



Heike Neuroth, Stefan Strathmann, Achim Oßwald,
Regine Scheffel, Jens Klump, Jens Ludwig [Hrsg.]

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

Vorwort



Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft



Universitätsverlag Göttingen

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme

Heike Neuroth, Stefan Strathmann, Achim Oßwald,
Regine Scheffel, Jens Klump, Jens Ludwig [Hrsg.]

Langzeitarchivierung von Forschungsdaten

Eine Bestandsaufnahme



Förderkennzeichen: 01 DL 001 B

Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme
hg. v. Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump,
Jens Ludwig
im Rahmen des Kooperationsverbundes nestor – Kompetenznetzwerk Langzeitarchivierung
und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Lektorat) und Sonja Neweling (Redaktion) sowie Martina Kerzel und Lajos Herpay (Gestaltung und Montage) für ihre unverzichtbare Unterstützung bei der Fertigstellung des Handbuchs.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Die Inhalte des Buches stehen auch als Onlineversion
(<http://nestor.sub.uni-goettingen.de/bestandsaufnahme/>)
sowie über den Göttinger Universitätskatalog
(<http://www.sub.uni-goettingen.de>) zur Verfügung.
Die URN lautet: <http://nbn-resolving.de/urn:nbn:de:0008-2012031401>.

Die digitale Version dieses Werkes steht unter einer Creative Commons Namensnennung-
Nicht-kommerziell-Weitergabe unter gleichen Bedingungen 3.0 Unported Lizenz.



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2012
<http://www.vwh-verlag.de>
In Kooperation mit dem Universitätsverlag Göttingen.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und
als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide
Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-86488-008-7

Vorwort

Matthias Kleiner

Gemeinsam mit ihren Partnern aus der Allianz der Wissenschaftsorganisationen setzt sich die Deutsche Forschungsgemeinschaft (DFG) seit Jahren engagiert dafür ein, dass Wissenschaftlerinnen und Wissenschaftler direkt von ihrem Arbeitsplatz aus einen möglichst umfassenden und komfortablen Zugriff auf elektronische Ressourcen haben. Eine Forschung, die zunehmend durch die kooperative Tätigkeit weltweit vernetzter Communities und durch den Einsatz Computerbasierter Verfahren bestimmt ist, erfordert nun einmal die kontinuierliche und vor allem langfristige Verfügbarkeit von Publikationen und Forschungsdaten über das Internet. Nicht nur die Notwendigkeit, Forschungsergebnisse durch den Rückgriff auf die diesen Ergebnissen zugrunde liegenden Daten verifizieren zu können, sondern auch die produktive Nachnutzung von Forschungsdaten in anderen Kontexten setzt voraus, dass digital kodierte Information über Jahrzehnte hinweg authentisch verfügbar bleibt.

Die Langzeitverfügbarkeit von Forschungsdaten zu gewährleisten, bedeutet in erster Linie, sich mit der Organisation und Standardisierung von Arbeitsschritten und Prozessen zu befassen, die zu digital kodierter Information führen. Um Integrität und Authentizität dieser Informationen zu bewahren, müssen Standards für die Erzeugung, Erfassung und Dokumentation von Forschungsdaten definiert und eingehalten, technische, rechtliche und inhaltliche Metadaten vergeben, Workflows definiert und Schnittstellen für das Einspeisen ebenso wie für das Auslesen von Forschungsdaten aus einem Archiv geschaffen und optimiert werden. Im Prozess des wissenschaftlichen Arbeitens muss die Langzeitarchivierung daher stets schon bei der Erzeugung der Inhalte ansetzen. Ein essentieller Schritt, die Prozesskette im Hinblick auf die langfristige Verfügbarkeit von Forschungsdaten zu optimieren, ist daher darin zu sehen, dass zu-

nehmend mehr Förderorganisationen – wie es auch die DFG tut – verlangen, mit jedem Antrag ausführlich zu beschreiben, in welcher Art und Weise Projektnehmer dafür Sorge tragen, die in einem Projekt erhobenen Forschungsdaten langfristig verfügbar und potenziell nachnutzbar zu halten.

Der eigentliche Sinn jedweder langfristigen Archivierung liegt darin, dass einmal erfasste Forschungsdaten erneut genutzt werden können. Die Effekte solcher Nachnutzung werden umso größer sein, wenn Forschungsdaten in dezidierte, in einer Community weithin bekannte und akzeptierte Repositorien eingepflegt und von dort abgerufen werden können. In unterschiedlichen Wissenschaftsbereichen entwickeln sich dergleichen Archive in verschiedenen Geschwindigkeiten: Während manche Fachgebiete noch vor der Aufgabe stehen, Standards für die Erfassung und Beschreibung von Forschungsdaten zu entwickeln und deren allgemeine Verbindlichkeit abzusichern, professionalisieren andere Disziplinen den Umgang mit bereits etablierten Datenarchiven. All diesen Anstrengungen ist gemeinsam, dass Konzepte für die Organisation des Workflows, für die Gestaltung von Schnittstellen, für die Beschreibung von Metadaten am effizientesten umgesetzt werden können, wenn sie im engen Dialog von Wissenschaftlerinnen und Wissenschaftlern mit Vertreterinnen und Vertretern von Einrichtungen der Informationsinfrastruktur entwickelt werden.

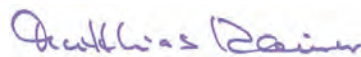
Im Wissen darum, dass Wissenschaftlerinnen und Wissenschaftler zunehmend auf den kontinuierlichen und langfristigen Zugang zu digitalen Ressourcen angewiesen sind, hat die DFG in Forschungsprojekten ebenso wie in ihren Förderprogrammen zum Aufbau einer modernen Informationsinfrastruktur sowohl den Bereich der Langzeitarchivierung als auch den der Forschungsdaten adressiert.

Dabei wird die oben angemahnte Zusammenarbeit von Wissenschaft und Informationsinfrastruktur in vielen der im Jahr 2011 bewilligten Projekte zum Auf und Ausbau von Forschungsdatenrepositorien realisiert. Ebenso zeigt sich in dezidiert der Langzeitarchivierung gewidmeten DFG geförderten Projekten, dass für die Langzeitarchivierung nicht allein technische Fragestellungen gelöst werden müssen, sondern vor allem organisatorische Herausforderungen zu bewältigen sind, die nach kooperativen Ansätzen verlangen: So projiziert das Vorhaben „LOCKSS und KOPAL

Infrastruktur und Interoperabilität“ eine technische Entwicklung, indem der Vorteil einer kostengünstigen Speicherung elektronischer Ressourcen mit effizienten Werkzeugen für die Migration kombiniert wird, während das ebenfalls DFG geförderte Projekt „Digital Preservation for Libraries“ eine Organisationsform für die Langzeitarchivierung entwickelt, die auf die kooperative Nachnutzung des KOPAL Systems abzielt und dabei auch Forschungsdaten im Blick behält.

Langzeitarchivierung, das belegt die aufmerksame Lektüre dieses Bandes, ist und bleibt ein Zukunftsthema, das sorgfältig gestaltet werden muss. Wie andere Organisationen wird sich auch die DFG weiterhin verantwortlich dieses Bereichs annehmen, wenn Anfang 2012 in einer Ausschreibung zur Weiterentwicklung überregionaler Informationsservices der Bereich der Langzeitarchivierung eigens ausgelobt werden wird. Allerdings kann die DFG nur Impulse dafür setzen, dass sich über die Projektförderung Infrastrukturen so entwickeln, wie sie von der Wissenschaft benötigt werden. Diese Strukturen müssen dann dauerhaft abgesichert werden, was – und auch das zeigt der vorliegende Band – nicht nur als rein nationale Aufgabe begriffen werden kann. Doch ist darauf zu dringen, dass im nationalen Rahmen die Träger von Universitäten und Forschungsinstituten, von Bibliotheken, Daten und Rechenzentren ihre Verantwortung dafür wahrnehmen, die Investitionen in gute Forschung auch durch eine auskömmliche und verlässliche Finanzierung der Langfristverfügbarkeit der Forschungsergebnisse abzusichern. Denn gerade die Langzeitverfügbarkeit ist es, die das Potenzial der Wissenschaft maßgeblich erhöht, zumal wissenschaftliche Ergebnisse heute häufig nur noch in digitaler Form dargestellt werden.

In diesen skizzierten Sinnen wünsche ich dieser Veröffentlichung lange Gültigkeit, weitreichende Verfügbarkeit und viele Leserinnen und Leser, die dazu greifen und die darin zusammengetragenen Erkenntnisse und Betrachtungen diskutieren und natürlich nutzen und weiterentwickeln.



– Präsident der Deutschen Forschungsgemeinschaft (DFG) –