

**Ausschuss für Wissenschaftliche Bibliotheken und
Informationssysteme**

Unterausschuss für Informationsmanagement

**Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung
digitaler Forschungsprimärdaten**

Januar 2009

Auf Einladung der Gruppe LIS fand am 17. Januar 2008 ein Rundgespräch zum Thema „Forschungsprimärdaten“ in der Geschäftsstelle der DFG statt (Protokoll s.: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/forschungsprimaerdaten_0108.pdf). Ziel des Rundgespräches war es, das Problemfeld „Speicherung und Bereitstellung von Forschungsprimärdaten“ zu diskutieren, um darauf aufbauend Eckpunkte für zukünftige Fördermaßnahmen der DFG in der Informationsinfrastruktur zu spezifizieren. Diese Eckpunkte sollten einen disziplinübergreifenden Rahmen schaffen, innerhalb dessen fachspezifisch weitere Regelungen und Maßnahmen entwickelt werden sollten.

Aufbauend auf die Ergebnisse des Rundgespräches hat der Unterausschuss für Informationsmanagement (UA-INF) „Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten“ erarbeitet und diese auf der Sitzung am 21. Mai 2008 verabschiedet. Der Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme (AWBI) hat die Empfehlungen des Unterausschusses in der Sitzung am 25. und 26. September 2008 bestätigt.

**Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler
Forschungsprimärdaten**
Stand 26.6. 2008

Forschungsprimärdaten bilden einen wertvollen Fundus an Informationen, die mit hohem finanziellem Aufwand erhoben werden. Je nach Fachgebiet und Methode sind sie replizierbar oder basieren auf nicht wiederholbaren Beobachtungen oder Messungen. In jedem Fall sollten die erhobenen Daten nach Abschluss der Forschungen öffentlich zugänglich und frei verfügbar sein. Dieses ist die wesentliche Voraussetzung dafür, dass Daten im Rahmen neuer Fragestellungen wieder genutzt werden können sowie dafür, dass im Falle von Zweifeln an der Publikation die Daten für die Überprüfung der publizierten Ergebnisse herangezogen werden können.

1997 veröffentlichte die DFG „Vorschläge zur Sicherung guter wissenschaftlicher Praxis“ mit 16 Empfehlungen. Die Empfehlung 7 lautet „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“ Die Forderung, Daten verfügbar zu halten, steht für die DFG somit schon seit über 10 Jahren im Raum.

Für die Speicherung und Bereitstellung von Forschungsprimärdaten müssen technische und organisatorische Voraussetzungen gegeben sein, die aus den einzelnen Fachdisziplinen heraus zu entwickeln sind. In einigen Disziplinen ist dies bereits in beispielhafter Weise gelöst, andere stehen noch am Anfang dieser Aufgabe. Zu den wesentlichen Zielen gehört, dass sich diese Strukturen an internationalen Maßstäben und Standards orientieren und in bereits vorhandene überregionale, internationale Strukturen einbetten. Für ihr nachhaltiges Wirken ist von Anfang an Sorge zu tragen.

Für die Schaffung der erforderlichen organisatorischen Voraussetzungen und Strukturen sollen fachspezifische Arbeitsgruppen eingesetzt werden, die für ihre Community Vorschläge zur Bestimmung der an der Speicherung und Bereitstellung beteiligten Institutionen sowie der Verfahren und Strukturen erarbeiten. Die Orientierung an bereits praktizierten Modellen mag hier hilfreich sein. Nach Möglichkeit soll das Zitieren der gespeicherten Datensätze und ihrer Fundstellen angestrebt werden. Es wäre dies eine Option, die dem Wissenschaftler Anreiz bietet, die Daten in eigener Initiative zu publizieren. Auch hierfür gibt es bereits erfolgreiche Beispiele. Ziel soll die freiwillige Übergabe der Daten durch den Wissenschaftler sein. Weiteres wesentliches Kriterium, das zu berücksichtigen wäre, ist die Qualitätssicherung der Daten.

Die nun vorliegenden Empfehlungen richten sich sowohl an die Wissenschaftler als auch die Informationsspezialisten. Es soll damit das Bewusstsein für die Dringlichkeit dieses Problems geweckt werden. Darüber hinaus werden die Ziele benannt, die bei der Umsetzung berücksichtigt werden sollten und deren Einhaltung von grundlegender Bedeutung ist. Schließlich sollen die Wissenschaftler dazu angeregt werden, dieser Aufgabe auf dem Wege der Selbstverpflichtung nachzukommen und für eine sachgemäße Umsetzung im Rahmen der gegebenen Strukturen in eigener Initiative Sorge zu tragen.

Auf der Grundlage dieser Empfehlungen wird die DFG im Rahmen ihrer Fördermöglichkeiten Mittel bereitstellen, um den Ausbau der Speicher- und Nachweisstrukturen zu unterstützen. Insbesondere regt sie zu Explorations-Projekten an, um für die einzelnen Disziplinen weitere Erfahrungen zu sammeln. Die Regelung der Aufbewahrung und Bereitstellung der Daten sollte von Anfang an im Arbeitsplan des Wissenschaftlers vorgesehen und im Rahmen der Förderverfahren bereits in den Anträgen dargelegt werden.

Die 7 Empfehlungen im einzelnen:

1. *Forschungsprimärdaten sind Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind. Sie stellen die Grundlagen für die wissenschaftlichen Publikationen dar.*

In Abhängigkeit von der jeweiligen Fachzugehörigkeit sind die Forschungsprimärdaten unterschiedlich zu definieren. Die Wissenschaftler sollen in ihren Fachcommunities selber entscheiden, ob bereits Rohdaten hierzu zählen oder ab welchem Grad der Aggregation die Daten langfristig aufzubewahren sind. Des weiteren soll die Granularität in groben Umrissen

vereinbart sein: wie viele Daten ergeben einen Datensatz, der mit einer stabilen Adresse (persistent identifier) ausgestattet wird?

2. *Es ist ein fachspezifisches Organisationskonzept festzulegen, das die nachhaltige Speicherung der Daten regelt.*

Die Wissenschaftler bestimmen jeweils innerhalb ihrer Disziplinen eine Institution oder organisatorische Struktur, die sowohl die rasche, komfortable Bereitstellung der Daten im laufenden Forschungsprozess als auch die langfristige Verfügbarkeit sicherstellt. Darüber hinaus ist zu vereinbaren, ob es eine maximale Aufbewahrungsfrist geben soll, nach deren Ablauf die Daten dem Forscher zurückgegeben oder gelöscht werden.

3. *Die Speicherung der Forschungsdaten erfolgt im Rahmen definierter Standards*

Wesentlich ist die Einhaltung von Standards. Entsprechende Regelungen sind von den Wissenschaftlern in Kooperation mit Informationsspezialisten zu treffen. Hier sind vor allem auch internationale Regelungen zu berücksichtigen, damit die Kompatibilität auch international gewährleistet ist.

4. *Die Daten sind persönlich gekennzeichnet und werden unter dem Namen des Forschers abgelegt.*

Die Speicherung der Daten muss die Rechteverwaltung mit umfassen. Die Regelungen sollen den Prinzipien des Open Access entsprechen.

5. *Jeder Wissenschaftler stellt seine Forschungsprimärdaten nach Möglichkeit überregional und frei zur Verfügung.*

In Abhängigkeit von der jeweiligen Fachkultur soll unter den Wissenschaftlern Konsens darüber hergestellt werden, dass die Daten unmittelbar nach Abschluss der Forschungen oder nach wenigen Monaten der Öffentlichkeit frei zur Verfügung gestellt werden. Sofern die Daten in Projekten erarbeitet wurden, die aus öffentlich-rechtlichen Mitteln finanziert wurden, stehen sie im Grundsatz der Öffentlichkeit frei zur Verfügung. Für Forschungsdaten, die der Patentierung oder anderer unmittelbarer wirtschaftlicher Nutzung unterliegen, sind gesonderte Regelungen zu treffen. Dieses soll im Konsens mit den betroffenen Wissenschaftlern geschehen.

6. *Die Daten werden durch Metadaten beschrieben.*

Mit den Metadaten (mindestens nach Dublin Core) werden zum einen die bibliographischen Fakten festgehalten. Es sind dies der Name des Forschers, der die Daten erhoben hat, die Benennung des Datensatzes, Ort und Jahr der Veröffentlichung sowie technische Daten (Format etc). In den inhaltsbezogenen Metadaten werden die Primärdaten umfassend beschrieben. Hier finden sich die Angaben zu den Rahmenbedingungen, unter denen sie erhoben bzw. gemessen wurden. Hier beschreibt der Autor auch die Fragestellung, unter der die Daten entstanden. Es sollen hier alle Informationen vorliegen, die für eine wiederholte Nutzung der Daten in anderen Fragestellungen erforderlich sind. Die Kriterien des Information Life Cycle Management sollen dabei berücksichtigt werden.

7. *Die Daten sind qualitätsgeprüft.*

Es sind in den Fachcommunities Kriterien und Verfahren zu entwickeln, die die Qualitätssicherung der Daten gewährleisten. In dem Zusammenhang ist auch festzulegen, ab welchem Grad der wissenschaftlichen Qualifikation die Forschungsdaten aufbewahrt und öffentlich bereitgestellt werden sollen.