

## **Arbeitskreis Wissenschaftliche Archive in Berlin Protokoll des vierten Treffens**

9. März 2016, 9:00-12:30 Uhr, Technische Universität Berlin

### **Anwesende Archive:**

Insgesamt 17 Teilnehmende aus 8 Einrichtungen:

Archiv der Helmholtz-Gemeinschaft, Deutsches Archäologisches Institut: Archiv Friedrich W. Hinkel, Deutsches Historisches Museum: Hausarchiv, Geheimes Staatsarchiv Preußischer Kulturbesitz, Ibero-Amerikanisches Institut Preußischer Kulturbesitz, Archiv der Max-Planck-Gesellschaft, Universitätsarchiv der Freien Universität Berlin, Universitätsarchiv der Technischen Universität Berlin

### **Getroffene Absprachen:**

- Das nächste Treffen wird voraussichtlich in der 1. oder 2. Woche im Juni 2016 zum Thema „Fotos, Videos, Filmmaterial, AV-Medien“ stattfinden. Ort und genauer Termin werden zeitnah bekanntgegeben.

## **Thema: Forschungsdatenmanagement**

### **Gastbeiträge:**

#### **1. Leonard Palm, FH Potsdam/UA FUB**

Zunächst ging Herr Palm auf die Definition von Forschungsdaten (FD) nach den Leitlinien der DFG ein. FD liegen in einem weiten Spektrum verschiedenster Formate vor, weil praktisch alles FD sein können. Daher kann die Form kein Entscheidungskriterium für die Bewertung dieser Daten sein. Vielmehr (hier verwies er auf einen Vortrag von Jens Ludwig vom nestor – Kompetenznetzwerk zur digitalen Langzeitarchivierung), sei es zweckmäßiger, sich diesen Daten vom Standpunkt der Forschung selbst zuzuwenden. Archive bzw. Repositorien müssen verstehen, was die Zielgruppe der Forschenden macht und welche Anforderungen dadurch entstehen. Forschungsdatenspeicherung geschieht nicht nur, damit der finanzielle Aufwand der Daten(neu)erhebung begrenzt wird, sondern auch, damit eine Replikation von Forschung ermöglicht wird, damit Forschungsergebnisse gesichert werden und damit Transparenz in der Forschung hergestellt bzw. die Reputation der Forschungseinrichtung geschützt wird. Der Wunsch der Wissenschaftler nach seriösem Arbeiten auch durch seriöse Dokumentation wird nach seinem Eindruck dabei jedoch nicht immer realisiert.

Herr Palm weist darauf hin, dass durch die Forschungsdatenspeicherung den Wissenschaftlerinnen und Wissenschaftlern ein zusätzlicher Arbeitsaufwand (Dokumentation) entsteht, der ihnen aber durch den Beruf des Embedded Data Manager (EDM) abgenommen werden kann/muss. Er zitiert dazu einen Vortrag von Fabian Krämer (EDM eines Forschungsprojekts), der die Aufgaben eines EDM in der Unterstützung des Wissenschaftlers/der Wissenschaftlerin sieht. Diese wollen frei sein in der Verwendung ihrer Daten und brauchen Datensicherheit, kennen sich aber nicht mit dem Teilen ihrer Daten aus. Schwierig sei es allerdings, dafür geeignetes

Personal zu finden. Voraussetzung muss ein wissenschaftliches Studium sein, um wissenschaftliche Workflows leiten und anbieten zu können.

Zur Bewertung von FD zitierte Herr Palm Rainer Mauer (Abteilung Datenarchiv für Sozialforschung, gesis Leibnitz-Institut für Sozialwissenschaften), der darauf hingewiesen habe, dass die meisten FD schon bei der Erhebung kassiert werden würden. Je nach Fragestellung der Forschenden werde nur ein kleiner Teil gespeichert und der Rest gelöscht. Andererseits werden die einmal übernommenen Daten später nicht mehr kassiert. Migration und Speicherkapazität sei kein technisches Problem (mehr).

Wer die FD bewertet, muss sinnvollerweise die Methodik des jeweiligen Projektes verstehen. Dazu gehören wissenschaftliche Güte, Nachnutzungsmöglichkeiten, methodische Zugehörigkeit (keine archivarische Terminologie), Einzigartigkeit und Innovativität. Formalisierte Bewertungskriterien stehen im Vordergrund. Es stellen sich dabei Fragen nach der zukünftigen Wichtigkeit der FD oder den Kriterien für die Nachnutzung.

Einer der wichtigsten Unterschiede zwischen der Bewertung im Archiv und der in einem Forschungsdatenrepositorium ist, dass Bewertung im Archiv nach Schließung der Unterlagen stattfindet, was so bei FD-Repositorien nicht der Fall ist.

## **2. Monika Kuberek, Leiterin HA Elektronische Dienste, Universitätsbibliothek der TUB**

Frau Kuberek stellte das Repositorium für Forschungsdaten und Publikationen der TU Berlin „DepositOnce“ vor. Dabei handelt es sich um ein institutionelles Repositorium, in dem Forschungsergebnisse aus vielen verschiedenen Fachdisziplinen abgelegt und gespeichert werden können. Somit wird eine Forschungsdatenstruktur aufgebaut. DepositOnce ist eine mit bestimmten Zugangsberechtigungen im Internet frei zugängliche Serviceplattform u. a. zur Nachnutzung bereits vorhandener Daten, bei der alle Metadaten frei recherchierbar sind. Das vorhandene Know-How der Zentralen Einrichtungen der TU Berlin wurde gebündelt, in dem die Leitung dieses Repositoriums bei der Universitätsbibliothek in Zusammenarbeit mit dem IT-Service-Center tubIT und der Forschungsabteilung der TU Berlin liegt. Zugänglich gemacht werden Forschungsdaten, wie z. B. Auswertungen und/oder Daten, die zur Verifizierung und Reproduktion eines Forschungsergebnisses notwendig sind, aber auch wissenschaftliche Publikationen, Dissertationen und Habilitationen sowie unter bestimmten Voraussetzungen auch Abschlussarbeiten von Studierenden der TU. Die Auswahl und Erschließung der Daten erfolgt durch die Wissenschaftler und Wissenschaftlerinnen selbst. Die UB prüft lediglich formale Kriterien. Sobald die UB die Daten frei gibt sind sie publiziert. Die Wissenschaftler und Wissenschaftlerinnen haben die Möglichkeit mit einer Embargofrist zunächst nur die Metadaten, nicht jedoch die Forschungsdaten öffentlich zugänglich zu machen. Nach Ablauf des Embargos werden die Daten automatisch freigeschaltet. Einmal veröffentlichte Daten werden nicht mehr verändert und nicht vom Server gelöscht. Für die formale und inhaltliche Erschließung werden national und international definierte Standards und Schnittstellen benutzt, wie die der Dublin Core Metadata Initiative, die Richtlinien der Open Archives Initiative (OAI) oder das xMetaDissPlus-Format der Deutschen Nationalbibliothek. Bisher gibt es keine Probleme mit Dateiformaten. Die Metadaten werden über ein Upload-Formular

erfasst. Die Daten werden von der TU Berlin langfristig gespeichert, entsprechend den Empfehlungen der DFG für eine Dauer von mindestens 10 Jahren.

Die Folien zum Vortrag finden Sie unter: [http://www.user.tu-berlin.de/kuberek/kuberek\\_2016-03-09.pdf](http://www.user.tu-berlin.de/kuberek/kuberek_2016-03-09.pdf).

In der **anschließenden Diskussion** ging es darum, dass Daten, für die bereits entsprechende Repositorien existieren, nicht in dieses Repository aufgenommen werden. Auf Nachfrage bestätigt Frau Kuberek, dass die Zusammenarbeit mit den Wissenschaftlern und Wissenschaftlerinnen der TU Berlin bisher positiv verläuft. Für die Eingabe der Metadaten gibt es einen einfachen Workflow, bei dem nur wenige Daten verlangt werden und der Forschende durch die Eingabemaske geleitet wird. Über den Service der Teaching Library der UB der TU Berlin können Nutzer und Nutzerinnen dieses Repository kennenlernen. Bisher ist die geführte Nutzungsstatistik zu DepositOnce anonym, so dass es keine Aussagen dazu gibt, ob Wissenschaftler und Wissenschaftlerinnen die Daten für ihre Arbeit nutzen oder sich nur informieren. Aus dem Teilnehmerkreis wird empfohlen, Werbung damit zu machen, dass die Veröffentlichung von Forschungsdaten auch den Vorteil der Zitierbarkeit mit sich bringt.

### **3. Maurice Heinrich, Leitung der wissenschaftlichen Fachsäule der IT im IT-Referat der Zentrale des DAI**

Das sich z. Zt. in der zweiten Förderphase der DFG befindliche IANUS-Projekt des DAI zur Sammlung, Beschreibung, Katalogisierung, Migration, Archivierung und Bereitstellung von Daten aus der Archäologie sowie den Altertumswissenschaften und benachbarten Forschungsdisziplinen, das Herr Heinrich vom IT-Referat des DAI vorstellte, wird im April 2016 erstmalig Forschungsdaten nachweisen, die derzeit noch aufbereitet werden.

In der langen Tradition dieser Wissenschaftsdisziplin sind alte Daten oft nicht mehr les- oder interpretierbar. Oftmals werden mit Erscheinen der Publikation zu einem Forschungsprojekt die Forschungsrohdaten Dritten zur Nachnutzung weitestgehend vorenthalten. Sie werden selten mit anderen Forschern geteilt und kaum veröffentlicht. Eine große Herausforderung bei der Aufbereitung stellen die vielen verschiedenen Dateiformate dar. Im IANUS-Projekt sollen die Daten so offen wie möglich und so geschlossen wie nötig präsentiert werden. Zur weiteren Nutzung der Daten ist es notwendig auch Metadaten anzugeben und zu dokumentieren. Hier setzt das Projekt u. a. auf den Einsatz von Datenkuratoren zur Datenpflege („Koalition der Willigen“). Beratungsbedarf besteht im Bereich der Metadaten, der Dateiformate und Methoden, um Wissenschaftler für den nachhaltigen Umgang mit digitalen Daten zu sensibilisieren. Ziel ist es Vertrauen zu schaffen, um Nachhaltigkeit zu fördern und damit alle Projektergebnisse in transparenter Form zu veröffentlichen.

Die Folien zum Vortrag finden Sie unter: <http://www.slideshare.net/IANUS-FDZ/forschungsdaten-nach-der-publikation-ist-vor-der-archivierung>

In **Fragen zum Vortrag** ging es u. a. um die Exklusivität der Daten in einem Repository, die nicht vollständig gewährleistet werden kann. Ebenso lässt sich die Klärung der Rechtsfragen und Haftbarkeit nur mit einigen Unschärfen eingrenzen und muss von Fall zu Fall entschieden werden. IANUS ist i. d. R. nicht haftbar. Die Frage nach der inhaltlichen Prüfung der Daten ergibt, dass man zum einen Experten

für fachspezifische Daten benötigt und zum anderen diese Prüfung sehr personalintensiv und somit nur teilweise möglich ist. Wenn die groben Kriterien für IANUS erfüllt sind, werden die Daten angenommen oder ansonsten an andere Datenbanken weiterverwiesen. Strukturen sollen soweit wie möglich abgebildet werden, teilweise sind jedoch gewisse Standardisierungen notwendig, die möglichst in Absprache mit der Forschungseinrichtung erfolgen sollten. Redundante Informationen kommen vor und können nur in Zusammenarbeit mit der Forschungseinrichtung identifiziert und anschließend entfernt werden. Aus der Sicht von IANUS gehören E-Mails nicht in den Bereich der Forschungsdaten, sondern eher in den Bereich der Projektdokumentation.

In der **abschließenden Diskussionsrunde** wurde mehrfach die Befürchtung geäußert, dass die Archive den Anschluss verpassen u. a. aufgrund der Personalsituation in den Einrichtungen und wie sich die Archive dazu positionieren können/werden. Für die Archive sind zudem nicht nur die „erfolgreichen“ Forschungsprojekte von Bedeutung, sondern die Archive als Datenmanager wollen auch die „gescheiterten“ Projekte archivieren. Deshalb ist eine Verbindung zu den Repositorien wichtig ebenso wie die Frage, was mit den Daten nach 10 Jahren geschehen wird, ob hier vom Konzept her auch auf langfristige Nachnutzung gesetzt wird. Forschungsdaten, die auf Papier vorliegen, können zumindest als gesichert betrachtet werden. Allgemeiner Konsens ist, dass sich die einzelnen Archive hier positionieren und im Verbund stärker zusammenschließen sollten. Die Frage bleibt, ob es dafür Kapazitäten gibt und ob die Archive eine Chance haben, Herr des Verfahrens zu bleiben, wenn sie frühzeitig eingebunden werden. Da die einzelnen Archive dies nicht alle erreichen können, sollte versucht werden, im Verbund Schnittstellen mit anderen zu finden, um so gemeinsame Stärke zu erlangen. Dazu ist es wichtig, das Leistungsspektrum der Archive deutlich hervorzuheben. Eventuell könnte es auch von Bedeutung sein, sich über Bewertungsmodelle zu verständigen. Die Archive interessieren sich auch für Daten, die im Vorfeld entstehen und nicht zwangsläufig in die Forschungsprojekte mit einfließen.

Eine weitere Problematik stellen die verschiedenen Formen der hybriden Arbeitswelt dar (E-Mails, Fachverfahren, Papierdokumente), wofür die Kernkompetenzen sprich die klassischen Bewertungsmodelle von Archiven anwendbar sind. Die Schriftgutverwaltung in der elektronischen Welt wird sich anders darstellen als vorher, u. a. auch, weil es parallele elektronische Plattformen gibt. Im wissenschaftlichen Bereich gibt es z. B. in der FU Berlin den Ansatz, dass jedes Fachverfahren, von denen es eine große Vielzahl gibt, dokumentiert wird. Auch hier muss Überzeugungsarbeit geleistet werden, damit die Fachverfahren benutzt werden (Stichwort: Koalition der Willigen). Die Archive befinden sich allerdings häufig in der Defensivposition. Um besser zusammenzuarbeiten sind Mittler und Koordinatoren sinnvoll. Ein Teil der Forschungsgelder könnte auch von den Archiven genutzt werden: Wie kann man hier Lobbyarbeit betreiben? Oft gibt es Parameter für ein gemeinsames Konzept, das man an die Institution anpassen kann, um zu zeigen, hier ist unsere Kompetenz/unser Produkt, das können wir bieten.

Um konkret zu werden, gibt es die übereinstimmende Idee eines Arbeitstreffens zu diesem Thema.