

Dirk Weisbrod

## Pflichtablieferung von Forschungsdaten

### Die Policy für dissertationsbezogene Forschungsdaten der Deutschen Nationalbibliothek (DNB)<sup>1</sup>

Seit einiger Zeit stehen Forschungsdaten auch im Fokus der bibliothekarischen Diskussion. Dabei geht es nicht nur um das Management und die Publikation dieser Daten, sondern auch um deren Langzeitarchivierung. Besondere Aufmerksamkeit erreichte das Thema 2016 durch die Veröffentlichung der Empfehlungen des Rates für Informationsinfrastruktur (RfII) zum Forschungsdatenmanagement<sup>2</sup>, in denen unter anderem der Aufbau einer immer noch fehlenden Nationalen Forschungsdateninfrastruktur (NFDI) angemahnt wird. Fast zeitgleich begann das von der Deutschen Forschungsgemeinschaft (DFG) geförderte Projekt »Elektronische Dissertationen Plus« (eDissPlus), das die DNB zusammen mit der Humboldt-Universität zu Berlin durchführt. eDissPlus konzentriert sich auf Forschungsdaten, die Promovierende im Rahmen ihres Dissertationsprojekts generieren und veröffentlichen. Unter anderem werden in diesem Projekt ein Ingest-Prozess für die Publikation solcher Daten entwickelt und der Pflichtablieferungsworkflow der DNB in Hinblick auf die neuen Sammelobjekte erweitert. Ein Ergebnis von eDissPlus ist die DNB-Policy für dissertationsbezogene Forschungsdaten, die in diesem Artikel vorgestellt wird.

### Sammelauftrag der DNB bezüglich Forschungsdaten

Warum beschränkt sich die DNB bei eDissPlus auf dissertationsbezogene Daten, die einen vergleichsweise kleinen Ausschnitt des gesamten Forschungsdaten-Aufkommens repräsentieren? Die Antwort ergibt sich aus den gesetzlichen Vorgaben. Demnach sammelt die DNB nur Daten, die Teil einer Veröffentlichung sind – oder wie es die Pflichtablieferungsverordnung (PflAV) in Paragraph 7, Absatz 2

formuliert: »Elemente (...), die in physischer oder in elektronischer Form erkennbar zu den ablieferungspflichtigen Netzpublikationen gehören, auch wenn sie für sich allein nicht der Ablieferungspflicht unterliegen« (Paragraph 7). Selbstständig veröffentlichte Primär-, Forschungs- und Rohdaten sind hingegen nicht ablieferungspflichtig (Paragraph 9, PflAV). Mit dem Projekt eDissPlus bietet sich folglich die Gelegenheit, den aus den gesetzlichen Bestimmungen hervorgehenden Sammelauftrag für unselbstständig publizierte Forschungsdaten am Beispiel von Dissertationen zu bearbeiten.

### Das Kriterium der Unverzichtbarkeit

Diesen Sachverhalt präzisiert die Policy, indem sie die Frage klärt, wann Forschungsdaten erkennbar zu einer Dissertation gehören. Dabei sind inhaltliche und strukturelle Aspekte zu beachten. Der inhaltliche Aspekt lässt sich sehr leicht aus den »Empfehlungen zur Sicherung guter wissenschaftlicher Praxis« der DFG ableiten, wonach Primärdaten als »Grundlagen für Veröffentlichungen zehn Jahre aufbewahrt werden sollen.«<sup>3</sup> Damit will man unter anderem die Verfügbarkeit der Daten für den Nachvollzug und die Nachprüfbarkeit der Forschungsergebnisse sicherstellen. Sieht man von dem für die Langzeitarchivierung viel zu knappen Zeitraum von zehn Jahren ab, lässt sich daraus das gewünschte Kriterium ableiten: Forschungsdaten gehören dann erkennbar zu einer Dissertation, wenn sie für die Nachprüfbarkeit der präsentierten Forschungsergebnisse und damit für Dissertationen selbst unverzichtbar sind. In diesem Fall können sie laut Policy mit der Dissertation an die DNB abgeliefert werden. Über die Unverzichtbarkeit von Daten entscheidet selbstverständlich immer der Promovend oder die abliefernde Institution, da nur die beteiligten Wissenschaftlerinnen und Wissenschaftler über das notwendige Wissen verfügen.

Um Redundanzen zu vermeiden, macht die Policy jedoch eine Einschränkung. Es sollen nach Möglichkeit keine Daten abgeliefert werden, die für die jeweilige Disziplin allgemeine Bedeutung haben und folglich auch in Dissertationen und anderen wissenschaftlichen Publikationen sehr oft nachgenutzt werden. Solche Daten sind zum Beispiel sozialwissenschaftliche Daten wie etwa der Mikrozensus, die von Instituten wie dem GESIS bereits vorbildlich archiviert werden.

## Verschiedene Anlagerungsformen

Strukturell drückt sich die Erkennbarkeit durch die Anlagerungsform aus. Hier unterscheidet die DNB zwischen Daten, die entweder in den Dissertations-text integriert sind (Typ 1), als Supplement beige-fügt (Typ 2) oder referenziert (Typ 3) werden.<sup>4</sup> Da Daten des Typs 1 immer schon als Bestandteil einer Dissertation gesammelt wurden, konzentriert sich die Policy auf Daten des Typs 2 und 3. Daten, die im oben genannten Sinne unverzichtbar sind, können somit bei der Abgabe einer Dissertation beige-fügt oder – wenn sie schon auf einem Fach-Repo-sitorium publiziert wurden – durch Angabe eines Persistent Identifiers (URN, DOI, handle) in den Dissertations-Metadaten referenziert werden.

## Verteilter Archivierungsansatz

Mit der Möglichkeit, Daten zu referenzieren, unterstützt die DNB zudem jenen verteilten Archivierungsansatz, der auch vom Rat für Informationsinfrastruktur gefordert wird. Dabei muss sie aber beachten, dass die betreffenden Repositorien nicht

immer vertrauenswürdig im Sinne der Langzeitarchivierung sind und die Daten oftmals nur über einen begrenzten Zeitraum aufbewahrt werden. In solchen Fällen behält sich die DNB laut Policy vor, die Daten zu harvesten, um deren Langzeitarchivierung in den eigenen Systemen sicherzustellen. Hierzu müsste zunächst anhand des Persistent Identifiers das Repositorium identifiziert und dann in einem zweiten Schritt abgeglichen werden, ob dieses Repositorium vertrauenswürdig ist. Dieser Abgleich kann zum Beispiel anhand einer Liste mit als vertrauenswürdig zertifizierten Repositorien geschehen, die regelmäßig aktualisiert wird. Die Entwicklung eines solchen Prozesses ist sehr aufwändig, weswegen das Harvesting der referenzierten Daten im Rahmen von eDissPlus nicht bearbeitet wird.

## Ausblick

Mit der Veröffentlichung der Policy definiert die DNB ihre Zuständigkeit für Forschungsdaten. In Zukunft wird der Katalog der DNB jene Forschungsdaten nachweisen, die mit Dissertationen abgeliefert wurden. Sie können entsprechend der Lizenzierung benutzt und nachgenutzt werden. Zugleich passt die DNB auch ihr Langzeitarchivierungskonzept an die Aufbewahrung von Forschungsdaten an. Dabei ist die Vielfalt und Komplexität der abgelieferten Daten und Formate zu beachten. Aus diesem Grund ist es derzeit nicht möglich, die Nachnutzbarkeit für alle abgelieferten Datentypen über lange Zeiträume zu garantieren. Auch auf diesen Umstand weist die Policy hin. Die Langzeitarchivierung wird aber selbstverständlich »State of the Art« durchgeführt. Zugleich aber gilt es, zukünftige Entwicklungen zu beobachten und die Policy neuen Rahmenbedingungen anzupassen.

## Anmerkungen

- 1 Die Policy für dissertationsbezogene Forschungsdaten finden Sie im Katalog der DNB. Ihr Persistent Identifier lautet: <urn:nbn:de:101-2017092701>
- 2 Rat für Informationsinfrastrukturen (Hrsg.), a. a. O., S. 11
- 3 Sicherung guter wissenschaftlicher Praxis: Denkschrift Hrsg. von der DFG, 2013, S. 21
- 4 ODE war ein Projekt, an dem auch die DNB beteiligt war. Reilly, Susan [u. a.]: Opportunities of Data Exchange: Report on Integration of Data and Publications, <<http://doi.org/10.5281/zenodo.8307>>, S. 18ff