

Datum 27. April 2011 / hn

Entwicklung

Von Dr. Roland Kamzelak

Literaturarchiv und Internet

Arbeitstreffen der Exekutivgruppe in Marbach am Neckar (DLA)

14./15.3.2011 - Protokoll

**Deutsche
Schillergesellschaft e.V.**

Schiller-
Nationalmuseum

Deutsches
Literaturarchiv

Literaturmuseum
der Moderne

Teilnehmer:

Silke **Becker** (Deutsches Literaturarchiv Marbach)
silke.becker@dla-marbach.de

Dr. Ulrike **Bischof** (Klassik Stiftung Weimar, Goethe- und Schillerarchiv)
ulrike.bischof@klassik-stiftung.de

Harald **Kaluza** (Deutsches Literaturarchiv Marbach)
harald.kaluza@dla-marbach.de

Dr. Roland **Kamzelak** (Deutsches Literaturarchiv Marbach)
roland.kamzelak@dla-marbach.de

Heinz Werner **Kramski** (Deutsches Literaturarchiv Marbach)
heinz.werner.kramski@dla-marbach.de

Dr. Manfred **Koltes** (Klassik Stiftung Weimar, Goethe- und Schillerarchiv)
Manfred.Koltes@klassik-stiftung.de

Christian **Liedtke** (Heinrich Heine Institut, Düsseldorf)
christian.liedtke@duesseldorf.de

Maximilian **Schreiber** (Bayerische Staatsbibliothek München)
Maximilian.Schreiber@bsb-muenchen.de

Uta **Simmons** (Akademie der Künste, Berlin)
simmons@adk.de

Entschuldigt:

Sebastian **Meyer** (Staats-, Landes- und Universitätsbibliothek Dresden)
sebastian.meyer@slub-dresden.de

Dr. Jutta **Weber** (Staatsbibliothek Berlin)
jutta.weber@sbb.spk-berlin.de

Eingegangene Stellungnahmen:

- E-Mail von Oliver Duntze vom 8.3.2011 (Staatsbibliothek Berlin, nach Aufforderung durch Frau Weber).
- E-Mail von Jutta Weber vom 25.3.2011.
- E-Mail von Sebastian Meyer vom 14.3.2011

**1. Begrüßung und Einführung, Feststellen der Tagesordnung
(Roland Kamzelak)**

Herr Kamzelak begrüßt alle Anwesenden und dankt für die Bereitschaft, sich tiefer mit den in Frage stehenden Formaten zu beschäftigen. Die Stellungnahmen der entschuldigten Mitglieder Jutta Weber und Sebastian Meyer werden als Tischvorlage verteilt (Lesepause).

Einführend stellt Herr Kamzelak fest, dass sieben Gruppensitzungen auf Einladung von Herrn Koltes vorausgegangen sind, die bis in den Juni 2008 zurückreichen. Damals galt es, nach den Weimarer Empfehlungen von 2004 und dem Open Access Gebot der DFG einen Erfahrungsaustausch in Sachen Digitalisierung zu ermöglichen. Einige Sitzungen hinweg blieb das Thema sehr vage, spitzte sich dann aber ab Ende 2009, Anfang 2010 zu auf die Frage nach einem verbindlichen Metadataset für Digitalisate von Handschriften, das die Formulierung von DFG-Praxisregeln "Digitalisierung"¹ auch für diesen Bereich erlauben würde. Aus Sicht von Herrn Kamzelak wurde die Frage des Formats zu rasch zu Gunsten von EAD² entschieden, ohne dass die möglichen Formate praktisch getestet wurden. Er hatte auch den Eindruck, dass die Arbeitsgruppe in der letzten Sitzung „abgehängt“ worden ist. Die Entscheidung, einen DFG-Antrag vorzubereiten, mit dem der DFG-Viewer⁵ EAD-fähig gemacht werden sollte, kam daher zu schnell. Diese Zweifel äußerte er in einer ausführlichen E-Mail am 22.11.2011. Nicht recht bedacht schien ihm dabei die Neuentwicklung eines TEI-Sets, um alte Drucke im DFG-Viewer anzuzeigen. Auch diese Entwicklung wurde in der Gruppe nicht genügend reflektiert.

Der Vorschlag, eine „Exekutivgruppe“ zu bilden, um alle Formate auf den Prüfstand zu stellen, fand Konsens in der AG. Aufgabe ist die Vorbereitung einer Ent-

¹ http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung.pdf

² <http://www.loc.gov/ead/>

⁵ <http://dfg-viewer.de/>

scheidung über das weitere Vorgehen bzw. das Erarbeiten von Argumenten für und wider möglicher Formate.

Voraussetzungslos ist die Arbeit der Gruppe jedoch nicht, denn es gibt bereits zwei akzeptierte und erarbeitete Formate, die mit dem DFG-Viewer funktionieren: METS⁴ / MODS für Buchdigitalisate und METS / TEI⁵ für mittelalterliche Drucke. Für beide Formate sind Aggregationsinstrumente verfügbar oder im Entstehen. Somit steht jede neue Lösung unter starkem Erläuterungsdruck. Die Exekutivgruppe hat daher eine wichtige Funktion für den Fortgang der Erarbeitung eines Formats für Handschriftendigitalisate.

Herr Liedtke stimmt zu, dass der Eindruck der Sitzung in Dresden der von Unsicherheit war. „Abgehängt“ scheint das richtige Wort zu sein. Sich auf ein Minimalset zu einigen, ist für die community sinnvoll, aber alle in ein einziges Format zu zwingen, scheint zunächst falsch bzw. bedarf der eingehenderen Begründung.

Herr Schreiber äußert Verständnis dafür, dass Berlin offensichtlich seinen Hausstandard durchsetzen möchte. Es muss trotzdem Mapping ermöglicht werden, schließlich gibt es evtl. andere Hausstandards. Er begrüßt daher die Arbeit, die Formate zu sichten und kritisch anzuwenden, damit alle klarer sehen.

Die Tagesordnung wird wie vorgelegt genehmigt.

2. Der Probestand Hermann Hauff (Silke Becker)

Frau Becker stellt den Probestand Hermann Hauff im Original und im Scan vor. Er ist einerseits klein, enthält andererseits aber nahezu alle relevanten Strukturen. Dabei finden besonders die Unteraufnahmen Berücksichtigung, da sie mit separaten Datensätzen und Verknüpfungen auf die Hauptaufnahme katalogisiert werden.

⁴ <http://www.loc.gov/standards/mets/>

⁵ <http://www.tei-c.org/>

Interessant ist dabei ein Brief, der die Schreiberhand wechselt und ein gezeichnetes Bild enthält. Im Katalog ergibt dies drei unabhängige Aufnahmen.

5. Hauff in EAD (Harald Kaluza)

Bevor Herr Kaluza seine Umsetzung mit EAD erläutert, erläutert Herr Kamzelak kurz das Grundprinzip der Arbeit mit XML. Die Grundbegriffe Element, Attribut, Wert und Text werden differenziert.

Tischvorlage: EAD-Beispiel Hauff und Tabelle zur Darstellung der Praxisregel.

Herr Kaluza erläutert die Datei ausführlich Zeile für Zeile.

In der Diskussion wird festgestellt, dass in physDesc statt genreform lieber materialspec Anwendung finden sollte. Genre ist doch zu sehr inhaltlich zu verstehen.

Interessant war die Entdeckung, dass das Element persname in unittitel möglich ist: Der „Widerstand“ von Herrn Pahl (Bundesarchiv) in der letzten Sitzung war also unbegründet. Die Möglichkeit einer Annäherung scheint in EAD angelegt zu sein.

Die Abbildung der hierarchischen Struktur des Bestandes wurde über verschachtelte Elemente „c“ (mit Level Attribut) realisiert, so dass auch Unteraufnahmen gut abgebildet werden konnten. Wenn der Bibliothekskatalog selbst eine Levelzählung liefert, dann wäre auch c01, c02 sinnvoll.

Herr Kaluza stellt abschließend fest, dass EAD als Format sehr gut und ohne erkennbare Einschränkungen oder Verrenkungen geeignet ist. Bei einzelnen Feldern im vorliegenden Beispiel wären sicherlich auch alternative Elemente möglich gewesen. Es müsste nun METS für die Strukturdaten hinzukommen.

In der Diskussion wird nochmals der Unterschied zwischen Aggregationsinstrument, Viewer und Datei erläutert:

Die strukturierte Datei soll möglichst aus dem jeweiligen Bibliothekskatalog generiert werden. Dabei muss mindestens das Minimal-Datenset ausgegeben wer-

den, weil diese Felder für den DFG-Viewer (Anzeige) bzw. das Aggregationstool (Suchmaschine) notwendig sind.

4. Hauff in MODS (Heinz Werner Kramski)

Herr Kramski erläutert, dass das Anwenderprofil des ZVDD für Buchdigitalisate eine Ausformulierung oder Verengung des MODS-Standards für diesen einen Zweck ist. Ein solches Anwenderprofil sollten auch wir für unsere Zwecke erstellen.

Der DFG-Viewer erwartet für die Darstellung der Digitalisate zwei Auflösungsstufen: default und min, optional zusätzlich max, thumbs und download.

Herr Kramki legt ebenfalls einen Ausdruck der strukturierten Datei und eine Anwendertabelle vor. Deklariert sind METS (Struktur der Digitalisate) und MODS (Inhalt). Durch die Mischung der Standards werden Elemente jeweils durch die Präfixe mets: bzw. mods: eingeleitet (namespace). Das Nebeneinander mehrerer Standards in einer Datei ist unproblematisch möglich.

Das Aggregationsinstrument ZVDD⁶ verlangt zwingend URNs⁷ als Identifier. Es stellt sich die Frage, auf welcher Ebene diese URN ansetzt: Bestands- oder Dokumentenebene?

Es scheint sinnvoll, pro Dokument eine XML-Datei zu verwenden. Dann benötigt man mindestens eine übergeordnete Datei, die den Gesamtbestand beschreibt (vergleichbar dem Reihentitel bei Druckwerken). Dadurch würde z. B. die Aktualisierung von einzelnen Dokumenten erleichtert. Alle Hierarchien würden in dieser Variante über verknüpfte externe XML-Dateien abgebildet, nicht über MODS-interne Schachtelungen. Die Binnenstrukturen des komplexen Dokumentmodells scheinen ohnehin nicht zu logisch eigenständigen Unteraufnahmen zu passen.

⁶ <http://www.zvdd.de/>

⁷ http://www.d-nb.de/netzpub/erschl_lza/np_urn.htm

Herr Kramski zeigt auch eine funktionierende Datei im DFG-Viewer, ein Manuskript aus dem Bestand Schiller, weil dort gute Scans zur Verfügung stehen.

Herr Kramki stellt abschließend fest, dass METS / MODS als Format sehr gut und ohne erkennbare Einschränkungen oder Verrenkungen geeignet ist. Das ZVDD-Anwendungsprofil selbst gibt bereits Hinweise, wie mit seiner Hilfe Sammlungen beschrieben werden können.

In der Diskussion wird festgestellt, dass es bereits eine Conversion (mapping) von MARC⁸ zu MODS gibt (siehe Webpage).

5. Hauff in TEI (Roland Kamzelak)

Herr Kamzelak beginnt seine Präsentation mit dem Aufruf der Webseiten von EAD und MODS. Dabei hält er fest, dass beide Formate von der Library of Congress entwickelt wurden und dort gehostet werden. Die Text Encoding Initiative (TEI) ist eine Entwicklung der scientific community und nicht an ein Land oder eine Institution gebunden. Das Grundprinzip von TEI ist der modulare Aufbau der document type definition (dtd) oder des Schemas, was dem zuvor erläuterten Anwenderprofil von MODS entspricht. Die Community hat dabei eine Reihe von Spezifizierungen vorgenommen, die in DTDs beschrieben sind. Diese lassen sich modular kombinieren oder geregelt verringern (ODD = one document does it all). Von den vorhandenen Modulen hat sich MSDESCRIPTION als tauglich herausgestellt. Es wurde zwar an der europäischen Tradition der mittelalterlichen Drucke entwickelt, dient aber der Beschreibung von Handschriften.

Herr Kamzelak verteilt einen Ausdruck der strukturierten Datei mit TEI und erläutert diese Zeile für Zeile. Ergänzend wird die Anwendungstabelle ausgeteilt. Das Beispiel wurde nur bis zu einer Ebene ausgeführt, weil sich dann prinzipiell verschachtelte Wiederholungen in weiteren div-Elementen ergeben. Es kann zur nächsten Sitzung ergänzt werden.

⁸ <http://www.loc.gov/marc/>

Herr Kamzelak stellt abschließend fest, dass auch TEI als Format sehr gut und ohne erkennbare Einschränkungen oder Verrenkungen geeignet ist.

Schließlich wird noch das DOM (document object model), eine Darstellung des Textes als Baum, der EAD und der TEI Datei verteilt und erläutert. Beide Bäume sind ganz ähnlich und bedienen sich ähnlicher Terminologie. Auch diese Darstellung verdeutlicht die Tauglichkeit beider Formate.

6. Ergänzende Beispiele

Herr Schreiber gibt zunächst den Hinweis auf die Hannah Ahrend Papers <http://memory.loc.gov/ammem/arendthtml/arendthome.html>. Dann erläutert er anhand von Powerpointfolien den DFG-Antrag: Digitalisierung und suchmaschinenbasierte Recherche in den Nachweisinstrumenten für Nachlässe und Autographen der BSB. Findmittel der BSB sollen in ein Webfindmittel integriert werden. Für die Texterkennung der gedruckten Bestandslisten soll OCR eingesetzt werden mit 99% Erkennungsgüte, die Bestandsstrukturen sollen durch Verfahren der BSB-internen IT erkannt werden. Ein großer Teil liegt auch als MS-Word-Dateien vor. Die unkorrigierten Volltexte werden durch eine FAST-Suchmaschine indexiert. Die Nachlass-Titelaufnahmen werden als mehrbändige Werke im BVB erfasst und lokal als digitale Sammlung (<http://digi20.digitale-sammlungen.de/de/fontsize.1/page/project.xhtml>) wie auch in Kalliope präsentiert. Scans werden als PDF-Dateien zur Verfügung gestellt. Ein lokales Portal mit eigenen Daten der BSB und Daten aus Kalliope soll aufgebaut werden.

Gesamtaussprache:

Herr Kramski erläutert sein Unbehagen bei der Lösung, die Strukturdaten im DFG-Viewer für Bestandshierarchien zu verwenden: Navigiert man mit der rechten Leiste, bleibt bisher die Hauptbeschreibung oben stehen, statt eine Beschreibung des angezeigten Dokumentes, also einer unteren Ebene, zu zeigen. Dies ist unbefriedigend, weil für den Benutzer verwirrend.

Es wird festgestellt, dass alle Formate geeignet sind. Der Aufwand für unsere Definitionsarbeit (Minimalset, Nomenklatur, Anwendungsprofil bzw. DOM) ist

überall gleich. Auch die Implementierungsarbeit dieser Definition (mapping aus dem Katalog in die strukturierte Datei) ist überall gleich. Aber beim DFG-Viewer gibt es bereits Spezifikationen für MODS (fertig) und für TEI (Erprobungsphase). Es müssen evtl. Anpassungen gemacht werden. Für EAD gibt es noch keine Vorleistung, der Aufwand ist somit am höchsten.

Vertiefte Beispiele aus anderen Häusern sollten erarbeitet werden. Jedes Beispiel sollte dabei im Gespann von EDV und Bibliothekar erarbeitet werden.

Frau Simmons spricht das Kempowski-Beispiel der AdK an: Dort gibt es Querverbindungen zwischen Tonbändern und (veränderten / bearbeiteten) Transkriptionen. Es handelt sich also nicht um eine 1:1 Ableitung. Wie soll / kann das abgebildet werden? Der Weg zwischen Beschreibung (Katalog) und Edition ist fließend. Unser Ziel ist die Beschreibung der digitalisierten Objekte. Allenfalls noch die Ermöglichung der Weiterarbeit mit den strukturierten Texten.

Die AdK ist im Begriff, vom System Augias auf easyDB umzustellen. Kapazitäten für die Erarbeitung von strukturierten Texten seien deshalb nicht vorhanden. Herr Kamzelak macht das Angebot, ein Beispiel gemeinsam zu bearbeiten.

Es wird die Frage aufgeworfen, wie der Wunsch nach Nähe zu den Archiven zu bewerten ist? Normdaten (Personen und Körperschaften) stehen im Zentrum bei der Nachlasserschließung, im Archiv leitet die Sachakte. Die Gruppe sieht bei bestehender Namensgleichheit (Archiv) eine regelrechte Kluft zwischen der Archivwelt und der Nachlasswelt, die sich auch darin ausdrücke, dass Herr Pahl der Gruppe nicht beigetreten ist.

Herr Liedtke äußert seine Beruhigung darüber, dass alle Formate funktionieren. Durch die Beispiele sehe man klarer, wie sie funktionieren. Man könne nun freier den besten Weg auswählen.

Frau Simmons spricht das Problem Urheberrechte an. Kann man timeflags setzen? – Auch Marbach hat das Urheberrechtsproblem in starkem Maße. Man wird wohl erst gar keine Digitalisate ablegen, wenn Sie nicht frei sind. Bislang sind Benutzungsbeschränkungen als empfohlenes Feld vorgesehen.

Die Frage nach der Auflösungsqualität wird aufgeworfen. Die Antwort muss in der AG aufgegriffen werden.

Viele Texte sind nicht entzifferbar (Handschrift, Steno, Korrekturgänge usw.), ohne Transkription sind sie als Digitalisat nicht brauchbar, bleiben bloßes Bild. Es scheint fast geboten, immer zumindest auch Transkriptionen mit zu liefern.

Es sollten weitere Informationen eingeholt werden:

- Konsortium der Landesarchive? Gerald Meier oder Kretschmar? – Herr Kamzelak wird versuchen, Herrn Meier nach Düsseldorf einzuladen.
- TEI-Anwenderbeispiel für den DFG-Viewer? – Herr Kamzelak wird versuchen, Herrn Stäcker oder Herrn Schaßan nach Düsseldorf einzuladen.

7. Beantwortung der Arbeitsfragen

Ad 1) Was sollen die Metadaten für Digitalisate leisten?

1. Wir wollen nicht katalogisieren, sondern Digitalisate angemessen beschreiben! Diese Beschreibung soll aus dem Bibliothekskatalog generiert werden.
2. Eine nationale Suchmaschine für digitalisierte, literarische Nachlässe ermöglichen.
3. Die Anzeige im DFG-Viewer ermöglichen.
Als Desiderat werden ein strukturierter Suchbaum und die Einzelanzeige für untergeordnete Dokumente empfunden.
4. Es soll der Sprung vom DFG-Viewer in den jeweiligen Katalogdatensatz ermöglicht werden, da dort alle Informationen zum Dokument jeweils aktuell vorliegen.

Ad 2) Mit welchem Format lassen sich die Zwecke realisieren?

1. Alle Formate sind möglich: EAD, MODS und TEI, jeweils in Verbindung mit METS.

Ad 3) Wenn es mehrere Formate geben sollte, welchem sollte man warum den Vorzug geben? (pro und contra).

Format	Pro	Contra
METS / EAD	<ul style="list-style-type: none"> • Speziell ausgerichtet auf Archivdaten 	<ul style="list-style-type: none"> • Es gibt keine Vorleistung für den DFG-Viewer • Es gibt keine Vorleistung für ein Aggregationssystem • Für die Weiterarbeit (z. B. Transkription) ist ein Formatwechsel notwendig.
METS / MODS	<ul style="list-style-type: none"> • Es gibt eine erprobte Vorleistung für den DFG-Viewer • Es gibt eine Vorleistung für ein Aggregationssystem • Es gibt eine conversion von MARC in MODS. 	<ul style="list-style-type: none"> • Für die Weiterarbeit (z. B. Transkription) ist ein Formatwechsel notwendig.
METS / TEI	<ul style="list-style-type: none"> • Open community Entwicklung • Speziell für Manuskripte entwickelte DTD • Modularer Aufbau der DTD • Es gibt eine Vorleistung für den DFG-Viewer • Es gibt eine Vorleistung für ein Aggregationssystem • Ergänzung durch Transkriptionen im selben System möglich. 	

Ad 4) Was soll das Aggregationssystem leisten?

1. Das Aggregationssystem soll alle Handschriftendigitalisate auffinden können (zentrale Sucheingabe).
2. Es soll die Struktur ansprechend darstellen.

Ad 5) Gibt es ein geeignetes Aggregationsinstrument? Falls nein, was fehlt?

1. Wäre ZVDD auch für Handschriften geeignet (dann: ZentralVerzeichnis Digitaler Dokumente)?
2. Kalliope wäre möglich, müsste erweitert werden (Einschränkung der Suche nur auf Digitalisate. Link in den DFG-Viewer).

8. Weiteres Vorgehen

Herr Liedtke bestätigt, dass der 2./3. Mai im Plenum für das Heine Institut möglich wäre. Er wird dazu einladen. Die Gruppe sollte sich vorher treffen, um letzte

Ergebnisse aus weiteren Gesprächen zu besprechen und evtl. zu bewerten. Dazu wird der Montagvormittag vorgesehen. Plenum wie gewohnt um 14 Uhr bis zum nächsten Tag um 15 Uhr. Die Gruppe müsste dann bereits am Sonntag anreisen. Auf dem Programm der AG stehen dann Kurzberichte zu den Format-Beispielen, dann ein Bericht über das Für und Wider (Antworten auf die Fragen).

Vertiefende Arbeitsgespräche mit Herrn Meyer in Dresden erscheinen sinnvoll. Die drei „Formatierer“ Kramski, Kaluza, Kamzelak sollten ihre Spezifizierungen mit Herrn Meyer besprechen und Unklarheiten oder Alternativen zu klären versuchen. Herr Kamzelak wird gebeten, einen Termin zu suchen. Wenn weitere Interessenten aus der Gruppe diesen Termin möglich machen können, sollten sie dazu stoßen. Ansonsten wird in Düsseldorf am 2. vormittags berichtet.

Gerald Meier und Thomas Stäcker oder Herr Schaßan sollen nach Düsseldorf eingeladen werden (Kamzelak), um von Ihren Bemühungen in diese Richtung zu berichten.

Im Vorfeld der nächsten Sitzung soll eine Umfrage per Formular (Kramski) unter der AG gemacht werden, mit welchen Katalogisierungs-Instrumenten gearbeitet wird.

Herr Kamzelak schließt die Sitzung um 15 Uhr mit dem Dank für die sehr konzentrierte Arbeit an dieser doch sehr schwierigen Materie. Die zwei Tage haben sicher zum besseren Verständnis der Formate beigetragen. – Herr Koltes bedankt sich als Initiator und stellvertretend für die Gruppe für die Gastfreundschaft, die Magazinführung und die geleistete Arbeit an den Formaten.