



Organisatorisch-technisches Konzept für eine Forschungsdaten-Infrastruktur in der TU Berlin

– Kurzfassung –

AG Forschungsdaten (M. Kuberek)
7. März 2012

Kontakt:

Monika Kuberek

Universitätsbibliothek der Technischen Universität Berlin

E-Mail: monika.kuberek@tu-berlin.de

Das Konzept steht in einer Kurzfassung und in einer Langfassung online auf der Website des „Servicezentrums Forschungsdaten und –publikationen“ zur Verfügung.

Kurzfassung:

http://www.szf.tu-berlin.de/fileadmin/f33_szf/TUB_Forschungsdaten_Konzept_kurz_20120307.pdf

Langfassung:

http://www.szf.tu-berlin.de/fileadmin/f33_szf/TUB_Forschungsdaten_Konzept_lang_20120315.pdf

Beide Versionen des Konzeptes stehen unter folgender Creative-Commons-Lizenz:

„Attribution-NonCommercial-ShareAlike 3.0 Unported“

<http://creativecommons.org/licenses/by-nc-sa/3.0/>



Inhalt

Management Summary	3
1. Ausgangssituation und Zielsetzung.....	4
1.1 Forschungsdaten in der TU Berlin	4
1.2 Services für die TU-Wissenschaftler/innen und Mehrwert für die TU Berlin	4
2. Organisatorisches Modell: „Servicezentrum Forschungsdaten und -publikationen“	4
2.1 Verteilte Zuständigkeiten und strukturierte Workflows.....	5
2.2 Organisationsstruktur	6
3. Umsetzung des Konzepts: Projekt „Forschungsdaten“	6
4. Perspektive: Gesamtnachweis der in der TU Berlin erbrachten Forschungsleistungen.....	7
4.1 Nachweis und Recherche der Forschungsergebnisse	7
4.2 Perspektive Gesamtnachweis	7

Management Summary

„Unter Forschungsdaten sind ... digitale und elektronisch speicherbare Daten zu verstehen, die im Zuge eines wissenschaftlichen Vorhabens z.B. durch Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“ (DFG 2010)¹

Das vorliegende organisatorisch-technische Konzept hat zum Ziel, an der TU Berlin eine Forschungsdaten-Infrastruktur aufzubauen, die die langfristige Sicherung und Verfügbarkeit der hier entstandenen Forschungsdaten und der darauf basierenden Forschungspublikationen² gewährleistet.

Um die Forschungsdaten-Infrastruktur organisatorisch in der TU Berlin zu verankern und die Nachhaltigkeit der getroffenen Maßnahmen sicherzustellen, sieht das Konzept vor, dass Universitätsbibliothek (UB), IT-Service-Center (tubIT) und Forschungsabteilung in einem „Servicezentrum Forschungsdaten und -publikationen“ kooperieren. Zur Unterstützung der Wissenschaftler/innen stellt das Servicezentrum ein umfassendes Dienstleistungs- und Beratungsangebot zu Forschungsdaten und -publikationen bereit. Es wird eine technische Infrastruktur aufgebaut, die dem großen Spektrum unterschiedlicher Fachdisziplinen an der TU Berlin gerecht wird und in der Lage ist, Forschungsdaten vieler Disziplinen zu verwalten und zu speichern. Dabei werden Forschungsdaten und -publikationen als Einheit angesehen, die gegenseitig aufeinander referenzieren; beide werden im institutionellen Forschungsdaten-Repository der TU Berlin gespeichert. In dem Konzept wird berücksichtigt, dass eine Reihe von Fach-Communities in den vergangenen Jahren weltweit disziplinspezifische Forschungsdaten-Infrastrukturen aufgebaut hat, an denen sich auch Wissenschaftler/innen der TU Berlin beteiligen. Zur Dokumentation der von Wissenschaftler/innen der TU Berlin erbrachten Forschungsleistungen ist perspektivisch ein Gesamtnachweis – inkl. Link zu den Forschungsdatensets und Volltexten – über die verteilten Forschungsdaten-Repositories (institutionell und weltweit) hinweg möglich.

Das Konzept soll im Rahmen eines Projektes ab der ersten Jahreshälfte 2012 umgesetzt werden.

Erstellt wurde das Konzept zum Umgang mit Forschungsdaten in der TU Berlin von einer Arbeitsgruppe, die – unter Federführung der UB und mit Beteiligung von tubIT und Forschungsabteilung – im September 2011 von der Kanzlerin eingesetzt wurde.³

¹ www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_forschungsdaten_1001.pdf, S. 1. Weitere Begriffe für Forschungsdaten: Rohdaten, Primärdaten, Forschungsprimärdaten.

² Forschungsdaten und -publikationen werden im Text zusammenfassend auch als Forschungsergebnisse bezeichnet.

³ Mitglieder der AG Forschungsdaten: M. Kuberek (UB/HA3), A. De Nobile (ZUV/PDM) und D. Otto (Abt. V/V C 4).

1. Ausgangssituation und Zielsetzung

Im letzten Jahrzehnt ist weltweit die Erkenntnis gewachsen, dass Forschungsdaten eine unschätzbare Ressource und einen Wert an sich darstellen, den es zu erhalten, disziplinübergreifend bereitzustellen und nachzunutzen gilt. Bereits 1998 hat die DFG auf den Wert von Forschungsdaten hingewiesen und die sichere Aufbewahrung durch die Institution, in der sie entstanden sind, in ihrer Denkschrift zur „Sicherung guter wissenschaftlicher Praxis“ formuliert. Inzwischen hat sie in ihren Förderrichtlinien entsprechende Anforderungen als Voraussetzung für die Mittelbewilligung aufgenommen.

1.1 Forschungsdaten in der TU Berlin

Der „guten wissenschaftlichen Praxis“ hat sich auch die TU Berlin verpflichtet und 2002 Richtlinien aufgestellt, in denen die Aufbewahrungspflicht von Primärdaten ausdrücklich verankert ist. Die TU Berlin ist eine traditionsreiche und renommierte Forschungsuniversität, an der fortlaufend eine Vielzahl von Forschungsprojekten durchgeführt wird, sowohl Drittmittelprojekte als auch durch die TU Berlin selbst grundfinanzierte Projekte. Eine TU-weite Infrastruktur zur Aufbewahrung von Forschungsdaten existiert bislang nicht, und es droht die Gefahr, dass die Daten verlorengehen, wenn ein Projekt zu Ende ist oder ein/e Wissenschaftler/in an eine andere Institution wechselt. In den verschiedenen Fachdisziplinen ist der Bedarf an einer institutionellen Forschungsdaten-Infrastruktur unterschiedlich ausgeprägt: So wurden in einer Reihe von Fachdisziplinen – zumeist initiiert von den jeweiligen wissenschaftlichen Fachgesellschaften – in den vergangenen Jahren weltweit verteilt disziplinspezifische Repositorien und Archive aufgebaut, an denen auch Wissenschaftler/innen der TU Berlin partizipieren. Nicht für alle Fachdisziplinen, die es an der TU Berlin gibt, existieren jedoch solche Infrastrukturen; auch bei fächerübergreifenden Forschungsvorhaben besteht Bedarf, zudem bei kleinen Projekten. Angesichts der stetig steigenden Zahl an Forschungsaktivitäten, von denen viele interdisziplinär angelegt sind, gilt es daher, in der TU Berlin eine leistungsfähige, verlässliche und nachhaltige Infrastruktur für Forschungsdaten aufzubauen. Sie umfasst neben einer technischen Komponente das Forschungsdatenmanagement, das alle Aspekte hinsichtlich der Forschungsdaten, ihrer Beschreibung, Veröffentlichung und Aufbewahrung organisiert und regelt.

1.2 Services für die TU-Wissenschaftler/innen und Mehrwert für die TU Berlin

Den Wissenschaftler/innen bringt die Infrastruktur mit ihrem umfangreichen Serviceangebot in vielerlei Hinsicht Vorteile: automatisierte Verfahren zum öffentlichen Nachweis ihrer Forschungsergebnisse (auf der eigenen Homepage, in fachspezifischen Repositorien usw.) und damit Aufwandsminimierung bei gleichzeitig erhöhter Sichtbarkeit der eigenen Forschung, eine hohe Transparenz und Vertrauenswürdigkeit der Publikationen durch Referenzierung auf supplementäre Forschungsdaten, die dauerhafte Zitierfähigkeit von Publikationen und Forschungsdaten durch konstante Internet-Adressen (Persistent Identifier), die nachhaltige Speicherung der Forschungsdaten und -publikationen sowie eine umfassende Beratung und Betreuung beim gesamten Prozess der Erstellung zitier- und recherchierfähiger Metadaten und der Publikation ihrer Forschungsergebnisse.

Die TU Berlin als Institution profitiert insbesondere davon, dass ihre Forschungsleistungen nach außen dokumentiert und sichtbar werden. Darüber hinaus stellt sie ihren Wissenschaftler/innen eine zukunftsgerichtete Infrastruktur und ein Dienstleistungsangebot zur Verfügung, das den Förderrichtlinien deutscher und internationaler Drittmittelgeber entspricht und ihren Ruf als renommierte Forschungsuniversität weiter fördert. Da dennoch eine Reihe von Wissenschaftler/innen dem neuen Dienstleistungsangebot zunächst zurückhaltend gegenüberstehen wird, sind in der TU Berlin Anreizsysteme zu schaffen, die die Akzeptanz erhöhen.

2. Organisatorisches Modell: „Servicezentrum Forschungsdaten und -publikationen“

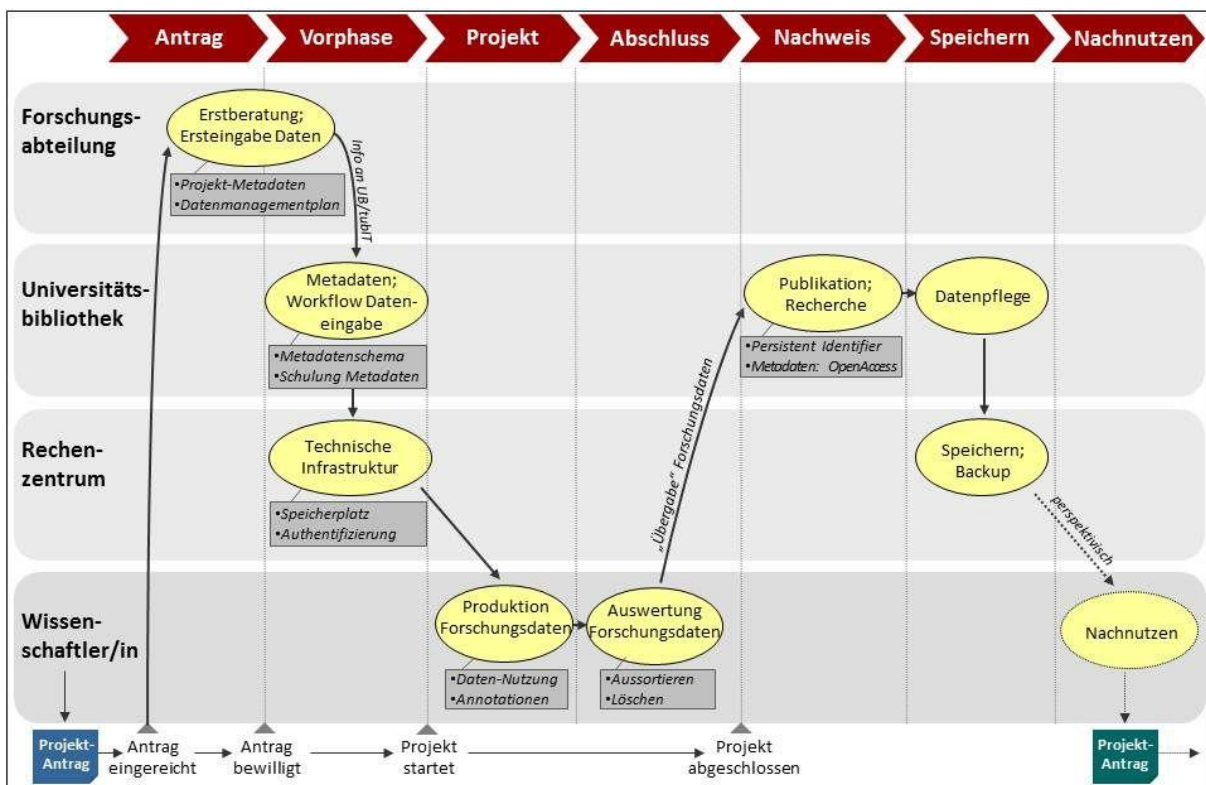
Um die Forschungsdaten-Infrastruktur organisatorisch nachhaltig in der TU Berlin zu verankern, sieht das Konzept vor, dass UB, tubIT und Forschungsabteilung in enger Kooperation ihre Kompetenzen in

einem gemeinsamen „Servicezentrum Forschungsdaten und -publikationen“ bündeln.

Das Servicezentrum ist während der gesamten Dauer eines Forschungsvorhabens – von der Antragstellung über die Bereitstellung der technischen Infrastruktur bis zu Veröffentlichung und Nachweis der Projektergebnisse – Ansprechpartner für die Wissenschaftler/innen und unterstützt sie mit einem umfassenden Beratungs- und Dienstleistungsangebot. Das Angebot enthält unter anderem die Bereitstellung und Pflege der technischen Forschungsdaten-Infrastruktur, standardisierte Metadatenschemata, Unterstützung bei Datenmanagementplänen und beim Erstellen recherchier- und zitierfähiger Metadaten, Datenpflege, Beratung beim Publizieren der Forschungsergebnisse und Open Access sowie Vermittlung grundlegender Informationen in Bezug auf Urheber- und Nutzungsrechtsfragen. Eine Informationsplattform im Internet gibt den Wissenschaftler/innen weitere Hilfsmittel zum Umgang mit Forschungsdaten an die Hand. Transparente Kommunikations- und Informationsstrukturen sorgen dafür, dass Wissenschaftler/innen in jeder Phase eines Forschungsprojektes wissen, wohin sie sich wenden können.

2.1 Verteilte Zuständigkeiten und strukturierte Workflows

Um die Zuverlässigkeit und Dauerhaftigkeit des Dienstleistungsangebotes zu gewährleisten, wird die Arbeitsteilung zwischen den drei beteiligten Einrichtungen verbindlich geregelt. Dazu werden – dem Forschungsdaten-Lebenszyklus⁴ entsprechend – für jede Phase eines Forschungsprojektes die Zuständigkeiten hinsichtlich der Forschungsdaten und strukturierte Workflows zwischen den drei beteiligten Infrastruktureinrichtungen definiert (s. Abb.):



„Forschungsdaten-Lifecycle“ – Zuständigkeiten und Workflows

© Monika Kuberek, Januar 2012



Abb.: Verteilte Zuständigkeiten und strukturierte Workflows im „Servicezentrum Forschungsdaten und -publikationen“ der TU Berlin

- Die Forschungsabteilung, die als erste der drei Einrichtungen von einem Projektvorhaben erfährt, führt die Erstberatung durch. Sie liefert in einem automatisierten Verfahren erste

⁴ Siehe z.B. S. Hodson: *The JISC Managing Research Data Programme: Helping UK Universities meet research data challenges and realise benefits*, 2011, Folie 21: Supporting the Research Data Lifecycle. www.escidoc.org/pdf/escidoc-days-2011/day1-hodson-keynote.pdf

formale Informationen (Titel, Wissenschaftler/in, ...) in den Projekt-Metadatenatz und beginnt den Datenmanagementplan, eine strukturierte Checkliste, in der sie Angaben zum Drittmittelgeber, seinen Anforderungen und weitere allgemeine Informationen zum Projekt dokumentiert. Sobald die Bewilligung vorliegt, werden von UB und tubIT vorbereitende Arbeiten für das Forschungsprojekt aufgenommen. Der gesamte Komplex Metadaten liegt – auch im weiteren Verlauf des Forschungsdatenmanagements – in der Zuständigkeit der UB. In der Phase der Projektvorbereitung geht es dabei u.a. um den Workflow für die Dateneingabe, der – ebenso wie das Metadatenchema – eventuell an die fachspezifischen Projekterfordernisse anzupassen ist, und die Schulung der Wissenschaftler/innen hinsichtlich der Erstellung strukturierter, recherchefähiger Metadaten. tubIT ist für die technische Infrastruktur zuständig, stellt u.a. den benötigten Speicherplatz zur Verfügung und passt die Rechteverwaltung an, entsprechend den Projektanforderungen.

- Auch den Wissenschaftler/innen kommt beim Forschungsdatenmanagement eine wesentliche Rolle zu, da sie die Forschungsdaten produzieren, diese inhaltlich kennen und sie beurteilen können. Inhaltliche Informationen zu den Daten, wie Annotationen, geben die Wissenschaftler/innen daher selbst in den Projekt-Metadatenatz ein. Vor dem Abschluss des Projektes begutachten sie die Daten und werten sie aus – auch im Hinblick auf eine eventuelle spätere Wiederverwendung. Dabei wählen sie die Forschungsdaten aus, die aufbewahrt werden sollen, und löschen die Forschungsdaten, die nicht mehr benötigt werden.
- Nach Abschluss des Projektes werden die Forschungsergebnisse an die UB „übergeben“: Sie unterstützt den öffentlichen Nachweis der in der TU Berlin erbrachten Forschungsleistungen durch elektronisches Publizieren der Forschungsergebnisse, sorgt unter anderem dafür, dass die Datensätze mit konstanten Internet-Adressen (Persistent Identifier) versehen und damit dauerhaft zitierfähig werden, berät die Wissenschaftler/innen hinsichtlich Open Access und vermittelt grundlegende Informationen zu Urheber- und Nutzungsrechtsfragen. Mit dem Wissensportal Primo betreibt die UB ein Rechercheinstrument, mit dem die Forschungsergebnisse im Internet weltweit nachgewiesen, suchbar und – falls lizenzrechtlich erlaubt – zugreifbar sind. Im weiteren Forschungsdaten-Lebenszyklus ist die UB für die Datenpflege zuständig, tubIT für das Speichern und regelmäßige Sichern der Daten.

2.2 Organisationsstruktur

Wesentlich für das Servicezentrum ist eine schlanke, effiziente Organisationsstruktur. Ein kleines Steuerungsgremium, dem jeweils eine Person aus der UB, tubIT und der Forschungsabteilung angehören, lenkt und koordiniert die Arbeiten und ist für die Weiterentwicklung der Forschungsdatenkonzepte in der TU Berlin zuständig. Die Leitung liegt bei der UB, deren Kerngeschäft als Informationsinfrastruktur-Einrichtung der TU Berlin eine hohe Affinität zum Themengebiet Forschungsdaten und deren Management aufweist.

3. Umsetzung des Konzepts: Projekt „Forschungsdaten“

Der Aufbau der Forschungsdaten-Infrastruktur soll im Rahmen eines zweijährigen Projektes erfolgen. Das Projekt soll in der ersten Jahreshälfte 2012 beginnen.

- Die Arbeitsgruppe wird das Konzept im Frühjahr 2012 in einem kleinen Workshop innerhalb der TU Berlin vorstellen und diskutieren. Zu dem Workshop sollen insbesondere – als potenzielle Pilotpartner – interessierte Wissenschaftler/innen eingeladen werden.
- Die Umsetzung des organisatorisch-technischen Konzeptes erfolgt in enger Zusammenarbeit mit einem Pilotpartner. Die AG Forschungsdaten, die das vorliegende Konzept erstellt hat, wird das Projekt begleiten; die Federführung liegt bei der UB.
 - Im Rahmen des Projektes wird das „Servicezentrum Forschungsdaten und -publikationen“ eingerichtet und die Kooperation zwischen UB, tubIT und Forschungsabteilung etabliert. Dazu werden die künftigen Kommunikationsstrukturen aufgebaut, die vorgesehene Aufgabenteilung und die Workflows verifiziert und im Detail ausgearbeitet.
 - Während des Projektes wird mit dem Aufbau der Internet-Informationenplattform begonnen. Dabei kann auf bereits erarbeitete Materialien, insbesondere aus Großbritannien und den

USA, zurückgegriffen werden, die unter der Creative Commons Licence im Internet zur freien Nachnutzung zur Verfügung stehen. Die Materialien gilt es zu sichten und ggf. in Zusammenarbeit mit den Wissenschaftler/innen an den konkreten Bedarf anzupassen.

Mit den Vorarbeiten – Auswahl einer Repository-Software, Aufbau einer Testumgebung, Testinstallation – soll nach dem Workshop möglichst umgehend begonnen werden. Für die Installation und erste Implementierungsarbeiten werden 2-3 Monate veranschlagt. Die Anpassung der Software erfolgt in Zusammenarbeit mit dem Pilotpartner.

4. Perspektive: Gesamtnachweis der in der TU Berlin erbrachten Forschungsleistungen

4.1 Nachweis und Recherche der Forschungsergebnisse

Über Standardprotokolle soll die ausgewählte Repository-Software einen umfassenden Nachweis der Forschungsergebnisse im Internet ermöglichen, der die Forschungsleistung der einzelnen TU-Wissenschaftler/innen und der TU Berlin als Institution nach außen dokumentiert und sichtbar macht. Dazu werden Verzeichnisse automatisiert aufgebaut und laufend ergänzt, so beispielsweise

- der Nachweis der wissenschaftlichen Tätigkeit auf der Homepage einer Wissenschaftlerin oder eines Wissenschaftlers,
- der Nachweis in den Web-Verzeichnissen der jeweiligen wissenschaftlichen Community,
- eine Hochschulbibliographie als öffentliches Nachweisinstrument der an der TU Berlin entstandenen Forschungsergebnisse,
- ein Preprint-Server zur Vorab-Veröffentlichung von Forschungsergebnissen,
- elektronische Publikationen mit referenzierten Forschungsdaten.

Die Recherche nach den Forschungsergebnissen wird über das Wissensportal Primo der UB realisiert. Primo ist ein "One-Stop-Portal" mit einer Suchmaschine, die gleichzeitig sämtliche Ressourcen der UB durchsucht: Online-Katalog, Dokumentenserver, Nationallizenzen, lizenzierte elektronische Zeitschriften und E-Books sowie zahlreiche freie und lizenzierte Datenbanken – und künftig auch das institutionelle Forschungsdaten-Repository.

4.2 Perspektive Gesamtnachweis

In einer Reihe von Fachdisziplinen wurden in den vergangenen Jahren weltweit verteilt disziplinspezifische Infrastrukturen für Forschungsdaten aufgebaut. Auf diese Weise sind einzigartige Wissensspeicher entstanden, an denen auch Wissenschaftler/innen der TU Berlin ihren Anteil haben. Der starken Verankerung der Wissenschaftler/innen in ihren Fachdisziplinen trägt das vorliegende Konzept Rechnung, indem es eine verteilte Datenhaltung für die von TU-Wissenschaftler/innen erbrachten Forschungsleistungen vorsieht: im institutionellen Forschungsdaten-Repository der TU Berlin und in den weltweit verteilten disziplinspezifischen Repositorien.

Um die gesamte in der TU Berlin erbrachte Forschungsleistung zu dokumentieren, ist perspektivisch ein Gesamtnachweis – inkl. Link zu den Datensets und Volltexten – aller von Wissenschaftler/innen der TU Berlin erstellten Forschungsdaten und der zugehörigen Publikationen grundsätzlich möglich. Mittels automatisierter Verfahren über offene Standardprotokolle wie OAI-PMH und SWORD können die Metadaten inkl. Persistent Identifier eingesammelt, im TU-Forschungsdaten-Repository gespeichert und über das Wissensportal Primo suchbar gemacht werden. Mit Hilfe von Primo kann die UB den zentralen Einstieg in die Metadaten-Recherche aller Forschungsergebnisse der TU Berlin organisieren. Über konstante Internet-Adressen (Persistent Identifier) sind darüber hinaus die entfernt liegenden Volltexte bzw. Forschungsdatensets prinzipiell zugreifbar: Entsprechend der Entscheidung der Wissenschaftler/innen können sie gesperrt, nur für bestimmte Gruppen oder frei im Netz zugänglich gemacht werden.