

Multimediale Digitale Archive und Oral History

Multimedia Digital Archives and Oral History

Prof. Nicholas Apostolopoulos, Jan Rietema, Wolfram Lippert
Center für Digitale Systeme (CeDiS)
Freie Universität Berlin
Hnestr. 24, 14195 Berlin
Tel.: 030/838-55735, Fax: 030/838-52843
E-Mail: jan.rietema@fu-berlin.de, Internet: www.cedis.fu-berlin.de
Archiv-Portal: www.zwangsarbeit-archiv.de

Zusammenfassung:

Das Center für Digitale Systeme der Freien Universität Berlin stellt mehrere Sammlungen von aufgezeichneten Interviews mit Zeitzeugen Nationalsozialistischer Verfolgung und Zwangsarbeit digital für ein breites Spektrum an Nutzern bereit. Für das im Auftrag der Stiftung Erinnerung, Verantwortung und Zukunft entwickelte multimediale Archiv „Zwangsarbeit 1939-1945. Erinnerungen und Geschichte“ wurde eine eigene Methodik in der Verbindung von Transkription und deren Übersetzung und den AV-Medien umgesetzt. Diese ermöglicht eine nonlineare Navigation innerhalb der Interviews über Volltextsuche sowie eine isochrone Darstellung der Transkription zum AV-Medium. Das Online-Archiv konnte in seiner ersten Version im Januar 2009 im Web einem breiten Publikum bereitgestellt werden und wird seitdem kontinuierlich weiter entwickelt.

Abstract:

The Center for Digital Systems at the Freie Universität Berlin provides digital access for a broad range of users to a number of collections of recorded interviews with witnesses of national socialist persecution and forced labor. For the Foundation Remembrance, Responsibility and Future the multimedia archive "Forced Labor 1939-1945. Memory and History" was developed using a unique methodology for the combined use of transcript and the translation thereof and the audiovisual media. This allows nonlinear navigation within the interviews via fulltext searching and an isochronous display of transcription during playback of the audiovisual media. The online archive was launched in a first version on the web in January 2009 to a broad audience and is in continuous development since.

Ausgangslage

Im Rahmen des Projekts "Dokumentation lebensgeschichtlicher Interviews mit ehemaligen Sklaven- und Zwangsarbeitern" entstand 2005 bis 2006 eine Sammlung von 590 Zeitzeugen-Erinnerungen, in denen Überlebende der nationalsozialistischen Zwangsarbeit ihre Lebensgeschichte in ausführlichen Audio- und Video-Interviews erzählten. Historisch versierte und persönlich engagierte Interview-Teams von 32 Institutionen in 26 Ländern führten die Gespräche und erarbeiteten die dazugehörigen Materialien. Die Koordination des von der Stiftung "Erinnerung, Verantwortung und Zukunft" finanzierten Kooperations-Projekts lag beim Institut für Geschichte und Biographie der FernUniversität Hagen¹.

¹ Vgl. Hitlers Sklaven: Lebensgeschichtliche Analysen zur Zwangsarbeit im internationalen Vergleich, hrsg. von Alexander von Plato, Almut Leh und Christoph Thonfeld, Wien: Böhlau 2008 (ISBN-13: 978-3-205-77753-3)

Ende 2007 begann die Aufbereitung der Sammlung für die Bereitstellung im multimedialen Online-Archiv "Zwangsarbeit 1939-1945" an der Freien Universität Berlin, während die Archivierung der analogen Originale und Begleitmaterialien dem Deutschen Historischen Museum obliegt. Das Center für Digitale Systeme digitalisierte die 190 Video-Interviews und 393 Audio-Interviews mit einer durchschnittlichen Länge von 3 h 15 min in ein Digitalisat als Referenz zur Bestandsicherung und in streamingfähige Formate (MPEG4, FLV, MP3).

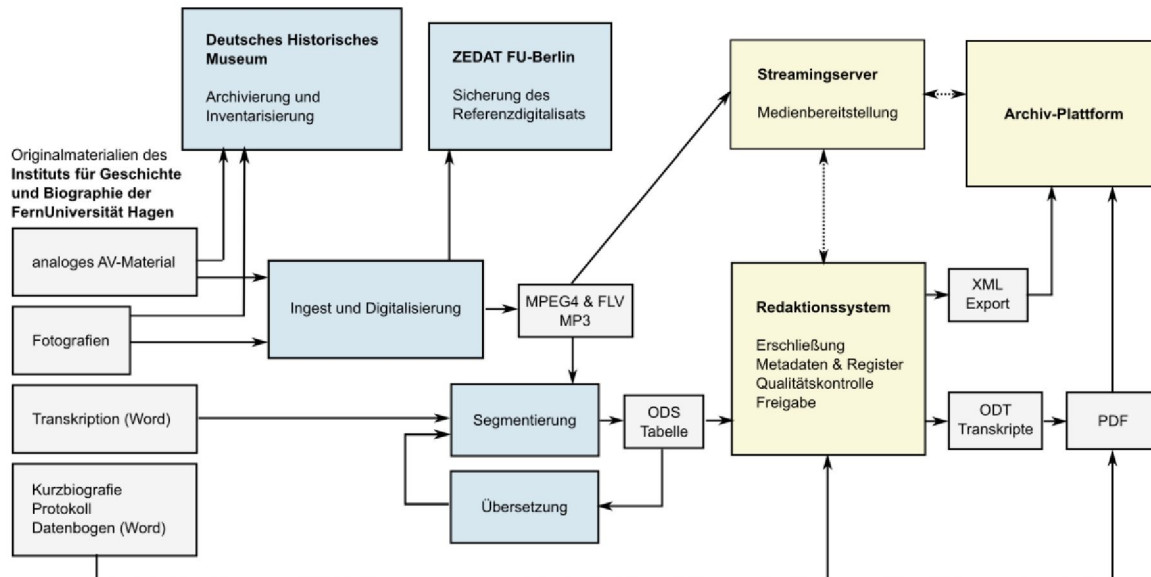


Abb.1 Arbeitsabläufe von den Originalmaterialien zur Online-Bereitstellung im Archiv

Die vorhandenen Excel-Metadaten aus dem Projekt des Instituts für Geschichte und Biographie der FernUniversität Hagen wurden in eine MySQL Datenbank importiert, die als Grundlage für eine kollaborative webbasierte Redaktionsumgebung dient. Diese Umgebung stellte im weiteren Verlauf des Projekts die Arbeitsschritte der Datenerfassung und -normalisierung, der chronologischen und biographischen Erschließung der Interviews sowie des Qualitätsmanagements in einer Webumgebung zur Verfügung. Die Programmierung dieser Webumgebung sowie der eigentlichen Archiv-Plattform wurde mittels des Webframeworks Ruby on Rails realisiert. Die Suchmaschine Solr wird in der Archiv-Plattform für alle Rechercheprozesse eingesetzt.

Transkript und Audiovisuelles Medium

Die vorhandenen Transkriptionen wurden mittels der Transkriptions- und Untertitelungssoftware InqScribe basierend auf Timecodes in ein tabellarisches Textformat übertragen. Diese wurden in OpenOffice importiert und im ODS-Format abgelegt. Dieses Format diente als Austauschformat für alle Qualitäts- und Übersetzungsprozesse und der Import der Redaktionsplattform wurde aus dem Grund auch auf ODS ausgelegt. OpenOffice wurde hier gewählt weil es eine für die Auftragsnehmer der unterschiedlichen Arbeitsprozesse kostenfreie Software ist und unabhängig von den Benutzereinstellungen mit UTF-8 Kodierung nativ umgehen kann. Der letzte Punkt ist wichtig, da Osteuropäische Sprachen sowie das Hebräische unter den Interviewsprachen eine große Rolle spielen.

Für die Segmentierung der Interviews bietet sich eine Vorgehensweise nach konstanten Zeitintervallen an, wie dies z.B. durch das USC Shoah Foundation Institute bei der Aufbereitung des Materials für das Visual History Archive erfolgt ist (welches ebenfalls an der FU in einem deutschen Portal zur Verfügung steht²). Für das Projekt „Zwangsarbeit 1939-1945“ wurde eine aufwändigere Methode gewählt: die Segmentierung sollte satzbasiert, also nach sprachlichen

² Die Sammlung ist nur von der Freien Universität und Partnerinstitutionen aus einsehbar. Das Portal findet sich im WWW unter <http://vha.fu-berlin.de>

Einheiten erfolgen, woraus sich variable Längen der einzelnen Segmente ergeben. Diese Vorgehensweise wurde gewählt, um eine möglichst genaue Ansteuerung im Interview sowie eine nutzerfreundliche, der Untertitelung nahe kommende, isochrone Darstellung der Transkription und deren Übersetzung zu ermöglichen.

In der Praxis ist die Zeichenlänge der satzbasierten Transkriptsegmente hoch variabel. Um eine vereinheitlichte, dem User-Interface angepasste Darstellung der Transkriptsegmente zu gewährleisten, wurde ein nachgelagerter, automatisierter Prozess eingeführt, der die Transkriptsegmente und ihre Übersetzung in Untertitelsegmente überführt. Die Aneinanderreihung oder Aufspaltung der Transkriptsegmente wird regelbasiert vorgenommen unter Berücksichtigung:

- der Zeichenlänge
- der Interpunktion (Kommata, Semikolon, Gedankenstrich)
- der Sprecherwechsel
- der durch die inhaltliche Erschließung gesetzten Kapitelüberschriften

Bei einer Aufspaltung eines Transkriptsegments entsteht ein neuer Timecode, der durch Interpolation auf der Basis der Zeichenlängen-Verhältnisse der Transkriptteile ermittelt wird.

Crossmediale Potentiale

Für die Wiedergabe der Interviews wurde der JW-Player ausgewählt, da er quelloffen entwickelt wird und zahlreiche Schnittstellen für die Interaktion mit Javascript bietet. Es wurde ein Modul für diesen Player entwickelt, dass aus XML-Untertiteldaten eine Darstellung des Transkriptsegments zum jeweiligen Zeitpunkt als HTML ermöglicht. Damit können größere Textabschnitte dargestellt werden als dies bei einer Untertitelung innerhalb des Players möglich wäre, und ohne dass der Interviewpartner überlagert wird. Der Text ist außerdem als HTML dem Nutzer zugänglich z.B. für die Erstellung von Zitaten mittels Copy&Paste. Der Nutzer kann dynamisch zwischen originalsprachlicher und deutscher Transkription wechseln.

The screenshot shows a web application titled "ZWANGSARBEIT 1939-1945 ERINNERUNGEN UND GESCHICHTE". It features a search bar with "Suche im Archiv" and "162 Interviews in der Auswahl". The main content area displays an interview with "Abkowitsch, Iossif Iwanowitsch" (geb. 1922 in Polen). The video player shows a timestamp of 11:58 out of 36:06. Below the video, there is a text snippet: "Uns trieb man ungefähr 100 km, dann setzte man uns in einen Zug und brachte uns nach Buchenwald. In Buchenwald waren wir ungefähr einen Monat lang, Januar-Februar." To the right, there is a search bar for "Suchen in 'Abkowitsch, I.'" and a table of contents with chapters 4.2 through 7. At the bottom, there are navigation links: "Stand:05.04.2011", "Nutzungsbedingungen", "Team", "Kontakt", and "Impressum".

Abb. 2: Isochrone Darstellung von Transkript (und Übersetzung) zum AV-Medium, sowie Navigation über Kapitelüberschriften (rechts im Bild).

Ist das Interview inhaltlich erschlossen, stehen dem Nutzer Kapitelüberschriften zur Verfügung, die nicht nur eine Zusammenfassung von Inhalt und Chronologie des Interviews bieten, sondern auch navigatorische Einstiegspunkte in die jeweiligen Interviewsequenzen darstellen. All diese Navigationsmöglichkeiten werden realisiert, ohne dass die Interviewmedien nachbearbeitet oder geschnitten werden müssen.

Eine Indizierung der segmentierten Transkriptionen erlaubt eine Volltextsuche über die Transkripte, die einzelne, timecodebasierte Passagen zurückliefert. Diese Transkriptsegmente sind verlinkt mit einer Ansicht des Interviews, in der die Wiedergabe des Interviews an der ausgewählten Stelle beginnt.

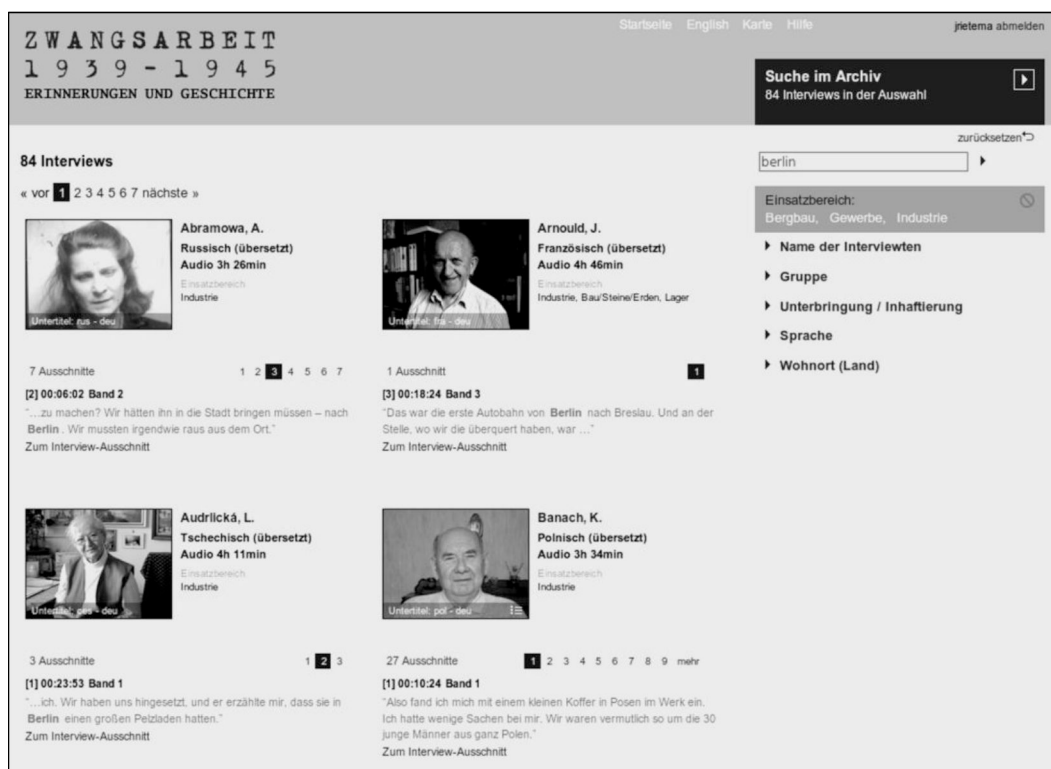


Abb. 3: Die Volltextsuche als Bestandteil der facettierten Suche bietet Navigationsmöglichkeiten an die entsprechenden Stellen im Interview.

Die Volltextsuche ist ein zentraler Bestandteil der Recherche im Archiv. Daneben stehen weitere klassifizierte Facetten zur Verfügung, mit der eine thematische Verfeinerung der Suche möglich ist. Die Volltextsuche fungiert dabei als eine Facette und ist mit diesen beliebig kombinierbar.

Andererseits steht auch in der Interview-Ansicht eine Volltextsuche zur Verfügung, die auf die Transkription dieses Interviews eingegrenzt ist. Diese Suche ermöglicht eine weitere Recherche im Interview, ohne die Wiedergabe des AV-Mediums zu unterbrechen. Dies gibt dem Betrachter die Möglichkeit, aus dem Kontext des gerade Gesehenen weitere Fragestellungen zu formulieren und andere, relevante Interviewsequenzen aufzuspüren.

Metadaten und Ortsindexierung

Das segmentierte Transkript erlaubt es ferner, während der Erschließung der Interviews auf der Ebene der Transkriptsegmente relevante Informationen zu indexieren. So werden für das Projekt „Zwangsarbeit 1939-1945“ die erwähnten Orte, Lager- und Haftstätten, Firmen und Personen vermerkt. Daraus können im späteren Verlauf Register entstehen, die einen strukturierten Zugang zu den Interviews ermöglichen – wie im Fall der Volltextsuche direkt an die entsprechenden Interviewsequenzen.

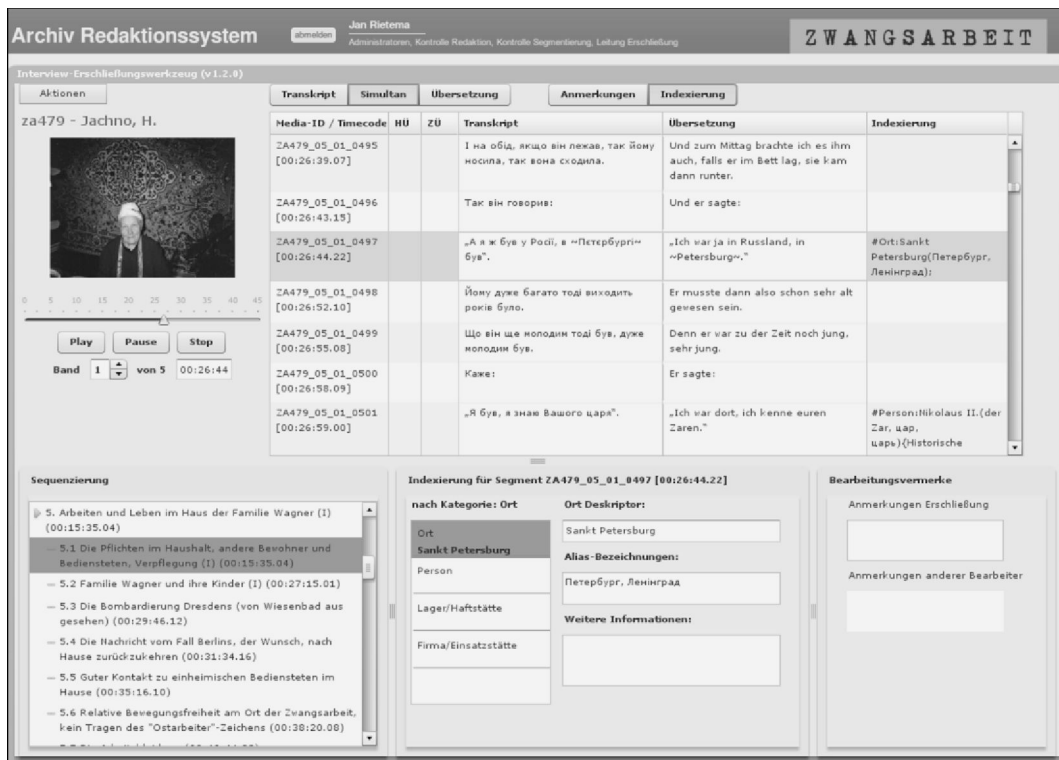


Abb. 4: Die Erschließungssoftware ermöglicht eine kollaborative, webbasierte Bearbeitung in Verbindung von segmentiertem Text und AV-Medien.

Sind die indexierten Begriffe von biographischer Bedeutung für den Interviewten, wie etwa der Ort der Deportation oder eine Firma, für die Zwangsarbeit geleistet wurde, so wird dies gekennzeichnet und diese Informationen gehen ein in die biographischen Metadaten zum Interview. So werden die bestehenden, aus den Teilprojekten erfassten Daten durch eine intensive Überprüfung des Interviews korrigiert und vervollständigt. Eine herausragende Rolle bei den ortsbezogenen Daten spielen etwa die Lager-, Haft- und Einsatzorte, unter denen die unterschiedlichen Stätten der Verfolgung und Zwangsarbeit zusammengefasst werden.

Für die Thematik der Zwangsarbeit ist kein kontrolliertes Vokabular als Schlagwortkatalog verfügbar. Übliche Metadatenstandards wie Dublin Core oder MARC21 bieten keine ausreichende Semantik für die biographischen Ereignisse oder deren Verknüpfung mit Transkriptsegmenten. Deshalb wurde für dieses Vorhaben ein eigenes Datenmodell entworfen, das Angaben unterschiedlicher Genauigkeit zulässt und auf den drei Ebenen Land, Region und Ort Zuordnungen zu einem inkrementell erstellten Vokabular ermöglicht. Dabei werden durch den Erschließer frei wählbare Deskriptoren in einem nachgelagerten Schritt klassifiziert. Die Software für die Registererschließung erleichtert die Zuordnung mit bereits standardisierten Deskriptoren und deren Aliassen. Somit entsteht im Laufe der Erschließung des Bestands ein fachspezifisches Vokabular von Orten, Lager- und Haftstätten und Firmen, die mit der Zwangsarbeit und Verfolgung in Zusammenhang stehen (im Rahmen des Umfangs der Sammlung der Zeitzeugenberichte).

Geolokalisierung und Vernetzung

Neben der Standardisierung und Klassifikation der eingehenden Ortsangaben erfolgt auch eine möglichst umfassende Geolokalisation der Angaben. Nicht alle Angaben sind eindeutig zuzuordnen, so werden häufig kleinere Zwangsarbeiterlager erwähnt, die nicht aktenkundig sind und vom Interviewten nicht näher lokalisiert werden (können).

Georeferenzierte Informationen erlauben eine lose Verknüpfung von Datensätzen, zwischen denen eine eindeutig verifizierbare Verknüpfung nicht möglich wäre. So könnte beispielsweise der heutige Lebensmittelpunkt eines Interviewpartners damals ein Ort der Zwangsarbeit eines anderen

Zeitzeugen gewesen sein. Diese Herangehensweise ermöglicht es der NutzerIn, eigenständig auf Indizien und Spurensuche im Bestand zu gehen und ihre eigenen Schlüsse zu ziehen. Regionalgeschichtliche Projekte können diese Lücke durch eigene Forschung füllen. Besonders im schuldidaktischen Bereich spielt dieser lokalgeschichtliche Ansatz eine bedeutende Rolle und bietet Möglichkeiten für weiterführende eigene Recherchen vor Ort und in lokalen Archiven, sowie der Suche nach lokalen Zeitzeugen.

Georeferenzierte Daten bieten ferner Verknüpfungsmöglichkeiten mit georeferenzierten Daten aus anderen Quellen. Dies wurde in einem Kooperationsprojekt mit dem NS-Zwangsarbeitsportal des Bundesarchivs³ umgesetzt. Über die Schnittstelle eines Webservice werden dort zusätzlich zu den Archivstandorten mit Material zur Zwangsarbeit auch Links zu Zeitzeugenberichten aus dem Zwangsarbeiter-Archiv in die Kartensuche nach Archivbeständen eingefügt.

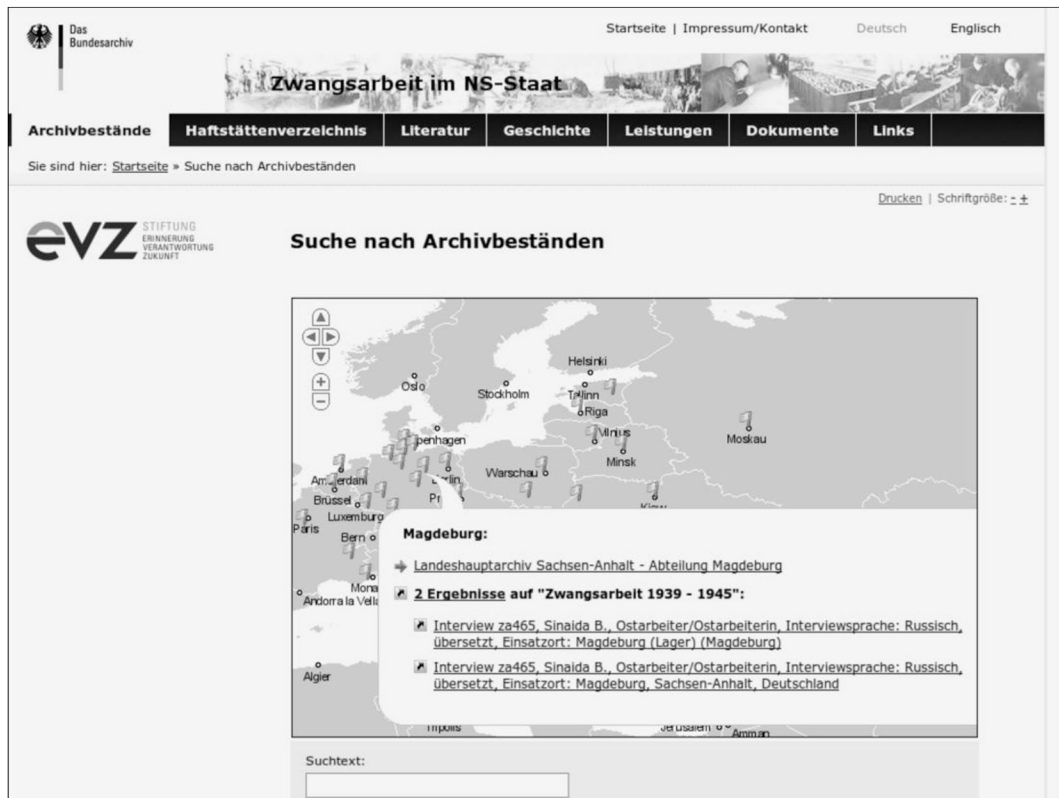


Abb. 5: Verbundsuche der Archivbestände im Portal „Zwangsarbeit im NS-Staat“ des Bundesarchivs.

Ausblick

In dem bis Ende 2011 geförderten Projekt „Zwangsarbeit 1939-1945“ soll die Erschließung von bis zu 400 der 590 Zeitzeugenberichte abgeschlossen werden. Ferner sollen noch weitere Funktionen im Archiv veröffentlicht werden. Eine interaktive Karte wird die Suche nach ortsbezogenen Zeitzeugenberichten ermöglichen und die durch Geolokalisierung entstandenen Querverweise visualisieren. Es wird eine Arbeitsmappe entwickelt, in dem die NutzerIn ihre Recherche-Ergebnisse kommentieren und mit anderen NutzerInnen teilen kann.

Die geplante Fortführung des Archivs nach 2011 sieht auch eine darüber hinaus gehende weitere Entwicklung der Software vor.

³ <http://www.bundesarchiv.de/zwangsarbeit/archiv/index.html>