

User-Experience von Spracharchiven: Eine Neubewertung der Interaktion von Archiv und Nutzern.

Rau, Felix

f.rau@uni-koeln.de
Universität zu Köln, Deutschland

Blumtritt, Jonathan

jonathan.blumtritt@uni-koeln.de
Universität zu Köln, Deutschland

In den letzten 15 Jahren hat sich weltweit eine aktive und diverse Landschaft digitaler Spracharchive entwickelt. Von diesen Archiven wird eine kontinuierlich wachsende Menge an digitalen Audio- und Videodaten gespeichert und zugänglich gemacht. Dabei wird immer deutlicher, dass die Nachnutzung der in den Spracharchiven archivierten Forschungsdaten noch hinter den Erwartungen zurück bleibt und im Moment die vielleicht größte Herausforderung für diese Institutionen darstellt. Eine geringe Nachnutzung von Forschungsdaten ist allerdings ein Problem, das nicht nur digitale Spracharchive betrifft.

In unserem Beitrag präsentieren wir unsere Erfahrungen aus Betrieb und Ausgestaltung eines kürzlich neugegründeten Spracharchivs sowie Ergebnisse unserer Untersuchung der User Experience von Spracharchiven und diskutieren den Einfluss der Konzeption und Ausgestaltung der Nutzererfahrung auf die Nachnutzung der Forschungsdaten. Unsere Präsentation fokussiert auf die konzeptuelle Modellierung des Prozesses der Archivierung, Annotation und Nutzung von audiovisuellen Sprachdaten. Die Probleme und Lösungen sind aus unserer Sicht aber ebenso relevant für andere Services und Plattformen in digitalen Forschungsinfrastrukturen.

Digitale Archive und Forschungsdatenzentren sind nicht zuletzt auch Einrichtungen, in denen der Gegenwert aufwändiger Forschungsförderung gesichert und vorgehalten wird. Nationale und internationale Förderer finanzieren seit Jahren direkt oder indirekt in allen fachgebundenen Förderlinien die Erhebung von Forschungsdaten. Gleichzeitig werden Millionen in den Aufbau von Kompetenzzentren, Datenzentren und Forschungsinfrastrukturen für die Geisteswissenschaften investiert. Mit dieser Förderung ist auch die Hoffnung verbunden, dass Datenarchivierung nicht nur Vorhaltung für die Nachwelt leistet, sondern die Bereitstellung von Forschungsdaten mittel- und kurzfristig vielfältige positive Effekte entfaltet. Explizites Ziel ist es unnötige

Redundanz bei der Datenerhebung zu verhindern, den Austausch unter den Forschern zu fördern und zu beschleunigen sowie Input für neue Forschungsfragen, Methoden und Verfahren zu schaffen. Jüngst hat die DFG mit der Förderlinie "Forschungsdaten in der Praxis" eine Ausschreibung veröffentlicht, die gezielt dazu anregen soll, "Forschungsfragen überwiegend durch eine Sekundär- bzw. Nachnutzung verfügbarer Forschungsdaten zu bearbeiten". Nach einer initialen Förderzeit fragen die Geldgeber damit nun vermehrt nach dem „return on investment“, wobei die Dynamik, die aus der Verfügbarmachung von Forschungsdaten erwachsen sollte, in vielen Fällen partiell hinter den Erwartungen zurückbleibt.

Digitale Spracharchive sind Teil dieser Landschaft an Forschungsdatenzentren und haben das spezifische Ziel audio-visuelle Sprachdaten und Dokumente zu sichern und auf dieser Basis Wissensgenerierung zu ermöglichen und zu unterstützen. Ein Spracharchiv ist in diesem Sinn eine Plattform, die zwischen Produzenten und Konsumenten von Primärdaten vermittelt, so dass diese direkt oder indirekt interagieren können. Den datenproduzierenden Forschern ermöglicht das Archiv Audio- und Videoaufnahmen menschlicher Kommunikation zu archivieren und idealerweise webbasiert zugänglich zu machen. Auf der anderen Seite werden Forscher, Sprachgemeinschaften und die weitere Öffentlichkeit in die Lage versetzt, diese Daten aufzufinden, zu betrachten, herunterzuladen und weiterzuverwenden und auf dieser Grundlage neues Wissen zu generieren. Um diesen Austausch zu unterstützen, haben die verschiedenen Spracharchive komplexe Webplattformen entwickelt.

Spracharchive in den verschiedenen Ländern sind aus unterschiedlichen Organisationen hervorgegangen und haben sich in sehr diversen Kontexten entwickelt. Einige digitale Spracharchive sind aus vor-digitalen Archiven entstanden, während andere Archive seit ihrer Gründung ausschließlich mit digitalen Daten umgehen. Darüberhinaus sind manche Archive relativ alleinstehende Institutionen, während andere Teil größerer Institutionen sind. Diese Einbettung in größere Netzwerke hat wiederum Einfluss auf die Ausgestaltung der Archive. In den letzten Jahren kam dazu noch eine Integration verschiedener Archive in nationale und übernationale Forschungsinfrastrukturen wie CLARIN-D hinzu. Dieser Prozess hatte Einfluss auf so diverse Aspekte wie Metadatenformate oder die Implementation von Webservices.

Unsere Erfahrungen aus Betrieb und Ausgestaltung eines kürzlich neugegründeten Spracharchivs legen nahe, die Interaktion zwischen Spracharchiven und ihren Nutzern neu zu überdenken und weiterzuentwickeln. Die vorherrschenden Konzepte der Interaktionsgestaltung sind das Produkt einer 15-jährigen Entwicklung und der Förderungskontexte, in denen diese Archive gewachsen sind. Unsere Präsentation nimmt die Webplattformen der verschiedenen Spracharchive als Ausgangspunkt

unserer Diskussion. Die unterschiedliche Gestaltung der User Interfaces, Schnittstellen und Funktionalitäten ist der konkrete Ausdruck unterschiedlicher Konzepte und Schwerpunkte. Letzendlich spiegeln die teilweise sehr unterschiedlichen Ausgestaltungen der Webplattformen unterschiedliche Annahmen über die Bedürfnisse der Forscher und Forscherinnen und die Funktion von Archiven im Forschungsprozess wieder. Die Diversität der Plattformen demonstriert somit große Unterschiede in den oft impliziten Annahmen, die wir diskutieren werden.

Im Rahmen der Planung eines Zentrums für audio-visuelle Daten haben wir nun begonnen die User-Experience-Strategie und das UX-Designs unseres Spracharchivs grundlegend zu überarbeiten. Dafür haben wir das User-Interface und die User-Experience sowohl unseres Web-Front-Ends, als auch der Benutzeroberflächen anderer Spracharchive analysiert. Darüber hinaus haben wir Interviews mit Nutzern unseres Archivs sowie potentiellen neuen Nutzern durchgeführt. In diesem Vortrag berichten wir vom Prozess und den Ergebnissen dieser Arbeit. Damit geben wir sowohl direkten Einblick in die Planung eines Zentrums für audio-visuelle Daten, als auch in die Probleme und Herausforderungen, die sich uns gestellt haben.

Unsere Untersuchungen zeigen, dass die Konzeption der Interaktion bestehender digitaler Spracharchive durch die Sichtweise der Archivbetreiber bestimmt wird. Interessen der Archivnutzer werden oft erst nachträglich berücksichtigt. Dabei dominiert durchgehend die Perspektive der datenproduzierenden Forscher, während sich die Interessen der Datenkonsumenten in der Struktur der Webportale kaum niederschlagen. Auf der anderen Seite zeigen unsere Interviews und die praktische Erfahrung aus dem Support von Archivnutzern, dass Produzenten und Konsumenten fundamental unterschiedliche Bedürfnisse haben und damit auch die Interaktion mit dem Archiv grundsätzlich anders konzeptualisieren.

Das entscheidende Ergebnis unserer Untersuchung ist somit, dass es keine allgemeine Nutzerrolle in Bezug auf Spracharchive gibt. Vielmehr muss der fundamentale Unterschied zwischen den Interessen und Erwartungen von Datenproduzenten und Datenkonsumenten anerkannt werden. Datenproduzenten möchten die Ergebnisse ihrer Arbeit präsentieren und den Zugang zu den von ihnen archivierten Daten kontrollieren. Letzendlich ist es das Ziel des Produzenten wissenschaftliche Anerkennung für die Erhebung und Kuration der Daten zu erhalten. Die Datenkonsumenten streben Zugang zu einem Datensatz an, der das Potenzial hat ihnen Informationen zu liefern, um eine bestimmte Fragestellung zu bearbeiten. Dies bedarf verlässlicher Findmechanismen, die die Identifikation entsprechender Datenmengen erlauben. Darüberhinaus müssen Datensätze möglichst offen zugänglich sein. Falls dies aus rechtlichen oder ethischen Gründen nicht möglich ist, müssen die Prozesse um Zugang zu erlangen gut dokumentiert, transparent und klar geregelt sein. Aus Sicht

des Konsumenten stellen Zugangsbeschränkungen Hürden dar, die einen Datensatz weniger attraktiv machen.

Produzenten und Konsumenten stellen somit grundverschiedene User-Typen dar, deren Interaktion mit dem Archiv durch unterschiedliche Bedürfnisse und Erwartungen geleitet wird.

Die Nichtberücksichtigung der Konsumenten-Perspektive beim UX-Design führt aus unserer Sicht zu einer geringeren Rezeption der archivierten Daten und stellt damit einen der Hauptgründe für eine fehlende Nachnutzung dar. Dadurch wird wiederum der niedrigen Stellenwert von Datenpublikationen in der Sprachforschung verfestigt. Eine stärkere Berücksichtigung der User-Experience von Datenkonsumenten scheint uns wesentlich für eine erfolgreiche Weiterentwicklung digitaler Spracharchive zu sein. Verbesserungen in der Auffindbarkeit und Zugänglichkeit von Daten und bei der Referenzierung und Zitierung von Datensätzen werden zu einer größeren Relevanz von Primärdaten in der wissenschaftlichen Praxis führen. Darüber hinaus ermöglicht eine Überarbeitung der User-Experience ein Öffnung der Archive zu einer breiteren Öffentlichkeit. Letzendlich müssen Spracharchive sich von Institutionen der reinen Kuration hin zu partizipatorischen Plattformen, die den direkten und einfachen Austausch zwischen Datenproduzenten und Datenkonsumenten ermöglichen, entwickeln.

Bibliographie

Austin, Peter K. (2011): „Who uses digital language archives?“, in: *Endangered Languages and Cultures (Blog)* <http://www.paradisec.org.au/blog/2011/04/who-uses-digital-language-archives/> [letzter Zugriff 13. Oktober 2015].

Beagrie, Neil / Lavoie, Brian / Woollard, Matthew (2010): *Keeping Research Data Safe 2. Final Report* <http://www.webarchive.org.uk/wayback/archive/20140615221405/http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf> [letzter Zugriff 12. Oktober 2015].

Bow, Catherine / Christie, Michael / Devlin, Brian (2014): „Developing a Living Archive of Aboriginal Languages“, in: *Language Documentation & Conservation* 8: 345-360.

DFG (2015): *Ausschreibung: Forschungsdaten in der Praxis*. Deutsche Forschungsgemeinschaft e.V. Bonn http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/ausschreibung_forschungsdaten/index.html [letzter Zugriff 14. Oktober 2015].

Holton, Gary (2012): „Language archives: They're not just for linguists any more“, in: Seifart, Frank / Haig, Geoffrey / Himmelmann, Nikolaus P. / Jung, Dagmar / Margetts, Anna / Trilsbeek, Paul (eds.): *Potentials*

of Language Documentation: Methods, Analyses, and Utilization. Honolulu: University of Hawai'i Press 111-117.

Nathan, David / Austin, Peter K. (2014) *Language Documentation and Description, 12: Special Issue on Language Documentation and Archiving*. London: SOAS 4-16.

Schwartz, Gabriele (2012): „Online presentation and accessibility of endangered languages data: The General Portal to the DoBeS Archive“, in: Seifart, Frank / Haig, Geoffrey / Himmelmann, Nikolaus P. / Jung, Dagmar / Margetts, Anna / Trilsbeek, Paul (eds.): *Potentials of Language Documentation: Methods, Analyses, and Utilization*. Honolulu: University of Hawai'i Press 126–128.

Trilsbeek, Paul / König, Alexander (2014): „Increasing the future usage of endangered language archives“, in: Nathan, David / Austin, Peter K. (eds.): *Language Documentation and Description 12: Special Issue on Language Documentation and Archiving*. London: SOAS 151-163.