Liliane Regamey

Schweizer Presse Online (SPOL): eine erfolgreiche Zusammenarbeit von Bibliotheken und Verlagen

Abstract: Die Schweizerische Nationalbibliothek (NB) hat in ihrer Strategie den Schwerpunkt definiert, ihre Sammlungen zu digitalisieren. Höchste Priorität genießt in diesem Rahmen die Digitalisierung der Schweizer Presse. Seit 2011 werden die Zeitungen, die die NB in Zusammenarbeit mit ihren kantonalen Partnern digitalisiert, auf der Plattform Schweizer Presse Online (SPOL) aufgeschaltet. Dies auf Anregung der Kantonsbibliotheken hin, die sich eine zentrale Infrastruktur wünschten, sowie wegen der Bedürfnisse der Nutzerschaft, die möglichst grosse Zeitungssammlungen auf einmal durchsuchen können wollte. 2017 hat die NB die bestehenden Angebote im Bereich Zeitungspräsentation und -erschliessung neu evaluiert und sich entschieden, SPOL weiterzuentwickeln, um einen grösseren Nutzen für das allgemeine Publikum und die Forscher zu erzielen. Ungefähr drei Millionen Seiten werden in den nächsten Jahren zu der Sammlung hinzukommen, die bereits jetzt 2,5 Millionen digitalisierte Zeitungsseiten umfasst.

Digitalisierung der Presse in der Schweizerischen Nationalbibliothek

Warum die Presse?

Die Schweizerische Nationalbibliothek (NB) wurde als Schweizerische Landesbibliothek 1895 gegründet. Sie erhielt den Auftrag, ab 1848 erschienene ‚schweizerische‘ Publikationen zu sammeln, zu erschliessen, zu erhalten und der Öffentlichkeit zugänglich zu machen. Darunter verstand – und versteht – der Gesetzgeber die Publikationen, welche von Schweizerinnen und Schweizern verfasst oder übersetzt wurden, aus einem Schweizer Verlag stammen oder einen thematischen

Anmerkung: Übersetzung von Andreas Berz (andreas.berz@nb.admin.ch).

Bezug zur Schweiz haben.¹ Im Vergleich mit anderen Nationalbibliotheken, Kantonsbibliotheken und weiteren Gedächtnisinstitutionen hat die NB ein geradezu jugendliches Alter. Ihre Sammlungen enthalten nur wenige alte und besonders wertvolle Dokumente; solche sind es, die typischerweise von Bibliotheken digitalisiert werden. Die NB besitzt jedoch eine einmalige Zeitungssammlung, die alle in der Schweiz erschienenen Titel umfasst und die intensiv genutzt wird.

Zeitungen sind fragile Objekte, die meist auf Papier bescheidener Qualität gedruckt werden, das sich innert kurzer Zeit abbaut: Die Blätter vergilben, werden brüchig und zerfallen zu Staub und dies auch bei besten konservatorischen Bedingungen. Die Bibliotheken sind sich schon lange bewusst, welchen Gefahren diese Dokumente ausgesetzt sind. Deshalb betreiben sie seit Mitte der 1940er Jahre Programme zur Mikroverfilmung, um den Inhalt dieser Zeitzeugen des Schweizer Alltagslebens zu bewahren.²

Zu den konservatorischen Aspekten kommt der Umstand, dass unterschiedliche Kundengruppen die Zeitungssammlung intensiv nutzen. Beides zusammen hat die NB dazu bewogen, Mittel und Wege zu suchen, um deren Konsultation zu erleichtern. Diese Materialien online zu stellen, erwies sich als geeignete Lösung. Im Jahr 2007 führte die NB eine Befragung bei ihren aktiven und potenziellen Benutzenden durch, welche das Interesse an einem Online-Zugang zu den Inhalten von Zeitungen bestätigte. Mit der Digitalisierung der Zeitungen werden die Originalausgaben auf Papier nicht etwa eliminiert; vielmehr werden sie weiterhin unter optimalen Bedingungen aufbewahrt.

Digitalisierungsleitlinie

Im Jahr 2009 hat die NB ihre erste Strategie für die Digitalisierung erarbeitet (die spätere *Digitalisierungsleitlinie*, DLL³), in welcher die Schweizer Presse oberste Priorität erhielt. Diese Stellung behielt die Presse auch in den Revisionen von 2014 und 2017 bei (Kap. 4, S. 6): „Presse: entspricht einem grossen Bedürfnis des Publikums und wird im Allgemeinen stark genutzt; die Digitalisierung bringt einen grossen Mehrwert, da sie einen Zugriff zu den Inhalten ermöglicht, der nur

1 Bundesgesetz über die Schweizerische Nationalbibliothek vom 18. Dezember 1992 (NBibG): SR 432.21. http://www.admin.ch/ch/d/rs/432_21/. Alle Links in diesem Beitrag wurden am 26.03.2018 überprüft.

2 Mehr Informationen über die Mikroverfilmung in der NB im KGS Forum 14 (2009), Publikationen: <http://www.babs.admin.ch/de/aufgabenbabs/kgs.html>.

3 Digitalisierungsleitlinie, Version 2017: <https://www.nb.admin.ch/snl/de/home/themen/die-digitalen-sammlungen/digitalisierte-sammlungen.html>.

mit den Metadaten des Kataloges allein nicht möglich wäre; die Zeitungssammlung der NB ist aussergewöhnlich, da sie nahezu vollständig ist.“

In der ersten Fassung der DLL werden zwei Bedingungen für die Digitalisierung von Zeitungen durch die NB besonders hervorgehoben: die Einhaltung des geltenden Rechts (Kap. 5) und die Partnerschaft mit anderen Institutionen (Kap. 10). Auf der Basis der festgelegten Grundsätze hat die NB Werkzeuge und Verfahren entwickelt, die es ihr in Zusammenarbeit mit den anderen Gedächtnisinstitutionen der Schweiz und mit den betroffenen Verlagen ermöglicht haben, Millionen von Zeitungsseiten zu digitalisieren.

Erste Erfahrungen

Im Jahr 2006 beschlossen die Bibliothèque de Genève (BGE) und der Verlag Le Temps, die Zeitung *Journal de Genève*, die von 1826 bis 1998 erschienen war und 550'000 Seiten umfasst, zu digitalisieren. Das Projekt wurde um die beiden Titel *Gazette de Lausanne* (1798–1991, 450'000 Seiten) und *Le nouveau quotidien* (1991–1998, 50'000 Seiten) ergänzt. Damit konnten alle Titel, welche Le Temps gehören, digital bereitgestellt werden. Die Aufschaltung im Web erfolgte 2009.

In diesem Projekt wurde ein grundlegendes Werkzeug für die Online-Stellung zeitgenössischen Materials entwickelt: eine Modellvereinbarung zur Regelung der urheberrechtlichen Fragen, in welcher der private Partner die Verantwortung gegenüber den Autorinnen und Autoren, die für die Zeitung geschrieben haben, übernimmt. Dieses Modell wird, mit einigen Anpassungen, auch bei den aktuellen Projekten verwendet. Eine Ergänzung betrifft eine Bestimmung zum Schutz der Persönlichkeitsrechte, namentlich das ‚Recht auf erschwertes Gefundenwerden‘ (Hans-Peter Thür): Auf Ersuchen der Betroffenen wird ihr Name aus dem Index für die Volltextsuche gelöscht.⁴

Im Rahmen dieses Projekts suchten die NB und ihre Partner auch nach einer technischen Lösung, um die digitalisierten Zeitungssammlungen online zu stellen. In einer Marktstudie bestimmten wir die Firmen, die in der Lage waren, die Konsultation von mehr als einer Million Seiten zu bewältigen, und wählten die Software aus, die das beste Preis-Leistungs-Verhältnis aufwies.

Die NB führte mit den Archivbibliotheken weitere Projekte durch. Dazu gehörte jenes mit den Neuenburger Bibliotheken und der Société Neuchâteloise

⁴ Die verschiedenen Vorlagen befinden sich auf der Plattform Digicoord: https://www.digicoord.ch/index.php/PSEL_-Juristische_Informationen. Den zitierten Begriff verwendete Thür als amtierender Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB) auf der infoclio-Tagung *Datenschutz und Geschichtswissenschaften* im November 2014 in Bern.

de Presse (SNP), das mit der Aufschaltung von *L'Express* (erscheint seit 1738) und *L'Impartial* (erscheint seit 1881), das heisst von mehr als eineinhalb Millionen Zeitungsseiten, erfolgreich abgeschlossen wurde. Die Sammlung wurde ab 2011 in Etappen auf einer eigenen Website online gestellt.⁵ Am 04. Juni 2013 wurde offiziell mitgeteilt, dass dieses Online-Archiv existiert.

Zu nennen ist auch die Digitalisierung der *Schaffhauser Nachrichten*, in Zusammenarbeit mit dem Verlag Meier+Cie AG, der Stadtbibliothek Schaffhausen und dem Staatsarchiv Schaffhausen. Für dieses Online-Angebot wurde eine eigene Website eingerichtet, auf welche ausschliesslich Abonnentinnen und Abonnenten der Zeitung sowie daran interessierte Bibliotheken zugreifen können.

Schweizer Presse Online

Eine Plattform entsteht

Bei Publikum und Forschenden lösten diese Angebote ein positives Echo aus. Doch bei aller Begeisterung wurde oft bedauert, dass die gleichzeitige Suche in mehreren Zeitungssammlungen nicht möglich war.

Manche Institutionen, die am Digitalisieren ihrer Zeitungen interessiert waren, äusserten den Wunsch nach einer zentralen Infrastruktur für die Präsentation der Digitalisate. Dafür waren sie bereit, die Mittel zusammenzulegen. Zwischen 2009 und 2010 erhielt die NB vier Anfragen, die in diese Richtung gingen. Eine Umfrage bei einigen weiteren Bibliotheken bestätigte das Bedürfnis.

Aufgrund dessen prüfte die NB die Vorteile einer zentralen Infrastruktur für Bibliotheken und Publikum. Sie ergriff die Gelegenheit zur Zusammenarbeit mit der Mediathek Wallis (MV), die 2010 die Digitalisierung des *Confédéré* und verwandter Titel mit dem Ziel in Angriff nahm, eine gemeinsame Plattform für den Zugriff einzurichten. In einer erneuten Marktstudie untersuchten wir die Angebote für die digitale Präsentation von Zeitungen und wählten dieselbe Firma aus, mit der wir bereits bei früheren Projekten zusammengearbeitet hatten. Ihre Software bot – zu vertretbaren Kosten – die von den Projektpartnern gewünschten Funktionen: die Volltextsuche, die Anzeige der Treffer in Form einer Liste, die Möglichkeit zum Verfeinern des erhaltenen Resultats (mittels zusätzlichem Suchbegriff oder Filter für den Zeitraum oder andere Merkmale), die Möglichkeit zum Blättern in den Ausgaben eines Titels, die Funktionen für das Exportieren und das Teilen von Daten.

5 <http://www.lexpressarchives.ch> oder <http://www.limpartialarchives.ch>.

Da die NB im Bundesinformatikumfeld nicht in der Lage war, eine preislich interessante Lösung anzubieten, strebten wir eine Kooperation mit der Zentrale des Westschweizer Bibliotheksverbundes (RERO) an. Zusammen haben wir ein Modell auf die Beine gestellt, das bis heute funktioniert: Jede Institution, die der Plattform beiträgt, unterzeichnet mit RERO und der NB einen Vertrag, der die Rechte und Pflichten aller Beteiligten festhält. Ein Anhang enthält die Liste der Titel und deren geschätztes Datenvolumen in Terabytes. Sie wird mit jedem neuen Projekt aktualisiert.

Im Pilotprojekt des *Confédéré* (mit insgesamt sechs Titeln) konnten alle Prozesse eingerichtet werden, die für das Online-Stellen einer Zeitung nötig sind: das Abklären der Urheberrechte, die Vorbereitung der Originaldokumente, die Digitalisierung, die Qualitätskontrolle, die Segmentierung, die Übertragung und die Bearbeitung der Dateien, die graphischen Aufbereitungen, die individuelle Adressierung der Titel, die Indizierung und die Schlusstests. Am 28. September 2011 wurde die Plattform Schweizer Presse Online (SPOL) im Rahmen der Feierlichkeiten zum 150-Jahr-Jubiläum des *Confédéré* offiziell lanciert. Im Jahr 2012 wurden erstmals Titel aus den Kantonen St. Gallen und Graubünden auf der Plattform veröffentlicht, 2014 gefolgt von Zeitungen aus Neuenburg, 2015 aus Zug und Freiburg.

Im Juni 2017 zählte SPOL 45 Titel aus sechs Kantonen (sowie den ‚nationalen‘ Titeln, die sich keinem Kanton zuordnen lassen), was rund 2,5 Millionen Seiten entspricht.

Evaluation

Die digitale Welt wandelt sich rasch und die Lösungen müssen regelmässig evaluiert werden, um auf dem neusten Stand zu bleiben.

2012 startete die Schweizerische Konferenz der Kantonsbibliotheken (SKKB) ein grosses Projekt zur Digitalisierung von Zeitungen. Ein Teilprojekt trug den Titel „Den Zugriff zu den digitalisierten Zeitungen vereinheitlichen“. Im Verlauf des Jahres 2013 definierte eine Arbeitsgruppe die Anforderungen an Funktionen, Architektur und Vertragsbedingungen, die ein System zur Online-Konsultation von Zeitungen erfüllen muss. Auf dieser Grundlage testete die Gruppe die Funktionen von fünf Systemen. Sie kam zu dem Schluss, dass ein Wechsel der Software keine signifikante Verbesserung bringen würde, womit eine Migration und der damit verbundene Arbeitsaufwand nicht zu rechtfertigen seien.⁶

⁶ Die Ergebnisse dieses Teilprojektes sind auf Digicoord zugänglich: https://www.digicoord.ch/index.php/PSEL_-_Zugriff.

2015 wurde eine HTML5-Version für SPOL installiert. Die Plattform kann seither auch auf Geräten mit iOS-Betriebssystemen konsultiert werden, die Antwortzeiten und der Komfort beim Navigieren wurden beträchtlich verbessert. Die neue Version zeigte uns, dass die Segmentierung der digitalisierten Seiten nicht optimal war. Das bewog uns zu einer neuen Marktstudie, aufgrund welcher wir uns nun für eine Migration der Plattform entschieden haben.

Perspektiven

Die Migration, an der wir zurzeit arbeiten, sollte uns erlauben, einige Erwartungen zu erfüllen, die namentlich aus der wissenschaftlichen Forschung an uns herangetragen werden.

Erkennung der Strukturelemente

Was regelmässig angefragt wird, ist die Möglichkeit, in den Zeitungssammlungen statistische Analysen vorzunehmen. Um relevante Ergebnisse zu erhalten, ist es ausserdem nötig, auf die Elemente, die eine Zeitung strukturieren (Artikel, Anzeige, Bild etc.), gezielt zugreifen zu können. Aus diesem Grund haben wir für die nächsten Titel, die für SPOL vorgesehen sind, eine feinere Strukturierung gewählt, welche die Erkennung dieser Elemente erlauben wird. So wird die Suche mittels eines oder mehrerer Stichwörter als Resultat die ‚Artikel‘ ergeben, in denen alle Elemente vorkommen, und nicht mehr die ‚Seiten‘, auf welchen sie – möglicherweise in mehreren Artikeln separat – genannt werden. Die Relevanz dieser Resultate wird damit wesentlich grösser sein, ist doch die logische Grundeinheit eines Presseerzeugnisses der einzelne Artikel.

Korrektur der automatisch bearbeiteten Zeitungstexte

Je weniger Fehler der automatisch erkannte Text enthält, desto zuverlässigere Ergebnisse ergibt die Suche. Doch die automatische Texterkennung (OCR), mit welcher die digitalisierten Zeitungssseiten bearbeitet werden, ist weit davon entfernt, perfekt zu sein: Viele Wörter werden gar nicht oder nur teilweise richtig erkannt. Mit der erwähnten Migration möchten wir zum einen den Benutzenden die Möglichkeit geben, Artikel zu korrigieren, und zum anderen Werkzeuge zur Verbesserung des OCR-Verfahrens anwenden, die in zahlreichen Forschungsprojekten entwickelt werden.

Mit diesen Massnahmen hoffen wir, die Qualität der angebotenen Digitalisate deutlich verbessern zu können.

Erweiterung des Korpus

Mit 45 Titeln und 2,5 Millionen Seiten ist SPOL eine sehr reichhaltige Informationsquelle. Sie leidet aber an einem Ungleichgewicht bezüglich der Sprachregionen, stammt doch die Mehrzahl der Titel aus der Romandie. Die Erklärung dafür liegt darin, dass der Verband der Westschweizer Presseverlage (Médias Suisses) seine Mitglieder stark ermuntert hat, den Bibliotheken die Digitalisierung und Online-Stellung der Zeitungstitel zu ermöglichen. Sein Deutschschweizer Pendant, Schweizer Medien, ist gegenüber solchen Aktivitäten wesentlich zurückhaltender. Dennoch wird mit den laufenden Projekten der Zentralbibliothek Zürich (im Rahmen von DigiTUR) und der Universitätsbibliothek Bern das deutschsprachige Publikum in naher Zukunft über ein Korpus verfügen, das seinen Informationsbedürfnissen besser entspricht. Die Kantone Jura und Genf sehen ebenfalls vor, in den nächsten vier Jahren zahlreiche Titel in SPOL zu integrieren.

Private Nutzung, Nutzung für die Forschung

Die Vereinbarungen zwischen den Bibliotheken und den Verlagen sehen eine strikt private Nutzung dieser Sammlungen vor. Auf dieser Grundlage ist es nicht möglich, den Forschenden das Material in Form eines Datenpools für die Analyse und Bearbeitung (Big Data, Digital Humanities) zur Verfügung zu stellen, ohne dass diese dafür erneut mit den Verlagen in Verhandlung treten müssen.

Le Temps hat diesen Schritt gemacht und hat dem DHLAB (Digital Humanities Laboratory) der ETH Lausanne erlaubt, die digitalisierten Versionen des *Journal de Genève*, der *Gazette de Lausanne* und des *Nouveau quotidien* für ein Forschungsprojekt zu verwenden. In diesem Projekt wurden neue Schnittstellen für die Konsultation dieser Zeitungen geschaffen,⁷ in welche die Forschungsergebnisse teilweise eingeflossen sind:

- Ein Graphikbetrachter vom Typ N-Gramm macht es möglich, die Entwicklung des Gebrauchs eines Wortes oder einer Wortgruppe innerhalb des Textkorpus darzustellen;
- ein Zugang für die Abfragesprache SPARQL (SPARQL Endpoint) ermöglicht, Suchanfragen über die 50 Millionen Entitäten, die während der Arbeiten definiert wurden, durchzuführen und die Ergebnisse zu visualisieren.

Für die Zeitungen, die aktuell auf SPOL sind, ist dieser Schritt noch zu tun. Angesichts der zunehmenden Nachfrage seitens der Forschenden sieht die NB vor, diese Art der Nutzung der Zeitungen in die nächste Modellvereinbarung mit den

⁷ <http://www.letempsarchives.ch/>.

Verlagen aufzunehmen. Auf jene Verlage, die ihre Einwilligung für den Online-Zugriff bereits erteilt haben, werden wir aktiv zugehen.

Indizierung mittels Suchmaschinen

SPOL wird von den Suchmaschinen zurzeit nicht indiziert. Andere Anwender der von uns verwendeten Software sind vor einigen Jahren erfolglos bei Google vorstellig geworden. Von sich aus haben die NB und ihre Partner noch nichts in dieser Richtung unternommen. Eine Recherche beginnt heute fast immer in Google. Was so nicht zu finden ist, „existiert nicht“. Es lohnt sich darum, anlässlich der Migration zu prüfen, welche Möglichkeiten – und auch Risiken – eine solche Indizierung bieten könnte.

Integration von SPOL in Portalen

Die Migration der Zeitungen von SPOL wird auch zur Folge haben, dass standardisierte Metadaten gemäss dem im Projekt Europeana Newspapers⁸ definierten Modell METS/ALTO vorhanden sind. Damit steht einer Zusammenarbeit mit diesem Portal technisch nichts mehr im Wege. Die Integration in andere – nationale oder internationale – Portale wird geprüft, sobald sich die Gelegenheit dazu ergibt.

Schluss

SPOL ist das sichtbare Ergebnis einer erfolgreichen Zusammenarbeit zwischen der NB und den Schweizer Gedächtnisinstitutionen. Sie ist getragen vom gemeinsamen Willen, Ressourcen und Kompetenzen zusammenzulegen, um einen einheitlichen Zugang zu den Zeitungssammlungen anbieten zu können.

SPOL ist auch ein beispielhaftes Modell einer Partnerschaft zwischen öffentlicher und privater Hand. Es hat ein beinahe einzigartiges Ergebnis hervorgebracht: das Angebot von jungen Zeitungen, von denen einige noch immer erscheinen, ohne Zugangsbeschränkungen im Internet. Ohne das Wohlwollen der Presseverlage, die der NB und ihren Partnern die Genehmigung erteilten, ihre Titel zu digitalisieren und online zu stellen, wäre (fast) nichts zu machen gewesen. Die frei zugänglichen Sammlungen würden sich auf einige alte Titel beschränken oder die Konsultation wäre, wie in einigen anderen Ländern, nur in bestimmten Institutionen (Universitäten, Nationalbibliothek) oder nur gegen Bezahlung möglich.

⁸ <http://www.europeana-newspapers.eu/>.

Die nächsten grossen Herausforderungen sind die Erweiterung der Nutzungsmöglichkeiten für Forschungsprojekte und die Aufnahme von Deutschschweizer Titeln in die Plattform SPOL.