



Bericht zur Untersuchung der DLR Institute zum Bedarf beim Forschungsdatenmanagement

Konzept: Projekt „Forschungsdatenmanagement (FDM)“



Dokumenteigenschaften

Titel Bericht zur Untersuchung der DLR Institute zum Bedarf beim Forschungsdatenmanagement

Betreff Erhebung im Rahmen des Konzepts: Projekt „Forschungsdatenmanagement (FDM)“

Institut IW

Erstellt von Dr.-Ing. Christian Langenbach

Beteiligte Carina Haupt (SC), Tobias Schlauch (SC)

Geprüft von _____

Freigabe von _____

Datum 11.2020

Version 1.1

Dateipfad Z:\FDM-Projekte\DLR-FDM_Projekt\Umfrage\20201104_Bericht Untersuchung DLR Institute zum Forschungsdatenmanagementbedarf-V1.1.docx

Der Text dieser Publikation wird unter der Lizenz Creative Commons Namensnennung 4.0 International (CC BY 4.0) veröffentlicht. Den vollständigen Lizenztext finden Sie unter: <https://creativecommons.org/licenses/by/4.0/legalcode>.

Lizenz



Kurzzusammenfassung

Im Projekt FDM wurde für die Umsetzung des Konzeptes Forschungsdatenmanagement im DLR eine Umfrage unter den Instituten und Einrichtungen durchgeführt, um die tatsächlichen Bedarfe im Bereich des Forschungsdatenmanagements besser einordnen und abschätzen zu können.

Es hat sich herausgestellt, dass eine Vielzahl von Datensammlungen und Forschungsanwendungen existiert. Teilweise lassen sie sich bereits in fachspezifischen Repositorien langfristig verfügbar halten, teilweise müssen neue Lösungen in Betracht gezogen werden.

Für die Bereitschaft, sich an der Befragung zu beteiligen, möchte ich mich noch einmal herzlich bedanken.

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für alle Geschlechter.

Der Projektleiter und Forschungsdatenmanager des DLR

Dr.-Ing. Christian Langenbach



Inhaltsverzeichnis

1. Einleitung	5
1.1. Untersuchungsdesign	5
1.2. Teilnahme und Rücklauf	5
2. Auswertung	6
2.1. Einleitende Fragen	6
2.2. Fragen zu Art und Umfang der forschungsbezogenen Daten	7
2.3. Fragen zum Stand der Praxis	11
2.4. Fragen zu Unterstützungsangeboten	16
2.5. Abschließende Frage	18
3. Schlussfolgerungen	19
Abbildungsverzeichnis	21
Anhang Fragebogen	22



1. Einleitung

Das DLR hat 2018 eine Richtlinie zum Umgang mit Forschungsdaten als Teil des Organisationshandbuchs verabschiedet. Zur Unterstützung der dort geforderten Praxis, hat die OE Wissenschaftliche Information (IW) unter der Leitung des DLR-Forschungsdatenmanagers ein Projekt im DLR initiiert. Mit dem Projekt FDM wird das Ziel verfolgt, eine dauerhafte Informationsinfrastruktur im Rahmen ihres Konzepts Forschungsdatenmanagement aufzubauen. Das Vorhaben strebt an, die langfristige Verfügbarkeit von Datenbanken, Repositorien und Services im DLR sicherzustellen.

Um die konkreten Bedarfe besser einschätzen zu können, fand zwischen September und November 2019 eine schriftliche Befragung aller Institute und Einrichtungen des DLR statt. Bei dieser Umfrage sollte unter anderem ermittelt werden, welche Forschungsdaten und -anwendungen an den Instituten und Einrichtungen bereits existieren, auf welcher technischen Basis sie basieren, welche Anwendungen und Datenerhebungen in Zukunft geplant sind und wie es um die langfristige Verfügbarkeit dieser Daten und Anwendungen bestellt ist.

Die Umfrage lieferte erste allgemeine Erkenntnisse über den Umgang mit Forschungsdaten im DLR, denn Forschungsdatenmanagement beginnt nicht erst am Ende des Forschungsprojektes mit der Datenarchivierung, sondern bereits in der Planungsphase. In vielen Gesprächen zum Datenhandling an den Instituten und Einrichtungen wurde deutlich, dass viele Forschende das Potenzial eines optimalen Datenmanagements offensichtlich noch nicht ausschöpfen.

1.1. Untersuchungsdesign

Die Umfrage gliederte sich in fünf Abschnitte: Zuerst wurde die Fach- bzw. Institutszugehörigkeit erfragt. Der zweite Abschnitt galt der Art und Umfang der forschungsbezogenen Daten im Forschungsprozess: Art der Daten, Datentypen, -formate und -volumen. Der Dritte enthielt Fragen zum aktuellen Stand der Praxis, wie Zuständigkeiten, Speicherung, Archivierung, Zugriff auf die Daten und die Nutzung von Datenmanagementplänen. Weiter wurden im vierten Abschnitt Fragen zum Wissen über und Problemen mit dem Datenmanagement sowie Unterstützungsangeboten gestellt. Abschließend konnten die Teilnehmer in einem freien Kommentarfeld weitere Hinweise oder Ideen zum Forschungsdatenmanagement (insbesondere hinsichtlich des künftigen Bedarfs) angeben.

Insgesamt umfasste der Fragebogen 19 thematische Fragen. Bei bestimmten Fragen waren auch Mehrfachnennungen zugelassen und fünf Fragen erforderten eine freie Textantwort. Die erhaltenen Fragebögen wurden anonym ausgewertet.

Der Fragebogen wurde unter der Leitung des Forschungsdatenmanagers in einer Arbeitsgruppe mit Vertretern des Instituts für Softwaretechnologie entwickelt und aufbereitet. In Zusammenarbeit mit VO-S wurde der Fragebogen allen Instituts- und Einrichtungsleiter persönlich per E-Mail mit Anschreiben übermittelt.

In den Kommentaren am Ende der Umfrage sowie in persönlichen Rückmeldungen zeigte sich ferner, dass die Initiative zur Entwicklung von zentralen IT- und Beratungsangeboten für das Forschungsdatenmanagement wiederholt als sehr positiv bewertet und befürwortet wird.

1.2. Teilnahme und Rücklauf

Insgesamt wurde der Fragebogen an alle 43 Institute und Einrichtungen (Stand August 2019) versandt. Davon haben knapp 47 % oder 20 Institute und Einrichtungen an der Befragung teilgenommen. Alle Teilnehmer haben den Fragebogen vollständig ausgefüllt. Die Verteilung auf die Fachbereiche zeigt, dass der Anteil der Teilnehmer an der Umfrage die Verteilung auf die Fachbereiche repräsentiert.

An der Umfrage zum Umgang mit Forschungsdaten im DLR haben insgesamt 20 Institute teilgenommen. Bezogen auf die einzelnen Schwerpunkte sieht die Verteilung der Umfrageteilnehmer folgenderweise aus:

- Energie 25 % oder 5 Institute und Einrichtungen (DI, TT, SF, VT, VE)
- Luftfahrt 30 % oder 6 Institute und Einrichtungen (AT, FL, TP, WF, MO, AS)
- Raumfahrt 45 % oder 9 Institute und Einrichtungen (DW, SC, MP, PA, PF, HR, KN, MF, DFD)
- Verkehr 0 % oder 0 Institute und Einrichtungen

Die Umfrage wurde geschlossen, nachdem auf eine erneute Erinnerung nahezu keine weiteren Rückläufe innerhalb einer weiteren Woche folgten. Es wurde davon ausgegangen, dass bereits alle an der Umfrage teilgenommen hatten, die einen konkreten Bedarf haben bzw. denen die Relevanz des Themas bereits bewusst ist.

2. Auswertung

Die Darstellung der Ergebnisse der Umfrage erfolgt in textueller und/oder grafischer Form. Die Fragen wurden gemäß ihren übergeordneten Kategorien, die auch im Fragebogen ersichtlich sind, strukturiert. Bei den Ergebnissen wären auch nur teilbeantwortete Fragebögen berücksichtigt worden. Ausgehend von der Gesamtmenge pro Frage wurde für jede Antwortmöglichkeit pro Frage ein Prozentsatz ermittelt, mit welcher Häufigkeit die jeweilige Antwort gegeben wurde. Es wurde also ermittelt, wie viele Teilnehmer bei jeder einzelnen Frage welche Antwort gaben. Hierbei wurden Mehrfachnennungen beachtet, genauso wie nicht beantwortete Fragen.

2.1. Einleitende Fragen

Frage 1: Für welches Institut bzw. welche Einrichtung des DLR beantworten Sie die folgenden Fragen?

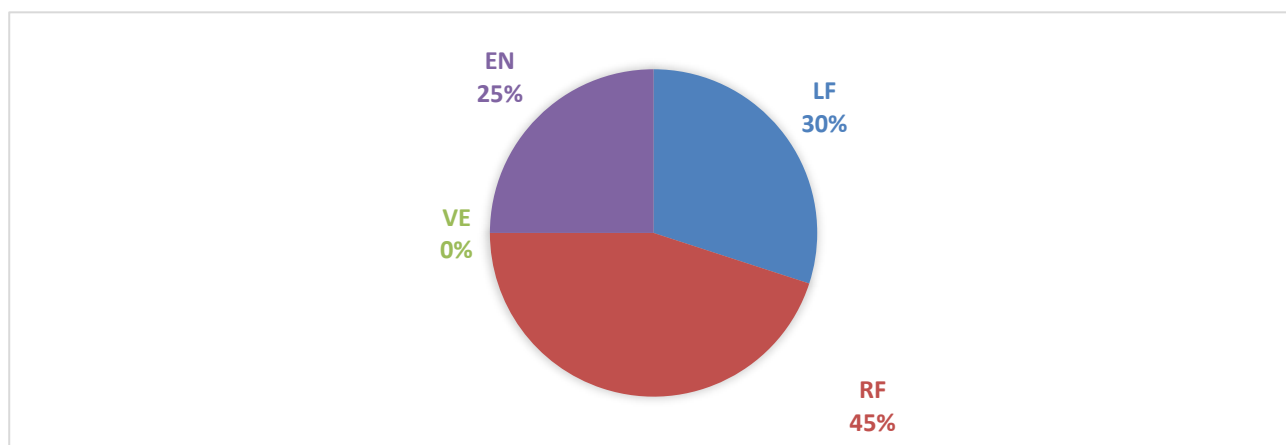


Abbildung 1: Verteilung der Umfrageteilnehmer bezogen auf die einzelnen Schwerpunkte

Frage 2: Welche Person kann man für das Institut/die Einrichtung bei ggf. weiteren Nachfragen zum Forschungsdatenmanagement ansprechen?

Hier wurden Daten von natürlichen Personen angegeben. Da es sich dabei um personenbezogene Daten handelt, wurde von einer Aufführung auf Grund des Datenschutzes abgesehen.

2.2. Fragen zu Art und Umfang der forschungsbezogenen Daten

Frage 3: Welche Arten von Forschungsdaten fallen im Rahmen der Forschung in Ihrem Institut an?

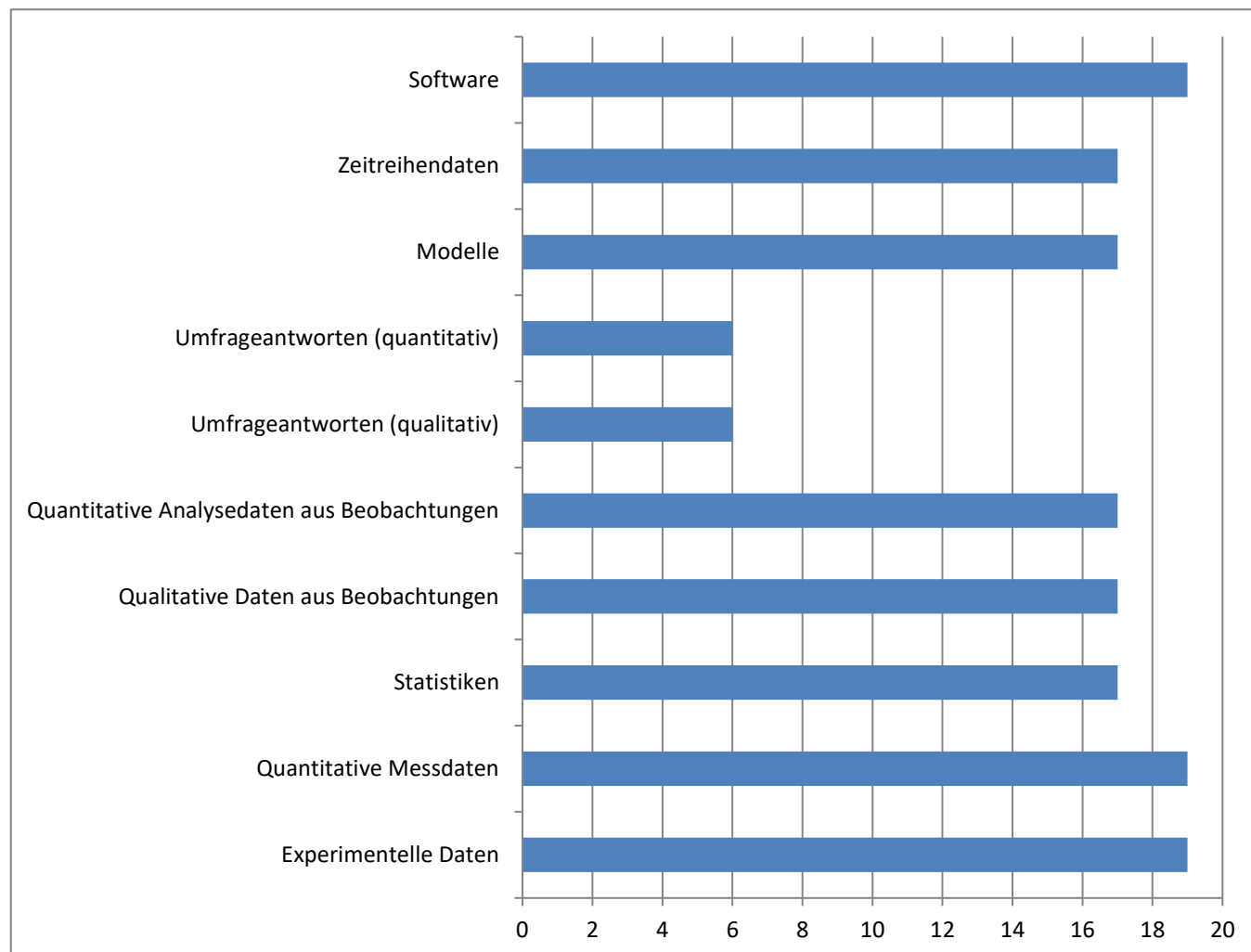


Abbildung 2: Gängige im DLR genutzte Arten von Forschungsdaten

Bei dieser Frage bestand die Möglichkeit weitere, nicht in der Abfrageliste aufgeführte, Datenarten zu nennen. Die Freitextantworten sind:

- Simulationsdaten und Performanzanalysedaten;
- Daten aus Strömungssimulationen und Designoptimierungen (Geometriedaten);
- Anlagendaten und Steuerungsdaten;
- Anhängende QM-Relevante Daten;
- Zertifikate und Reports;
- Dokumentsammlungen und Log Files;
- Simulationsergebnisse;
- Bilddaten (Satellit & Wolkenkamera);
- Eingangsdaten und abgeleitete Produkte von Satelliten, Geoinformationsprodukte;
- Eingangsdaten von Satelliten. Dokumente und Zwischenergebnisse (heterogen).

Frage 4: In welchen Datentypen liegen die Forschungsdaten in Ihrem Institut vor?

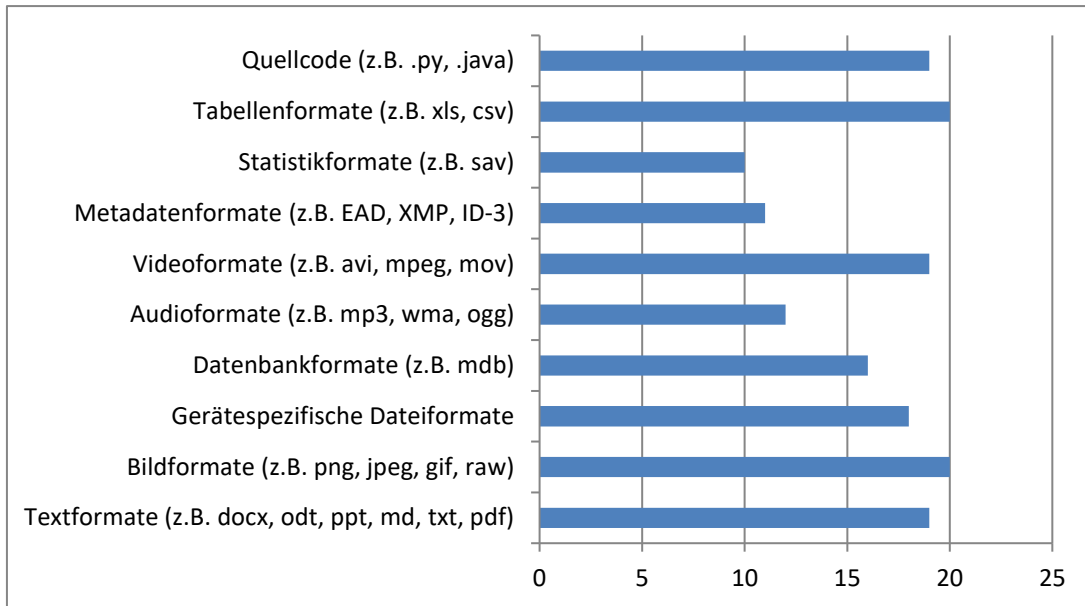


Abbildung 3: Gängige genutzte Datentypen im DLR

Bei dieser Frage bestand die Möglichkeit weitere, nicht in der Abfrageliste aufgeführte, Datentypen zu nennen. Die Freitextantworten sind:

- Simulationsdatenformate (.nc, vtk, etc.), Geometrieformate, Formate für Performanzanalysedaten (abhängig von den verwendeten Tools);
- CGNS, Tecplot, Catia, Labview, ASCII, Binär, Delphin;
- Xml; rdf, JSON, LAS, LAZ, XYZ, Log Files;
- NetCDF, HDF5; GRIB; BUFR;
- Eigene Formate (binär, ASCII, etwa Pixelgrafik und Experimentlogs), proprietäre Format wie IDL, Matlab, Mathematica, u.ä.;
- ISO; VDH;
- Große Matrizen (Binärdaten) in diversen, auch nicht genormten Formaten;
- Zip; rar.

Frage 5: Wie hoch ist derzeit typischerweise die Größe der Forschungsdaten, die für eine einzelne wissenschaftliche Publikation archiviert werden müssen?

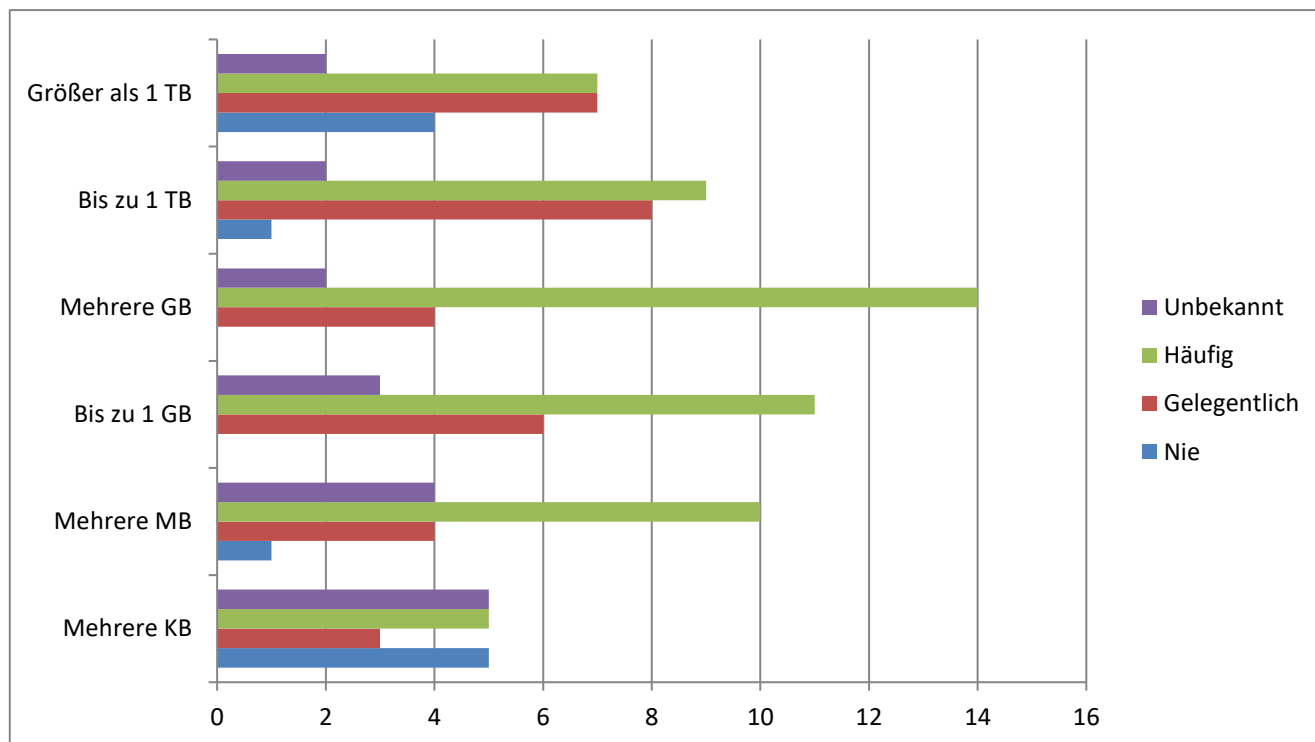


Abbildung 4: Einschätzung des Speichervolumens der gegenwärtigen Tätigkeit.

Aus den Antworten ist ersichtlich, dass die Majorität der zu archivierenden Datensätze derzeit typischerweise zwischen einem Gigabyte und einem Terrabyte liegt. Darauf sind die gängigen Repositoriensysteme ausgelegt.

Frage 6: In welcher Größenordnung entstehen pro Jahr Forschungsdaten in ihrem Institut?

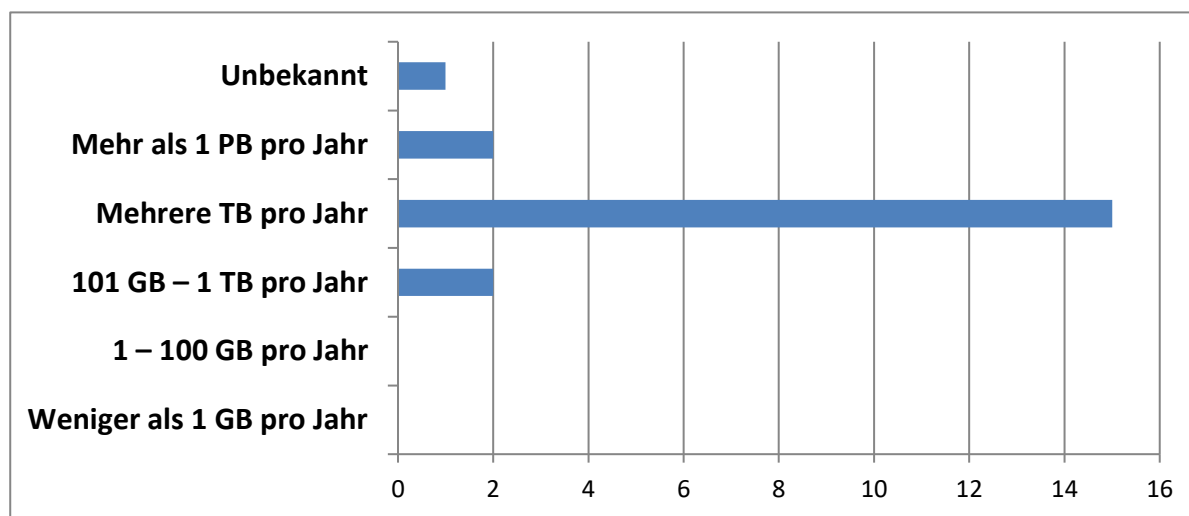


Abbildung 5: Abschätzung der Größenordnung des gesamten Speichervolumens pro Jahr

Frage 7: Welchen Umfang hat die Gesamtgröße der an Ihrem Institut vorliegenden Forschungsdaten ungefähr?

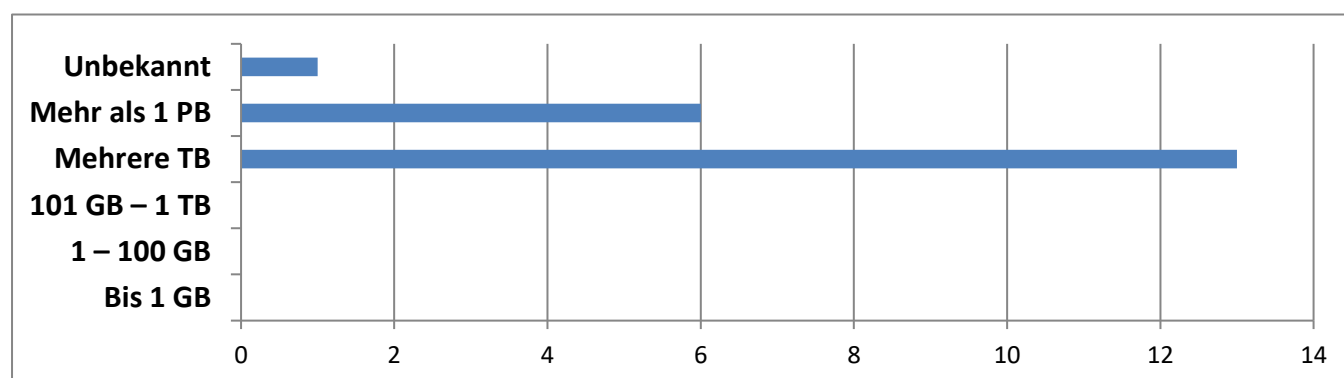


Abbildung 6: Gesamtgröße der bisher gespeicherten Daten

Die Antworten zu den Fragen 6 und 7 spiegeln die Einschätzung der Größenordnung der einzelnen Datensätze aus Frage 5 wieder. Aufsummiert über das Jahr liegt die breite Mitte der Institute bei mehreren Terrabyte pro Jahr. Das Ergebnis spricht für ein zentrales System, das so auch wirtschaftliche Vorteile bieten kann. Weiterhin existieren im DLR aber auch sogenannte „Datenelefanten“ mit mehr als ein Petabyte pro Jahr. Diese Institute erfordern eine eigene disziplinäre Lösung, da sie also einen weit höheren zu deckenden Bedarf haben.

2.3. Fragen zum Stand der Praxis

Frage 8: Wer bzw. welche Rollen sind typischerweise beim Thema Forschungsdatenmanagement im Institut beteiligt?

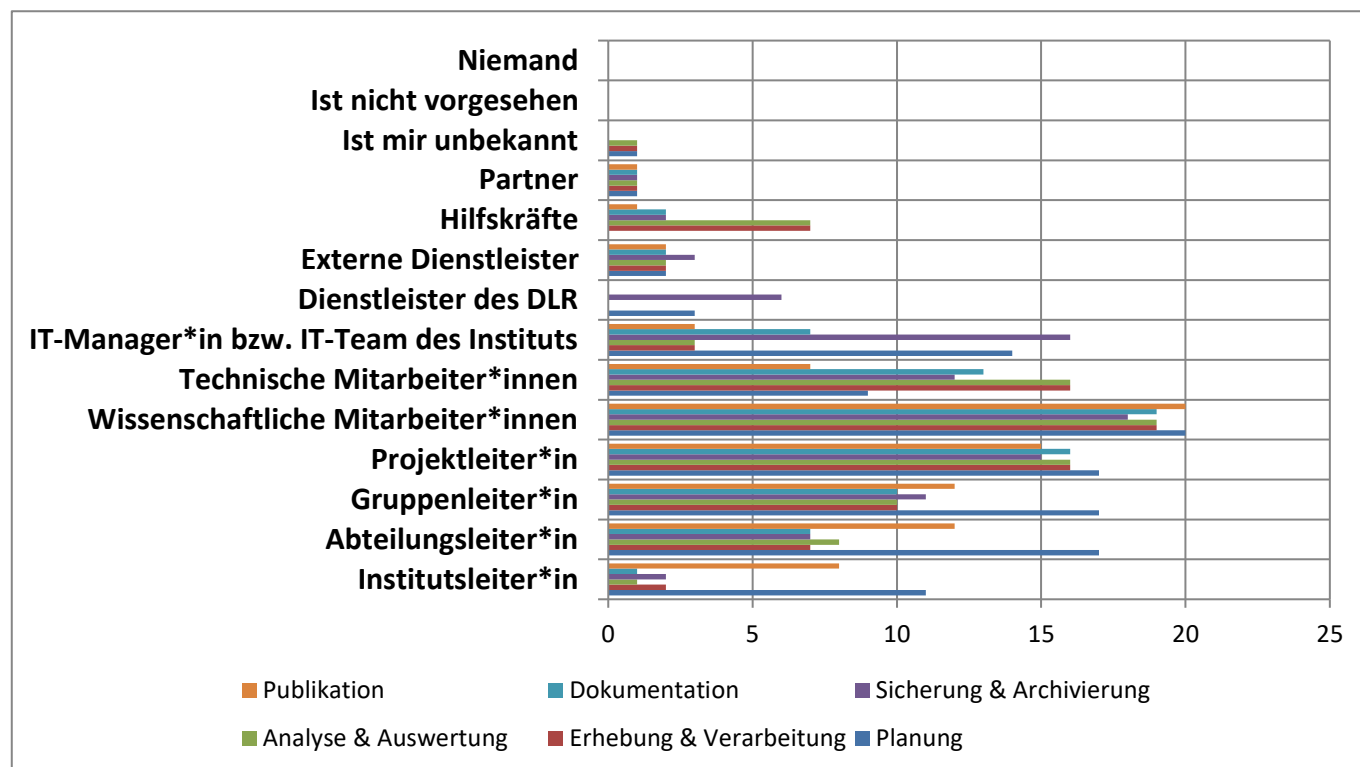


Abbildung 7: Wer im Institut ist im Datenzyklus beteiligt

Mit der Frage, wer während der verschiedenen Phasen des Datenzyklus für die Forschungsdaten verantwortlich ist, sollte in Erfahrung gebracht werden, in welchem Ausmaß die Forschenden selbst für die Forschungsdaten zuständig sind, bzw. ob bereits Datenarchivierungen außerhalb des Verantwortungsbereichs der Datenerzeuger erfolgen.

Während des Forschungsprozesses sind in der Regel ausschließlich die Datenerzeuger im engeren Sinne, Wissenschaftler und Projektleiter, für die Forschungsdaten im Datenzyklus verantwortlich. Daraus wird ersichtlich, dass es bislang nur sehr wenige Verfahren für die Verwaltung archivierter Forschungsdaten gibt, die sicherstellen, dass diese Daten unabhängig von der Erreichbarkeit der Urheber aufgefunden werden können. Ist der Datenurheber nicht mehr erreichbar, z. B. aufgrund eines Institutswechsels oder wegen Ausscheidens aus dem Forschungsbetrieb, verlassen auch seine Daten das Institut, an dem sie erhoben wurden und können nicht mehr aufgefunden werden.

Frage 9: Wo speichert das Institut die Forschungsdaten, die z.B. im Rahmen von Projekten anfallen?

Mit der Frage nach den Speicherorten für die Forschungsdaten sollte ermittelt werden, welche Optionen der vielfältigen Speichermöglichkeiten am meisten genutzt und in wieweit dienstliche Infrastrukturen verwendet werden. Es stellte sich heraus, dass die Forschungsdaten hauptsächlich auf dienstlichen bzw. wissenschaftlichen Infrastrukturen abgelegt werden. Allerdings werden die Daten ebenfalls in erheblichem Umfang auf externen Datenträgern gespeichert. Ferner werden, wenn auch in geringem Maße, Forschungsdaten bei ex-

ternen Anbietern hinterlegt. Hierbei werden u. a. Cloud-Anbieter und die Repositorysysteme von Zenodo und GitHub genannt

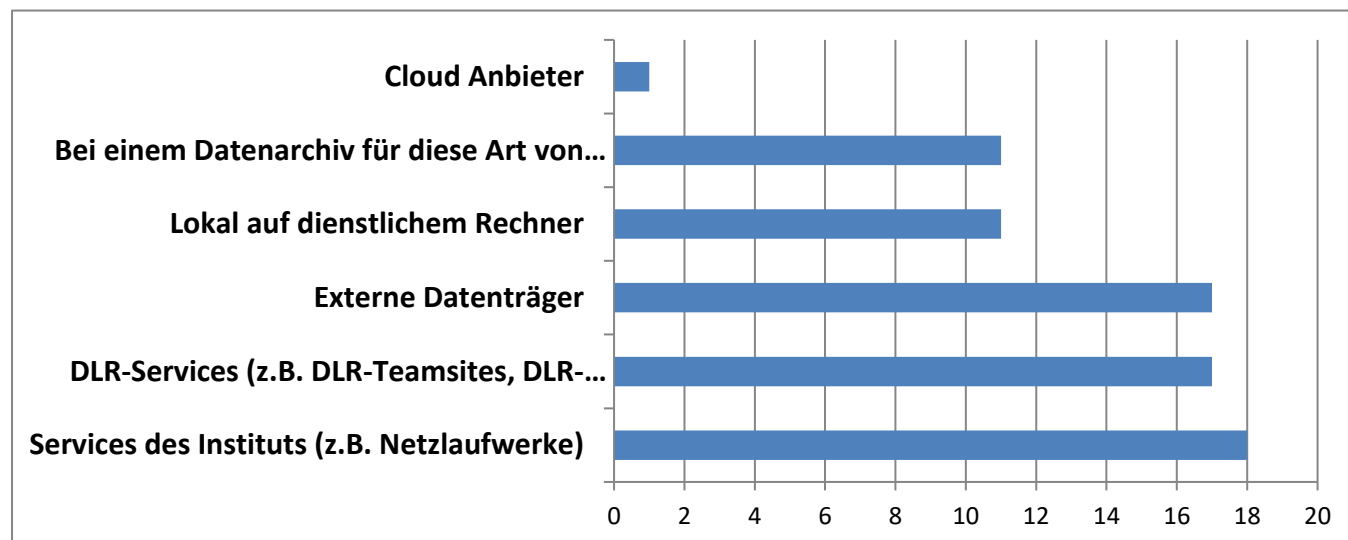


Abbildung 8: Speicherorte während der Forschungsphase

Bei dieser Frage bestand die Möglichkeit weitere, nicht in der Abfrageliste aufgeführte, Datentypen zu nennen. Die Freitextantworten sind:

- Server von externen Projektpartnern; Tivoli; Eigene Serveranlagen in Trauen;
- Veröffentlichung auf GitHub oder Zenodo;
- Wurde noch nicht festgelegt;
- Speicher der HPC-Cluster;
- Forschungsdaten liegen in der Regel auf Netzlaufwerken der wissenschaftlichen Infrastruktur (Linux);
- Hypertest; MUSC, BUSOC; zfs Dateisystem.

Frage 10: Wie lange bewahrt das Institut Forschungsdaten in der Regel nach Abschluss des jeweiligen Projekts auf?

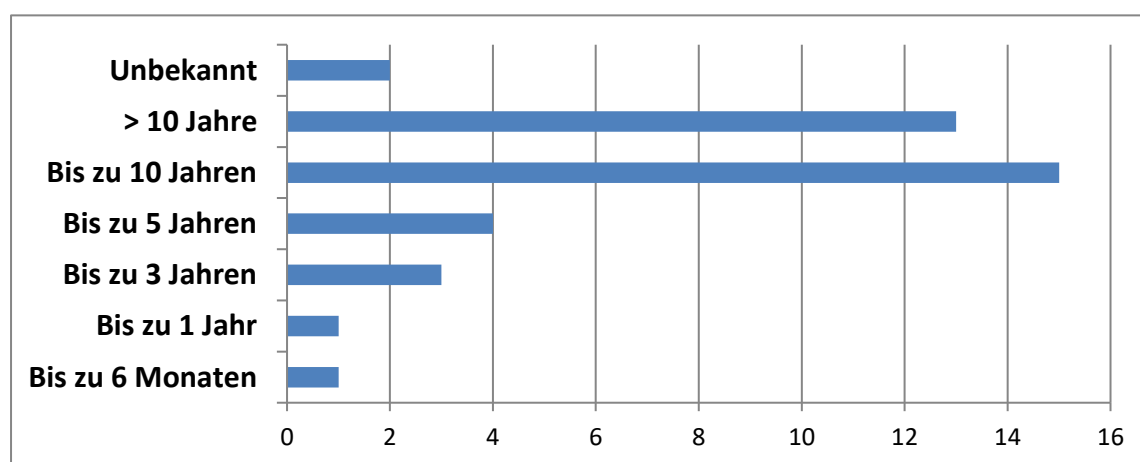


Abbildung 9: Derzeitige Aufbewahrungszeiträume für archivierte Forschungsdaten

Um indirekt abzufragen, inwieweit die Wissenschaftler ihrer zehnjährigen Aufbewahrungsfrist für Forschungsdaten nachkommen, wurde danach gefragt, wie lange die Forscher ihre derzeitig archivierten Forschungsdaten aufbewahren. 33,3 % der Rückmeldungen gaben als Zeitraum „mehr als 10 Jahre“ an. Bei 38,5 % der Antwortenden werden die Daten 5-10 Jahre archiviert, während bei 23,1 % die Daten weniger als 5 Jahre vorgehalten werden. Demzufolge bewahren knapp ein Drittel der Befragten ihre Daten für die vorgegebene Frist und sogar darüber hinaus auf.

Frage 11: Zu welchem Zweck bewahrt das Institut Forschungsdaten auf?

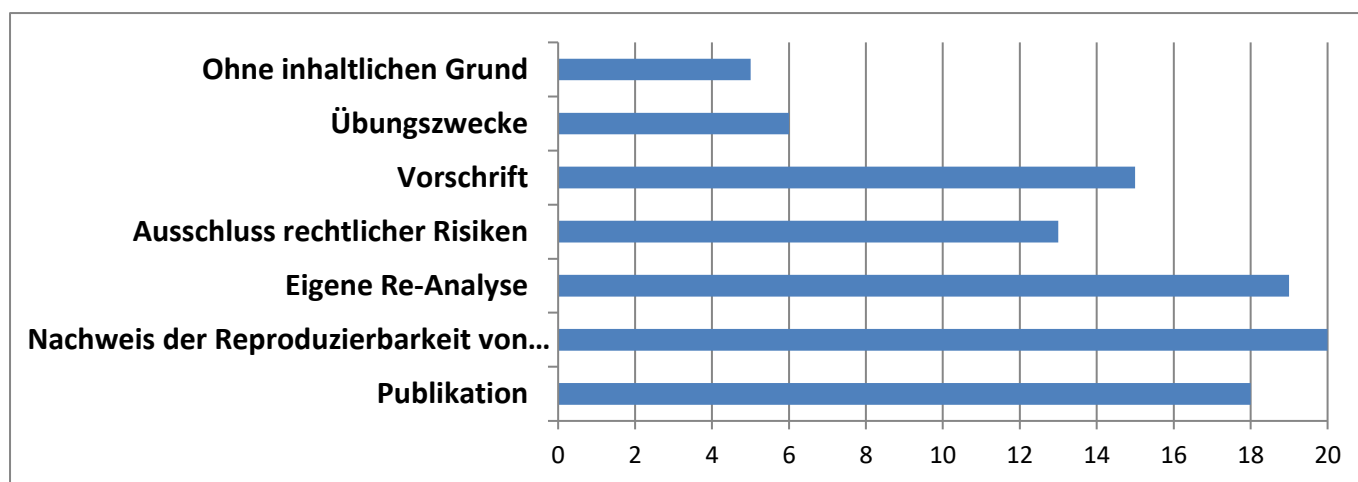


Abbildung 10: Rahmenbedingungen für die Archivierung der Forschungsdaten

Die Frage nach den Rahmenbedingungen diente sowohl der Erhebung des Status quo als auch der Bewusstseinsbildung für diese Problematik. Wie zu erwarten war, unterliegen die Forschungsdaten vielfältigen gesetzlichen Regelungen bzw. internen und externen Nutzungsbedingungen. Interessant ist, dass über 50 % der Antwortenden die Reproduzierbarkeit der Forschungsergebnisse in den Fokus der Archivierung stellen.

Bei dieser Frage bestand die Möglichkeit weitere, nicht in der Abfrageliste aufgeführte, Datentypen zu nennen. Die Freitextantworten sind:

- Kundenforderungen bei Drittmitteldaten (z.B. Dokumente zur Luftfahrttüchtigkeit & Zulassung. Forderungen bis zu LOP +10 = 30 und mehr Jahre);
- Nachvollziehbarkeit der Ergebnisse;
- Internationale Vereinbarungen;
- Und später Vergleiche zu aktuellen Experimenten zu ziehen oder auch, um neue Erkenntnisse aus alten Daten zu ziehen;
- Gute wissenschaftliche Praxis.

Frage 12: Wer darf auf die Forschungsdaten zugreifen?

Des Weiteren sollten Erkenntnisse darüber gewonnen werden, wer Zugang zu den archivierten Daten hat, um herauszufinden, ob die Daten derzeitig von anderen Wissenschaftlern nachgenutzt werden könnten. Es zeigt sich, dass die Bereitschaft, Forschungsdaten anderen, über das Institut hinaus, zur Verfügung zu stellen wenig ausgeprägt ist. Hieran zeigt sich, dass meist nur die Datenerzeuger Zugang zu den archivierten Daten haben

und bestenfalls noch für universitätsinterne und -externe Arbeitsgruppen begrenzter Zugang besteht. Zusätzlich fällt auf, dass für die Fachcommunity bzw. die Öffentlichkeit die Existenz der Daten so gut wie nicht sichtbar ist, aber immer noch besser, als der DLR weiter Zugriff.

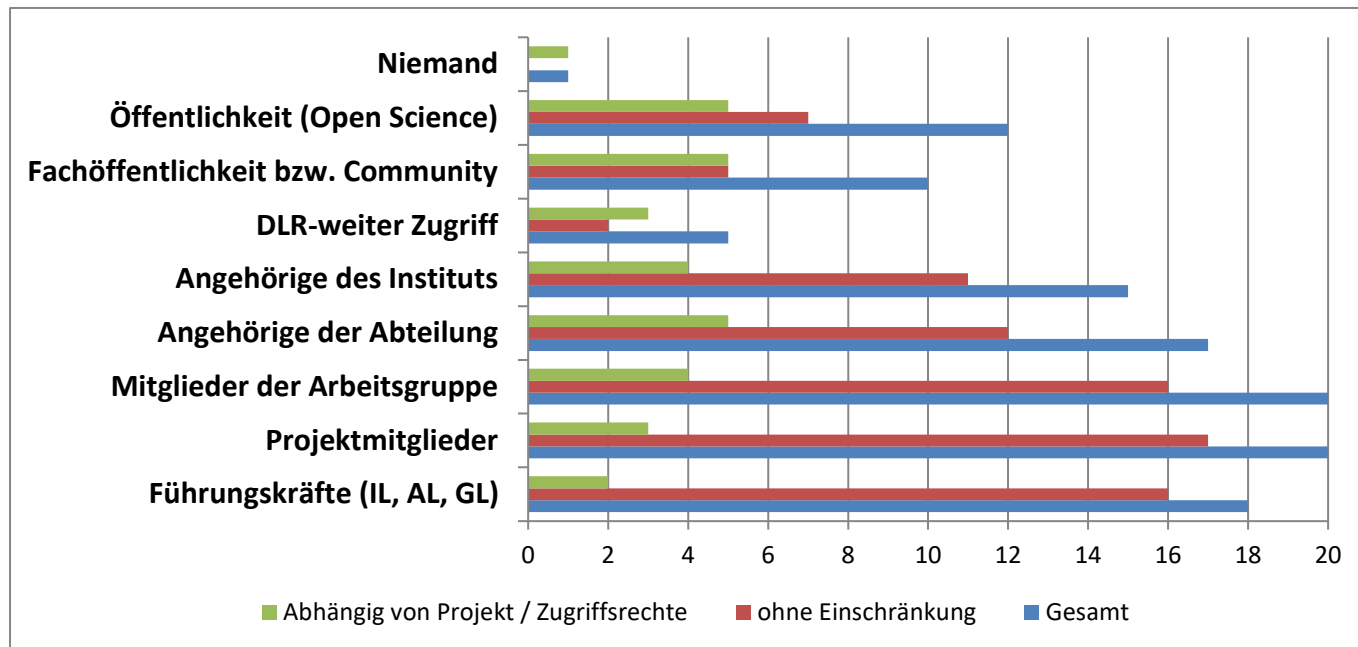


Abbildung 11: Gegenwärtiger Zugang zu archivierten Forschungsdaten

Frage 13: Werden für Projekte bereits Datenmanagementpläne im Institut erstellt?

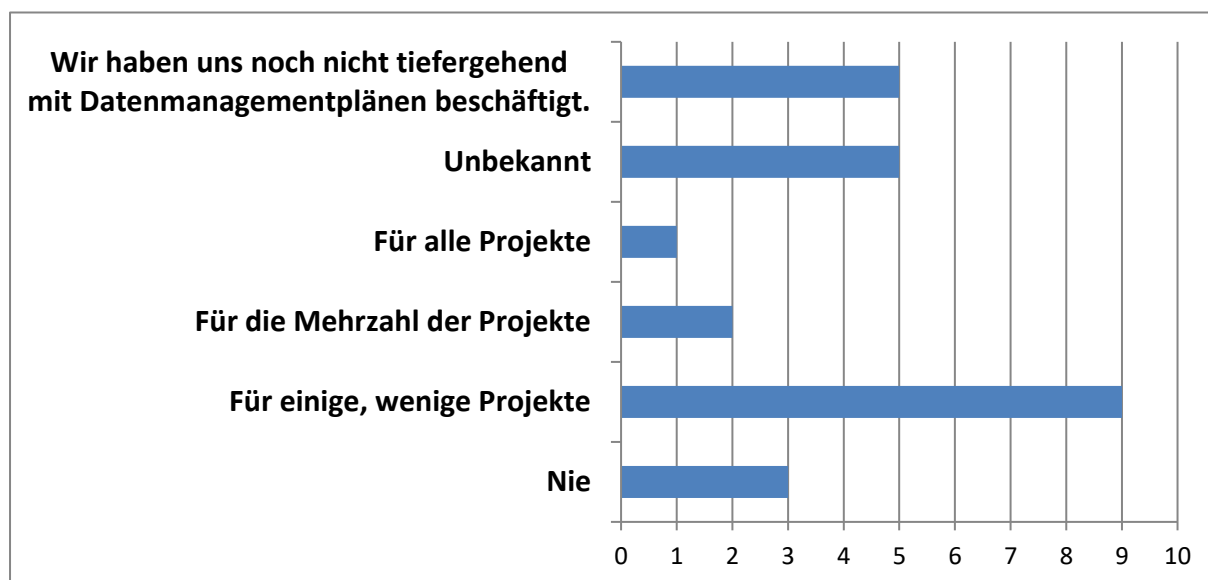


Abbildung 12: Nutzungsgrad von Datenmanagementplänen

Die Frage nach der Erstellung und somit Nutzung von Datenmanagementplänen dient sowohl der Erhebung des Status quo als auch der Bewusstseinsbildung für diese Problematik, um daraus einen entsprechenden Aufklärungs- und Schulungsbedarf abzuleiten. Wie zu erwarten war, ist der größte Anteil bei den Nichtnutzern

zu finden. Obwohl immer mehr Zuwendungsgeber einen Datenmanagementplan einfordern, gehört er noch nicht zum normalen Workflow in den Instituten. Es wird wahrscheinlich befürchtet, dass der Dokumentationsaufwand für die Daten zu hoch wird und nicht genügend Zeit für die eigentliche Forschungsarbeit verbleibe. Aufwand und Nutzen scheinen den Wissenschaftlern in diesem Punkt in einem Missverhältnis zu stehen.

Frage 14: Wenn Sie bereits Datenmanagementpläne erstellen, welche Werkzeuge setzen Sie dafür ein?

Mit Hilfe der geschlossenen Frage nach dem Datenmanagementplänen sollte ermittelt werden, inwieweit bereits gängige Unterstützungstools mit entsprechenden vorkonfigurierten Fragenkatalogen zur Hilfe herangezogen werden. 30 % der Befragten gaben an, keine Werkzeuge einzusetzen. Bei den Antworten mit JA wurde um die Angabe der genutzten Tools gebeten:

- MS Excel, MS Project, Derzeit wird überlegt / getestet, die Datenmanagementpläne im RDMO zu erstellen;
- Es werden keine Werkzeuge verwendet. Die Dokumentation erfolgt innerhalb von Projektrepositorien. Die Datenmanagementpläne folgen allerdings keinem vordefinierten Prozess/Muster;
- Absprachen mündlich & schriftlich;
- Netzlaufwerke; Indexing; Repository; Langzeitarchiv;
- Office Produkte;
- Hirnschmalz.

Frage 15: Existieren für die Forschungsgebiete des Instituts spezielle Workflows und/oder Leitlinien, die den Umgang mit den Forschungsdaten regeln? Falls ja, bitte konkretisieren Sie, um welche Workflows und/oder Leitlinien es sich dabei handelt.

Neben den Regeln zur guten wissenschaftlichen Praxis gibt es für einzelne Fachdisziplinen bereits spezifische Regelungen zum Umgang und zur Bereitstellung von Forschungsdaten. Einige Forschungsinstitute, Drittmittelgeber und Fachorganisationen haben bereits Richtlinien, sogenannte Data Policies, erstellt. Um zu ermitteln, inwieweit Richtlinien zum Umgang mit Forschungsdaten bekannt sind, existieren und ob diese angewendet werden, wurde gefragt, welche Workflows und/oder Leitlinien umgesetzt werden:

- Es gibt Leitlinien der EC, des BMWi und der DFG; Teilweise über AT-QM geregelt; Abteilungsinterne Organisation der Datenablage, Allgemeine Archivierungsrichtlinien sind in Arbeit (QMS), Definierte Ordnerstrukturen auf den Netzlaufwerk vorhanden;
- Qualitätsmanagement Handbuch;
- Bestandteil des Institutshandbuches (und zertifiziert) ist ein Prozess zur Dokumentation von Daten im Institut incl. Workflows und gute wiss. Praxis. Dieser Prozess wird derzeit angepasst, da sich naturgemäß die Wissenschaftsumgebung, Art und Menge der Daten und die technischen Anforderungen / Möglichkeiten verändern. Diese Gelegenheit wird genutzt, um den Stand im Forschungsdatenmanagement in die Institutsprozesse aufzunehmen;
- Das Institut ist an einer Zusammenarbeit im DLR und an DLR – Lösungen sehr interessiert;
- Es gibt in den einzelnen Abteilungen/Gruppen vereinbarte Workflows zum Datenmanagement;
- Das Institut verfolgt eine eigene Leitlinie der guten wissenschaftlichen Praxis. Aufbewahrung von Forschungsdaten beträgt bspw. mindestens 10 Jahre. Daten sind grundsätzlich auf sicheren Netzlaufwerken und nicht lokal zu speichern. Workflows werden projektspezifisch durch die Projektleiter festgelegt. Es gilt jedoch, dass Daten in Netzlaufwerken innerhalb von Projektordnern gespeichert werden;
- Eingabe Daten in Hypertext ist fester Bestandteil von Kampagnendurchführung;
- Projektdaten werden in Projektordnern (DLR-Ressource) gespeichert, nach Ablauf des Projektes werden diese Daten in einem Langzeitarchiv für 10 Jahre archiviert. Nach 10 Jahren werden die Archive gelöscht;



- Basierend auf den Anforderungen der Zuwendungsgeber;
- DFG Richtlinie zur guten wissenschaftlichen Praxis;
- HGF Richtlinie zum Umgang mit Forschungsdaten;
- DLR Grundlagen zur Forschungsdatenhaltung;
- Instituts-interne Verfahren zu Publikation und DOI Vergabe;
- Kapitel im Management-Handbuch des Instituts: Management von dokumentierten Informationen;
- IT-Richtlinie zu geeigneten Datenformaten für die Archivierung von Dokumenten;
- Im Institut besteht ein QM-Prozess, der die definiert, nach welcher Ordner-Struktur Projekt- und Vorhabensdaten auf dem gesicherten Institutsserver abzulegen sind. Diese Struktur ist nach Projektende ins Datenarchiv des Instituts zu überführen;
- Es existiert eine Archivierungsrichtlinie im Institut

2.4. Fragen zu Unterstützungsangeboten

Frage 16: Wie schätzen Sie das vorhandene Wissen zum Forschungsdatenmanagement in Ihrem Institut ein?



Abbildung 13: Einschätzung des eigenen Kenntnisstandes zum Forschungsdatenmanagement

Anhand der Antworten zu der Fragestellung zeigt sich, dass für viele Forschenden die Themen rund um Forschungsdaten und Forschungsdatenmanagement ein neuartiges Phänomen sind. Die Einschätzung ist sehr subjektiv und richtet sich entscheidend nach den Vorkenntnissen der Forschenden sowie den in der jeweiligen Wissenschaftsdisziplin vorherrschenden Werkzeugen. Hier deutet sich ein wesentliches Problemfeld an, wenn WissenschaftlerInnen zwar grundsätzlich offen für die Erprobung neuer Verfahren sind, es aber häufig an einem konkreten Wissen um bestehende Angebote mangelt.

Frage 17: Waren Sie in Bezug auf Forschungsdaten schon einmal mit folgenden Problemen konfrontiert bzw. befürchten Sie diese für die Zukunft?

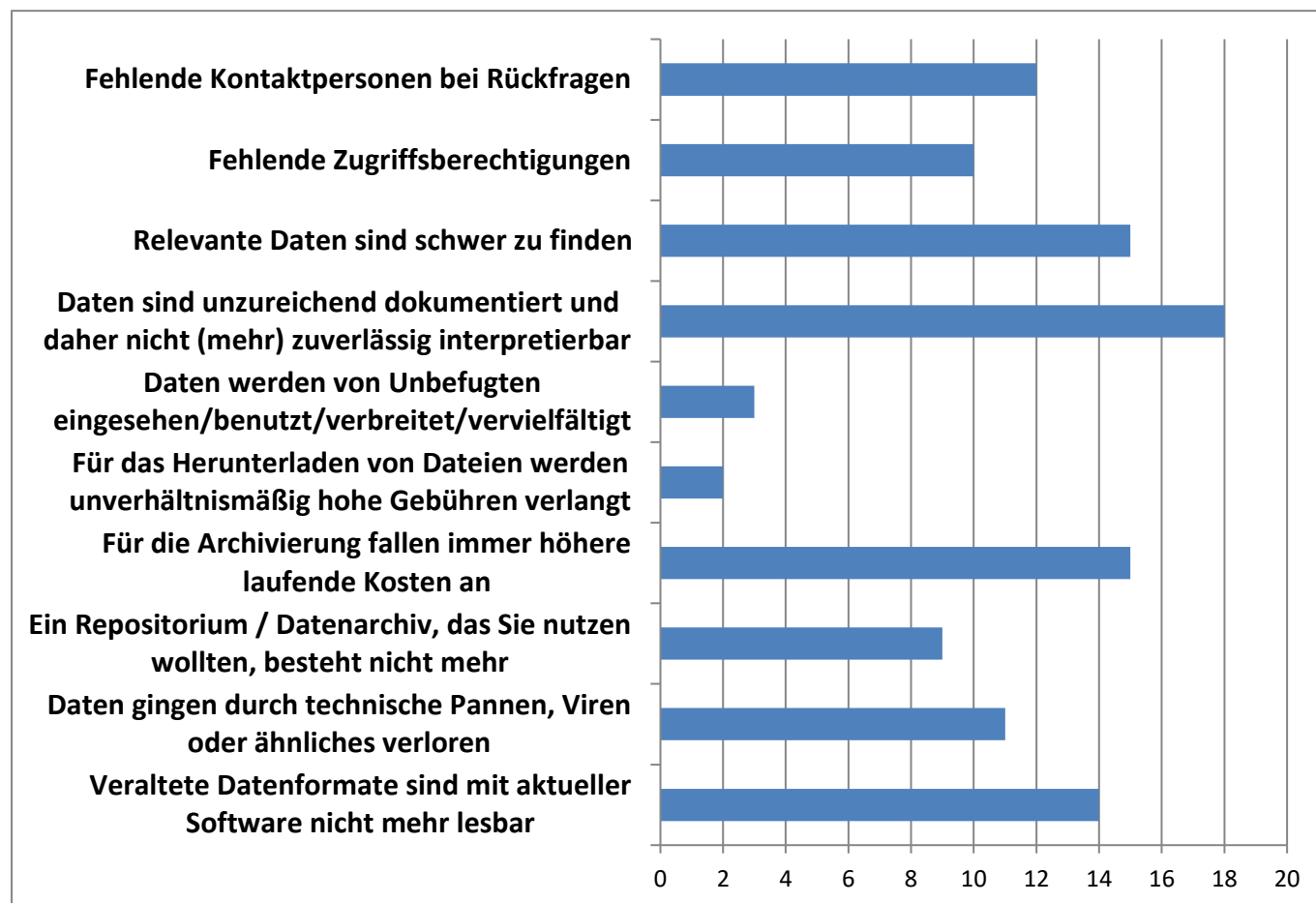


Abbildung 14: Problemfelder in Bezug auf Forschungsdaten

Eine Perspektivierung des Themas in dieser Richtung ermöglicht es nun, konkretere Problemfelder zu benennen und diese in geeigneten Formaten aufzubereiten, bereitzustellen und zu vermitteln.

Bei dieser Frage bestand die Möglichkeit weitere, nicht in der Abfrageliste aufgeführte, Datentypen zu nennen. Die Freitextantworten sind:

- Probleme bei ausgeschiedenen MA;
- Wachsende Datenmengen;
- Proprietäre Software (z.B. veraltete Windows Software), proprietäre Formate (Excel), umständliche/aufwändige Datenzugriffe wegen IT-Security-Policies (Firewall);
- Urheberrecht unbekannt.

2.5. Abschließende Frage

Frage 18: Welche Unterstützungsangebote würden Sie sich im DLR wünschen?



Abbildung 15: Gewünschte Unterstützungsangebote für das Forschungsdatenmanagement

Ziel der Frage ist, eine Angebots-, Informations- und Beratungsstruktur zu Forschungsdaten zu ermitteln. Auf dieser Basis sollen Lösungsansätze zum Forschungsdatenmanagement für die Wissenschaften entwickelt werden. Nicht zuletzt sollen so WissenschaftlerInnen dabei unterstützt werden, beispielsweise einen geeigneten Ort zur Veröffentlichung ihrer Forschungsdaten zu finden.

Bei dieser Frage bestand die Möglichkeit weitere, nicht in der Abfrageliste aufgeführte, Datentypen zu nennen. Die Freitextantworten sind:

- Externe Bereitstellung bereits verwerteter Datensätze;
- Zentrale Bereitstellung von DOIs zur Veröffentlichung;
- Unterstützung bei Langzeitarchivierung mit entsprechenden Zusicherungen zur Laufzeit und Zugriff von extern;
- Einbindung von Datenpublikationen in elib (warum soll mit edata ein separates, weitgehend funktionsgleiches Portal aufgebaut werden?)
- Tools bzw. Datenbanken, die möglichst leicht Open Access ohne Aufwand im Institut ermöglichen;

- Der Zugriff auf den DLR Speicherplatz sollte über Open Source Technologien und nicht über proprietäre Schnittstellen erfolgen.
 - Implementierung eines praktikablen FDM unter Berücksichtigung von großen Datenmengen, Nutzerunabhängigkeit (Zeitverträge, Langfristigkeit) und Transparenz (Was liegt wo?) mit Zugriffskontrolle (z.B. zeitbeschränkte Schlüssel)
 - Zugang zu / Umgang mit öffentlichen Datenarchiven

3. Schlussfolgerungen

Die vorgestellte Umfrage liefert eine Informationsbasis für viele Punkte des Arbeitsprogrammes des FDM-Projektes. Es ist schwierig, alle Aspekte zum Umgang mit Forschungsdaten in nur einer Umfrage zu adressieren, weil sich Forschungsdatenmanagement immer am Forschungskontext (Einzelprojekt, Projektverbund, etc.), der Fachdisziplin und dem konkreten Forschungsvorhaben ausrichtet. Der Fragebogen stellt einige Aspekte deshalb sehr abstrakt dar.

Trotz der Einschränkungen, die eine so breit angelegte Umfrage an eine heterogene Zielgruppe bzgl. des Themas mit sich bringt, kann festgehalten werden, dass das Forschungsdatenmanagement und insbesondere die Datenarchivierung bei den DLR Instituten relevant und präsent sind. Dies belegt schon die gute Rücklaufquote von 47 %. Das Thema ist also gegenwärtig, wenn auch eher mit einem engen Blick auf die geforderte Datenarchivierung bei Projektabschluss. Insbesondere wird der Bedarf nach Infrastrukturen zur Archivierung der Forschungsdaten explizit von den Befragten artikuliert.

Des Weiteren wurde deutlich, dass die Teilnehmer der Umfrage bereit sind, ihre Forschungsdaten zur Nachnutzung zur Verfügung zu stellen. Sie sind aber zurückhaltend und unsicher, was dies für ihren Forschungsalltag bedeutet und ob die notwendigen Infrastrukturen vorhanden sind. Eine zentrale Annahme hat sich bestätigt: Der Bedarf für ein umfangreiches und niederschwelliges Beratungs- und Informationsangebot ist groß. Dies hatte sich schon im Vorfeld in der Beratungspraxis und dem Austausch mit anderen Serviceeinrichtungen für das Forschungsdatenmanagement abgezeichnet.

Die Möglichkeiten des Forschungsdatenmanagements sind vielen Forschenden noch nicht vollumfänglich bekannt. In diesem Zusammenhang müssen die Begrifflichkeiten wie z. B. Repositorium, Metadaten, Langzeitarchivierung, Nachnutzung von Forschungsdaten, etc. breiter vermittelt werden. Vor allem aber sind allgemeine Richtlinien zum Umgang mit Forschungsdaten als auch fach- und datenspezifischen Standards verständlich und leicht zugänglich darzustellen. Ergänzend dazu sollten Best-Practice-Beispiele für die konkrete Umsetzung des Forschungsdatenmanagements angeboten werden.

Ferner werden Informationen über die rechtlichen Rahmenbedingungen benötigt, die einfach und anschaulich darstellen, welche Auswirkungen diese auf das Forschungsdatenmanagement im eigenen Projekt haben und wie den daraus erwachsenden Anforderungen begegnet werden kann, um der Verunsicherung, ob und wie man Daten Dritten zur Verfügung stellen kann, entgegenzuwirken.

Demnach müssen Arbeitsprozesse und Werkzeuge so gestaltet sein, dass bspw. die Beschreibung der Forschungsdaten (Metadaten) keinen übermäßigen Mehraufwand bedeuten. Aber auch die Anerkennung der Bereitstellung nachnutzbarer Forschungsdaten als wissenschaftliche Leistung wird hierzu beitragen. Ebenso könnten institutionelle Rahmenleitlinien, Organisationsmodelle und zentrale IT-Infrastrukturen helfen, die Abläufe ressourcenschonender zu organisieren, um der eigentlichen Forschung genügend Raum zu lassen. Dazu zählt auch die Entscheidung darüber, welche Forschungsdaten der Forschende wann, zu welchen Bedingungen und wo archiviert oder bereitstellt.

All dies stellt hohe Anforderungen an IT-Infrastrukturen und an die mit ihnen verbundenen Services: Sie sollten einfach zu bedienen, aber auch individuell auf die unterschiedlichsten Anforderungen anpassbar sein. Darüber hinaus sollten sie sowohl einfache als auch komplexe Daten aufnehmen können, und nicht zuletzt ist es erforderlich, dass der Datenzugang individuell festgelegt werden kann, weil zwischen der einfachen Veröffentlichung



lichung eines kompletten Datensatzes und der vollkommenden Sperrung des Datenzugriffs für andere auf einen archivierten Datensatz mannigfaltige Varianten liegen. Ebenso müssen Organisationsmodelle gefunden werden, die die Verantwortlichkeiten für die Daten in diesem Repositorium definieren, um z. B. Regelungen für Datenbestände vorzuhalten, deren Datengeber nicht mehr erreichbar sind.

Zusammenfassend muss die Implementierung des Forschungsdatenmanagements als ein tiefgreifender Veränderungsprozess in der Wissenschaft gesehen werden, der alle am Wissenschaftssystem beteiligten Akteure herausfordert.



Abbildungsverzeichnis

Abbildung 1: Verteilung der Umfrageteilnehmer bezogen auf die einzelnen Schwerpunkte.....	6
Abbildung 2: Gängige im DLR genutzte Arten von Forschungsdaten	7
Abbildung 3: Gängige genutzte Datentypen im DLR	8
Abbildung 4: Einschätzung des Speichervolumens der gegenwärtigen Tätigkeit.....	9
Abbildung 5: Abschätzung der Größenordnung der gesamten Speichervolumens pro Jahr	10
Abbildung 6: Gesamtgröße der bisher gespeicherten Daten	10
Abbildung 7: Wer im Institut ist im Datenzyklus beteiligt.....	11
Abbildung 8: Speicherorte während der Forschungsphase.....	12
Abbildung 9: Derzeitige Aufbewahrungszeiträume für archivierte Forschungsdaten	12
Abbildung 10: Rahmenbedingungen für die Archivierung der Forschungsdaten	13
Abbildung 11: Gegenwärtiger Zugang zu archivierten Forschungsdaten	14
Abbildung 12: Nutzungsgrad von Datenmanagementplänen.....	14
Abbildung 13: Einschätzung des eigenen Kenntnisstandes zum Forschungsdatenmanagement	16
Abbildung 14: Problemfelder in Bezug auf Forschungsdaten	17
Abbildung 15: Gewünschte Unterstützungsangebote für das Forschungsdatenmanagement.....	18



Anhang Fragebogen

Mit Hilfe der folgenden Fragen versuchte das Projektteam einen Überblick über die aktuelle Praxis in Bezug auf das Forschungsdatenmanagement im DLR sowie zu gewünschten Unterstützungsangeboten erhalten. Die Ergebnisse werden zur Ausgestaltung der Projektinhalte genutzt und so unterstützte die Beteiligung der Institute bzw. Einrichtungen aktiv den Ausbau des Forschungsdatenmanagements im DLR.

Untersuchung zum Forschungsdatenmanagementbedarf im DLR

Das DLR hat 2018 eine [Richtlinie zum Umgang mit Forschungsdaten](#) als Teil des Organisationshandbuchs verabschiedet. Zur Unterstützung der dort geforderten Praxis, hat das DLR unter der Leitung des DLR-Forschungsdatenmanager Dr.-Ing. Christian Langenbach ein Projekt zum Aufbau von Unterstützungsangeboten im DLR initiiert. Mit Hilfe der folgenden Fragen möchte das Projektteam einen Überblick über die aktuelle Praxis in Bezug auf das Forschungsdatenmanagement im DLR sowie zu gewünschten Unterstützungsangeboten erhalten. Die Ergebnisse nutzen wir zur Ausgestaltung der Projektinhalte. Sie unterstützen durch die Beteiligung ihres Instituts bzw. ihrer Einrichtung aktiv den Ausbau des Forschungsdatenmanagements im DLR.

Die maximal 19 Fragen benötigen ca. 20 min reine Bearbeitungszeit. Pro Institut bzw. Einrichtung des DLR soll ein Untersuchungsbogen für das Institut gesamt ausgefüllt werden. Die zur Beantwortung erforderlichen Informationen benötigen ggf. die Zuarbeit weiterer Personen. Sprechen Sie sich bitte intern dazu in ihrem Institut bzw. ihrer Einrichtung ab.

Den ausgefüllten Untersuchungsbogen senden Sie bitte an die E-Mail-Adresse fdm@dlr.de zurück.

Im Fall von Fragen, wenden Sie sich bitte direkt an Herrn Langenbach (E-Mail: Christian.Langenbach@dlr.de, Telefon: 02203 / 601-2704).



Einleitende Fragen

Für welches Institut bzw. welche Einrichtung des DLR beantworten Sie die folgenden Fragen?

Klicken Sie hier, um Text einzugeben.

Welche Person kann man für das Institut/die Einrichtung bei ggf. weiteren Nachfragen zum Forschungsdatenmanagement ansprechen?

Klicken Sie hier, um Text einzugeben.



Fragen zu Art und Umfang der forschungsbezogenen Daten

Welche Arten von Forschungsdaten fallen im Rahmen der Forschung in Ihrem Institut an?

Bitte geben Sie alle zutreffenden Optionen an.

Experimentelle Daten	
Quantitative Messdaten	
Statistiken	
Qualitative Daten aus Beobachtungen	
Quantitative Analysedaten aus Beobachtungen	
Umfrageantworten (qualitativ)	
Umfrageantworten (quantitativ)	
Modelle	
Zeitreihendaten	
Software	
Sonstiges (Bitte explizit nennen!)	

In welchen Datentypen liegen die Forschungsdaten in Ihrem Institut vor?

Bitte geben Sie alle zutreffenden Optionen an.

Textformate (z.B. docx, odt, ppt, md, txt, pdf)	
Bildformate (z.B. png, jpeg, gif, raw)	
Gerätespezifische Dateiformate	
Datenbankformate (z.B. mdb)	
Audioformate (z.B. mp3, wma, ogg)	
Videoformate (z.B. avi, mpeg, mov)	
Metadatenformate (z.B. EAD, XMP, ID-3)	
Statistikformate (z.B. sav)	
Tabellenformate (z.B. xls, csv)	
Quellcode (z.B. .py, .java)	
Sonstiges (Bitte explizit nennen!)	



Wie hoch ist derzeit typischerweise die Größe der Forschungsdaten, die für eine einzelne wissenschaftliche Publikation archiviert werden müssen?

Bitte kennzeichnen Sie pro Zeile die zutreffende Option.

	Nie	Gelegentlich	Häufig	Unbekannt
Mehrere KB				
Mehrere MB				
Bis zu 1 GB				
Mehrere GB				
Bis zu 1 TB				
Größer als 1 TB				

In welcher Größenordnung entstehen pro Jahr Forschungsdaten in ihrem Institut?

Bitte kennzeichnen Sie genau die für ihr Institut zutreffende Option.

Weniger als 1 GB pro Jahr	
1 – 100 GB pro Jahr	
101 GB – 1 TB pro Jahr	
Mehrere TB pro Jahr	
Mehr als 1 PB pro Jahr	
Unbekannt	

Welchen Umfang hat die Gesamtgröße der an Ihrem Institut vorliegenden Forschungsdaten ungefähr?

Bitte kennzeichnen Sie genau die für ihr Institut zutreffende Option.

Bis 1 GB	
1 – 100 GB	
101 GB – 1 TB	
Mehrere TB	
Mehr als 1 PB	
Unbekannt	



Fragen zum Stand der Praxis

Wer bzw. welche Rollen sind typischerweise beim Thema Forschungsdatenmanagement im Institut beteiligt?

Bitte kennzeichnen Sie alle zutreffenden Optionen.

Rolle	Planung	Erhebung & Verarbeitung	Analyse & Auswertung	Sicherung & Archivierung	Dokumentation	Publikation
Institutsleiter*in						
Abteilungsleiter*in						
Gruppenleiter*in						
Projektleiter*in						
Wissenschaftliche Mitarbeiter*innen						
Technische Mitarbeiter*innen						
IT-Manager*in bzw. IT-Team des Instituts						
Dienstleister des DLR						
Externe Dienstleister						
Hilfskräfte						
Ist mir unbekannt						
Ist nicht vorgesehen						
Niemand						

Wo speichert das Institut die Forschungsdaten, die z.B. im Rahmen von Projekten anfallen?

Bitte geben Sie alle zutreffenden Optionen an.

Services des Instituts (z.B. Netzlaufwerke)	
DLR-Services (z. B. DLR-Teamsites, DLR-Subversion)	
Externe Datenträger	
Lokal auf dienstlichem Rechner	
Bei einem Datenarchiv für diese Art von Forschungsdaten	
Cloud Anbieter	
Sonstiges (Bitte explizit nennen!)	

Wie lange bewahrt das Institut Forschungsdaten in der Regel nach Abschluss des jeweiligen Projekts auf?

Bitte geben Sie alle zutreffenden Optionen an.

Bis zu 6 Monaten	
Bis zu 1 Jahr	
Bis zu 3 Jahren	
Bis zu 5 Jahren	
Bis zu 10 Jahren	
> 10 Jahre	
Unbekannt	

Zu welchem Zweck bewahrt das Institut Forschungsdaten auf?

Bitte geben Sie alle zutreffenden Optionen an.

Publikation	
Nachweis der Reproduzierbarkeit von Ergebnissen	
Eigene Re-Analyse	
Ausschluss rechtlicher Risiken	
Vorschrift	
Übungszwecke	
Ohne inhaltlichen Grund	
Sonstiges (Bitte explizit nennen!)	

Wer darf auf die Forschungsdaten zugreifen?

Bitte geben Sie alle zutreffenden Optionen an.

Führungskräfte (IL, AL, GL)	
Projektmitglieder	
Mitglieder der Arbeitsgruppe	
Angehörige der Abteilung	
Angehörige des Instituts	
DLR-weiter Zugriff	
Fachöffentlichkeit bzw. Community	
Öffentlichkeit (Open Science)	
Niemand	



Werden für Projekte bereits Datenmanagementpläne im Institut erstellt?

Bitte kennzeichnen Sie genau die für ihr Institut zutreffende Option.

Nie	
Für einige, wenige Projekte	
Für die Mehrzahl der Projekte	
Für alle Projekte	
Unbekannt	
Wir haben uns noch nicht tiefergehend mit Datenmanagementplänen beschäftigt.	

Wenn Sie bereits Datenmanagementpläne erstellen, welche Werkzeuge setzen Sie dafür ein?

Klicken Sie hier, um Text einzugeben.

Existieren für die Forschungsgebiete des Instituts spezielle Workflows und/oder Leitlinien, die den Umgang mit den Forschungsdaten regeln? Falls ja, bitte konkretisieren Sie, um welche Workflows und/oder Leitlinien es sich dabei handelt.

Klicken Sie hier, um Text einzugeben.



Fragen zu Unterstützungsangeboten

Wie schätzen Sie das vorhandene Wissen zum Forschungsdatenmanagement in Ihrem Institut ein?

Bitte kennzeichnen Sie genau die für ihr Institut zutreffende Option.

Gut	
Durchschnittlich	
Gering	

Waren Sie in Bezug auf Forschungsdaten schon einmal mit folgenden Problemen konfrontiert bzw. befürchten Sie diese für die Zukunft?

Bitte geben Sie alle zutreffenden Optionen an.

Veraltete Datenformate sind mit aktueller Software nicht mehr lesbar	
Daten gingen durch technische Pannen, Viren oder ähnliches verloren	
Ein Repository / Datenarchiv, das Sie nutzen wollten, besteht nicht mehr	
Für die Archivierung fallen immer höhere laufende Kosten an	
Für das Herunterladen von Dateien werden unverhältnismäßig hohe Gebühren verlangt	
Daten werden von Unbefugten eingesehen/benutzt/verbreitet/vervielfältigt	
Daten sind unzureichend dokumentiert und daher nicht (mehr) zuverlässig interpretierbar	
Relevante Daten sind schwer zu finden	
Fehlende Zugriffsberechtigungen	
Fehlende Kontaktpersonen bei Rückfragen	
Sonstiges (Bitte explizit nennen!)	



Welche Unterstützungsangebote würden Sie sich im DLR wünschen?

Bitte geben Sie alle zutreffenden Optionen an.

Gesicherten, verlässlichen und leicht zugänglichen Speicherplatz für Forschungsdaten	<input type="checkbox"/>
Beratung bei rechtlichen Fragen (z.B. Zugangseinschränkungen, Umgang mit sensiblen Daten, Nutzung von Lizenzen)	<input type="checkbox"/>
Beratung bei technischen Fragen (z.B. Metadaten, Standards, langfristige Archivierung)	<input type="checkbox"/>
Beratung zum Publizieren und Zitieren von Forschungsdaten	<input type="checkbox"/>
Beratung in allgemeinen Fragen zum Umgang mit Forschungsdaten	<input type="checkbox"/>
Unterstützung bei konkreten Angelegenheiten (z.B. beim Einreichen eines Manuskripts in einer Zeitschrift, die Veröffentlichung von Forschungsdaten vorgibt)	<input type="checkbox"/>
Unterstützung beim Erstellen eines Datenmanagementplans	<input type="checkbox"/>
Wir brauchen keine Unterstützung.	<input type="checkbox"/>
Sonstiges (Bitte explizit nennen!)	<input type="checkbox"/>



Abschließende Fragen

Haben Sie weitere Hinweise oder Ideen für uns?

Klicken Sie hier, um Text einzugeben.

Bitte senden Sie den ausgefüllten Fragebogen bis zum 16.09.2019 an fdm@dlr.de!

Wir bedanken uns recht herzlich für Ihre Mitarbeit!

