



Mitteilungsheft Frühjahr 2015

**DRV** DEUTSCHER  
ROMANISTEN  
VERBAND

# **Ergebnisse der Umfrage der AG Digitale Romanistik zur Langzeitarchivierung von digitalen Forschungsdaten für die Romanistik**

## *1. Einleitung*

Editionen, Korpora der gesprochenen und geschriebenen Sprache, Bildarchive: Datenbasierte Forschung wird in allen romanistischen Teildisziplinen zunehmend üblich. Aus zahlreichen Projekten gehen Daten hervor, ohne dass jedoch deren

langfristige Speicherung (Langzeitarchivierung, LZA) und Nutzungsmöglichkeit sichergestellt würde. Auch fehlen Standards, die bereits bei der Planung von Forschungsprojekten die spätere Publikation von Daten berücksichtigen. Die Arbeitsgruppe „Digitale Romanistik“ widmet sich derzeit diesem Thema. Erster Schritt war eine Umfrage zur Bedarfserhebung, auch um den Austausch zwischen DFG, den Fachverbänden und den Universitäten zu erleichtern. Wir danken allen Kolleginnen und Kollegen, die an der Umfrage teilgenommen haben.

Ein aktueller Anlass ist die Neuausrichtung der Informationsversorgung für die Romanistik: die DFG hat die Förderung von Sondersammelgebieten (SSG) eingestellt, zu Gunsten von neu zu beantragenden sogenannten „Fachinformationsdiensten für die Wissenschaft“ (FID). Die AG Digitale Romanistik unterstreicht in diesem Zusammenhang auf Basis der vorliegenden Umfrage die Notwendigkeit, die Frage des langfristigen Zugriffs auf Forschungsdaten mit der allgemeinen Publikationsversorgung zu verbinden.<sup>1</sup> Im Folgenden werden die wichtigsten Ergebnisse in Kürze zusammengefasst. Eine ausführlichere Darstellung – insbesondere eine grafische Aufbereitung – sowie weiterführende Informationen zum Thema können auf der Webseite der AG Digitale Romanistik konsultiert werden – <http://www.deutscher-romanistenverband.de/der-drv/ag-digitale-romanistik/>.

## 2. Zentrale Ergebnisse

An der Umfrage nahmen 72 KollegInnen mit vollständig ausgefüllten Fragebögen teil, so dass aussagekräftige Ergebnisse erzielt werden konnten. Positiv ist zu vermerken, dass alle Fachgebiete vertreten sind und dass sich VertreterInnen aus Linguistik und Literaturwissenschaft in etwa gleich großen Anteilen beteiligt haben. In vielen Bereichen zeigt sich Übereinstimmung unter den einzelnen romanistischen Fachgebieten. Nur in einzelnen Punkten sind unterschiedliche Auffassungen zu erkennen. Alle Statusgruppen sind in der Umfrage gut vertreten. Die insgesamt geringe Zahl der Antworten deutet darauf hin, dass das Bewusstsein für die Bedeutung des Themas innerhalb der Romanistik noch nicht weit verbreitet ist. Viele KollegInnen fühlten sich offenbar nicht angesprochen. Die Ergebnisse der Umfrage sind insofern repräsentativ nur für die – überschaubare – Gruppe derjenigen KollegInnen, die für die anstehenden Probleme bereits sensibilisiert sind.

Die Relevanz der Langzeitarchivierung von digitalen Forschungsdaten (im Folgenden kurz als LZA bezeichnet) in verschiedenen Fachgebieten wird unterschiedlich eingeschätzt. Die überwiegende Mehrheit aller Befragten findet das Thema für die Linguistik wichtig oder sehr wichtig (82%) und eine große Mehrheit sieht dies auch für die Kulturwissenschaften und Literaturwissenschaften so (72% und 71%). Gleichzeitig wird das Thema nur von knapp der Hälfte der Befragten auch für die Fachdidaktik als wichtig oder sehr wichtig erachtet (49%). Die Frage wurde von allen Befragten für alle – nicht nur für die eigenen – Fachgebiete beantwortet, d.h. es drückt sich hier eventuell Unsicherheit darüber aus, welche Rolle die LZA in anderen als dem eigenen Fachgebiet haben mag. Deutlich wird die große Bedeutung der LZA aber auch in der Frage nach der persönlichen Bedeutung des Themas für die Befragten. Schon jetzt empfinden 62% der Befragten LZA für sich selbst wichtig oder sehr wichtig, für die Zukunft

---

<sup>1</sup> Allgemeine Informationen zum FID für die Romanistik sind auf den Seiten der ULB Bonn einsehbar – <http://www.ulb.uni-bonn.de/fid-blog/>

rechnen sogar 83% der Befragten damit, dass das Thema wichtig oder sehr wichtig sein wird. Im Kontrast hierzu steht allerdings der Befund, dass das Thema LZA insgesamt eher selten im Kreis der KollegInnen diskutiert wird (regelmäßig oder häufig nur 26%, gar nicht oder fast gar nicht dagegen 42%). Das Thema ist demnach noch kein etablierter Gesprächsgegenstand, sondern offenbar eher etwas, das bislang noch den Spezialisten oder technisch Versierten überlassen bleibt.

Im Bereich der derzeitigen Nutzung von Forschungsdaten zeigt sich zunächst, dass Forschungsdaten erwartungsgemäß häufiger genutzt als selbst produziert werden. Etwas unterdurchschnittlich ausgeprägt ist dies bei Korpora (fast jede/r Zweite produziert diese auch), sehr stark ausgeprägt bei Katalogen (nur jede/r Zehnte produziert sie auch). Dennoch ist festzuhalten, dass immerhin 22% der Befragten selbst digitale Forschungsdaten produzieren. Für eine Disziplin wie die Romanistik, die keinen „nationalen Editions-auftrag“ hat, ist das nicht wenig. Zudem geben rund 71% der Befragten an, digitale Forschungsdaten zu nutzen. Bemerkenswert ist auch, dass dieser Anteil bei den (linguistischen) Korpora und den (literaturwissenschaftlichen) Textsammlungen fast gleich hoch ist. Für die Speicherung der eigenen Daten nutzen fast alle Befragten die eigene Festplatte oder USB-Stick, rund die Hälfte nutzen kommerzielle Anbieter oder lokale universitäre Speicherdienste, fast niemand nutzt Speicherangebote bei nicht-lokalen Einrichtungen. Unter den häufig genutzten Arten von Daten ist überraschend, dass insbesondere Bilder von sehr vielen Befragten häufig oder sehr häufig genutzt werden (59%). Auch digitale Sprachdaten, Daten in Tabellenform, strukturierte Textdaten und Videodaten werden genutzt, jedoch deutlich weniger häufig (36%, 35%, 29% resp. 27% der Befragten nutzen diese Datentypen häufig oder sehr häufig). An und für sich stark genutzte Textdaten werden offenbar nicht für die programmgestützte Analyse strukturiert aufbereitet, sondern ohne Rückgriff auf Auszeichnungssprachen in klassischen Textverarbeitungsformaten gespeichert (\*.docx oder \*.txt).

Ebenfalls erfragt wurden Einschätzungen zu den Anforderungen an ein Repository für Forschungsdaten der Romanistik. Es wurde ein breites Spektrum an Anforderungen angeboten, von denen alle von gut 60% der Befragten für wichtig oder sehr wichtig erachtet werden. Fast einstimmig (über 95%) werden die spätere Nutzbarkeit und Zitierbarkeit der Daten, die Nutzerfreundlichkeit der Plattform und die zuverlässige Speicherung der Daten für wichtig oder sehr wichtig erachtet. Auch Aspekte wie die Sicherstellung der langfristigen Finanzierung des Angebots, der Nachweis der Einträge in Katalogen und die Kontextualisierung der Daten finden über 80% der Befragten wichtig oder sehr wichtig. Die drei Themen mit den geringsten Werten sind Beratung und Schulung zur Nutzung des Repositoriums, geeignete Lizenzmodelle für die Daten und differenzierte Zugriffsrechte auf die Daten. Aus Sicht der AG Digitale Romanistik handelt es sich hier allerdings eher um unterschätzte Themenfelder, denn genau diese Bereiche erweisen sich erfahrungsgemäß bei der konkreten Nutzung von Angeboten als besonders herausforderungsreich (und diese Aspekte sind im Freitextteil der Umfrage prominent vertreten, siehe unten). Neben den Anforderungen wurde auch erfragt, was Repositorien für die Forschung leisten. Hier sind keine überraschenden Ergebnisse zu verzeichnen. Als positive Leistungen eines Forschungsdatenrepositoriums werden insbesondere der langfristige Zugriff auf die Daten, die Nachvollziehbarkeit von Forschungsergebnissen, die erhöhte Sichtbarkeit der Daten bzw. der eigenen Forschung und die Möglichkeit der erneuten Auswertung der Daten durch andere Forscher/innen genannt (zwischen 80% und 96% mei-

nen, dass Forschungsdatenrepositorien diese Dinge leisten). Erfreulich ist, dass die Risiken offenbar erkannt, aber nicht überbewertet werden (bspw. Mehraufwand für die Aufbereitung der Daten) und auch keine allzu optimistische Erwartungen bestehen (bspw. Hoffnung auf reduzierte Kosten).

Der geeignete Zuschnitt eines Forschungsdatenrepositoriums für die Romanistik wird kontrovers beurteilt. Insgesamt lässt sich jedoch feststellen, dass die Befragten tendenziell eine mittelgroße Lösung bevorzugen: So halten 59% der Befragten einen disziplinären Zuschnitt für geeignet oder sehr geeignet, der ein Angebot für alle Geisteswissenschaften darstellt. Eine Lösung nur die Romanistik als Ganzes oder für einzelne Teilbereiche der Romanistik erscheint aber auch noch 54% bzw. 49% der Befragten als geeignet oder sehr geeignet (allerdings mit deutlich geringerem Gewicht der „sehr geeignet“-Anteile). Noch größere, beispielsweise verbundbasierte, deutschlandweite oder europaweite Lösungen werden weniger favorisiert. Allerdings ist gerade bei dieser letzten Gruppe von Antwortmöglichkeiten auch der Anteil der Befragten besonders hoch, die sich hier keine Einschätzung zutrauen (rund 25%).

Die Umfrage gab abschließend Gelegenheit, weitere Anmerkungen zu verschiedenen Aspekten im Freitext zu formulieren. Mit Blick auf den Bedarf für Beratung und Information im Kontext von Forschungsdatenrepositorien wurden mehrere Themen wiederholt angesprochen, darunter insbesondere die technischen Möglichkeiten und ihre Umsetzung (sowohl bzgl. Standards, Formaten, Werkzeugen bei der Erstellung eigener Daten, als auch bzgl. der Nutzung der Daten Dritter) sowie rechtliche Aspekte (u.a. Urheberrecht, Nutzungsverträge, Datenschutz). Mit Blick auf die möglichen Hindernisse für die erfolgreiche Etablierung von Forschungsdatenrepositorien wurde eine noch größere Bandbreite an Aspekten genannt, wovon die folgenden herausstechen: einerseits der (dauerhafte) technische, zeitliche und finanzielle Aufwand für den Betrieb eines solchen Repositoriums und andererseits die notwendige Koordination und Kooperationsbereitschaft, aber auch die – legitimer Weise – breit divergierenden Forschungsinteressen innerhalb der Romanistik, bei gleichzeitig hohem (inhaltlichem, technischem und organisatorischem) Abstimmungsbedarf für ein solches Unterfangen. Unter den sonstigen Anmerkungen fallen mehrere Punkte ins Auge: zum einen die Aufforderung an die Bibliotheken, sich als Serviceeinrichtung auch für digital vorliegende Daten zu sehen, zum anderen der Wunsch, vielfältige Bestände aus dem Bereich des kulturellen Erbes (nicht nur Texte, sondern auch Bilder, Videos, Tonaufnahmen, etc.) in digitaler Form nutzen zu können. Zugleich wird hier auch deutlich, dass die in der Umfrage verwendeten Begriffe „Forschungsdaten“ bzw. „Primärdaten“ von einigen Befragten enger ausgelegt wurden, als in der Umfrage beabsichtigt. Dies weist im Hinblick auf das Thema nicht zuletzt auf den Bedarf geeigneter Kommunikationsformen und Terminologien hin. Zukünftig muss noch deutlicher werden, dass es hier um jede Form digital vorliegender Daten aus dem Bereich der romanistischen Forschung geht.

### 3. Schlussfolgerungen

Aus den Ergebnissen der Umfrage leiten sich aus Sicht der AG Digitale Romanistik mehrere Feststellungen und Schlussfolgerungen ab.

- Texteditionen, Korpora und andere digital vorliegende Forschungsdaten sind bereits heute intensiv und auf vielfältige Weise in Gebrauch.
- Formate der strukturierten Textdatenaufbereitung, die eine Überführung der Ressourcen in digitale (LZA-)Infrastrukturen ermöglichen, werden derzeit noch wenig genutzt.

- Eine große Mehrheit sieht Handlungsbedarf für den Aufbau von Speichermöglichkeiten, nicht zuletzt da abgestimmte und tragfähige Konzepte erst noch entwickelt werden müssen.
- Als Ort für die LZA werden eine disziplinäre (gesamtgeisteswissenschaftliche) oder eine fachliche (romanistische) Lösung präferiert, weniger eine globale (alle Disziplinen übergreifende).
- Der Wissensstand ist, den Umfrageergebnissen zufolge, zu niedrig, um die Forschenden in wünschenswerter Weise in den Entwicklungsprozess der LZA einzubeziehen. Es besteht noch beachtlicher Informationsbedarf.

Die Arbeitsgruppe Digitale Romanistik wird die hier zusammengefassten Umfrageergebnisse und Schlussfolgerungen in die laufenden und zukünftigen Diskussionen im Bereich der Datensicherung und Informationsversorgung einbringen.

AG Digitale Romanistik