

Francia – Forschungen zur westeuropäischen Geschichte

Bd. 40

2013

DOI: 10.11588/fr.2013.0.40967

---

#### Copyright

Das Digitalisat wird Ihnen von perspectivia.net, der Online-Publikationsplattform der Max Weber Stiftung - Deutsche Geisteswissenschaftliche Institute im Ausland, zur Verfügung gestellt. Bitte beachten Sie, dass das Digitalisat urheberrechtlich geschützt ist. Erlaubt ist aber das Lesen, das Ausdrucken des Textes, das Herunterladen, das Speichern der Daten auf einem eigenen Datenträger soweit die vorgenannten Handlungen ausschließlich zu privaten und nicht-kommerziellen Zwecken erfolgen. Eine darüber hinausgehende unerlaubte Verwendung, Reproduktion oder Weitergabe einzelner Inhalte oder Bilder können sowohl zivil- als auch strafrechtlich verfolgt werden.

TORSTEN SCHASSAN

## WIE KÖNNEN DIE BIBLIOTHEKEN DAS EDIEREN DIPLOMATISCHER QUELLEN UNTERSTÜTZEN?

### Was bedeutet »Edieren«?

»Digitales Edieren«<sup>1</sup> bedeutet, in einem ersten Schritt eine Ansammlung von Materialien verschiedenster Art in Ebenen zu organisieren und auf vielfältigste Weise miteinander zu verknüpfen. Im besten Fall erlaubt bereits die Organisation, das In-Beziehung-Setzen und die Präsentation dieser Daten die Gewinnung von Wissen aus den Editionsmaterialien, im Zweifelsfall werden aber in einem zweiten Schritt das Wissen des Editors über die Quellen und deren Zusammenhang sowie erste Interpretationen explizit in das Material eingearbeitet und in Form von (klassischen) Apparaten oder als Einleitungstexte verfügbar gemacht<sup>2</sup>.

Betrachtet man die Materialarten, die beim Edieren versammelt werden, so sind dies heutzutage typischerweise

- das Digitalisat der Quelle(n),
- Strukturdaten<sup>3</sup> als Inhaltsverzeichnis zur Masse der Abbildungen,

- 1 Zu unterscheiden wären die »digitalisierte Edition« und die »digitale Edition«. Der Unterschied zwischen beiden ergäbe sich mindestens dadurch, dass die digitalisierte Edition aus dem Buchdruck ins Digitale überführt worden ist und dabei alleine die Funktionalität der Edition im Buchdruck nachzuahmen versucht, während die digitale Edition Funktionalitäten besitzt, die nur unter Verlusten in die gedruckte Form überführt werden können. Vgl. dazu Patrick SAHLE, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Bd. 2: Befunde, Theorie und Methodik, Norderstedt 2013, S. 125–155, bes. 148ff. Schon Robinson verweist auf die erweiterten Nutzungsmöglichkeiten der digitalen Edition als einer, die »must offer all [...] in a manner which enriches reading«. Peter ROBINSON, *What is an Electronic Critical Edition?*, in: *Variants. The Journal of the European Society for Textual Scholarship* 1 (2002), S. 51–57.
- 2 Im schlimmsten Fall verbleibt das Material der Edition in einem Zustand der Materialansammlung, ohne noch mehr eine Synthese oder Auswertung zu erfahren.
- 3 Während die meisten der vorgenannten Materialarten ziemlich selbsterklärend sein dürften, wird der Begriff »Strukturdaten« mit durchaus unterschiedlichen Konnotationen gebraucht. Ich möchte ihn daher kurz näher beleuchten bzw. diskutieren. In Bezug auf die Funktion von Strukturdaten besteht Einigkeit darin, dass Strukturdaten die Navigation in den Digitalisaten ermöglichen und durch Verlinkung zu einzelnen Bildern schnelle Zugriffsmöglichkeiten auf das Digitalisat bieten. Die Strukturdaten sollten zu diesem Zweck mindestens die Konkordanz von Bildnamen zu Folierung, Paginierung bzw. anderen Benennungsschemata für einzelne Seiten enthalten. Darüber hinaus sind folgende Informationen wünschenswert: Nachweis textueller Einheiten wie Bücher, Kapitel etc., vorzugsweise mit Überschriften; Auflistung von Abbildungen, gegebenenfalls mit Beschreibung des Bildinhaltes oder einer kunsthistorischen Klassifikation; Verzeichnung sonstiger Besonderheiten, wie beispielsweise Annotationen, Glossierung, etc. Während diese Kategorien vor allem für Bücher, insbesondere Handschriften, gedacht sind, spielen sie bei der Erschließung diplomatischer Quellen oft eine eher untergeordnete Rolle. Durchaus unterschiedliche Vorstellungen gibt es aber davon, welche Qualität diese

- die Beschreibung der Quelle(n) samt Einordnung in die Überlieferung,
- Transkription(en) bzw. Editionstexte,
- kritischer, Sach- und andere Apparate sowie
- Übersetzung(en),
- sonstige Materialien wie Einleitung, Bibliographie und Register.
- Im Falle der Edition diplomatischer Quellen kommt noch das Regest hinzu.

Das Ergebnis des »Edierens«, die Edition, ist zugleich eine Momentaufnahme der wissenschaftlichen Auseinandersetzung mit, Bearbeitung und Erschließung von einer Quelle<sup>4</sup>.

### Was ist das Besondere an der Edition diplomatischer Quellen?

Welche Besonderheiten weist die Edition diplomatischer Quellen gegenüber der Edition anderer, handschriftlicher oder gedruckter Quellen auf?

- Die Materialarten sind weitgehend gleich, in der diplomatischen Edition wird das Regest weitaus häufiger verwendet.
- Die textuellen Strukturen sind unterschiedlich und liegen oftmals unterhalb der in Bibliotheken gewöhnten Textebenen. Dies erfordert eigene Referenzierungsziele.
- Die Gewichtung einzelner Informationseinheiten ist deutlich unterschiedlich. Diplomatische Quellen müssen weitaus stärker kontextualisiert werden, da sie anders als »literarische« Texte abhängig sind von bzw. Bezug nehmen auf Informationen, die außerhalb des Textes liegen, beispielsweise Personen, Institutionen etc.
- Die Relationen zwischen Quellen und Kontexten sind – durch den in der Regel gegebenen Realitätsbezug – vielfach deutlich anders.
- Die Überlieferungssituation dürfte bei diplomatischen Quellen noch kritischer sein als die anderer Quellen: Durch die faktische Bedeutung sind die Zeugnisse wichtiger, für Betroffene ist entweder die Bewahrung, in vielen Fällen aber auch explizit die Vernichtung von größtem Interesse gewesen.

Daten haben sollen. Eine Position ist die, dass möglichst viele dieser Informationen aus den beschreibenden, bibliographischen Metadaten in die Strukturdaten zu überführen seien. Das bedeutet, dass die Gewinnung von Strukturdaten mithin keine wissenschaftliche Tätigkeit wäre. Dies trifft tatsächlich vor allem dann zu, wenn einfache Klassifikationssysteme wie das Vokabular des DFG-Viewers angewendet werden: <http://dfg-viewer.de/strukturdatenset>. Strukturdaten wären in diesem Sinne eine die Komplexität reduzierende Beschreibung von (digitalen) Objekten. Die andere Position sieht in den Strukturdaten einen Vorgriff auf oder ein Extrakt aus dem Volltext der digitalisierten und edierten Quelle, da der Volltext Verweise auf Seitenumbrüche, die Textstrukturen, Überschriften etc. bereits enthält. Die Frage ist darum, ob Strukturdaten in normalisierter Form nach den Katalogen bzw. vorgegebenen Vokabularen gemäß oder als quellennahe Transkription enthalten sein sollen. Die Entscheidung für die eine oder andere Position muss wohl dem zu edierenden Material und der Projektökonomie gemäß gefällt werden. Im Falle der Erschließung diplomatischer Quellen entspricht die Problemstellung in Bezug auf die Strukturdaten den unterschiedlichen Ansätzen der Transkription vs. der Regestierung.

4 Dieser Zusammenhang ist im englischen Begriff für die Ausgabe (*edition*) deutlicher zu sehen als im deutschen.

## Wie können Bibliotheken das Edieren unterstützen?

### *Quellen zugänglich machen*

Die Bibliothek ist ein Ort der Quellen: handgeschriebener, gedruckter, in Mikroformen. Auch diplomatische Quellen werden nicht selten in Bibliotheken aufbewahrt. Diese Quellen zugleich mit der Forschungsliteratur zugänglich zu machen, ist eine erste Leistung, welche Bibliotheken erbringen können.

Zunehmend werden die Bibliotheken auch der Ort digitaler Quellen. Die Leistungen der Bibliotheken im Bereich der Digitalisierung beschränken sich dabei nicht auf die reine Image-Digitalisierung, sondern erweitert sich zunehmend auch in den Bereich digitaler Volltexte, die mindestens als Suchindizes angeboten werden und die reine OPAC-Suche ergänzen. Und so wie Bibliotheken immer schon die eigenen Bestände bewahrt, betreut und ergänzt haben, so setzt die Herzog August Bibliothek Wolfenbüttel (HAB) dies für digitale Quellen und Literatur fort. Die digitalen Ressourcen, welche einen Bezug zu den eigenen Beständen haben, werden wenn möglich gespeichert und in der Wolfenbütteler Digitalen Bibliothek (WDB)<sup>5</sup> für alle Benutzer zugänglich gemacht. Bibliotheken, die bereits eigene digitale und digitalisierte Quellen zugänglich machen, bieten immer häufiger auch anderen, kleineren Institutionen Platz an, deren Dokumente bei sich abzulegen.

Für die Publikation digitalisierter Quellen des historischen Kulturerbes ist eine Eigenschaft der digitalen Ressource, der digitalen Bibliothek, von entscheidender Wichtigkeit: die Adresse. Um als Ressource auch von anderen Webseiten und Angeboten weitergenutzt werden zu können, sind einerseits stabile Adressen in Form von URNs<sup>6</sup> oder auch PURLs<sup>7</sup> notwendig, andererseits sollten aber auch einzelne Seiten nicht nur stabil erreichbar, sondern die Adresse auch vorhersehbar oder errechenbar aufgebaut sein. Wünschenswert ist daher die Gestaltung der URN/URL unter Bezugnahme auf die Art der »kanonischen« Adressierung, die jedem historischen Material ohnehin innewohnt<sup>8</sup>.

Nur wenn die URN/URL vorhersehbar bzw. berechenbar ist, indem beispielsweise die Signatur zum Bestandteil wird, ist es möglich, automatisiert aus einer eigenen Ressource auf ein Digitalisat zu verweisen und ein Ziel zu erreichen. Datenbanken, welche mit lokalen IDs organisiert sind und keine Suche ermöglichen, an der ein kanonischer Identifikator zu einem unmittelbaren Ergebnis führt, erreichen nicht dasselbe Ergebnis, sondern zwingen den Benutzer, eine händische Suche abzusetzen oder einen browsenden Zugang zu wählen. Damit sind Maschinen von der Nutzung des eigenen Angebots ausgeschlossen.

5 <http://diglib.hab.de>.

6 URN = Unified Resource Name. Die URN realisiert die Idee, dass einer Ressource eine eindeutige Identifikation zugewiesen wird, welche sich dann über einen Resolver-Mechanismus in eine reale Webadresse auflösen lässt. Die URN wird im Bibliotheksbereich an die Deutsche Nationalbibliothek (DNB) gemeldet und kann von dort über einen Resolver aufgerufen werden.

7 PURL = Permanent Uniform Resource Locator. Die PURL zielt auf die dauerhafte Verfügbarkeit einer Ressource am angegebenen Ort. Die HAB garantiert in einer eigenen Erklärung die Zugänglichkeit der Drucke und Handschriften im Internet durch persistente URLs und stellt damit die Zitierbarkeit der elektronischen Dokumente sicher: <http://diglib.hab.de/?link=033>.

8 Beispielsweise werden Handschriften in der Regel mit dem Namen der Institution und gegebenenfalls dem Ort der Institution sowie der Signatur zitiert. Die Signatur wird bei sehr berühmten Handschriften gelegentlich durch einen Namen ersetzt. Als weitere Bestandteile können der Name einer Sammlung und eine einzelne Seitenzahl hinzukommen. In der HAB werden folgende Regeln der Normalisierung einer Signatur angewendet: Punkte und Leerzeichen werden durch Bindestriche ersetzt; mehrfach hintereinander auftretende Bindestriche werden zu einem einzigen zusammengefasst, Bindestriche am Ende werden weggelassen; der Name der Sammlung wird in abgekürzter Form verwendet; das hochgestellte »o« in der Größenangabe (2° etc.) wird durch ein »f« ersetzt.

### *Erschließungsinstrumente bereitstellen*

Bibliotheken haben große Erfahrung in der Erschließung und Katalogisierung. Und Bibliotheken haben eine große Menge an Daten als Ergebnisse dieser Erschließung anzubieten. Diese Daten zur weiteren Nutzung zugänglich zu machen, bleibt eine der größeren Aufgaben der (mittelbaren) Zukunft. Unter dem Stichwort »Open Linked Data« werden z.B. Kataloginformationen freigegeben<sup>9</sup>.

Von besonderer Bedeutung für die Erschließung diplomatischer Quellen, insbesondere in Hinblick auf die Kontextualisierung, sind personenbezogene Informationen. Die Identifikation einer Person, die in den Quellen beispielsweise als Zeuge, unterzeichnend oder Streitpartei vorkommt, kann von entscheidender Wichtigkeit für die Interpretation der Quelle sein. Personen zu individualisieren, also von allen anderen Personen gleichen Namens zu unterscheiden, ist das Ziel der Personennamendatei (PND)<sup>10</sup>. Da die Individualisierung möglichst wenig Fehler produzieren sollte, hat man seitens der Betreiber dieser Datenbank relativ hohe Hürden aufgebaut, um Informationen hinzufügen zu können. Beiträge zu dieser Ressource sind überwiegend Bibliotheken.

Allerdings gibt es einen immer stärker werdenden Trend, personenbezogene Daten zu historischen Persönlichkeiten, die in Forschungsprojekten erhoben werden, zu sammeln und auch von außerhalb des Bibliothekswesens in die PND einzuspielen. Dies trifft auch auf Bibliotheken wie die HAB zu, in denen solche Forschungsprojekte durchgeführt werden. Die PND, die zu Beginn eine Sammlung nur zu Personen gewesen ist, die als Autoren in Erscheinung getreten sind, wird auf diese Weise zu einer wertvollen Ressource von Personennamen und zur Identifizierung der dahinterstehenden Personen.

### *Standards setzen und betreuen*

Voraussetzung zur Nachnutzung der genannten Ressourcen sind standardisierte Formate und Schnittstellen. So beruht beispielsweise der Erfolg der Handschriftenkatalogisierung zu weiten Teilen auf Standardisierung, welche die Vergleichbarkeit und Ähnlichkeit der Katalogisate gewährleistet. Die Standardisierungsbemühungen betreffen im Fall der HAB aber nicht nur die Katalogisierungsrichtlinien, sondern auch die Kodierung der Katalogisate im Format, zum Beispiel im Format TEI<sup>11</sup>. Dies zielt zum einen auf die Weiterentwicklung des Standards<sup>12</sup>, aber auch auf die Betreuung von Schnittstellen, welche das Format TEI als Austauschformat nutzen. Solche Schnittstellen gibt es für die Handschriftendatenbank Manuscripta Mediaevalia<sup>13</sup>,

- 9 Was »frei« tatsächlich bedeutet, wird je nach Institution immer noch unterschiedlich interpretiert und ist vielfach mit großen Vorbehalten belegt. Die Tatsache, dass Drittmittelgeber wie die DFG verstärkt eine »freie« Zugänglichkeit der Daten, beispielsweise über das OAI-Protokoll einfordern, bedeutet nicht, dass die OAI-Schnittstelle selbst dann zugänglich ist, dass also die Adresse, an der man die Daten abholen könnte, bekannt gemacht wird.
- 10 PND ist die historische Bezeichnung für die Ressource. Im Bestreben, verschiedene Normdaten zusammenzuführen, hat die DNB u.a. die PND in der Gemeinsamen Normdatei (GND) aufgehen lassen. Dennoch wird häufig das Kürzel PND weiter verwendet.
- 11 Das Akronym TEI steht sowohl für die Organisation Text Encoding Initiative als auch für die Richtlinien Text Encoding und Interchange, die zurzeit in der Version P5 vorliegen und ständig weiterentwickelt werden: <http://www.tei-c.org> bzw. <http://www.tei-c.org/Guidelines/P5/>.
- 12 Gemeint ist hier insbesondere das Modul »msdescription«. Die Angebote der TEI, zunächst Handschriftenbeschreibungen kodieren zu können, sind erst 2004ff. integriert worden. Zu diesem Zweck wurde auf »Vorarbeiten« aus dem EU-Projekt MASTER – Manuscript Access through Standards for Electronic Records zurückgegriffen und für die TEI adaptiert.
- 13 <http://www.manuscripta-mediaevalia.de>. Die Schnittstelle ist sowohl für den Editor ManuscriptumXML (MXML) als auch für die Präsentationsdatenbank Advanced Publishing System (APS)

zur Europeana<sup>14</sup> und für die Darstellung im DFG-Viewer<sup>15</sup>. Nicht zuletzt wird auch die Verwendung der Standards standardisiert: In der HAB gibt es eine Arbeitsgruppe »Digitale Editionen«, welche die Verwendung der Formate projektübergreifend zu koordinieren versucht und die Diskussionsergebnisse auch dokumentiert<sup>16</sup>.

Wenn über Standards und deren Nutzen gesprochen wird, dann kommt die Rede meist und schnell auf Langzeitsicherheit<sup>17</sup>. Standardisierte Daten sind im Sinne der Interpretierbarkeit langzeitsicherer als nicht-standardisierte, doch Langzeitsicherheit entsteht in erster Linie durch eine Institution, welche sich zu der Aufgabe bekennt und sie dadurch gewährleistet. Zum anderen ist für die Langzeitsicherheit unter anderem die Benutzung der Daten entscheidend; dies betrifft sowohl die Sicherheit der Bits als auch der Datenstrukturen. Ähnlich wie analoge Medien, die durch Nutzung »abnutzen« und für die die Nutzung potentiell eine Gefahr darstellt, können digitale Daten durch Nutzung zerstört werden. Aber: genutzte, mithin nachgefragte Daten werden besser gepflegt, früher migriert und mehr angereichert, genutzte Daten werden häufiger gespiegelt und dadurch gesichert. Nutzung der Daten kann man aber dadurch generieren, dass die Daten als Services zur Verfügung gestellt und zugänglich gemacht werden.

Bibliotheken sind von allen Gedächtniseinrichtungen die im Schnitt vielleicht langlebigsten und deshalb prädestiniert, bestimmte Arten von Informationen wie Normdaten zu sammeln und zu pflegen. Bibliotheken sind auch im Schnitt die am besten ausgestatteten Einrichtungen und haben damit am ehesten die Ressourcen, Daten auch zu generieren und zur Verfügung zu stellen, die von anderen gebraucht werden: Digitalisate, Literatur, Metadaten, Kontextinformationen. Die Leistung der Bibliothek wäre damit das Beitragen zur kollaborativen Tätigkeit im Prozess des Edierens. Wenn damit der Beitrag der Bibliotheken charakterisiert wäre, muss es zugleich im Interesse aller sein, wenn Editionen diplomatischer Quellen diese Angebote nutzen, Normdaten verwenden und nach Standards arbeiten. Das Mindeste (und zugleich das Beste), was Bibliotheken für das Edieren diplomatischer Quellen dann leisten könnten, in eigenen Projekten mit gutem Beispiel voranzugehen!<sup>18</sup>

erstellt worden. Für MXML gibt es eine Im- und Exportversion, während für APS lediglich eine Import-Schnittstelle gebraucht wird.

14 Diese wurde im Rahmen des EU-Projektes Europeana Regia in Zusammenarbeit mit The European Library (TEL) entwickelt. Vgl. <http://www.europeanaregia.eu> sowie <http://diglib.hab.de/rules/documentation/ER/>. Siehe auch <http://sourceforge.net/projects/eregia2eseedm/>.

15 <http://www.dfg-viewer.de>. Um den Datenaustausch, aber auch die Publikation digitalisierter Quellen zu standardisieren, wurden für die Formate METS/MODS zur Anzeige alter Drucke und aktuell das Format METS/TEI für die Präsentation von Handschriften Anforderungen an Schnittstellen formuliert, Kodierungsbeispiele zur Verfügung gestellt und im Falle der TEI ein Prototyp eines Konversionsskripts für den Export von TEI-Daten in das Format METS/TEI beigetragen. Die TEI wird in einen METS-Container eingebettet, da METS bedeutend besser für die Repräsentation komplexer digitaler Ressourcen und der Beziehungen ihrer Teile untereinander geeignet ist.

16 <http://diglib.hab.de/rules/documentation/>.

17 Gemeinhin werden zwei Arten der Langzeitsicherheit unterschieden, in der Diskussion aber gerne vermengt: Zum einen ist die reine Speichersicherheit gemeint, die auf den Erhalt des Bitstreams abzielt. Zum anderen ist die Sicherheit der Nutzbarkeit der Daten gemeint, mithin Datenformate, Austauschbarkeit etc.

18 Die HAB veröffentlicht seit 1. März 2013 als erste deutsche Bibliothek alle Daten der WDB sowie die an der OAI-Schnittstelle zur Verfügung gestellten Daten unter der Creative Commons Lizenz CC-BY-SA. Vgl. <http://diglib.hab.de/?link=012>.

