

Sabine Schrimpf

PARSE.Insight



Das EU-Projekt »Insight into issues of Permanent Access to the Records of Science in Europe« (Parse.Insight) untersucht den aktuellen Stand der Langzeitarchivierung von wissenschaftlichen Forschungsdaten in Europa. Dabei handelt es sich z. B. um Messergebnisse von Instrumenten, Daten aus statistischen Erhebungen, aber auch Tonaufnahmen und vieles mehr. Solche Daten werden zunehmend, in einigen Disziplinen fast ausschließlich, digital erhoben. Mehr noch: Der gesamte Wissenschaftsprozess, von der Datenerhebung über die Analyse bis zur Publikation der Ergebnisse, findet zunehmend digital statt. Einerseits bietet die Digitalisierung Forschern neue und verbesserte Arbeitsbedingungen: Virtuelle Zusammenarbeit mit entfernten Kollegen, Datenaustausch und Zugriff auf Publikationen per Mausclick. Andererseits besteht das Risiko, dass digitale Daten kurz- oder langfristig unlesbar werden oder verloren gehen, wenn nicht aktiv für ihre dauerhafte Bewahrung gesorgt wird.

Dabei arbeiten einzelne Forschungsdisziplinen in Projekten und Initiativen bereits daran, die Gefahr eines Datenverlusts abzuwenden. Im Bereich der Erdbeobachtung ist die Verständigung auf europäische Leitlinien zur Datenarchivierung fast abgeschlossen.¹⁾ In den Sozialwissenschaften haben die meisten Archive, die dem Council of European Social Science Data Archives (CESSDA), dem Zusammenschluss der westeuropäischen Datenarchive angehören, Langzeitarchivierungsstrategien implementiert.²⁾ Solche Bemühungen sind jedoch fragmentiert und häufig voneinander isoliert. Ziel des Projektes PARSE.Insight ist es, auf diese Situation aufmerksam zu machen und Empfehlungen zum Aufbau einer europäischen Infrastruktur zur digitalen Langzeitarchivierung wissenschaft-

licher Forschungsdaten zu erarbeiten. Dazu wird das Projekt innerhalb des 7. Rahmenprogramms von der Europäischen Kommission von März 2008 bis Februar 2010 mit 1,25 Mio. Euro gefördert. Im PARSE.Insight-Konsortium sind Forschungseinrichtungen wie CERN (European Organization for Nuclear Research) und ESA (European Space Agency) vertreten, die beide in großem Umfang digitale Messdaten erzeugen, die Max-Planck-Gesellschaft als große nationale Wissenschaftsorganisation, die europäische Dachorganisation der Wissenschaftsverleger STM Publishers (International Association of Scientific, Technical & Medical Publishers), die Nationalbibliotheken Deutschlands und der Niederlande sowie die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) und die FernUniversität Hagen. Die Projektleitung hat der Science and Technology Facilities Council (STFC) in Großbritannien übernommen. Alle Mitglieder des Projektkonsortiums beteiligen sich gemeinsam mit weiteren Partnern an der »European Alliance for Permanent Access«³⁾, die auf eine nachhaltige Infrastruktur für digitale wissenschaftliche Informationen in Europa hinarbeitet.

Im ersten Projektjahr konzentrierten sich die Hauptanstrengungen des Konsortiums auf die Organisation und Durchführung einer breit angelegten Befragung der unterschiedlichen Akteure innerhalb der europäischen Forschungslandschaft. Dabei wurden vier Zielgruppen unterschieden: Die Forscher selbst, Verleger, »Datenbewahrer« (dieser Begriff umfasst Datenbankmanager, Spezialisten in Rechenzentren und Bibliotheken usw.) und Forschungsförderer. Gefragt wurde nach Gegenwart und Zukunft digitaler Archivierung in der Forschung, die interdisziplinäre Nutzung von Forschungsdaten und die erwartete Unterstützung und Verantwortlichkeiten im Archivierungsprozess. Außerdem wurden gezielte Befragungen einzelner Forschungscommunities im Rahmen von drei Fallstudien durchgeführt: In der Hochenergiephysik (koordiniert vom CERN), in der Erdbeobachtung

Projektpartner

Befragung unterschiedlicher Zielgruppen

Stand der Langzeitarchivierung von Wissenschaftsdaten in Europa

Datenverlust abwenden

(koordiniert von der ESA) und in den Geistes- und Sozialwissenschaften am Beispiel Psycholinguistik und Buchwissenschaft, koordiniert von der SUB Göttingen und der Deutschen Nationalbibliothek. Insgesamt beteiligten sich über 3.500 Personen an den Umfragen des Projekts.

Die Auswertung der Befragungen ist noch nicht vollständig abgeschlossen. Erste Ergebnisse flossen jedoch bereits in den Entwurf einer Roadmap für eine europäische Forschungsdaten-Infrastruktur ein, der zweiten Hauptleistung in der ersten Projektphase. Die Roadmap definiert zunächst, was die PARSE.Insight-Partner unter einer Forschungsdaten-Infrastruktur verstehen: Darunter fallen alle technischen, organisatorischen und finanziellen Komponenten, die in den Communities zur Unterstützung von Langzeitarchivierung und (Wieder-) Zugänglichmachung von Forschungsdaten gebräuchlich sind. Die Forschungsdaten-Infrastruktur muss mit vorhandenen Rechen- und Gridinfrastrukturen integrierbar sein und analoge Funktionalitäten in dem Sinne, dass sie bestehende, nun voneinander isolierte, Datenvorräte (»islands of resources«) vernetzt, vorweisen. Die Roadmap präsentiert anschließend eine Reihe Komponenten, die notwendig sind, um die vorhandenen Forschungsdaten, die zum Teil in spezialisierten

Umgebungen vorgehalten und genutzt werden, in einer übergeordneten Infrastruktur zu vernetzen. Die Roadmap steht auf der Projekthomepage zum Download bereit.⁴⁾ Sobald die Auswertung abgeschlossen ist, werden die Umfrageergebnisse dort veröffentlicht werden.

In der zweiten Hälfte der Projektlaufzeit sollen die Umfrageergebnisse in Bezug zur Roadmap gesetzt werden. Der Vergleich soll die Bereiche aufzeigen, in denen verstärkt Unterstützung notwendig ist, um eine vernetzte Infrastruktur für die Langzeitarchivierung digitaler Forschungsdaten und Publikationen aufzubauen. Zur Evaluierung werden die ermittelten Lücken in Expertenworkshops diskutiert.

In einem großen Workshop wurden im September in Darmstadt die vorläufigen Projektergebnisse und ermittelten Lücken präsentiert und mit den Zielgruppen der Befragung (Forschern, Verlegern, »Datenbewahrern« und Forschungsförderern) diskutiert. Allen interessierten Akteuren wurde so Gelegenheit gegeben, ihre Wünsche und Bedürfnisse an eine Forschungsdaten-Infrastruktur zu äußern. Die so gewonnenen Erkenntnisse fließen in ein finales Empfehlungspapier für eine Forschungsdaten-Infrastruktur ein, das zum Projekteende der EU-Kommission übergeben wird.

Roadmap-
Entwurf für euro-
päische For-
schungsdaten-
Infrastruktur

Workshop
September
2009

Anmerkungen

1 <<http://earth.esa.int/gscb/ltdp/>>

2 <http://www.cessda.org/sharing/archival_process/3/>

3 <<http://www.alliancepermanentaccess.eu/>>

4 <<http://www.parse-insight.eu/>>