

## Abschlussbericht Zentral- und Landesbibliothek

### Projekt „DigiTABBerlin – Digitalisierung der historischen Adress- und Telefonbücher Berlin“

#### Ausgangssituation

Die Berliner Telefonbücher (1936-1992) sind wie die Adressbücher (1799-1943) von großem zeitgeschichtlichem und dokumentarischem Wert und bilden eine Einheit. Der Quellenwert von Adress- und Telefonbüchern ist nicht nur von allgemeinem, sondern auch von historischem Interesse für alle Fragen der biographischen und historischen sowie soziologischen Forschung, die besonders mit dem Verbleib der ns-verfolgungsbedingt entzogenen Kulturgüter zu tun hat, zu der sich die ZLB nach den Washington Principles von 1998 verpflichtet hat. Erst wenn Eigentümer anhand der Spuren in den Büchern und Gemälden identifiziert sind, kann sich der nächste Schritt der Suche nach Nachfahren und Erben anschließen.

Für diese Recherchen stellen Adress- und Telefonbücher eine unverzichtbare, wertvolle Quelle dar. Die digitale Nutzung dieser Nachschlagewerke ist nicht nur von großem gesellschaftlichem Interesse, sondern bietet darüber hinaus ein ideales Nachweisinstrument für die Familienforschung. So sind in den Telefonbüchern neben den privaten Fernsprechteilnehmern auch Behörden, öffentliche Einrichtungen und die Berliner Wirtschaft aufgeführt.

Die Telefonbücher stellen ein vergleichbares öffentliches Nachweisinstrument für Adressen dar. Nach 1945 misslang der Versuch, das Berliner Adressbuch wieder zu etablieren. Lediglich von 1957-1967 erschien vereinzelt das Berliner Stadt-Adressbuch für den Westteil der Stadt. Die Berliner Telefonbücher Berlin-Ost und Berlin-West schließen somit eine Informationslücke für Adressdaten vom Ende des 2. Weltkrieges an bis zum heutigen Tage.

In den Telefonbüchern spiegelt sich zu dem Berlins wechselvolle Geschichte der letzten 60 Jahre wieder: Die durch die immense Zerstörung vorherrschende Wohnungsnot in den frühen Nachkriegsjahren, die im Zuge der Teilung der Stadt einsetzenden Fluchtwellen nach Westberlin und auch die Wandlung Berlins zur Bundeshauptstadt infolge der deutschen Wiedervereinigung.

Der Informationsbedarf nach diesen Daten beschränkt sich verständlicherweise nicht nur auf Berliner, sondern hat große Bedeutung für Berlin-Interessierte weltweit. Die große Nachfrage lässt sich an den Zugriffszahlen für die bereits digitalisiert vorliegenden 128 Berliner Adressbüchern (1799-1943) ablesen. (So wurden allein im Jahr 2012 die Seiten [adressbuch.zlb.de](http://adressbuch.zlb.de) ca. 7,8 Mio. Mal abgerufen.) Die 158 Telefonbücher lagen ebenfalls bereits als Scans vor, bei einigen bedurfte es einer Nachbearbeitung und einer OCR-Behandlung, die mit allen Digitalisaten (Scans) durchgeführt wurde. Die Mehrzahl der bereits online gestellten Digitalisate der ZLB werden über den beim KOBV gehosteten OPUS-Server angeboten. Insgesamt sind lediglich knapp 25% der digitalisierten Daten online verfügbar.

### **Adressbücher**

Die Adressbücher (1799-1943) liegen in Fraktur-Schrift vor. Auf sie entfallen 385.000 digitalisierte bildbearbeitete Seiten mit Index-Daten, die noch als Roh-OCR zu bearbeiten sind. Sie werden über den eigenen ZLB-Webserver angeboten und sind mit dem VÖBB verlinkt.

### **Telefonbücher**

Die Telefonbücher liegen in Antiqua vor. Es handelt sich insgesamt um 115.000 Seiten. Folgende Bände der Berliner Telefonbücher liegen als Scans mit entsprechenden Index-Daten vor:

- :: Berliner Telefonbücher (Personennamen)
  - 1936, 1940/41, 1943, 1945, 1948, 1950-1991/92
- :: Berliner Telefonbücher (Branchen)
  - 1940, 1941, 1945/46, 1948, 1950-1991/92
- :: DDR-Telefonbuch (Personennamen)
  - 1951, 1955, 1957, 1959-1961, 1963, 1965, 1967, 1969, 1972, 1973, 1975, 1977, 1979, 1981, 1984, 1986, 1989, 1991/92
- :: DDR-Telefonbuch (Branchen)
  - 1956-1957/58, 1961, 1964, 1966, 1969-1970/71, 1974/75, 1978/79, 1981/82, 1988

### **Projektziele**

Das Ziel des Projekts „DigiTABBerlin“ ist zum einen die zeitgemäße Aufbereitung und Online-Bereitstellung der digitalisierten Berliner Telefon- und Adressbücher, sowie die Anreicherung der Daten mittels einer Crowdsourcing-Funktionalität.

Die Scans lagen bereits in Graustufen als TIFF-Bilddatei vor. Die Bild- und OCR-Bearbeitung war noch im Rahmen des Vorhabens zu leisten.

Das Projekt zur Aufbereitung und Online-Bereitstellung der beiden digitalisierten Quellen sowie das Einrichten einer Crowdsourcing-Funktionalität erfolgte als Pilotprojekt in der ZLB in Zusammenarbeit mit digiS, der Servicestelle Digitalisierung beim Konrad-Zuse-Institut. Die Projektkoordination fand durch die ZLB, Abteilung 1 „Benutzungsdienste“ im Referat 1F „Digitale Dienste“, statt. Die Projektaufgaben werden mit Unterstützung externer Dienstleister umgesetzt.

## Projektergebnisse

### Schritt 1:

Für die Telefonbücher war eine Bildbearbeitung erforderlich. Um ein akzeptables Ergebnis zu erreichen, mussten folgende Arbeiten im Massenverfahren durch Stapelverarbeitung erfolgen: Ausrichten (Geraderücken) der Scanbilder, Freistellen (Entfernen des schwarzen Rahmens), auf einheitliche Bildgröße skalieren. Das Ziel war, eine möglichst gut lesbare Version. Dafür wurden die Scans als Graustufenbilder belassen.

### Schritt 2:

Die Telefonbücher mussten mit einer OCR-Texterkennung versehen werden. Die Adressbücher wurden mit einer Roh-OCR (mehr aufgrund der Frakturschrift und der Scanqualität nicht möglich) besser erschlossen.

### Schritt 3:

Sowohl für die Adress- als auch die Telefonbücher war eine gemeinsame Datenbankstruktur zur Erhöhung der Recherchequalität und eine recherchefreundlichen Web-Oberfläche (mit GOOBI) zu generieren. Die Entscheidung zugunsten von GOOBI erklärt sich daher, dass es sich um eine verbreitete Opensource-Anwendung handelt, die in Berlin auch eine nennenswerte Nutzer-Community hat (Staatsbibliothek und Ibero-Amerikanisches Institut - GOOBI-Presentation mit Workflow von MyBib eDoc; Humboldt Universität und Bibl. für Bildungsgeschichtliche Forschung - GOOBI Workflow und Intrantra Viewer).

Das Repository wird beim Zuse-Zentrum gehostet, wie die ZLB dies bereits für OPUS und ein kleineres, nicht ausbaufähiges Mudora-Repository erfolgreich praktiziert. Das entsprechende Repository hierzu wurde erstellt. Für Adressbücher und Telefonbücher ist der direkte Zugriff auf einzelne Buchabschnitte und auch die alphabetische Ordnung der Personennamen bzw. Branchenbereiche sinnvoll. Die hierfür benötigten Metadaten für diese Funktion lagen für alle gescannten Bände vor. Darüber hinaus haben wir eine Volltextrecherchefunktion eingerichtet. Diese kann durch die Texterkennung (OCR) gewährleistet werden.

### Schritt 4:

Für beide Quellen wurde eine gemeinsame Crowdsourcing-Funktionalität mit zwei Funktionen (Bearbeitung der OCR- im Volltext und Erfassung von Inhalten) installiert, um die OCR-Ergebnisse zu verbessern (<http://digital.zlb.de/viewer/>).

Bei allen Projektschritten wurde auf die Einhaltung von Standards, sowie Vorgaben und Richtlinien seitens der Deutschen Forschungsgemeinschaft und dem Nestor-Kompetenznetzwerk zur Langzeitarchivierung großen Wert gelegt. Die Digitalisate sind als unkomprimierte TIFFs gespeichert, die Metadaten dazu im METS/MODS-Format. Darüber hinaus können alle Scans als PDF heruntergeladen werden.

Die Digitalisate der Adressbücher liegen als bitonale Scans vor, die der Telefonbücher in Graustufen. Da für die OCR Graustufen-Bilder besser geeignet sind, haben wir diese so belassen. Die OCR wurde von einem externen Dienstleister mit der Software *Abby Fine Reader Professional 11* gemacht. Die mit der OCR erstellten Texte sind im XML-Format gespeichert.

Bezüglich der Online-Nutzung sowie der Crowdsourcing-Funktionalität der Adress- und Telefonbücher bestehen keine urheberrechtlichen Bedenken.

## Aussichten

Die ZLB plant mithilfe der eingekauften Software die digitalen Bestände nach und nach nutzerfreundlich in der neuen Präsentationssoftware dem Intraneta Viewer bereit zu stellen. Des Weiteren wurden Grundlagen für künftige Digitalisierungsprojekte geschaffen. Die Crowdsourcing-Funktionalität wird weitergeführt, erweitert und kann für folgende Projekte weitergenutzt werden. Folgeprojekte sind in Planung. Derzeit werden mögliche Themen und Kooperationspartner gesucht. Sicher ist schon jetzt, dass die bereits digitalisierten Daten der ZLB in die neue Präsentationsoberfläche importiert werden müssen. Ebenso steht die Migration der Daten vom OPUS-Server an.

Im Zuge des Projekts DigiTABBerlin wurde in der ZLB das Referat 1F „Digitale Dienste“ neu strukturiert. Es fungiert als ZLB-interne Kompetenz- und Koordinationsstelle. Bei regelmäßigen Treffen mit anderen Abteilungen im Rahmen der „Kompetenzgruppe Digitalisierung“ werden interne Absprachen und Vorbereitungen für weitere Digitalisierungs-Aktivitäten innerhalb unserer Institution geplant. Für die eingekaufte Software und die Fortführung des Projektes sind die Mitarbeiter des Referats 1F die Ansprechpersonen („Verfahrensverantwortliche“).

Zukünftige Aufgaben werden sich vor allem in der Öffentlichkeitsarbeit für das Projekt fokussieren. Alle etwaigen Kosten für technische Probleme in der Nachprojektphase können durch bezahlte Supportverträgen abgedeckt werden. Weitere Kosten sind derzeit unwahrscheinlich. Bei den Treffen der Förderprogramm-Teilnehmer wurden wir oft auf die Software, insbesondere zur Steuerung des Digitalisierungs-Workflows, angesprochen. Für diese und weitere Komponenten stehen wir gerne als Ansprechpartner auch in Zukunft zur Verfügung und können ggf. zu erfahrenen Institutionen den Kontakt herstellen.