

**2019** startet mit Förderung der Deutschen Forschungsgemeinschaft der Aufbau eines zentralen Forschungsdatendienstes für die Ost-, Ostmittel- und Südosteuropaforschung. *OstData* wird von der Bayerischen Staatsbibliothek gemeinsam mit vier in dem Bereich ausgewiesenen wissenschaftlichen Institutionen getragen. Es soll eine Infrastruktur entstehen, die den jeweiligen Bedürfnissen und technischen Voraussetzungen gemäß in verschiedenen Repositorien vorgehaltene Daten verknüpft und über einen zentralen Suchindex recherchierbar macht. Intensive Maßnahmen der Beratung und der Öffentlichkeitsarbeit in der Community werden Wahrnehmung und Nutzung durch Forschende gewährleisten. Durch die Einbindung aller Metadaten in überregionale Recherchertools wird *OstData* voll anschlussfähig an die NFDI und andere (inter-)nationale Plattformen sein. Damit soll zugleich die internationale Sichtbarkeit der deutschen Forschung zum östlichen Europa befördert werden.

Funded by the German Research Foundation (DFG), work is starting in 2019 on establishing a central research data service for eastern, eastern-central and southeastern Europe research. *OstData* is being implemented by the Bayerische Staatsbibliothek together with four scientific institutions in this field. The aim is to create an infrastructure which, reflecting the respective needs and technical requirements, links the data stored in different repositories and makes it retrievable via a central search index. Intensive consulting and PR measures in the community will raise the awareness and usage levels among researchers. By integrating all metadata into transregional research tools, *OstData* will be fully compatible with the NFDI (National Research Data Infrastructure) and other (inter-)national platforms. A further aim is to promote the international visibility of German research on eastern Europe.

ARNOŠT ŠTANZEL, GUDRUN WIRTZ

# OstData – ein Forschungsdatendienst für die Ost-, Ostmittel- und Südosteuropastudien

**D**er Umgang mit Forschungsdaten hat sich in den vergangenen fünf Jahren zu einem wichtigen wissenschaftspolitischen Thema entwickelt, das hohe Erwartungen an Infrastruktureinrichtungen wie Bibliotheken mit sich bringt. Forschungsdaten können je nach Fachkontext Grundlage oder Gegenstand eines Forschungsprozesses sein, während des Prozesses entstehen oder dessen Ergebnis darstellen.<sup>1</sup> Die Publikation von Forschungsdaten soll Forschungsergebnisse einerseits transparenter und nachvollziehbarer machen, andererseits für andere Forschungsprojekte die Nachnutzung ermöglichen. Um dies zu erreichen, ist ein durchdachtes Management dieser Daten durch Autorinnen und Autoren sowie Herausgeberinnen und Herausgeber in enger Koordination mit denjenigen Institutionen, die die Daten bereitstellen und auffindbar machen, unabdingbar.

Welch große Bedeutung Politik und wissenschaftliche Selbstverwaltungsorgane diesem Thema beimessen, wurde zum Jahresende 2018 deutlich, als die Gemeinsame Wissenschaftskonferenz (GWK) von Bund und Ländern den Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI) in Deutschland mit geplanten 90 Millionen Euro jährlicher Förderung beschloss. Das Thema ist freilich schon länger virulent, weshalb wissenschaftsfördernde Einrichtungen bereits seit einigen

Jahren Wert auf einen bewussten Umgang mit solchen Daten in Projektanträgen legen. Auch wissenschaftliche Institutionen sowie Autorinnen und Autoren suchen, wenngleich nicht flächendeckend, bereits seit dem Aufkommen des digitalen Publizierens nach Lösungen für die Aufbereitung und Bereitstellung von Forschungsdaten, durchaus auch solcher, die noch nicht in digitaler Form erhoben wurden. Ein Beispiel hierfür ist die *Amburger Datenbank*<sup>2</sup>: Sie beruht auf dem umfangreichen Personenarchiv des Osteuropahistorikers Erik Amburger (1907–2001) zu rund 100.000 Ausländern im Russischen Reich bis zum Jahr 1917. Amburger hatte dem Osteuropa-Institut München diese Kartei, die er im Zuge seiner Forschungen erstellt hatte, testamentarisch vermacht. 1991–1999 wurden die Daten im Osteuropa-Institut in digitale Form gebracht und 2006–2007 mit Mitteln der Deutschen Forschungsgemeinschaft (DFG) im Rahmen der *Virtuellen Fachbibliothek Osteuropa* ([www.ViFaOst.de](http://www.ViFaOst.de)) zu einer Online-Datenbank konvertiert. Sie stellt heute ein einzigartiges, von Wissenschaftlerinnen und Wissenschaftlern zahlreicher Länder reges genutztes Nachschlagewerk dar.

Die *Virtuelle Fachbibliothek Osteuropa*, in deren Rahmen auch andere Forschungsdaten – gleichsam nicht unter diesem Label – publiziert wurden, ging 2002 online und war ein Gemeinschaftsprojekt der Ludwig-

Maximilians-Universität (LMU), des Osteuropa-Instituts München (heute Leibniz Institut für Ost- und Südosteuropaforschung (IOS) in Regensburg), des Herder-Instituts für historische Ostmitteleuropaforschung (HI) und der Bayerischen Staatsbibliothek (BSB). Sie wurde seither kontinuierlich weiterentwickelt und bietet – seit Mitte 2018 unter dem Namen *osmikon*<sup>3</sup> – Zugang zu einer Vielzahl an Informationsangeboten, Services und Recherchemöglichkeiten aus und zu den Zielländern der Osteuropastudien. 2008 wurde sodann mit *OstDok* durch die BSB gemeinsam mit dem Collegium Carolinum (CC), HI und IOS ein Fachrepositorium für die Ost-, Ostmittel und Südosteuropaforschung aufgebaut, in dem aktuelle urheberrechtlich geschützte Publikationen der deutschen Forschung zum östlichen Europa sowie urheberrechtsfreie Osteuropaliteratur aus aller Welt im Open Access zur Verfügung gestellt werden. Die Erfahrungen aus *ViFaOst* und *OstDok* waren wichtige Voraussetzungen, das elektronische Publizieren weiterzudenken und systematisch auf den Bereich der Forschungsdaten auszuweiten. Dementsprechend diskutierten die *OstDok*-Partner unter Hinzuziehung des forschungstarken Leibniz-Instituts für Geschichte und Kultur des östlichen Europas (GWZO) das Thema ab 2015 intensiv, begleitet durch Umfragen, Workshops und Vorträge sowie institutsinterne Bestandsaufnahmen. Im Rahmen des DFG-geförderten Fachinformationsdienstes (FID) Ost-, Ostmittel- und Südosteuropa wurde sodann mit den Partnern ein eigenständiger Projektantrag (*OstData*) erarbeitet, welcher Ende 2018 durch die DFG bewilligt wurde.

### **OstData – Projektziele und Rahmenbedingungen**

*OstData* spiegelt gewissermaßen den aktuellen Stand zum Thema Forschungsdatenmanagement in der deutschen Ost-, Ostmittel- und Südosteuropaforschung wider: Auf der einen Seite stehen relativ gut ausgestattete Infrastruktureinrichtungen und außeruniversitäre Forschungsinstitutionen (davon drei Leibniz-Institute), welche über die Ressourcen verfügen, sich mit der neuen Thematik intensiv auseinanderzusetzen. Auf der anderen Seite ist die Forschungslandschaft von vielen kleinen Instituten sowie universitären Seminaren mit einzelnen Lehrstühlen geprägt, welche kaum über die Kapazitäten verfügen, sich intensiv mit dem Forschungsdatenmanagement zu beschäftigen. Dies hat zur Folge, dass in weiten Teilen der Fachcommunity das Interesse am Forschungsdatenmanagement noch relativ gering ist, und eine Publikationskultur, die solche Daten selbstverständlich miteinschließt, in weiter Ferne scheint.

*OstData* versteht sich als Projekt für beide Seiten. Zentrale Ziele sind einerseits der Aufbau einer technischen Infrastruktur zur Speicherung, Veröffentlichung, Langzeitarchivierung (LZA) und Suche von Forschungsdaten, andererseits vielfältige Beratungsleistungen für die wissenschaftliche Community. Beide

Projektbereiche müssen dabei so konzipiert sein, dass sie unterschiedlichsten Voraussetzungen seitens der Autorinnen und Autoren und ihren Institutionen Rechnung tragen. Zudem wird es vor dem Hintergrund des beschlossenen Aufbaus einer Nationalen Forschungsdateninfrastruktur darum gehen, *OstData* mittels Schnittstellen an nationale und internationale Forschungsdateninfrastrukturen anzubinden.

*OstData* ist zunächst auf drei Jahre angelegt, wobei die Verlängerung um weitere drei Jahre in der Konzeption mitgedacht ist: In den Jahren eins bis drei wird eine deutschlandweite Infrastruktur entworfen und technisch wie organisatorisch unter den Projektpartnern umgesetzt. Sie richtet sich gleichermaßen an Forschungsdaten produzierende Wissenschaftlerinnen, Wissenschaftler und Institutionen sowie an Nutzerinnen und Nutzer weltweit, die Forschungsdaten als Quelle suchen und verwenden wollen. In dieser Projektphase werden zudem exemplarisch verschiedene Typen von Forschungsdaten aufbereitet und die Verfahren in Form von »Policies« und »Best Practices« dokumentiert. In den Jahren vier bis sechs soll die etablierte Infrastruktur sodann für andere Institutionen geöffnet werden, d.h. Daten der deutschen Forschung zum östlichen Europa werden möglichst flächendeckend eingeworben, archiviert und in das Rechercheportal eingebunden.

Die Koordination von *OstData* liegt an der BSB (mit einer vollen Koordinations- sowie einer halben Technikerstelle). Die vier Partnerinstitutionen, jeweils mit halben wissenschaftlichen Mitarbeiterstellen ausgestattet, zeichnen für ausgewählte Teilbereiche verantwortlich, zu welchen sie ihr jeweiliges spezialisiertes institutionelles Wissen einbringen. Trotz beträchtlicher Förderung werden alle Partnerinstitutionen *OstData* mit einem hohen Maß an Eigenleistung unterstützen, um die geplanten Projektziele zu erreichen.

### **Aufbau einer technischen Infrastruktur für Speicherung und Veröffentlichung**

Eine zentrale Herausforderung von *OstData* ist es, über Deutschland verstreut und unter stark divergierenden institutionellen Rahmenbedingungen entstehende Forschungsdaten zu sichern und für zentrale und dezentrale Recherchen zu indexieren. Dementsprechend ist ein *modulares und netzwerkartiges System der Datenspeicherung* geplant, das folgende Konstellationen bzw. technische Rahmenbedingungen bedienen soll:

#### **1. Forschungsdaten im BSB-Repositorium**

Forschungsinstitutionen, die keine eigene Infrastruktur für Forschungsdaten aufbauen wollen oder können sowie einzelne Wissenschaftlerinnen und Wissenschaftler geben ihre publikationsfertig aufbereiteten Forschungsdaten an die BSB. Diese übernimmt die inhaltliche Erschließung, legt sie im Repositorium der BSB (Dokumentenserver BibDoRM) ab und übernimmt alle notwendigen Maßnahmen zur

Langzeitarchivierung der Daten. Sodann indexiert sie die Daten, ermöglicht die Recherche und stellt sie zur Nutzung und zum Download zur Verfügung.

## 2. Forschungsdaten in institutseigenen Repositorien

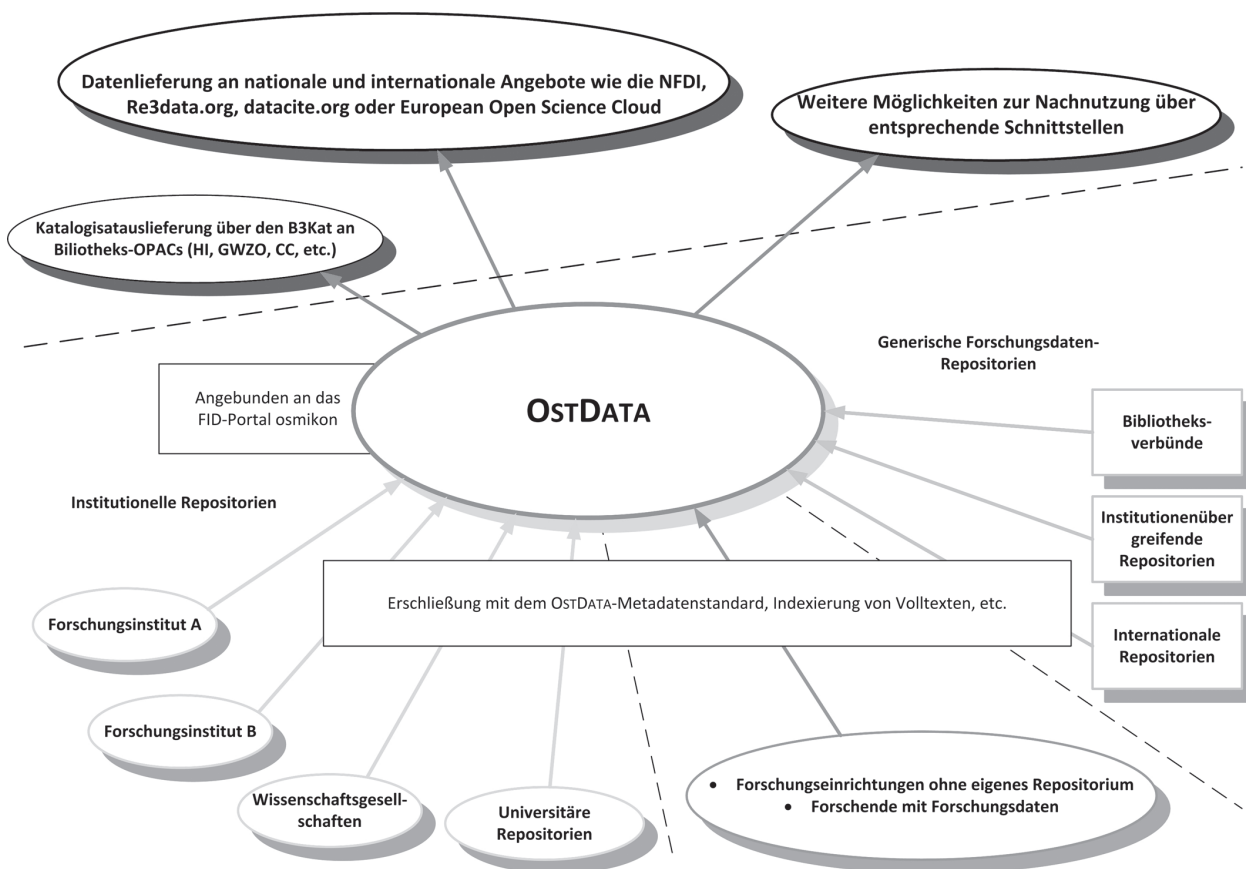
Sofern sich Institutionen entscheiden, ihre Forschungsdaten in einem eigenen Repository abulegen, werden die dazugehörigen Metadaten in den Suchindex von *OstData* samt zusätzlicher Informationen wie (Schatten-)Volltexten aufgenommen, um die einfache Auffindbarkeit zu gewährleisten. Die Bereitstellung und Nachnutzung geschieht im jeweiligen institutionellen Repository. Institutionen, die sich für dieses Modell entscheiden, suchen sich für die LZA ihrer Daten Partnerinstitutionen vor Ort – universitäre Rechenzentren, Landesinstitutionen oder Bibliotheksverbände.

## 3. Forschungsdaten in fachübergreifenden Repositorien

Entscheiden sich Autorinnen, Autoren oder Forschungsinstitutionen, ihre Daten in Universitätsrepositorien, Repositorien anderer Landesbibliotheken oder fachübergreifenden Forschungsdatendiensten wie RADAR oder ZENODO abulegen, werden die Daten dort jeweils auch langzeitarchiviert. Wie in Modell 2 werden sie von der BSB dann in den *OstData*-Suchindex übernommen.

Jede der Konstellationen wird während der ersten Projektphase von einem der wissenschaftlichen Institute gemeinsam mit der BSB erprobt und dokumentiert. Zugleich bereitet *OstData* den Austausch mit anderen Repositorien technisch vor, um die Forschungsdaten mittels ihrer Metadaten etwa in die vom Rat für Informationsinfrastrukturen vorgeschlagene Nationale Forschungsdateninfrastruktur<sup>4</sup> oder bestehende Projekte wie GERDI und die DARIAH-DE Collection Registry einspeisen zu können. Kommunikationsmöglichkeiten mit anderen Forschungsdatenrepositorien werden voraussichtlich über eine OAI-PMH-Schnittstelle<sup>5</sup> angeboten. Dem Paradigma von Open Data folgend sind weitere Schnittstellen vorgesehen, um perspektivisch Nutzungsszenarien im Kontext des Semantic Web, von Linked Open Data oder des Text- und Datamining vorzubereiten.

Ein besonders wichtiger Aspekt bei der Speicherung von Forschungsdaten ist deren langfristige Archivierung. Sie gewährleistet die Persistenz, welche für deren Zitation und Nachnutzung sowie für die Nachvollziehbarkeit von datenbasierten Forschungen notwendig ist. Basierend auf Erfahrungen der BSB im Rahmen von BABS<sup>6</sup> wird für *OstData* ein spezielles Konzept zur LZA erarbeitet. Die grundlegendste Ebene der LZA



1 Einbettung von *OstData* in der Forschungsdaten-Landschaft

Quelle: Eigene Bearbeitung

bilden die Bitstreams der Forschungsdaten. Dabei handelt es sich um den Binärcode der Dateien aus Nullen und Einsen. Dieser kann langfristig archiviert und zur Verfügung gestellt werden. Solche Erhaltungsmaßnahmen stellen die Grundlage für die weitere Nutzung der Daten dar. Sie sichern aber noch nicht die tatsächliche Interpretierbarkeit der Bitstreams, wie sie etwa dem Öffnen, Abspielen oder sonstigen Benutzen der Dateien zugrunde liegt. Hierfür sind weitere technische Metadaten sowie die Verfügbarkeit entsprechender Betrachtungs-Software erforderlich. Ausgehend von diesem technischen Hintergrund werden im Rahmen des Projekts Konzepte für weitere Datentypen mit Relevanz für die Osteuropaforschung erarbeitet.<sup>7</sup> Ihre Auswahl, die Priorisierung und die Ausgestaltung des Prozesses erfolgen in enger Rückbindung an die wissenschaftliche Zielgruppe. Für die Konzeption und Umsetzung von Maßnahmen im Bereich der LZA werden die Erhaltungs- und Nutzungsperspektive der Daten anhand von sogenannten Preservation Levels transparent und nachvollziehbar dokumentiert. Die Einstufung von Dateitypen in Preservation Levels gibt Auskunft darüber, in welchem Umfang die langfristige Interpretierbarkeit der Bitstreams im Rahmen der LZA sichergestellt wird – etwa die Zusage der fortlaufenden Konvertierung von Dateien in neuere Dateitypen-Standards oder die Begrenzung auf eine reine Speicherung der Bitstreams. Für Dateiformate, zu denen keine detaillierteren LZA-Verfahren verfügbar oder im Projektrahmen realisierbar sind, können zunächst nur die Bitstreams archiviert werden. Hier wird geprüft, wie auch die Nachnutzbarkeit langfristig gewährleistet werden kann.

### Umsetzung einer zentralen Rechercheanwendung

Damit Forschungsdaten sinnvoll recherchiert werden können, müssen sie in geeigneter Weise beschrieben und in Nachweisinstrumenten verzeichnet sein, deren Suchmöglichkeiten über die herkömmlichen hinausgehen. Die Beschreibung der Inhalte mit standardisierten Metadaten ist also auch in *OstData* grundlegend. Mit ihrer Hilfe werden unter anderem technische, administrative, rechtliche und inhaltliche Aspekte wie Dateigröße, Zugriffsberechtigungen und inhaltliche Beschreibungen erfasst. Zudem stellen sie die Nachnutzbarkeit über automatisierte Datenimporte und -exporte sowie über Schnittstellen sicher. Wesentlich für *OstData* ist dabei, der Vielfalt der Forschungsdaten generierenden Disziplinen in der Ost-, Ostmittel- und Südosteuropaforschung mit unterschiedlichen methodischen Ansätzen (beispielsweise qualitative oder quantitative) Rechnung zu tragen. Das *OstData*-Metadaten-Set muss folglich in der Lage sein, etablierte Sets aus verschiedenen Disziplinen abzubilden. Als Basisformat wird voraussichtlich eines der beiden am weitesten verbreiteten Metadaten-Schemata, Dublin Core oder Data Cite, gewählt und als »Kerndatensatz« zur Beschreibung der Forschungs-

daten verwendet. Hinzu kommen muss die Integration disziplinen- und formattypischer Metadaten-Schemata wie TEI für Texteditionen, LIDO für die Beschreibung musealer Objekte oder SDMX für die Beschreibung statistischer Daten. Für die verbale Sacherschließung soll die Gemeinsame Normdatei (GND) genutzt werden, zusätzlich auch Fachthesauri und freie Schlagwörter in anderen Sprachen. Um ein bequemes Browsing in den Forschungsdaten zu erlauben, wird darüber hinaus eine klassifikatorische Sacherschließung auf Basis einer facettierten Version der Dewey Decimal Classification (DDC) zur Anwendung kommen. Handlungsleitende Anforderung für das *OstData*-Metadaten-Schema ist, dass es den FAIR Data Principles<sup>8</sup> (Forschungsdaten sollen findable, accessible, interoperable und reusable sein) entspricht. Eine Verknüpfung des *OstData*-Metadaten-Schemas mit dem »Kerndatensatz Forschung«<sup>9</sup> wird den Entstehungskontext der Forschungsdaten dokumentieren und die veröffentlichten Forschungsdatensätze in Forschungsinformationssystemen nachnutzbar machen. In den Sozialwissenschaften, perspektivisch auch in den Digital Humanities, ist die Möglichkeit, theoretische Zugänge zu suchen, von Bedeutung. Zu diesem Zweck wird *OstData* den Einsatz Phänomen-basierter Wissensorganisation gemäß dem León Manifesto<sup>10</sup> erproben, das Facettenklassifikation von Phänomenen, Methoden, Theorien und Disziplinen erlaubt. Diese wären kombinierbar und ermöglichen gegenüber Schlagwortketten verbesserte Suchen.

Technisch wird das *OstData*-Metadaten-Schema mithilfe des Metadatenmodells Resource Description Framework (RDF) auf Basis der Auszeichnungssprache XML umgesetzt. Das Management der Metadaten aus verschiedenen Quellen soll mithilfe von Mapping-Tools<sup>11</sup> und anderen Software-Werkzeugen erleichtert werden.

Im Rahmen des in den Osteuropastudien etablierten Forschungsportals *osmikon* soll für *OstData* eine spezielle Suchoberfläche verankert werden, die erweiterte Suchmöglichkeiten wie Facetten-Navigation oder geografische, zeitliche, format- und quellenspezifische Sucheinstiege erlaubt. Den heutigen Anforderungen an die Abfrage digitaler Datenbestände entsprechend werden zudem maschinenlesbare Schnittstellen für Zugriff auf und Nachnutzung der Metadaten eingerichtet. Hierbei wird das REST-Programmierparadigma<sup>12</sup> unter Berücksichtigung des Schnittstellendesignmodells HATEOAS<sup>13</sup> ins Auge gefasst. Weiterführende Recherche- und Bereitstellungsverfahren wie beispielsweise die Abfragesyntax SPARQL oder RDF werden bei entsprechend vorhandenen Forschungsdaten erprobt werden.

Die stark diversifizierte Forschungsdatenlandschaft im Bereich der Osteuropastudien legt nahe, die in *OstData* nachgewiesenen Forschungsdaten möglichst breit, aber zugleich sehr gezielt zu streuen. Schließlich soll *OstData* auch zur nationalen und internationalen Sichtbarkeit der deutschen Osteuropaforschung beitra-

gen. Entsprechend wird *OstData* nicht nur als eigene Datenquelle in die bestehende Literaturrecherche von *osmikron* eingebunden, sondern durch die Speicherung der Kerndaten der Katalogisate (Titel, Autor, etc.) im B3Kat über die Verbünde auch in lokalen Bibliothekskatalogen suchbar gemacht. In Deutschland wird vor allem die Zusammenarbeit mit den Konsortien der NFDI eine wichtige Rolle spielen. Durch die Einrichtung weit verbreiteter Schnittstellen wie OAI-PMH wird *OstData* in der Lage sein, Forschungsdaten aus verschiedenen Quellen zu aggregieren und gleichzeitig in nationale und internationale Strukturen wie die NFDI, die German Research Data Infrastructure (Gerdi), die Helmholtz Data Federation, Leibniz Data, DataCite, OPENAire, European Open Science Cloud, EUDAT, re3data.org oder ZENODO einzuspeisen.

### Publikationskultur für Forschungsdaten fördern

Eine funktionierende technische Infrastruktur für Speicherung, Veröffentlichung und Recherche von Forschungsdaten ist selbstverständlich die Voraussetzung für ein erfolgreiches Management dieser Daten. Mindestens genauso wichtig, wenn nicht sogar bedeutender ist es aber, zugleich eine neue Publikationskultur für Forschungsdaten in der Ost-, Ostmittel- und Südosteuropaforschung anzustoßen. Hierzu plant *OstData* eine Reihe an Maßnahmen, um Forschende und weitere Stakeholder in der Wissenschaft über das Forschungs-

datenmanagement zu informieren und zu beraten und zu diesem Zweck konkrete Beispiele zur Veranschaulichung des Themas bereitzustellen.

Erfahrungen zeigen, dass publizierbare Forschungsdaten vor allem dann entstehen, wenn bereits am Beginn eines Forschungsprojekts deren Veröffentlichung mitbedacht wurde. Dem oben skizzierten Forschungsdatenlebenszyklus folgend wird *OstData* daher Forschenden Orientierungshilfen zur Unterstützung ihrer internen Projektorganisation aktiv anbieten. Diese Informationen werden über verschiedene Kanäle wie Webseiten, Wikis, Downloads, in der persönlichen Beratung, Online-Tutorials, Webinaren, etc. zugänglich gemacht. Für regelmäßig auftretende Rechtsfragen sollen Informationsmaterialien bereitgestellt werden. Für all dies kann auf umfangreiche Vorarbeiten durch andere Projekte zurückgegriffen werden.<sup>14</sup> Die Einführung eines Forschungsdatenmanagements ist für Institutionen grundsätzlich mit hohem administrativem, technischem und inhaltlichem Aufwand verbunden, stellt also in ihrem ohnehin unter hohem Zeit- und Erfolgsdruck stehenden Alltag eine recht große Hürde dar. Um den infrastrukturellen Wandel unter diesen Bedingungen zu unterstützen, wird *OstData* flexibel anpassbare Umsetzungsstrategien erarbeiten. Diese beinhalten Hilfen und Vorlagen für die Erstellung von Forschungsdatenpolicies sowie konkrete Maßnahmenkataloge zur Einführung eines institutionellen Forschungsdatenmanagements. Ferner werden für die Fort- und Weiterbildung von Mitarbeiterinnen und Mitarbeitern, die schwerpunktmäßig mit dem Forschungsdatenmanagement betraut sind (z.B. Bibliothekarinnen und Bibliothekare, Datenkuratorinnen und -kuratoren, etc.) spezielle Informationsmaterialien entwickelt.

Ein weiterer Schwerpunkt dieses Projektteils liegt auf Einwerbungsstrategien für Forschungsdaten. Dabei wird beispielhaft in der Bohemistik erprobt, wie mit Berührungsängsten, Widerständen und Vorbehalten der Forschenden am besten umzugehen ist. Im Ergebnis sollen frei nutzbare Informationsmaterialien sowie aussagekräftige Empfehlungen zum Umgang mit rechtlichen Hürden zusammengestellt werden. Denn die Erfahrung zeigt, dass nicht selten etwa fehlende Rechteeinholungen zur Nutzung und Veröffentlichung von Interviewdaten, Datenschutz oder forschungsethische Aspekte der Veröffentlichung und Nachnutzung wertvoller Forschungsdaten im Wege stehen. Auch Informationen zu Aspekten des Urheber-, Verwertungs- und Persönlichkeitsrechts, die bei Datenkonvertierung, Aufbereitung und Archivierung bedacht werden müssen, ebenso Fragen der Haftung und des Haftungsausschlusses, sind in diesem Kontext relevant.

Neben Maßnahmen der unmittelbaren Beratung stehen solche der breiteren Öffentlichkeitsarbeit: Zielgruppen sind Forschende, Verantwortliche an Forschungsinstituten und Lehrstühlen sowie Bibliotheksmitar-



2 Forschungsdatenlebenszyklus

Quelle: Eigene Bearbeitung

beiterinnen und -mitarbeiter. Genutzt werden dafür Informationskanäle wie geistes- und sozialwissenschaftliche Mailinglisten, Newsletter (u. a. jene der beteiligten Einrichtungen), persönliche Mailings, Informationen auf Fachportalen sowie Social Media. Erfahrungsgemäß ist gerade im Bereich der Forschungsdaten die persönliche Kommunikation besonders effektiv, so dass auch Vortragstätigkeit und Workshops an wissenschaftlichen Instituten, Lehrstühlen, Graduiertenschulen und -kollegs sowie bei Fachverbandstagen wesentlicher Teil des Projekts sind.

Die *OstData*-Projektverantwortlichen sind überzeugt, dass die Veränderung der Publikationskultur mithilfe konkreten Anschauungsmaterials besonders wirksam vorangebracht werden kann. Daher geht *OstData* auch die Aufbereitung von solchen Forschungsdaten an, welche in prä- und fröhdigitalen Zeiten an den beteiligten Forschungsinstituten entstanden und in den dortigen Archiven und Schriftgutsammlungen in analoger Form erhalten oder in heute nur schwer nutzbaren Dateiformaten und Anwendungen gespeichert sind. Jeder der Antragsteller widmet sich einem bestimmten Typ von Alt-Daten, der in der Forschung zum östlichen Europa verbreitet ist: Anvisiert sind strukturierte analoge Daten aus der biografischen Forschung; auf Karteikarten und in Zettelkatalogen festgehaltene Daten; Bild- und Geodaten; Audio- und Videodateien; quantitative Bestandsdaten sowie unselbstständige Tabellen und MS-Access-Datenbanken.

Zur Publikationskultur gehören Qualitätsstandards für die Auswahl, Übernahme und Speicherung von Forschungsdaten, welche die inhaltliche, formale und technische Qualität gewährleisten: Welche inhaltlichen Kriterien müssen Forschungsdaten erfüllen, welche administrativen Punkte müssen beachtet werden, in welchen Formaten werden diese im Forschungsdatenrepositorium gespeichert und wie wird die technisch einwandfreie Speicherung und Archivierung sichergestellt?<sup>15</sup> Auf einer gemeinsamen Definition von Forschungsdaten aufbauend und orientiert an den Empfehlungen des Wissenschaftsrates für die Qualitätssicherung im Forschungsdatenmanagement<sup>16</sup> wird *OstData* Kriterien für die inhaltlichen Anforderungen an Forschungsdaten in den Osteuropawissenschaften erarbeiten. Leitgedanke ist, dass veröffentlichte Forschungsdatensätze einen substantiellen Beitrag zur Wissenschaft leisten und folgende Bedingungen erfüllen:

- sie erlauben neue wissenschaftliche Zugänge oder bilden die Grundlage für neue wissenschaftliche Erkenntnisse;
- sie entsprechen dem aktuellen Stand der fachwissenschaftlichen Debatten;
- sie respektieren datenschutzrechtliche und ethische Aspekte;
- sie entsprechen den Regeln guter wissenschaftlicher Praxis.

Idealerweise wird ein allgemein formalisierbarer und operationalisierbarer Kriterienkatalog entstehen, der Bibliothekare oder Datenkuratorinnen<sup>17</sup> eine weitgehend standardisierte inhaltliche Qualitätsprüfung der Forschungsdaten ermöglicht.<sup>18</sup>

Zu den Qualitätsstandards gehören auch die Nutzung des *OstData*-Metadaten-Schemas und Persistenter Identifikatoren wie DOI sowie die notwendigen Rechtklärungen (ein Übernahmevertrag, der die Rechte und Pflichten zwischen Repositorium und Datengeberinnen und -gebern klärt, bei Forschungsdaten wie Interviews auch Einwilligungs- oder Datenschutzerklärungen usw.). Standards für die einheitliche Dateibenennung und Zitation von Forschungsdaten sollen Unsicherheiten bei Forschenden entgegenwirken und die Nachnutzung verbessern. Auf der Grundlage existierender Guidelines und Zitationsempfehlungen<sup>19</sup> werden entsprechende Richtlinien für *OstData* erstellt. Die formale Qualitätssicherung wird über eine zu entwickelnde Checkliste standardisiert.

Für die Nachnutzung von Forschungsdaten ist es zudem wichtig, dass gewisse technische Spezifikationen eingehalten werden. Generell bietet es sich an, offene Dateiformate zu nutzen, da diese nicht vom Fortbestehen einer Firma als Anbieterin proprietärer Formate abhängen. Das Dateiformat einer Datei sollte zudem auch technisch überprüft werden. Vor diesem Hintergrund kommt im Rahmen der technischen und formatspezifischen Qualitätssicherung zwei Punkten besondere Bedeutung zu: Zum einen die Wahl nachhaltig nutzbarer Dateiformate bereits bei Forschungsprojektbeginn, zum anderen technische Validierungsmaßnahmen. Dazu gehört beispielsweise die Frage, ob die Daten mit den richtigen Dateieindungen abgespeichert sind und der Code der Datei entsprechend der Dateispezifikationen aufgebaut ist, sodass auch zukünftige Programmversionen diese lesen und darstellen können. Zu diesem Zweck ist zu prüfen, wie die Dateiintegrität bei Übertragung und Einspeisung in das Repositorium sowie im Zuge der Bereitstellung und LZA vor nachträglicher Veränderung geschützt werden kann.

#### **Fazit: Ein Forschungsdatendienst für die Ost-, Ostmittel- und Südosteuropaforschung**

Ein erfolgreiches Forschungsdatenmanagement in der deutschen Ost-, Ostmittel- und Südosteuropaforschung zu verankern und Forschende zur Veröffentlichung und Nachnutzung von Forschungsdaten zu motivieren ist ein durchaus ambitioniertes Vorhaben. Die *OstData*-Projektpartner sind jedoch davon überzeugt, den richtigen Mix aus technischen Maßnahmen und Capacity-Building-Ansätzen gewählt zu haben, um den Forschungsdatendienst erfolgreich voranzubringen. Von Vorteil wird dabei sein, dass die Projektpartner jeweils auf unterschiedlichen Feldern wertvolle Erfahrungen und Kenntnisse besitzen, welche sie in einzelne

Arbeitsbereiche von *OstData* federführend einbringen werden: Schwerpunkte sind bei der BSB die Rechercheanwendungen, beim CC die Einwerbungsstrategien, am GWZO die Qualitätssicherung, am HI die Orientierungshilfen zur Einführung des Forschungsdatenmanagements und am IOS disziplinspezifische Aspekte bei der Konzeptionierung des *OstData*-Metadaten-Sets. Dabei erfolgt selbstverständlich in allen Bereichen eine sehr enge Abstimmung unter den fünf Partnern.

Insgesamt wird sicherlich auch das derzeit günstige wissenschaftspolitische und gesellschaftliche Umfeld für *OstData* förderlich sein, welches durch den beschlossenen Aufbau der NFDI und breitere Debatten über »Daten als Öl des 21. Jahrhunderts« geprägt ist.

### Anmerkungen

- 1 Vgl. Kindling, M.; Schirnbacher, P., »Die digitale Forschungswelt« als Gegenstand der Forschung. In: *Information – Wissenschaft und Praxis* 64, Nr. 2–3 (2013), S. 127–136.
- 2 [http://dbis.uni-regensburg.de/frontdoor.php?titel\\_id=10680](http://dbis.uni-regensburg.de/frontdoor.php?titel_id=10680)
- 3 [www.osmikon.de](http://www.osmikon.de) [Zugriff am: 31.1.2019].
- 4 Rat für Informationsinfrastrukturen (RFII): Schritt für Schritt – was bringt wer mit? Ein Diskussionsimpuls zu Zielstellung und Voraussetzungen für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI) (2017). Verfügbar unter: [www.rfii.de/download/rfii-diskussionspapier-2017/](http://www.rfii.de/download/rfii-diskussionspapier-2017/) [Zugriff am: 14.1.2019]. – Jüngst: Ders.: Zusammenarbeit als Chance Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland (März 2018). Verfügbar unter: [www.rfii.de/?wpdmdl=2529](http://www.rfii.de/?wpdmdl=2529) [Zugriff am: 15.3.2018].
- 5 Open Archives Initiative Protocol for Metadata Harvesting.
- 6 Siehe auch <https://www.babs-muenchen.de/> [Zugriff am: 14.1.2019].
- 7 Siehe auch Projekte zur LZA wie Nestor: [www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home\\_node.html](http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html) [Zugriff am: 14.1.2019].
- 8 <https://www.force11.org/group/fairgroup/fairprinciples> [Zugriff am: 14.1.2019].
- 9 Der Kerndatensatz Forschung ist ein vom Wissenschaftsrat angestoßener Standard für die sachangemessene Dokumentation von Forschungsaktivitäten und deren Kontext. Verfügbar unter: <https://kerndatensatz-forschung.de/> [Zugriff am: 14.1.2019].

- 10 Anstatt anhand von Disziplinen soll Wissen durch Phänomene organisiert werden, um so der wachsenden Interdisziplinarität der Wissenschaft Rechnung zu tragen. Vgl. [www.iskoi.org/ilc/leon.php](http://www.iskoi.org/ilc/leon.php) [Zugriff am: 14.1.2019].
- 11 Z. B. Coli-Conc: <https://coli-conc.gbv.de/> [Zugriff am: 14.1.2019].
- 12 REST steht für »Representational state transfer« und bezeichnet einen Architektur- und Funktionsstil für den Aufbau modularer computergestützter Systeme für die Verarbeitung und den Austausch von Daten.
- 13 HATEOAS beschreibt ein Verfahren der Datenanforderung und Übermittlung zwischen computergestützten Systemen, welches einen dynamischeren Zugang zu flexibleren Nutzungs- und Recherchemöglichkeiten der digitalen Objekte bietet, als es in herkömmlichen Vorgehensweisen möglich ist.
- 14 Unter Einbezug bestehender Informationsangebote wie [www.forschungsdaten.info](http://www.forschungsdaten.info), [www.forschungsdaten.org](http://www.forschungsdaten.org), laufender und abgeschlossener DFG-Projekte wie »DataWiz – Ein Assistenzsystem für das Management psychologischer Forschungsdaten«, »Discuss Data: offene Plattform für die interaktive Diskussion der Qualität von Forschungsdaten (am Beispiel der Regionalstudien zur post-sowjetischen Region)«, »GeRDI: Generic Research Data Infrastructure« und Literatur wie zum Beispiel Andorfer, Peter: *Forschungsdaten in den (digitalen) Geisteswissenschaften – Versuch einer Konkretisierung*. DARIAH-DE working papers 14 (2015). Verfügbar unter: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-7> [Zugriff jeweils am: 14.1.2019].
- 15 Vgl. Kindling, Maxi. »Qualitätssicherung im Umgang mit digitalen Forschungsdaten / Quality assurance of digital research data / La garantie de la qualité des données numériques de recherche«. In: *Information – Wissenschaft und Praxis* 64, Nr. 2–3 (2013), S. 137–148. Verfügbar unter: <https://doi.org/10.1515/iwp-2013-0020> [Zugriff am: 14.1.2019].
- 16 Wissenschaftsrat: *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020*, S. 56 f. Verfügbar unter: <https://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> [Zugriff am: 14.1.2019].
- 17 Siehe Kindling: *Qualitätssicherung im Umgang mit digitalen Forschungsdaten*, S. 140.
- 18 Operationalisiert sowohl für die interne Verarbeitung bei der Evaluation der ins Repositorium eingespeisten Daten (z. B. ein Plausibilitäts-Check ob Datenformate mit Daten übereinstimmen (z. B. csv Datei für Umfragen, Messdaten oder Datenbanken, Textformate (docx, txt, pdf) für Texte), als auch extern zur Kenntlichmachung verschiedener Qualitätseinstufungen durch das Qualitätssicherungssystem für die Nutzerin/den Nutzer.
- 19 Z. B. Data Citation Synthesis Group: *Joint Declaration of Data Citation Principles*. Martone M. (ed.) San Diego CA: FORCE11; 2014 [datacitation]. Endorsement List.

## Die Verfasser



Dr. Arnošt Štanzel, Fachinformationsdienst  
Ost-, Ostmittel- und Südosteuropa –  
Koordination Fachportal und Forschungsdaten,  
Osteuropaabteilung, Bayerische Staatsbibliothek,  
Ludwigstraße 16, 80539 München,  
Telefon +49 89 28638-2832,  
arnost.stanzel@bsb-muenchen.de  
Foto: privat



Dr. Gudrun Wirtz, Leiterin der Osteuropa-  
abteilung, Bayerische Staatsbibliothek,  
Ludwigstraße 16, 80539 München,  
Telefon +49 89 28638-2992,  
gudrun.wirtz@bsb-muenchen.de  
Foto: privat