen werden [16]. Darüber hinaus sollten bei der Erstellung von Passwörtern keine länderspezifischen Sonderzeichen wie bspw. ä, ö, ü, ß verwendet werden, da diese Zeichen auf Tastaturlayouts anderer Sprachen nicht zur Verfügung stehen und das Passwort somit nicht eingegeben werden kann.

Rechtemanagement

Einige Systeme zur Datenverwaltung bieten die Möglichkeit Zugriffsrechte festzulegen (z. B. Cloud-Seafile der LUH; derzeit nicht für Projekt-Seafile der LUH möglich). Dabei kann Bezug

auf einzelne Dateien, Ordner oder ganze Ordnerbäume und Datensätze in einer Datenbank genommen werden. Die einfachste Ebene ist die Vergabe von

- kombinierten Lese- und Schreibrechten,
- reinen Leserechten oder
- keinen Zugriffsrechten.

Grundsätzlich ist es für das zuverlässige Einrichten und Verwalten von Berechtigungen ratsam, IT-Fachpersonal einzusetzen. Damit werden fehlerhafte Konfigurationen vermieden, die zu Datenverlust oder Sicherheitslücken führen können. Das Rechtemanagement ist vor allem dann zu berücksichtigen, wenn Externe Zugriff auf Daten erhalten sollen. Da die Zugriffsrechte für Projekt-Seafile nicht festgelegt werden können, und Cloud-Seafile nicht für die Projektbearbeitung genutzt werden soll, sind Zugriffsrechte vor allem für die Institutsserver wichtig, da hier viele Mitarbeiter potenziell den Zugriff auf die Daten haben, der ggf. nicht erforderlich ist.

Verschlüsselung von Dateien und Datenträgern

Eine Verschlüsselung von Daten ist ggf. aus rechtlichen und ethischen Gründen erforderlich. Dies ist bspw. bei personenbezogenen oder patentrelevanten Daten der Fall, aber auch bei Daten, die einer Geheimhaltungspflicht unterliegen. Es gilt das Prinzip, dass je sensibler die Daten sind, desto wichtiger ist eine durchgehende Verschlüsselung.

Für die Sicherstellung dieses Prinzips sind folgende Tipps zu beachten:

- Verschlüsselung von USB-Sticks / externen Festplatten, da sie schnell verloren gehen
- Verschlüsselung von den in Rechnern integrierten Festplatten (Login-Passwort des Betriebssystems reicht nicht aus, da die Festplatte ausgebaut werden könnte)
- sensible Daten, die auf Institutsservern oder Rechenzentrumsservern gespeichert werden, sind i. d. R. sicher, da die Server in gesicherten Räumen stehen
- kein Versenden von sensiblen Ordnern oder Dateien per E-Mail oder Ablage auf kommerziellen Filehostern und Cloud-Speichern (z. B. Dropbox, Microsoft OneDrive etc.)

Für die Verschlüsselung von Laptops oder externen Festplatten sollte eine softwareseitige Verschlüsselung verwendet werden. Diese ist tendenziell sicherer als eine hardwareseitige Verschlüsselung. Die LUH empfiehlt das kostenlose Open Source Programm *VeraCrypt*, welches auf den meisten Betriebssystemen funktioniert [17], aber auch die auf das Betriebssystem angepassten Verschlüsselungstools (z. B. *BitLocker* bei *Windows*) können genutzt werden. Falls das Verschicken sensibler Daten per E-Mail unumgänglich ist, sollten die Dateien vorab verschlüsselt werden (z. B. mit 7-Zip).

Die meisten Datenmissbrauchsversuche erfolgen über das Internet. Dennoch ist auch der physische Zugang zu Datenträgern zu berücksichtigen. Dabei sind folgende Fragen zu beantworten:

- Auf welchen Datenträgern existieren Originale oder Kopien der Daten (z. B. E-Mail-Anhänge, Papierausdrucke, externe Speichermedien oder Backup-Server)?
- Wo befinden sich die Datenträger?
- Wer hat legal physischen Zugang (Kollegen, Hausmeister, Putzdienst etc.)?
- Wie einfach ist ein illegaler Zugang zu den Datenträgern möglich (Einbruch)?

Es sind Aufbewahrungsorte der Datenträger zu wählen, deren Sicherheit und Zugangsbeschränkung der Sensibilität der Daten angemessen ist.

Weitere organisatorische Maßnahmen

Um alle Projektbeteiligten für den Schutz vor Datenmissbrauch zu sensibilisieren, ist eine Schulung der datenverarbeitenden Personen sinnvoll.

Nicht mehr benötigte oder defekte Datenträger sind professionell physisch zu vernichten, um eine nachträgliche Wiederherstellung der Daten zu verhindern. Hierzu können die Datenträger z. B. im Welfenschloss der LUH (R318A) beim Sachgebiet 12 abgegeben werden. Die Datenträger werden entsprechend professionell vernichtet.

Es sollen regelmäßige Kontrollen stattfinden, die das Einhalten der Regeln überwachen. Der Workflow und die Verantwortlichkeiten legt jedes Teilprojekt selbst fest. Die festgelegten Randbedingungen sind im DMP schriftlich festzuhalten.

§9 Auswahl aufhebenswerter Daten

Im Sinne der guten wissenschaftlichen Praxis sind relevante Daten langfristig aufzubewahren und wenn möglich auch öffentlich zur Verfügung zu stellen. Nicht alle Daten müssen aufgehoben werden. Es ist aber bewusst zu entscheiden und zu begründen, welche Daten aufgehoben werden müssen und welche nicht. Im Folgenden werden Szenarien genannt, die als Beispiele für aufhebenswerte Daten fungieren.

Wenn die Daten Grundlage einer wissenschaftlichen Veröffentlichung sind, dann sind diese Daten i. d. R. mindestens 10 Jahre aufzubewahren.

Wenn eine Nachnutzung der Forschungsdaten wahrscheinlich ist, sind diese ebenfalls zu archivieren.

Einzigartige Daten, d. h. Daten, die nicht in derselben Form erhoben werden können (z. B. Wetteraufzeichnungen oder Meinungsumfragen) sollten aufgehoben werden.

Wenn Daten nicht in einem vertretbaren Aufwand reproduzierbar sind, dann sind die Daten aufzubewahren. Bspw. sollten Simulationen immer dasselbe Ergebnis produzieren, wenn genutzte Software, verwendete Algorithmen und die Eingangsdaten dieselben sind. In diesem Fall können die Daten gelöscht werden. Es sind aber Metadateninformationen zu archivieren, in diesem Beispiel Aussagen zur verwendeten Softwareumgebung und Eingangsdaten. Es sollte darauf geachtet werden, dass bspw. Berechnungsergebnisse ausreichend dokumentiert werden, um zu vermeiden, dass sämtliche Softwareversionen aufbewahrt werden müssen.

Qualitativ hochwertige Daten sind zu archivieren. Ausgangsdaten und Endprodukte sollten daher immer aufgehoben werden. Dateiversionen, die Zwischenstadien darstellen und teilweise fehlerhaft sind, können tendenziell gelöscht werden. Zu den Dateiversionen sollten dennoch Metadaten archiviert werden, um zu zeigen, dass diese Daten existiert haben und warum sie gelöscht wurden.

Bei sehr großen Dateien ist teilweise der Speicherplatzbedarf zu groß, um diese Daten zu archivieren. Zunächst ist zu versuchen, die Speicherkapazität zu erhöhen oder die Dateigröße abzumindern. Erst im nächsten Schritt sollten die weniger wichtigen und älteren Dateien zugunsten der wichtigeren und aktuellen Dateien gelöscht werden.

§10 Langfristige Archivierung und Publikation von Forschungsdaten nach den FAIR-Prinzipien

Für die langfristige Archivierung und Publikation von Forschungsdaten sind die Daten nach den FAIR-Prinzipien aufzubereiten. FAIR ist ein Akronym und steht für Daten, die auffindbar

(**f**indable), zugänglich (**a**ccessible), kompatibel zu anderen Systemen (**i**nteroperable) und nachnutzbar (**r**e-usable) sind.

Datenaufbereitung nach den FAIR-Prinzipien

Mit der Auffindbarkeit (**f**indable) soll anderen Forschenden das Finden der jeweiligen Daten ermöglicht werden. Es ist darauf zu achten, dass reichhaltige Metadaten in durchsuchbaren öffentlichen Verzeichnissen (z. B. Suchdatenbanken von Datenrepositorien) vorhanden sind und diese Daten über einen eindeutigen Identifier (z. B. DOI) abrufbar sind.

Die Zugänglichkeit (**a**ccessible) ermöglicht es anderen Forschenden auf die erstellten Daten zuzugreifen. Es sollte möglichst auf einen Online-Zugang über Standard-Protokolle (z. B. http(s)) geachtet werden. Darüber hinaus sind transparente Zugangsbedingungen zu schaffen, in welchen begründet wird, unter welchen Bedingungen Zugriff auf die Daten gewährt wird (sofern eine Zugriffsbeschränkung vorliegt).

Die Kompatibilität (**i**nteroperable) zu anderen Systemen stellt sicher, dass die generierten Daten mit anderen Daten kombiniert werden können und maschinell verarbeitbar sind. Die Daten sind demnach nach fachspezifischen Standards systematisch aufbereitet und dokumentiert. Ferner sind die maschinenlesbaren Daten und Metadaten in verbreiteten und möglichst offenen Dateiformaten zu archivieren sowie auf verwandte Daten zu referenzieren.

Die Nachnutzbarkeit (**r**e-usable) ermöglicht es anderen Forschenden, die Daten für ihre Zwecke nachzunutzen. Zusätzlich zu allen vorherigen Punkten gilt hier zusätzlich, dass eine gute Dokumentation wichtig ist und eine eindeutige Definition über Nachnutzungsbedingungen vorliegen.

Archivierung von Daten

Die richtige Archivierung gilt für Daten, Metadaten und weitere Informationen. Grundsätzlich gilt, dass

- Daten in langfristig lesbaren Dateiformaten vorliegen,
- Daten nach fach- oder datenspezifischen Standards strukturiert sind und
- Daten möglichst maschinenlesbar sind [10, 18].

Demnach sollten bspw. Bilder möglichst in das *.TIFF oder *.TIF-Format, Dokumente in *.TXT, *.PDF/A oder *.XML und Audiodateien in *.WAV abgespeichert werden. In Anlehnung an §7 sind für die langfristige Archivierung von Daten folgende Anforderungen wichtig:

- geeignete Serverräume (Brandschutz- und Sicherheitsanforderungen)
- redundante Speicherung
- regelmäßiger Austausch der Speichermedien (infolge von Verschleiß der Datenträger)
- garantierte Bitstream Preservation (fehlerfreier Datenerhalt) für mind. 10 Jahre
- Wartung durch IT-Fachpersonal
- Datenzugriff nur für autorisierte Personen

Für eine langfristige Archivierung sind CDs, externe Festplatten oder USB-Sticks nicht geeignet!

Publikation hochwertiger Daten

Wenn es rechtlich möglich ist, dann sollten die nach den FAIR-Prinzipien aufbereiteten Daten publiziert werden. Dies liegt nicht nur im Interesse von Förderern und Wissenschafts-

organisationen, sondern auch im Interesse des Forschenden, da relevante und hochwertige Daten zusätzlichen Einfluss und Aufmerksamkeit generieren.

Publizierte Daten müssen genauso zitiert werden wie Textpublikationen.

Die zu veröffentlichenden Daten sind über ein sog. Datenrepositorium (Datenzentrum) zur Verfügung zu stellen. Dieses Repositorium speichert und verwaltet Daten und die zugehörigen Metadaten. Über Suchkriterien sind die Daten dann zu finden und häufig auch herunterladbar. Hierbei ist darauf zu achten, dass das Repositorium auf das Fach oder die Art der jeweiligen Daten spezialisiert ist. Geeignete Repositorien lassen sich bspw. in [19] finden. Für die Auswahl eines geeigneten Repositoriums sollten folgende Aspekte berücksichtigt werden:

- Vergabe eines Persistent Identifiers (PI, z. B. DOI)
- Eindeutig geregelte und verbindliche Datennutzung
- Wer darf Daten einstellen?
- Welche Formate für Daten und Metadaten werden akzeptiert?
- Welche Services bietet das Repositorium (z. B. Qualitätssicherung)?
- Was kostet das Einstellen von Daten?

Idealerweise sind diese Punkte bereits vor Projektbeginn zu klären, um nötige finanzielle Mittel beantragen und die richtigen Dateiformate generieren zu können. Falls kein geeignetes Datenrepositorium gefunden werden kann, bietet die LUH über das eigene institutionelle Datenrepositorium eine kostenlose Möglichkeit der Datenpublikation [20].

Erfolgt mit der Datenpublikation auch eine Fachartikelpublikation, in der die Datenauswertung und –bewertung beschrieben wird, so ist in den Metadaten jeweils auf die jeweils andere Publikation zu verweisen.

Literaturverzeichnis

[1] Leibniz Universität Hannover, *Richtlinie zum Umgang mit Forschungsdaten an der Leibniz Universität Hannover*, 2020.

[2] Deutsche Forschungsgemeinschaft, *Leitlinien zum Umgang mit Forschungsdaten*, 2015.

[3] FAIRplus, *FAIR Cookbook.* [Online]. Verfügbar unter: https://fairplus.github.io/the-fair-cookbook/content/home.html (Zugriff am: 8. Juni 2021).

[4] Deutsche Forschungsgemeinschaft, *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Denkschrift; Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft"*. Weinheim: Wiley-VCH, 2013.

[5] Leibniz Universität Hannover, *Ordnung der Gottfried Wilhelm Leibniz Universität Hannover zur Sicherung guter wissenschaftlicher Praxis*, 2015.

[6] Leibniz Universität Hannover, *Open Access-Resolution der Leibniz Universität Hannover.* [Online]. Verfügbar unter: https://www.uni-hannover.de/de/universitaet/profil/ziele-strategien/open-access/open-access-resolution/ (Zugriff am: 8. Juni 2021).

[7] Datenschutz-Grundverordnung, *Datenschutz-Grundverordnung: DSGVO als übersichtliche Seite.* [Online]. Verfügbar unter: https://dsgvo-gesetz.de/ (Zugriff am: 10. Juli 2021).

[8] Bundesministerium der Justiz und für Verbraucherschutz, *UrhG.* [Online]. Verfügbar unter: https://www.gesetze-im-internet.de/urhg/ (Zugriff am: 11. Juli 2021).

[9] University of Sydney, *Research Data Management Policy 2014.* [Online]. Verfügbar unter: https://www.sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2013/337%20https://blog.imb.bau.tu-dresden.de (Zugriff am: 10. Juli 2021).

[10] Research Data Alliance, *Metadata RDA | Metadata Directory*. [Online]. Verfügbar unter: https://rd-alliance.github.io/metadata-directory/standards/ (Zugriff am: 22. Juni 2021).

[11] Leibniz Universität Hannover, *Projektverwaltung Seafile.* [Online]. Verfügbar unter: https://admin.projekt.uni-hannover.de/projekte/projects/view/project/sfb-megastrukturen (Zugriff am: 11. Juli 2021).

[12] Leibniz Universität Hannover, *Download Ticket Service – Leibniz Universität IT Services – Leibniz Universität Hannover.* [Online]. Verfügbar unter: https://www.luis.uni-hannover.de/de/services/speichersysteme/dateiservice/download-ticket-service/ (Zugriff am: 10. Juli 2021).

[13] Leibniz Universität Hannover, *GitLab – Leibniz Universität IT Services – Leibniz Universität Hannover.* [Online]. Verfügbar unter: https://www.luis.uni-hannover.de/de/services/speichersysteme/dateiservice/projektablage/teildienste-der-projektablage/gitlab/ (Zugriff am: 2. Juli 2021).

[14] Leibniz Universität IT Services, *Archivierung.* [Online]. Verfügbar unter: https://www.luis.uni-hannover.de/de/services/speichersysteme/archivierung/ (Zugriff am: 21. Juni 2021).

[15] Leibniz Universität IT Services, *Forschungsdaten-Repositorium.* [Online]. Verfügbar unter: https://www.luis.uni-hannover.de/de/services/speichersysteme/forschungsdaten-repositorium/ (Zugriff am: 21. Juni 2021).

[16] Bundesamt für Sicherheit in der Informationstechnik, *Empfehlungen des Bundesamtes für Sicherheit in der Informationstechnik.* [Online]. Verfügbar unter: https://www.bsi.bund.de/DE/Home/home_node.html (Zugriff am: 21. Juni 2021).

[17] VeraCrypt, *Download VeraCrypt.* [Online]. Verfügbar unter: https://www.veracrypt.fr/en/Downloads.html (Zugriff am: 21. Juni 2021).

[18] forschungsdaten.info, *Formate erhalten.* [Online]. Verfügbar unter: https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/formate-erhalten/ (Zugriff am: 22. Juni 2021).

[19] re3data, *re3data.org.* [Online]. Verfügbar unter: https://www.re3data.org/ (Zugriff am: 11. Juli 2021).

[20] Leibniz Universität Hannover, *Research Data Repository.* [Online]. Verfügbar unter: https://data.uni-hannover.de/ (Zugriff am: 22. Juni 2021).