

RatSWD Working Paper Series

www.ratswd.de

RatSWD ■

Rat für Sozial- und
WirtschaftsDaten

213

Herausforderungen der
Archivierung sozial-, verhaltens-
und wirtschaftswissenschaftlicher
Datenbestände

Reinhard Altenhöner
Gert G. Wagner

Januar 2013

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Herausforderungen der Archivierung sozial-, verhaltens- und wirtschaftswissenschaftlicher Datenbestände^{*}

Reinhard Altenhöner (DNB) und Gert G. Wagner (RatSWD)

Januar 2013

Der Umfang von im Prinzip mehrfach analysierbaren bzw. re-analysierbaren Forschungsdaten nimmt generell beständig zu und somit auch die Herausforderung, die Daten langfristig für die Wissenschaft verfügbar zu machen und zu archivieren. Der Zugang zu Forschungsdaten für Sekundäranalysen entspricht nicht nur den Regeln guter wissenschaftlicher Praxis, sondern stellt eine Voraussetzung für gleichermaßen innovative wie kostengünstige Wissenschaft dar. Die zunehmende Relevanz eines persistenten Zugangs zu Forschungsdaten wird auch dadurch sichtbar, dass – erleichtert durch die technischen Möglichkeiten der digitalen Publikation – Umfang und Dichte der direkten Verknüpfungen von wissenschaftlichen Publikationen zu den jeweiligen Forschungsdatenbeständen wachsen und damit der nahtlose Rückgriff auf die Quellen, auf denen eine wissenschaftliche Publikation aufsetzt, an Bedeutung gewinnt.

Damit Forschungsdaten für Sekundäranalysen dauerhaft zur Verfügung stehen, müssen diese archiviert und für potentielle Nutzer auffindbar sein. Dies

* Bei dem Beitrag handelt es sich um die Einleitung des Buches „Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen.“ Hrsg. v. Reinhard Altenhöner und Claudia Oellers. Scivero Verlag 2012. (ISBN 978-3-944417-00-4). Erhältlich über die Geschäftsstelle des RatSWD.

umfasst neben der bloßen Substanzerhaltung (Technik und Software veralten bekanntlich schnell) auch die Frage, wie die oftmals komplexen Datensätze von späteren Nutzern richtig interpretiert werden können.

Entsprechende Aufgabenstellungen im Bereich der Zusammenführung und Erschließung und in der Entwicklung geeigneter technischer Werkzeuge sind bislang breit verteilt und unterschiedlich weitreichend geregelt: Während beispielsweise für die unmittelbare Verbreitung in der wissenschaftlichen Community publizierte Materialien wie Zeitschriftenaufsätze und Bücher – seien sie digital oder als Print produziert – in den Sammelauftrag der Deutschen Nationalbibliothek fallen, ist die Situation bei Forschungsdaten sehr viel offener: Hier gibt es neben den Selbst- oder Förderverpflichtungen, die die Bewahrung und die öffentliche Zugreifbarkeit für einen begrenzten Zeitraum von 10 Jahren durch die betreuende Einrichtung statuieren, überregionale, meist fachlich definierte Einrichtungen, die Archivierungsaufgaben übernehmen und in unterschiedlicher Tiefe und Kompetenz die Verantwortung für einzelne Ausschnitte aus der Gesamtmenge entstehender Daten übernehmen.

Dieses Nebeneinander von Sammlungsprinzipien wie der Gattung der Publikation (eine Veröffentlichung in einer Zeitschrift oder ein Set von Forschungsdaten) oder die fachliche Abgrenzung (ein fachliches Repository gegenüber einer allgemein sammelnden Einrichtung wie zum Beispiel auch einem Hochschul-Repository) erschwert nicht nur aus der Nutzungsperspektive den Zugriff auf eng miteinander zusammenhängendes Material (wo recherchiere ich?), sondern stellt auch für die beteiligten Einrichtungen in zunehmendem Maße ein Problem dar, denn Such- und Verknüpfungsstrukturen müssen übergreifend funktionieren und vor allem dauerhaft verfügbar sein (also nicht nur die digitalen Objekte selbst).

So wie die materialspezifischen Aufteilungen immer weniger greifen (was schon daran deutlich wird, dass auch Verlage dazu übergehen, eigene Angebote zur Speicherung und Bereitstellung zumindest der aggregierten, für die Publikation wesentlichen Mess- und Datenreihen anzubieten), ist auch die fachliche Auftrennung von Beständen immer schwieriger zu vermitteln und durchzuhalten. Umso wichtiger wird es, gemeinsame Vorgehensweisen für die Erschließung und für die Sicherung der Langzeitverfügbarkeit des Materials zu verfolgen. Nur unter diesen Bedingungen kann ein/e forschende/r Wissenschaftler/in alle miteinander zusammenhängenden Materialien herbeiziehen, Literatur rezipieren, Erhebungsreihen sichten und ggf. mit

geeigneten Werkzeugen validieren, dann aber auch Vergleichsreihen bilden und Schritt um Schritt in eine eigene Forschungsarbeit überwechseln. Eine wesentliche Voraussetzung dafür ist, dass die Forschungsdaten regelmäßig auf ihre Benutzbarkeit hin überprüft werden, ggf. verlustfrei und dokumentiert in eine mit einem moderneren Werkzeug nutzbare Form migriert werden oder aber geeignete Softwarewerkzeuge bereitstehen, die in einer aktuellen technischen Umgebung ablauffähig sind, aber den Zugang und die Arbeit mit „alten“ Daten (hier sind 10 Jahre oft schon eine relevante Hürde) erlauben.

Was für die Kooperation und das Ziel einheitlicher Verzeichnungs- und Verweisungsstrukturen klar formuliert werden kann, gilt also in ähnlicher Weise für die Frage, wie denn die einzelnen archivierten Materialgruppen zugänglich gehalten werden: Hier sind sparten- und institutionenübergreifende Kooperationsstrukturen gefordert, die es Mitwirkenden auch auf pragmatischer Ebene erleichtern, Know-how und Expertise wechselseitig zur Verfügung zu halten.

Es gibt mittlerweile – variierend von Disziplin zu Disziplin – eine Vielzahl an Ansätzen und Initiativen, die den Prozess der Archivierung unterstützen und Standards bereitstellen. Was allerdings fehlt, ist eine einheitliche Strategie, in welcher Form Forschungsdaten archiviert, zugänglich, auffindbar, aber auch zitiert werden. Ohne die Möglichkeit der Zitation und der damit verbundenen wissenschaftlichen Reputation fehlen wichtige Anreizstrukturen für die Wissenschaftler, diese Daten für die Nachnutzung bereit zu stellen. Dies wird besonders in den Sozial- und Wirtschaftswissenschaften deutlich. In diesen Disziplinen gibt es seit langem Datenarchive (länger als zum Beispiel in den Verhaltenswissenschaften), aber viele Datensätze sind nach wie vor nicht zugänglich oder nur schwer zu finden. Hier bedarf es gemeinsamer Festlegungen auf einheitliche Beschreibungs- und Datenaustauschstrukturen, also der Formulierung von allen Beteiligten aktiv getragener Vereinbarungen, die sich zudem an international gängigen Standards orientieren sollten.

Ein großes Problem ist die angemessene Würdigung der Datenproduktion und -archivierung in den wissenschaftlichen Fachdisziplinen. Auch dieses Problem ist in den Sozial- und Wirtschaftswissenschaften besonders deutlich, da es dort – trotz längerer Tradition – für Nachwuchswissenschaftler riskant ist, sich der Datenproduktion und -archivierung zu widmen, denn mit dieser zeitaufwendigen Arbeit sind keine wissenschaftlichen Lorbeeren zu ernten. Deswegen besteht im Zeitalter der

Explosion von Forschungsdaten eine der großen Aufgaben darin, innerhalb der Fachdisziplinen Datenproduktion und -archivierung so zu würdigen, dass Spitzen-nachwuchskräfte bereit sind, sich in diesem Feld zu engagieren. Das Zitierbar- und Auffindbarmachen von Forschungsdaten und weltweit eindeutige „Researcher Identifier“, also eine eindeutige Identifikationsnummer für einzelne Forscher und ihre Ergebnisse (Daten und Schriften), könnten hier entscheidend weiterhelfen. Denn erst derartige Referenzsysteme würden es im Zeitalter der Messung von Forschungsleistung ermöglichen, Datenproduktion und -archivierung zitierbar und damit messbar zu machen.

Die Bedeutung der Zitierbarkeit von Forschungsdaten und ihrer Produzenten verweist unmittelbar auf die Bedeutung von Bibliotheken für die Welt der Forschungsdaten. Zugleich sind Bibliotheken Spezialisten für Langzeitarchivierung. Insofern ist es geradezu überfällig, dass endlich Datenproduzenten (aus dem Bereich der Sozial-, Verhaltens- und Wirtschaftswissenschaften, einschließlich der amtlichen Statistik), Datenarchive und Bibliotheken zu einem intensiven Gedankenaustausch zusammengebracht werden. Eine wesentliche Erkenntnis ist, dass Archive (gleich welcher Art) nur so gut sein können wie die Qualität ihrer Zusammenarbeit mit den Datenproduzenten. Das Know-how, wie die Daten entstanden sind, gepaart mit Methodenwissen zu ihrer Verzeichnung und Beschreibung, öffnet das Tor zu einer nachhaltigen Archivierung, deren Qualität sich in der Benutzbarkeit des Datensets auch für den Wissenschaftler und die Wissenschaftlerin aus der Nachbardisziplin erweist.



Reinhard Altenhöner und Claudia Oellers
(Hrsg.):

**LANGZEITARCHIVIERUNG VON
FORSCHUNGSDATEN**
STANDARDS UND
DISZIPLINSPEZIFISCHE LÖSUNGEN

Scivero Verlag 2012
ISBN 978-3-944417-00-4
16,70 €

Die Langzeitarchivierung von Forschungsdaten ist eine Voraussetzung für gute wissenschaftliche Praxis. Sie weist drei zentrale Bereiche auf: die Dokumentation der Forschungsdaten, deren langfristige Aufbewahrung sowie die Bereitstellung eines Zugangs zu den Daten. Ohne diese infrastrukturellen wie organisatorischen Voraussetzungen sind die Daten für die wissenschaftliche Sekundärnutzung, also für die Überprüfung von Ergebnissen und auch für die Beantwortung neuer Forschungsfragen nur eingeschränkt verfügbar. Das vorliegende Buch gibt einen Überblick über bestehende Standards und liefert einen Beitrag zur Diskussion über Voraussetzungen zur Archivierung von Datenbeständen. Es ist somit gleichermaßen für Infrastruktureinrichtungen, Fachbibliotheken, Archive, Wissenschaftler und alle, die im weitesten Sinne mit der Verfügbarmachung von Forschungsdaten betraut sind, lesenswert.

Bestellungen können über unsere Website www.ratswd.de aufgegeben werden.