

DIE HÄFTLINGSDATENBANK DER MAHN- UND GEDENKSTÄTTE RAVENSBRÜCK - VOM DOKUMENTENINHALT ZUR PRÄSENTATION

THE DATA BASE OF PRISONERS AT THE MEMORIAL PLACE RAVENSBRÜCK - FROM DOCUMENT`S CONTENTS TO PRESENTATION

Dr. Wolfgang Schade
WIDIS GmbH Berlin
Albert-Einstein-Str. 16

Tel.: +49-30-6392 1605, Fax: +49-30-6392 1650
Email: schade@widis.de, Internet:www.widis.de

Zusammenfassung:

Der Vortrag gibt einen Einblick in Untersuchungen und Arbeitsergebnisse zur Problematik der Erfassung, Aufbereitung und Auswertung des Inhalts alter Archivdokumente. Dabei werden die Resultate von Methoden der Bildvorverarbeitung, OCR-Anwendung, Datenbankaufbereitung und der Präsentation von Ortsdaten im historisch exakten Kontext anhand der Verarbeitung von Dokumenten des ehemaligen Konzentrationslagers Ravensbrück vorgestellt.

Abstract:

The paper gives an overview of successfully applied methods for data recording, processing and presentation from old archive documents. Results of background clearance, OCR-applications, data based processing methods and representation of geographic data in a historical context are presented. For testing the results documents of the former concentration camp Ravensbrück were used.

Einleitung

Im Herbst 1997 wandten sich Vertreter der Mahn- und Gedenkstätte Ravensbrück an die WIDIS mit der Bitte, ihnen bei der Erfassung von Informationen über ehemalige Häftlinge behilflich zu sein. Vorhandene, in unterschiedlichster Form vorliegende Angaben sollten erfaßt und in Datenbanken zusammengeführt werden. Die so gespeicherten Daten sollten dann genutzt werden, um daraus ein Gedenkbuch über die Toten von Ravensbrück zu erstellen.

Diese Aufgabe erwies sich als besonders langwierig und kompliziert, denn im KZ-Lager Ravensbrück waren nach dessen Räumung im April 1945 keine verwertbaren Dokumente verblieben. Diese grundsätzliche Situation ist in nahezu allen Archiven der Mahn- und Gedenkstätten gleich schlecht und der gründlichen Vernichtung der Unterlagen durch die SS geschuldet. Die Notwendigkeit der Auswertung zusammengetragener Dokumente ergibt sich aus dem Bedürfnis, Opfer jener Zeit zu ehren. Aktuell kommt der benötigte Nachweis der Überlebenden für den Empfang der Zwangsarbeiterentschädigung hinzu.

Im Konzentrationslager Ravensbrück waren von 1939 bis 1945 130 000 Frauen und Kinder sowie 20 000 Männer aus 20 Nationen inhaftiert. Informationen über sie werden in mühevoller Kleinarbeit aus den unterschiedlichsten Quellen zusammengetragen: dazu gehören u.a. die standesamtlichen Beurkundungen des Standesamtes Ravensbrück, die nachträglichen Beurkundungen des Sonderstandesamtes Arolson, Häftlingszugangs- und -transportlisten, -überstellungslisten, Angaben von Überlebenden, Veröffentlichungen in Gedenkbüchern (z.B. die Auschwitz-Gedenkbücher oder das der Berliner Holocaust-Opfer), aber auch Datenbanken von ukrainischen Häftlingen, die in Kiew bereits angelegt wurden. In diesen Quellen sind die Namen der in Ravensbrück inhaftierten Männer zu 90 - 95 % überliefert, die Namen der Frauen zu etwa 60%.

Ein erstes Problem ist die Erfassung der auf Dokumenten vorliegenden Informationen, ein zweites die möglichst widerspruchsfreie Zusammenführung und Auswertung dieser Informationen. Letzteres war nur zu bewältigen, weil dazu heute verfügbare Methoden von Datenbanken genutzt werden konnten.

Zur Erfassung der Informationen wurden ABM-Kräfte hinzugezogen. Eine Unterstützung dieser Arbeit durch den Einsatz von computergestützten Schrifterkennungs(OCR)-systemen war wegen der Beschaffenheit der Dokumente zum Zeitpunkt der Anfrage nicht möglich: zwar liegen (neben handschriftlichen Aufzeichnungen und Urkunden, in Sütterlin ausgestellt) viele maschinegeschriebene Dokumente vor, jedoch weisen diese entweder großflächige Störungen auf oder sind auf vorgedruckten Karteikarten aufgebracht, so daß der Hintergrund (unterschiedliche Färbungen, störende Feldlinien) den sofortigen Einsatz kommerzieller OCR-Systeme unmöglich machte. Dieses Problem trifft auf nahezu alle alten, selbst maschinegeschriebenen Dokumente zu und ist ein generelles Problem bei der Inhaltserfassung von Archivadokumenten, nicht nur bei denen der Mahn- und Gedenkstätten.

Um diese Ausgangslage zu verbessern, wurde beim Berliner Senat für Wirtschaft und Technologie unter Federführung der WIDIS ein Forschungs- Verbundprojekt („ASPA“) zur Erarbeitung und Anwendung von diesem Anliegen unterstützenden verfahren eingeworben.

Bildvorverarbeitungsverfahren und OCR für die Inhaltserfassung

Ein erster Schwerpunkt dieses Verbundprojektes war es, durch den Einsatz von Methoden der modernen Bildverarbeitung die images von eingescannten Dokumenten so aufzubereiten, daß ein Einsatz von OCR für die **Inhaltserfassung der alten Dokumente** sinnvoll war

Zu diesem Zweck waren Berliner Wissenschaftler in dieses Projekt integriert, die durch ihre bisherigen Arbeiten auf dem Gebiet der Bildverarbeitung bereits große Erfahrungen besaßen

Mit speziellen Fragestellungen der Dokumentenvorverarbeitung waren die GFal und das Fraunhoferinstitut IPK befaßt, während Probleme der Vektorisierung von Zeichnungen und Landkarten bei der Fa. Graphikon bearbeitet wurden. WIDIS oblag in diesem Projekt die Beschaffung von Archivmaterial, die Zusammenführung der einzelnen Bildvorverarbeitungstools mit OCR-Systemen, der umfangreiche Test, die Einbringung der Informationen in Datenbanken, der Datenabgleich und die Präsentation ausgewählter Daten im historischen Kontext. Das Projekt wurde im Zeitraum 1998 – 2000 durchgeführt. Als Ergebnis liegen eine Reihe von Tools vor, die in einem Demonstrator zusammengefaßt wurden.

Der Einsatz der Bildvorverarbeitungstools für die Inhaltserfassung soll anhand von zwei Beispielen im folgenden erläutert werden.

1. die Arolsen-Kartei

Beim Internationalen Roten Kreuz in Bad Arolsen befindet sich der größte Bestand an Originaldokumenten, Kopien und Mikrofiches in Deutschland, die die Vorgänge in den ehemaligen Konzentrationslagern dokumentieren. Hier werden auf Wunsch enger Angehöriger nachträglich Sterbeurkunden von KZ-Opfern ausgestellt, wenn sich deren Tod belegen läßt. Zu jeder dieser Sterbeurkunde wird gleichzeitig eine Karteikarte angelegt, auf der die wichtigsten Daten der Urkunde aufgeführt werden.

Der Ravensbrück-Anteil dieser Kartei besteht aus ca. 3700 Karten A6 unterschiedlicher Färbung, der Hintergrund ist liniert. Charakteristisch ist die regelmäßige Struktur der eingetragenen Informationen, wobei diese mit Schreibmaschine jeweils hinter aufgedruckten Schlüsselwörtern (Name, Vorname, Geburtsdatum..) eingetragen wurden.

Durch die Tools der GFal wurden die aufgedruckten Schlüsselwörter gesucht, die dahinter eingetragenen Informationen extrahiert, eine Hintergrundbereinigung und Kontrastverstärkung durchgeführt und auf das Ergebnis herkömmliche OCR angewandt.

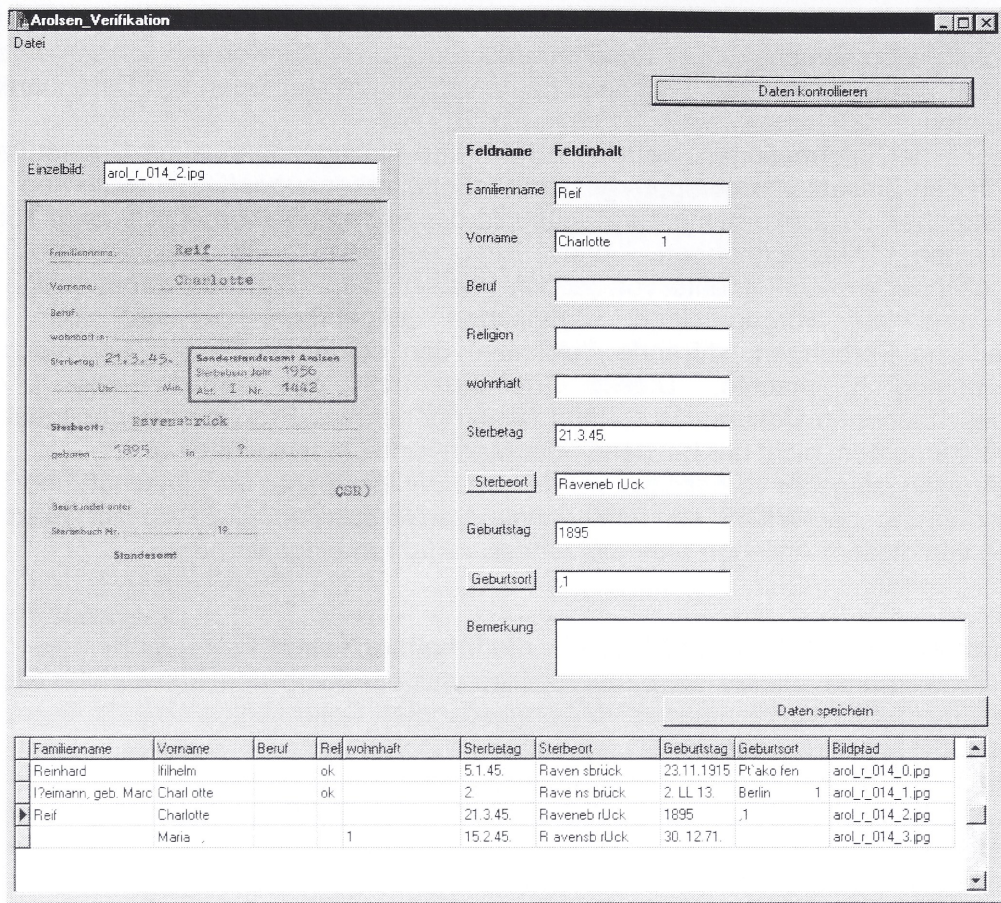


Bild 1: Beispiel für die Erkennung der Arolson-Kartei

2. die WVHA-Datei Hoppegarten

Die Kartei – bestehend aus ca. 120000 Karten - wurde wahrscheinlich auf Verlangen des Wirtschaftsverwaltungshauptamtes (WVHA) der SS Oranienburg in den einzelnen Lagern geführt. Auf diesen Karteikarten (A5, unterschiedliche Färbung, strukturierter Hintergrund) waren Angaben zu den arbeitsfähigen Häftlingen der entsprechenden Lager vermerkt (z.B. Verhaftungsort und -datum, Haftgrund, Geburtstag, Geschlecht, Beruf, das Arbeitseinsatzgebiet, die Häftlingsnummer in dem entsprechenden KZ sowie „Abgangsart“ und „-datum“ aus dem Lager; der Name des Häftlings fehlt). Die Eintragungen wurden in den dafür vorgesehenen Feldern in unterschiedlichster Form vorgenommen: mit Schreibmaschine, handschriftlich, mit Stempel, also zur Anwendung von OCR-Methoden ungeeignet.

Jedoch wurden diese Angaben codiert (Eintragung der entsprechenden Code-Nummern in dafür vorgesehenen Feldern handschriftlich mit Bleistift), um diese Codierung dann auf Hollerith-Lochkarten zu übertragen. Damit standen die Informationen dann zur statistischen und verwaltungstechnischen Auswertung mit Hollerith-Maschinen zur Verfügung (diese Hollerith-Vorkartei wird deshalb auch als Beleg für die enge Verflechtung von IBM mit dem System des III. Reiches in dem Buch „IBM und der Holocaust“ angeführt.). Zur Rückgewinnung der wesentlichen Informationen genügte es also, „nur“ die handschriftlichen Codierungen zu erkennen, um die Daten zu erfassen.

Leider waren die Aufdrucke auf den Karteikarten nicht vollständig identisch, so daß das Aufsuchen der Felder mit den herkömmlichen OCR-Programmen nicht zu bewältigen war. Im Rahmen des Projektes wurden Verfahren entwickelt, die ein sicheres Auffinden der entsprechenden Codierungsfelder erlaubte.

Häftlingskarte																		
Einlieferungsstelle		Einlieferungsdatum			Häftlingsart			Geburts-			Geschlecht		Familienstand		Kinder			
Stapo		Tag	Monat	Jahr	AZA			Tag	Monat	Jahr	männlich 1 <input checked="" type="checkbox"/>	ledig 1 <input checked="" type="checkbox"/>	verheiratet 2 <input type="checkbox"/>	verwitwet 3 <input type="checkbox"/>	geschieden 4 <input type="checkbox"/>	keine		
30		03	44				22	03	26									
Staatsangehörigkeit			Hauptberuf			1. Nebenberuf			2. Nebenberuf			3. Nebenberuf						
Russland			Tischler			-			-			-						
Wehrdienstverhältnis		T.-Grad		Anzahl Vorstrafen		Gefängnis Monate		Zuchthaus Monate		Eingeliefert in KL								
-		-		00		00		000		Auschwitz								
Zugangsart		Überstellung an KL			Häftlings-Nr.			eingesetzt als			Abgangs-		Holl. Verm.					
1		Sachsenhausen			034 329			Tischler			16.10.44		7					
Bemerkungen:													Kontrollvermerk					
													Ausgestellt		verschlossen		Loch gedrückt	

Bild 2: Beispiel einer Karte der Hoppegartenkartei

Nach Auffinden der entsprechenden Felder, dem Entnehmen der Teilbilder und der Anwendung von Hintergrundbereinigungsverfahren konnten dann OCR-Verfahren zur handschriftlichen Erkennung eingesetzt werden, wodurch sich eine Zeiteinsparung für die Inhaltserfassung von ca. 50% ergab.

Häftlingskarte																		
Einlieferungsstelle		Einlieferungsdatum			Häftlingsart			Geburts-			Geschlecht		Familienstand		Kinder			
Stapo		Tag	Monat	Jahr	AZA			Tag	Monat	Jahr	männlich 1 <input checked="" type="checkbox"/>	ledig 1 <input checked="" type="checkbox"/>	verheiratet 2 <input type="checkbox"/>	verwitwet 3 <input type="checkbox"/>	geschieden 4 <input type="checkbox"/>	keine		
01		01	01	44	010766													
Massenverifizierung																		
OK													Abbrechen					
Staatsangehörigkeit			Hauptberuf			1. Nebenberuf			2. Nebenberuf			3. Nebenberuf						
-			-			-			-			-						
Wehrdienstverhältnis		T.-Grad		Anzahl Vorstrafen		Gefängnis Monate		Zuchthaus Monate		Eingeliefert in KL								
-		-		00		00		000		Auschwitz								
Zugangsart		Überstellung an KL			Häftlings-Nr.			eingesetzt als			Abgangs-		Holl. Verm.					
1		Sachsenhausen			010766			Tischler			16.10.44		7					
Bemerkungen:													Kontrollvermerk					
													Ausgestellt		verschlossen		Loch gedrückt	

Bild 3: Beispiel der Erkennung einer handschriftlichen Codierung

Wegen der großen Anzahl der Karteikarten und den durch das Archiv auferlegten Restriktionen bei ihrer Erfassung (Einscannen nur vor Ort während der Arbeitszeit innerhalb von höchstens zwei Tagen) war eine Erfassung der Gesamtmenge nur als bitonales Image möglich. Mit fortschreitender Technik ist inzwischen auch eine Massenscannung von Farbbildern möglich. Farbscans bieten den Vorteil, daß sich Dokumentenhintergrund und aufgetragene Informationen besser unterscheiden lassen. Damit lassen sich Hintergrundstörungen (z. B.

auch durch lange Lagerung hervorgerufene Verfärbungen) besser erkennen und beseitigen (Hintergrund wird auf einheitliches Grau gesetzt), womit sich das IPK umfassend beschäftigte.

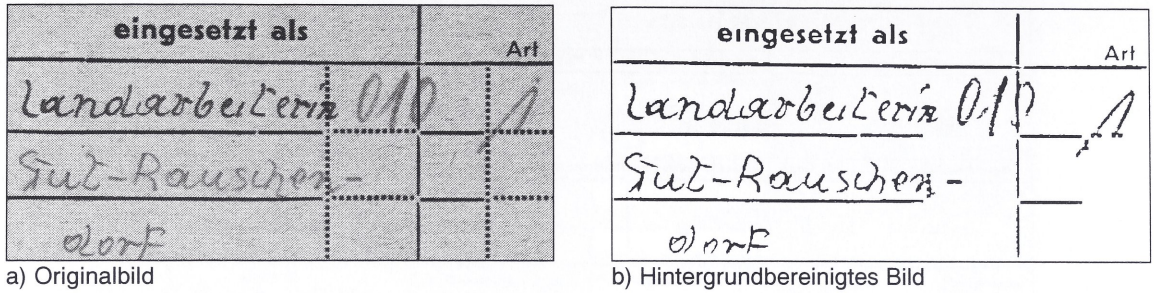


Bild 4: Beispiel der Hintergrundbereinigung bei historischen Dokumenten (IPK)

Auf diese Weise liegt eine Technologie vor, mit der sich Informationen aus maschinell (oder z.T. auch handschriftlich) ausgefüllten Formularen automatisch erfassen und verwalten lassen. (Abb.)

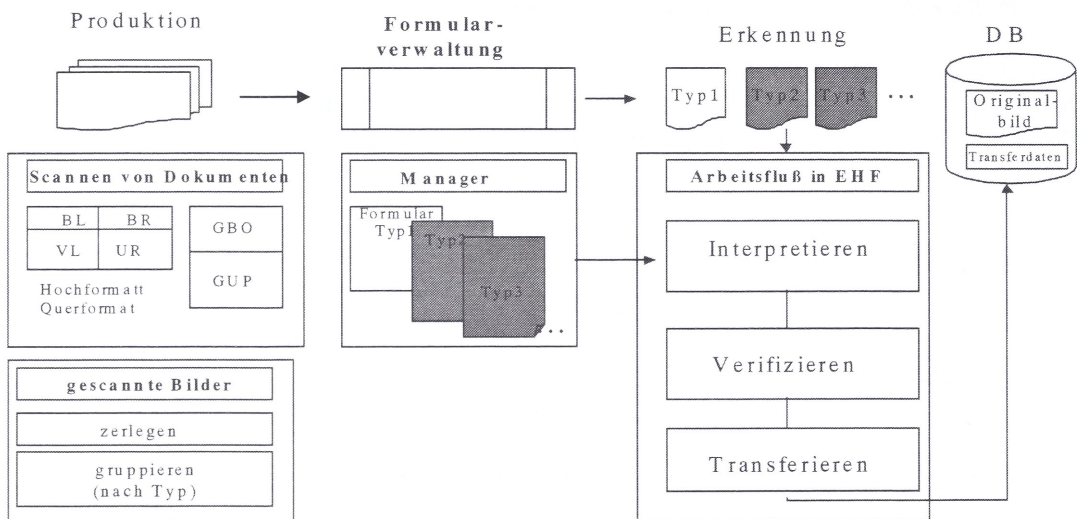


Bild 5: Produktion: Scannen – Formularverwaltung – Erkennung - Datenbank

Speicherung und Informationsverarbeitung in Datenbanken

Auf Anforderung der involvierten Historiker mußten die erfaßten Angaben quellennah abgelegt werden, so daß für jede Dokumentenklasse (über 100!) eine spezielle Datenbank notwendig ist. Zusammen mit den manuell erfaßten Informationen umfassen diese Datenbanken ca. 200 000 Datensätze

Die hierin insgesamt gespeicherten Angaben sind natürlich weder vollständig noch ohne Widersprüche (nicht konsistent), einige Angaben sind mehrfach erfaßt (Häftlingsnamen tauchen z.B. in Transportlisten oder in Zuganglisten auf). Um alle Daten, die in den unterschiedlichen Datenbanken erfaßt sind, zusammenführen zu können, wurden Hilfsprogramme entwickelt, die einen Datenabgleich ermöglichen.

Man denke z.B. an unterschiedliche Schreibweisen von Namen. Nicht nur, daß es z.B. mehrere Möglichkeiten der Erfassung des Namens in den entsprechenden Feldern gibt: Jozef van der Broek / van der Broek, Jozef / v.d. Broek, Jozef / Broek, van der, Jozef Sondern zusätzlich noch Abweichungen in der Schreibweise des Namens möglich sind.

Hier wurden Möglichkeiten geschaffen, um über die phonetische Sprechweise des Namens mögliche Identitäten aufzuspüren, um dann durch Vergleich weiterer Angaben zu den Personen ein doppeltes Vorkommen der Person in dem Datenbestand vermeiden zu können. Dazu wurden spezielle „Soundex-Algorithmen“ eingebunden oder auch ein Vergleich mit fehlerhafter bzw. unvollständiger Namenserfassung realisiert.

Adam	Adam	A4m
Adamska	Adam3ka	A4m3k
Bednarkiewicz	Bednarkiewicz	Be4nrkiewi3

Bild 6: Prinzip des „Soundex-Algorithmus“

Präsentation der Inhalte

Als Resultat dieser Arbeit konnte nach Beendigung des Projektes der Mahn- und Gedenkstätte Ravensbrück ein Gedenkbuch mit den Namen von über 10 000 Opfern übergeben werden. Dieses Gedenkbuch ist als vorläufig zu betrachten, weil ungefähr erst die Hälfte aller inzwischen zur Verfügung stehenden Dokumente digital erfaßt und aufbereitet wurden.

Um die gewonnenen Daten auf Info-Points in den Gedenkstätten präsentieren zu können, wurde ein Tool entwickelt, das die Geburts- bzw. Wohnorte der Häftlinge auf historischen Landkarten darstellen konnte. Dabei werden die unterschiedlichen Ortsnamen (z.B. die deutschen Namen für heute in Polen, Tschechien oder Rußland liegende Orte) ausgewertet, die Anzahl der Häftlinge wird dann durch verschieden große Ortsmarkierungen gekennzeichnet.

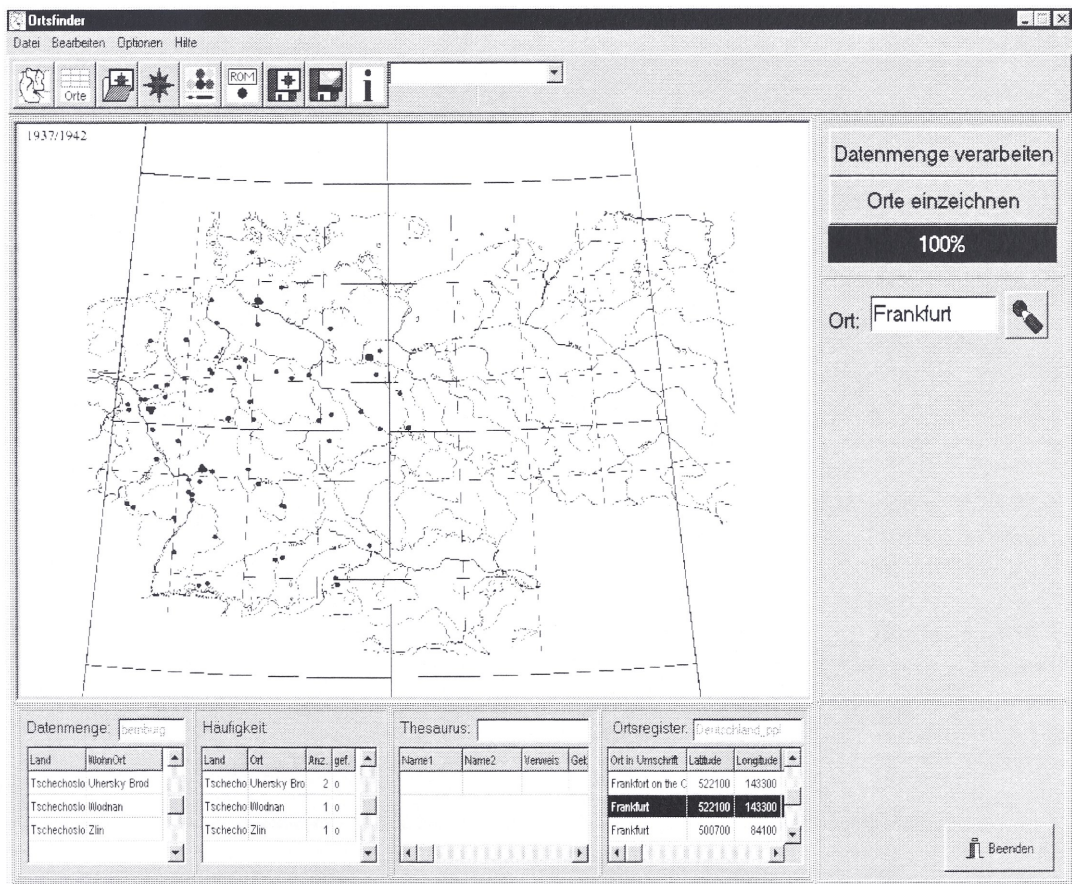


Bild 7: Ortsdarstellung im historischen Kontext

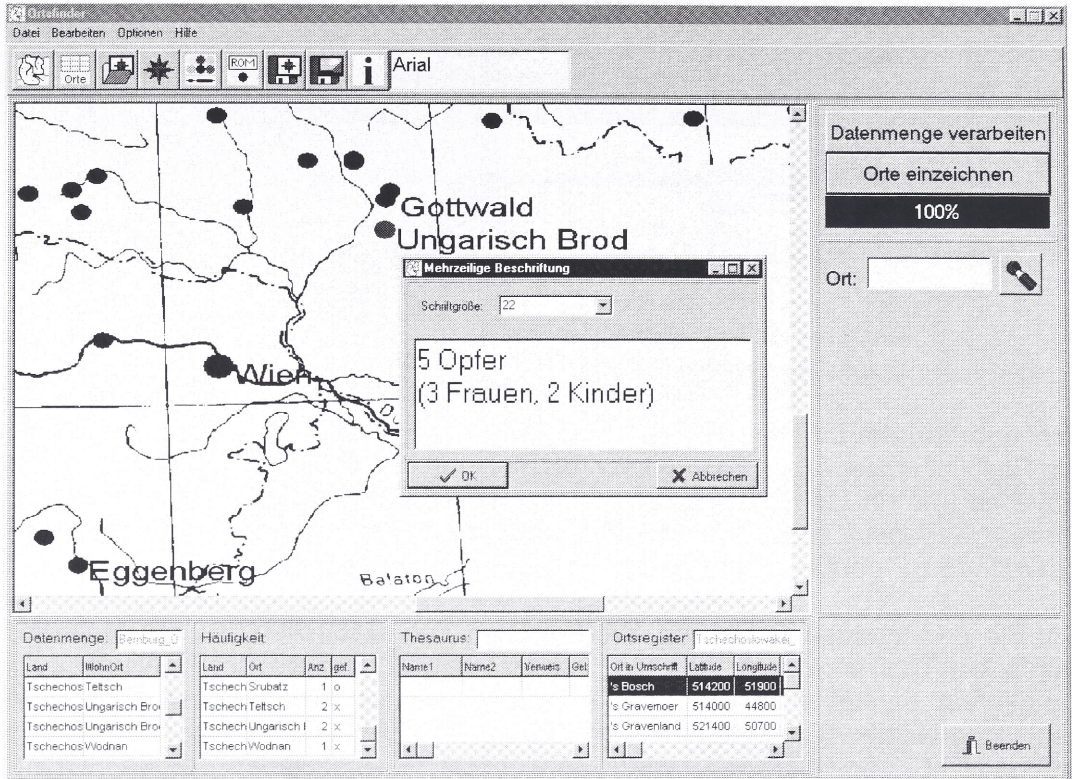


Bild 8: Die Orte können mit Zusatzinformationen hinterlegt werden

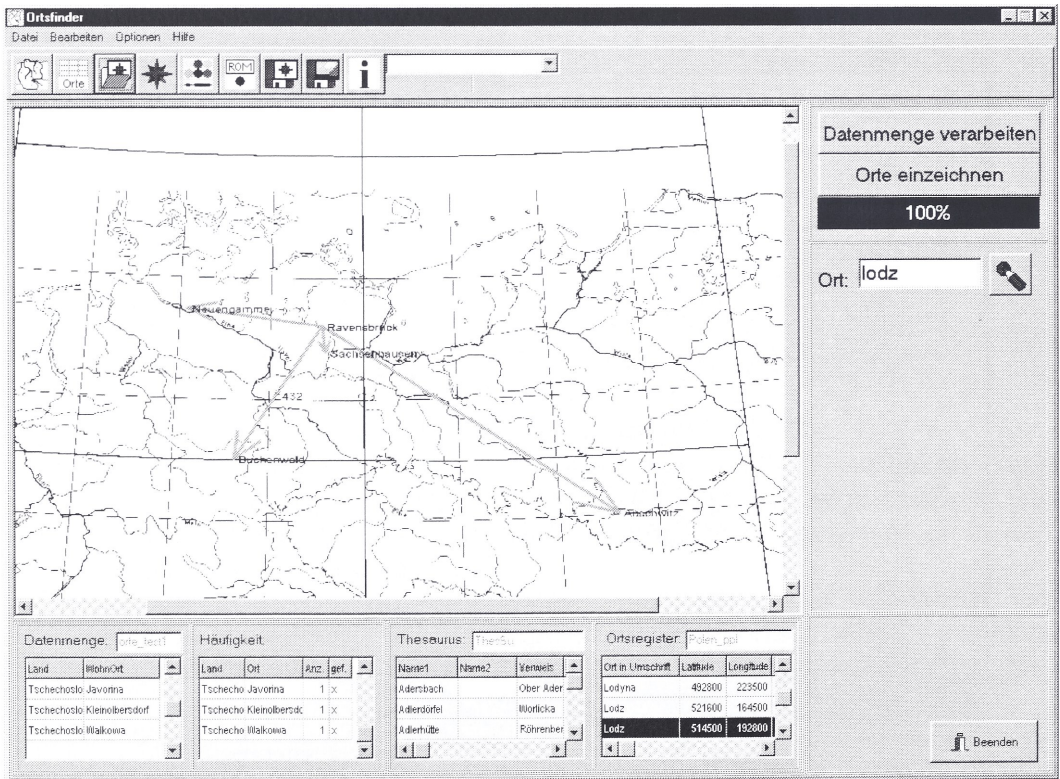


Bild 9: Darstellung von Häftlingstransporten