

Kathrin Jockel

4,7 Millionen Seiten Exilmonografien digital

Ein virtueller Fundus für
»Künste im Exil«

Die Aufgabenstellung

Für das Haushaltsjahr 2013 hat der damalige Staatsminister für Kultur und Medien, Bernd Neumann, der Deutschen Nationalbibliothek (DNB) 2 Millionen Euro allein für die Digitalisierung von Objekten zum Thema Exil und Emigration zur Verfügung gestellt. Eines der daraus finanzierten Projekte war die Digitalisierung aller Bücher aus den Exilsammlungen der DNB in Leipzig und Frankfurt am Main, die auch als Fundus für Exponate für die virtuelle Ausstellung »Künste im Exil« dienen sollten. Material genug gab es, da eine besondere Aufgabe der DNB darin liegt, die gedruckten und ungedruckten Zeugnisse der deutschsprachigen Emigration und des Exils der Jahre 1933 bis 1945 zu sammeln, zu erschließen und für die Öffentlichkeit bereitzustellen. Aus historischen Gründen besitzt die DNB zwei Exilsammlungen: die Sammlung Exil-Literatur 1933 - 1945 der DNB in Leipzig und das Deutsche Exilarchiv 1933 - 1945 am Standort Frankfurt am Main. Alle Bücher der beiden Sammlungen, die sogenannten Exilmonografien, sollten digitalisiert werden.

den kann (was 2013 glücklicherweise Mitte Januar der Fall war) sowie die Zeit vom Kassenschluss bis Jahresende (zwei Wochen), so standen tatsächlich nur acht Monate für die Digitalisierungsarbeiten zur Verfügung, von denen wiederum mindestens ein Monat für den Aufbau der Geräte und die Inbetriebnahme des Bearbeitungsworkflows einge-rechnet werden musste. Blieben also nur sieben Monate für die eigentliche Digitalisierung.

Auch mit dem Risiko der schieren Masse war umzugehen: Um 22.275 Werke mit 4,7 Millionen Seiten in sieben Monaten zu digitalisieren, sind bei einer Sechs-Tage-Woche 670.000 Seiten im Monat beziehungsweise 28.000 Seiten am Tag zu bearbeiten. Hierfür sind massentaugliche Verfahren, ein Mehrschichtbetrieb und eine hohe Ausfallsicherheit unabdingbare Voraussetzung, um ein solches Volumen in der vorgegebenen Zeit bewältigen zu können.

Eine weitere Unsicherheit steckte in den Sammlungen selbst: Die Bücher der beiden Exilsammlungen waren bis dahin noch keinem Dublettencheck unterzogen worden; es war somit in der knappen Vorbereitungszeit nur schwer abzuschätzen, wie viele der rund 40.000 Bücher beider Sammlungen dublett waren und mit welcher Qualität der Katalogsituation zu rechnen war.

Digitalisierung
von 28.000
Seiten pro Tag

Unklare Dublet-
tensituation

Die Rahmenbedingungen

Enger Zeitplan
größte Heraus-
forderung

Die größte Herausforderung dieses Projekts war der enge Zeitplan, der sich aus den Rahmenbedingungen der Mittelvergabe für dieses 2013 durchzuführende Vorhaben ergab. Da die DNB selbst über kein Digitalisierungszentrum verfügt, sondern die Aufträge projektbezogen an Dienstleister vergibt, musste zunächst eine EU-weite Ausschreibung vorbereitet werden. Zieht man von einem Haushaltsjahr (zwölf Monate) die Ausschreibungszeit (rund drei Monate) ab und die Zeit, bis der Haushalt verabschiedet ist und überhaupt ausgeschrieben wer-



Aufstellung aller zu digitalisierender Werke in einem separaten Magazin
Foto: Deutsche Nationalbibliothek, Stephan Jockel

Der Bestand

Bei den Büchern der Exilsammlungen handelt es sich vorwiegend um im Ausland verfasste oder veröffentlichte Druckwerke der Erscheinungsjahre 1933 - 1950. Aus drei Quellen speiste sich das Projekt: Zum einen aus der größten Sammlung, den Exilmonografien des Deutschen Exilarchivs am Frankfurter Standort (hier stehen rund 19.000 Werke in einem separaten Magazin, wobei es sich nicht nur um Monografien, sondern auch um Zeitschriften handeln kann, die nicht Projektinhalt sein sollten), zum anderen aus dem Bestand der Leipziger Sammlung Exil-Literatur, die 17.000 Werke aus den Jahren 1933 - 1945 separat aufgestellt hat. Drittens sollten die Leipziger Exilmonografien der Jahre 1946 - 1950 digitalisiert werden. Sie stehen allerdings nicht mit den Büchern der Sammlung Exil-Literatur bis 1945 zusammen, sondern im DNB-Hauptbestand verteilt. Die Zuordnung zur Sammlung wurde bislang lediglich durch einen Zettelkatalog dokumentiert. Dennoch sollten auch die nicht dubletten Werke aus den Jahren 1946 - 1950 aus diesem Bestandssegment in das Digitalisierungsprojekt mit einbezogen werden. Das bedeutete, dass sie gezielt aus dem Hauptbestand gezogen, katalogseitig nachbearbeitet und gesondert aufgestellt werden mussten, damit sie analog zu den übrigen Exilmonografien im Digitalisierungsworkflow bearbeitet werden konnten.

Erhaltungszustand der Werke

Der Zustand der Werke war ihrem Alter entsprechend. Durchgängig war das Papier sauer und in verschiedener Intensität vergilbt, oft waren die Ränder angegriffen, sodass beispielsweise Ecken fehlten. Teilweise waren auch Buchrücken oder einzelne Seiten lose.

Die Katalogdaten zu den Werken stammten größtenteils aus verschiedenen Retrokatalogisierungsprojekten. Da ein Dublettencheck bislang nur teilweise realisiert worden war, konnten vor Projektbeginn keine genauen Zahlen zum tatsächlich zu digitalisierenden Bestand ermittelt werden. Praktisch jede denkbare Situation konnte im Katalog auftreten. Es gab die Fälle: ein Datensatz für mehrere Exemplare, mehrere (dublette) Datensätze für ein Exemplar, mehrere (nicht verknüpfte) Datensätze für mehrere Exemplare eines Titels, Exemplare ohne Datensätze.

Somit war eine wichtige Aufgabe in der Vorbereitungsphase klar: Die Katalogsituation musste bereinigt werden. Innerhalb von drei Monaten führten an beiden Standorten Katalogisierungsteams eine Bestands- und Katalogrevision durch; dabei musste jeder Band der etwa 40.000 Werke in die Hand genommen, jeder Katalogeintrag überprüft und im Fall von Unstimmigkeiten bereinigt werden. Darüber hinaus waren alle Bände, die nicht digitalisiert werden sollten, also alle Dubletten und Zeitschriftenbände, optisch zu kennzeichnen, weil sie eben nicht in den Digitalisierungsworkflow gegeben werden sollten.

Um für die Ausschreibung verlässliche Zahlen zum zu digitalisierenden Bestand zu bekommen, musste außerdem am Regal gearbeitet werden – mit dem Zollstock wurden Regalmeter vermessen, dazu über mehrere Jahrgänge durchschnittliche Exemplarzahlen je Meter sowie die durchschnittlichen Seitenzahlen ermittelt und hochgerechnet. Zeitschriftenhefte wurden aus der Berechnung wieder herausgenommen. Ausgeschrieben wurde anhand dieser Berechnungen die Digitalisierung von 24.000 Monografien mit rund 4 Millionen Seiten – wie sich bei Projektende zeigte: eine durchaus gute Schätzung.

Das Ergebnis: Für die zu digitalisierenden Werke gab es immer einen – und zwar nur einen – Datensatz. Alle dubletten Datensätze waren identifiziert und bereinigt sowie die dubletten Exemplare und Zeitschriftenbände, die nicht digitalisiert werden sollten, sichtbar für die Magazinmitarbeiter gekennzeichnet. Darüber hinaus wurden die Werke der Sammlung Exil-Literatur aus den Jahren 1946 - 1950, die bislang nur in einer Zettelkartei zusammengeführt waren, auch in der Katalogdatenbank als zur Exilsammlung gehörend codiert, sodass sie heute als zur Sammlung gehörig im Portal unter dem Reiter »Exilsammlungen« recherchiert und gefunden werden können.

Bereinigung der Katalogsituation innerhalb von 3 Monaten

Bestandsermittlung mit Zollstock am Regal

Weitere Projektvorarbeiten

Digitalisiert werden sollte am Frankfurter Standort. Rund 8.000 »singuläre« (also nicht dublette) Leipziger Werke mussten deshalb nach Frankfurt gebracht werden. Um die Logistik der Werke so transparent wie möglich zu gestalten und eine laufend reibungs-



Eines der digitalisierten Werke: Zero hour for Germany - destruction or revolution?: come and hear about the present situation inside Germany / Free German Movement in Great Britain. Weltweit frei verfügbar unter <http://d-nb.info/1032297093>

freie Beschickung des Digitalisierungsworkflows mit Werken sicherzustellen, wurden alle Werke – die Leipziger ebenso wie die Frankfurter – in einem separaten Magazin zusammen aufgestellt. Außerdem wurden alle Exilmonografien vorab im Workflow zur Retro-Kataloganreicherung bearbeitet. Dabei wurden nicht nur 15.664 Inhaltsverzeichnisse digitalisiert, die über den Katalog umgehend weltweit für Recherchezwecke genutzt werden konnten, da sie urheberrechtsfrei sind und deshalb keine Einschränkungen bei der Bereitstellung bestehen. Der Bestand wurde außerdem gleichzeitig optimal für die Automatisierung des Digitalisierungsworkflows vorbereitet, indem alle Bände mit Barcodes versehen wurden und der Katalog um fehlende Elemente ergänzt werden konnte (z. B. Titel- und Bandaufnahmen, Sprachencodes) – entscheidende Voraussetzung für die Digitalisierung der Exilwerke in derart kurzer Zeit. Ende März 2013 waren diese Vorarbeiten abgeschlossen.

Vorarbeiten
Ende März 2013
abgeschlossen

Eine weitere wesentliche Voraussetzung für einen effizienten Workflow wurde dadurch geschaffen, dass noch vor Beginn der Digitalisierungsarbeiten die benötigten Katalogdatensätze für die Sekundärformen automatisch erzeugt werden konnten. Dabei wurden aus den vorliegenden, im Rahmen des Vorprojekts bereinigten Datensätzen für die gedruckten Werke mittels eines Skripts Katalogdatensätze für die Digitalisate mit eigener IDN (Identifikationsnummer) erstellt. 24.000 neue Datensätze sind dabei entstanden, die mit den Datensätzen der Druckausgabe verlinkt sind.

Das Ergebnis: Für jedes Digitalisat lag ein vorbereiteter Katalogdatensatz vor; die IDN wurde später vom Digitalisierungsdienstleister genutzt, um Dateien korrekt zu benennen. Auch der Import der Digitalisate sowie der OCR- und der Metadaten ins Repository der DNB konnte somit automatisiert ablaufen, da anhand der IDN das Digitalisat dem Katalogdatensatz zugeordnet werden konnte. Die Importroutine passte beim Einspielen den Katalogdatensatz an und ergänzte ihn um Angaben, die erst bei der Ablieferung entstehen konnten (wie beispielsweise eine URN).

24.000 neue
Katalogdaten-
sätze für Sekun-
därformen

Die Rechteklärung

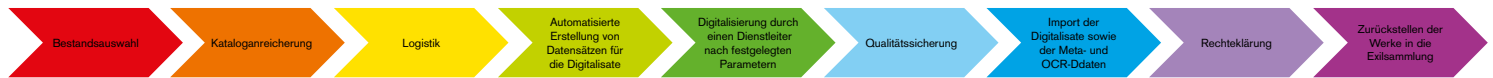
Es fand eine Rechteklärung statt, bei der der urheberrechtliche Status jedes Werkes ermittelt wurde. Die Rechteklärung wurde der Digitalisierung nachgelagert, um keinen »Flaschenhals« bei der Durchführung der Digitalisierungsarbeiten zu bilden. Anhand des Katalogdatensatzes sowie des gedruckten Werkes wurden die Urheber sowie deren Sterbedaten ermittelt – soweit das überhaupt möglich war – und für alle Werke der urheberrechtliche Status in der Katalogdatenbank in Form von Codes erfasst. Ziel der Rechteklärung ist natürlich, rechtfreies Material zu identifizieren und gegebenenfalls den freien Zugang zu den digitalen Objekten zu ermöglichen beziehungsweise den Zeitpunkt der Gemeinfreiheit zu ermitteln, um diese Objekte später für die weltweite Nutzung freizugeben. Rund 3 Prozent der Exilmonografien sind heute erst gemeinfrei. Die Größenordnung entspricht den Erwartungen und dem, was bereits vorab anhand von Stichproben ermittelt werden konnte.

Ziel: zukünftige
Freigabe zur welt-
weiten Nutzung
ermöglichen

Ebenso wichtig aber ist, dass nach erfolgter Rech-
teklärung sämtliche Urheber im Katalogdatensatz
verankert und, soweit möglich, mit Datensätzen
der Gemeinsamen Normdatei (GND) verknüpft
sind sowie der urheberrechtliche Status dokumen-
tiert ist. Damit ist die Ausgangsbasis dafür geschaf-
fen, beispielsweise im Rahmen der gesetzlichen

Der Digitalisierungsworkflow der DNB

Nachdem die Vorarbeiten realisiert waren, konnten
die Digitalisierungsarbeiten ausgeschrieben werden.
Dabei konnte sich die DNB auf die Erfahrungen
früherer Projekte stützen, u. a. aus dem Projekt zur



Der Digitalisierungsworkflow

Regelung zu verwaisten oder vergriffenen Werken
gezielt aktiv werden zu können, oder aber auch
dafür, dass Kuratoren einer virtuellen Ausstellung
Nutzungsrechte aktiv bei den Urhebern oder ande-
ren Berechtigten einholen können. Das derzeitige
Engagement für die Rechteklärung ist somit eine
wichtige Investition in die Zukunft.

Das Ergebnis: Zu jedem Werk ist der urheberrecht-
liche Status dokumentiert (beispielsweise »gemein-
frei«, »gemeinfrei ab dem Jahr XXXX«, »urheber-
rechtlicher Status derzeit nicht ermittelbar«). Die
Digitalisate gemeinfreier Werke wurden umgehend
für den weltweiten Zugriff freigegeben, alle anderen
Digitalisate stehen bislang nur in den Lesesälen der
Bibliothek zur Verfügung.

Dokumentation
des urheberrecht-
lichen Status

Digitalisierung beschädigter Buchmedien¹⁾, das seit
2011 am Leipziger Standort durchgeführt wird,
wenn auch nur mit einem Zehntel des für dieses
Projekt geplanten Durchsatzes.

Ausgeschrieben wurde die Digitalisierungsleistung
samt OCR- und Metadatenerstellung. Die Parame-
ter dafür waren in der Leistungsbeschreibung genau
definiert. Dem Dienstleister wurde ein Raum in der
DNB in Frankfurt am Main zur Verfügung gestellt,
die Scan- und PC-Arbeitsplätze mussten mitge-
bracht werden. Die Digitalisierung sollte so me-
dienschonend wie möglich durchgeführt werden.
Eine »Boutique-Digitalisierung mit weißen Hand-
schuhen« konnte und sollte das Projekt unter den
gegebenen Rahmenbedingungen aber nicht leisten
und war trotz des oben beschriebenen Zustandes
der Sammlung auch nicht erforderlich.

Ausschreibung
der Digitalisie-
rungsleistung

Die Ausschreibung für die Digitalisierung konn-
te der Harald Fischer Verlag, Erlangen, für sich
entscheiden. Umgehend nach Zuschlagserteilung
begann er im April mit dem Aufbau und ersten
Testläufen. Ab Anfang Mai konnten dann die Pro-
duktionszahlen von Woche zu Woche gesteigert
werden; ihren Spitzenwert erreichte die Produktion
im August mit 975.000 digitalisierten Seiten von
rund 4.100 Werken.

Natürlich ist ein Dienstleister verpflichtet, hin-
reichende Maßnahmen zur Qualitätssicherung
laufend durchzuführen und entsprechend nachzu-
weisen. Das bezieht sich sowohl auf die Qualität
des Digitalisats als auch auf die Vollständigkeit der
Bearbeitung, die Einhaltung der technischen Pa-
rameter, die fehlerfreie Bereitstellung aller Daten
sowie auf Plausibilitätskontrollen hinsichtlich der
Validität der gelieferten Daten.



Ein weiteres der digitalisierten Werke: Joseph Fouché: Bildnis eines politischen Menschen / Stefan Zweig. Weltweit verfügbar unter <http://d-nb.info/103233228X>

Zusätzlich führte die DNB laufend und zeitlich eng an der Produktion eigene Qualitätssicherungsmaßnahmen durch, um die Qualität aller Daten zu überprüfen, gegebenenfalls zu reklamieren beziehungsweise einwandfreie Daten abzunehmen und für den Import der Daten in die Systeme der DNB freizugeben. Daher war Teil der Ausschreibung auch, der DNB ein Monitoringsystem zur Verfügung zu stellen, das diese Arbeiten einfach und schnell ermöglicht. Von den Mitarbeiterinnen und Mitarbeitern in der Qualitätssicherung wurde stichprobenartig etwa ein Drittel der Produktion überprüft, gegebenenfalls reklamiert oder ansonsten freigegeben. Werke, die nicht geprüft wurden beziehungsweise alle Werke, die nicht innerhalb von acht Tagen von der DNB reklamiert wurden, galten als abgenommen und konnten importiert werden.



DNB-Mitarbeiterin bei der Qualitätssicherung
Foto: Deutsche Nationalbibliothek, Stephan Jockel

Alle Digitalisate wurden anfangs noch beim Digitalisierungsdienstleister zwischengespeichert, bis in der DNB auch für den laufenden Import dieser zusätzlichen großen Datenmengen die entsprechenden Voraussetzungen geschaffen waren. Ab Beginn der Digitalisierungsarbeiten konnte aber bereits selektiv importiert werden. Ab Juli fand der Trans-

fer der Daten auf die DNB-Systeme dann laufend statt, um die Digitalisate schon kurz nach ihrer Erstellung verfügbar zu machen und dabei gleichzeitig die Belastbarkeit der Schnittstellen zu testen und gegebenenfalls weiterzuentwickeln. Das Ergebnis war insgesamt sehr positiv: Jede Nacht konnten nun zusätzlich zu den sonstigen Importroutinen der DNB bis zu 300 Digitalisate von Exilmonografien mit rund 450 GB Speichervolumen in die DNB-Systeme überführt werden.

Und was macht die Projektleitung?

Die Projektleitung muss den Workflow permanent überwachen und eingreifen, sobald eine der folgenden Fragen auf ein Problem hinweist: Stimmt der tägliche Durchsatz in den verschiedenen Workflowschritten? Stimmt die Qualität? Gibt es systematische Fehler? Wird der Zeitplan eingehalten? Passt die Budgetplanung? Wurden die Mengen richtig kalkuliert, sowohl was die Exemplare als auch was die Seitenzahlen angeht? Sind Krankheitsmeldungen einzelner Projektmitarbeiter auszugleichen, weil sonst vereinbarte Durchsatzzahlen nicht zu schaffen sind?

Rund 1.000 Werke gingen im Rahmen der Beschickung fälschlicherweise an den Dienstleister, weil sie vorab nicht als Dubletten oder Zeitschriftenstücktitel erkannt wurden. Zudem lagen Werke noch in den unbeschnittenen Druckbögen vor oder mussten entklammert werden. Andere Werke waren durch das Raster der Sekundärformenerstellung gerutscht und hatten keinen Datensatz für das Digitalisat erhalten. Alle diese Werke und noch eine ganze Menge fraglicher Fälle mehr wurden vom Dienstleister täglich, in Stoßzeiten sogar zweimal täglich an eine DNB-Mitarbeiterin übergeben, von ihr nachbearbeitet und möglichst umgehend wieder an den Dienstleister zurückgeliefert.

Gerade bei einem Projekt, das – siehe oben – zahlreiche Risiken birgt und für seine Größenordnung eine extrem knappe Vorlaufzeit hat, braucht es eine enge und tagesaktuelle Kommunikation mit dem Dienstleister, um Problemfälle und Rückläufer umgehend zu besprechen beziehungsweise erneut durch einzelne Schritte des Workflows zu schieben, also beispielsweise erneut in die Erschließung

Enge Zusammenarbeit zwischen DNB und Dienstleister

zu geben, falls ein Fehler im Datensatz festgestellt wird. Dafür ist die Projektleitung erster Ansprechpartner. Reibungsverluste in der Kommunikation oder gar Stillstand der Produktion darf es bei einem solch engen Zeitplan nicht geben. Und es gab sie auch nicht: Pünktlich am 30. November 2013 wurde das 22.275ste Digitalisat freigegeben und das Projekt damit beendet.



Exilmonografien
Foto: Deutsche Nationalbibliothek, Stephan Jockel

Das Ergebnis

Mit diesem Projekt hat die DNB ihren Bestand an Sekundärformen deutlich erhöhen und damit den Anforderungen des Bestandsschutzes (Inhalts-sicherung sowie Schutz vor Beschädigungen durch

künftige Benutzung der Werke) Rechnung tragen können, gerade auch für diese exklusiven Sammlungsbestände.

Der anlässlich des Projekts aufgebaute Workflow zur Rechteklärung war lange schon ein Desiderat und wird für künftige Digitalisierungsprojekte als eigenständiger Workflowschritt je nach Anforderung des Digitalisierungsvorhabens vor- oder nachgelagert werden können.

Alle Digitalisate können genutzt werden, je nach ihrem urheberrechtlichen Status weltweit oder in den Lesesälen der Bibliothek. Zumindest die Digitalisate der rechtefreien Werke können zudem über die Deutsche Digitale Bibliothek und Europeanana angeboten werden. Die virtuelle Ausstellung »Künste im Exil²⁾« verfügt nun über einen großen Fundus an digitalen Objekten, deren urheberrechtlicher Status ermittelt wurde. Rechtefreies kann direkt in die Ausstellung übernommen werden, und zudem gibt die Verzeichnung der Rechtesituation den Kuratoren der Ausstellung die Möglichkeit, bei der Präsentation eines Werkes auf urheberrechtlich geschützte Teile (wie zum Beispiel einen Umschlag mit Illustration) zu verzichten und nur die rechtefreien Teile einzubinden – oder aber mit den Urhebern beziehungsweise deren Erben in Kontakt zu treten und die Nutzungsrechte einzuholen.

Und auch einige bis dahin nicht in den Datensätzen verzeichnete Besonderheiten der Werke konnten entdeckt werden, wie beispielsweise eine in einem Buch enthaltene handschriftliche Widmung Lion Feuchtwangers. Aber genau genommen ist jeder einzelne Band ein Schatz, der durch die Digitalisierung gehoben wurde und nun noch besser nutzbar ist.

Bessere Nutzbarkeit der Werke weltweit durch Digitalisierung

Anmerkungen

1 Vgl. Jockel, Kathrin: Eine Million Buchseiten digital. In: Dialog mit Bibliotheken, 24 (2012) 2, S. 4 - 7

2 <<http://www.kuenste-im-exil.de>>