

EAD-RLG-Profil im Vergleich zu EAD-BArch-Profil

Durchführung von Tests einiger EAD-Dokumente des Bundesarchivs (BArch) mit der von der RLG zur Verfügung gestellten Testsoftware „EAD Report Card“

Kerstin Arnold, Andres Imhof

EINLEITUNG.....	2
TESTSZENARIO.....	3
FEHLERMELDUNGEN.....	3
ZUSAMMENFASSUNG.....	5

Einleitung

Das Bundesarchiv setzt für die technische Umsetzung der Onlinefindbücher im DFG-Projekt „Ausbau des Portals „Netzwerk SED-/FDGB-Archivgut“ zu einer Referenzanwendung für ein Archivportal Deutschland“ – im Folgenden Verbundfindmittel genannt – den internationalen Standard EAD (Encoded Archival Description) für die Beschreibung von Archivgut ein¹. Bei EAD handelt es sich um einen Rahmen an definierten Elementen zur Beschreibung von Archivgut, das in XML-Syntax (eXtensible Markup Language) abgebildet wird. Neben den definierten Elementen werden auch Regeln der Wiederholbarkeit, der verpflichtenden oder freiwilligen Verwendung etc. aufgestellt, an die sich EAD-Anwendungen zu halten haben und die mit Hilfe des EAD-Schemas² auf ihre Korrektheit validiert werden. Trotz alledem lässt der EAD-Rahmen über diese Regeln hinaus viel Interpretationsspielraum, so dass sich Unterschiede zwischen einzelnen Anwendungen zwangsläufig ergeben, ja sogar ergeben müssen. Die jeweiligen nationalen und internationalen Umsetzungen werden darum als „EAD-Profile“ bezeichnet.

Die RLG (Research Libraries Group)³ ist eine der ersten Institutionen, die das EAD-Format in ihrem Portal „ArchiveGrid“⁴ umgesetzt hat. Das ArchiveGrid enthält laut eigener Aussage EAD-Findbücher von annähernd tausend internationalen Archiveinrichtungen, u.a. auch vom Bundesarchiv. In den RLG Best Practice Guidelines for Encoded Archival Description⁵, die von Oktober 2001 bis August 2002 von der RLG EAD Advisory Group entwickelt und 2002 veröffentlicht wurden, ist das RLG-eigene EAD-Profil näher erläutert.

Das Bundesarchiv hat im Jahr 2007 mit Unterstützung der Andrew W. Mellon-Stiftung, New York, im Projekt <daofind+>⁶ einen frei verfügbaren Editor namens MEX (MidosoEditor für XML-Standards) für die Erstellung integrierter Internetpräsentationen von Onlinefindbüchern mit digitalisierten Reproduktionen entwickelt. Hierfür wurde das EAD-BArch-Profil im MEX abgebildet. Vor der Programmierung des MEX wurde bereits der MidosoXML-Editor von der Archivschule in Marburg mit Unterstützung der DFG entwickelt⁷, der in der Lage ist, die darin verzeichneten Findbücher ebenfalls im EAD-BArch-Format abzuspeichern.

Bis Ende 2007 konnten EAD-Anwender die eigenen EAD-Findbücher über eine Online-Anwendung gegen das RLG-Profil testen. Mittlerweile wurde RLG von OCLC (Online Computer Library Center) übernommen⁸. In diesem Zusammenhang ist das Online-Testwerkzeug im Zuge der Migration der RLG-Website in die Website von OCLC bis zum aktuellen Zeitpunkt nicht mehr verfügbar. Dafür wird seit dem 04. April 2008 bis zur Wiedereinrichtung des Onlineangebotes die lokal einzusetzende Testsoftware „EAD Report Card“ im Internet zum Download angeboten⁹. Es handelt sich bei dieser Testsoftware zum einen um eine Serverversion und zum anderen um eine Version, die auf dem eigenen PC installiert werden kann. Im Rahmen des zu Anfang genannten DFG-Projektes wurde nun diese Testsoftware auf das EAD-BArch-Profil angewendet, um gegebenenfalls auf Fehler oder lediglich Unterschiede in den Profilen hingewiesen zu werden.

¹ <http://www.loc.gov/ead/>

² <http://www.loc.gov/ead/ead.xsd>

³ <http://www.oclc.org/programs/default.htm>

⁴ <http://archivegrid.org/web/index.jsp>

⁵ <http://www.oclc.org/programs/ourwork/past/ead/bpg.pdf>

⁶ <http://www.daofind.de/>

⁷ <http://www.archivschule.de/content/26.html>

⁸ <http://www.oclc.org>

⁹ <http://www.oclc.org/programs/ourwork/past/ead/eaddesktop1.0.zip>

TestszENARIO

Als Testdateien wurden EAD-Beispielfindbücher aus dem Bundesarchiv und dem Sächsischen Staatsarchiv, Hauptstaatsarchiv Dresden, das den MidosaxML-Editor für seine Onlinefindbücher einsetzt, gegen das EAD-RLG-Profil getestet. Die Onlinefindbücher aus beiden Archiven sollen zukünftig entsprechend in das Verbundfindmittel importiert werden. Zum Zeitpunkt des Tests waren noch keine originären, in MEX erschlossenen EAD-Findbücher oder bereits im für das Verbundfindmittel weiterzuentwickelnden MIDEX konvertierte Findbücher aus den Partnerarchiven verfügbar. Stattdessen wurden

- eine EAD-Feldbelegungsdatei, die im MEX neu angelegt wurde und in der alle geforderten und gewünschten Elemente im EAD-BArch-Profil mit deutschsprachigen, erklärenden Erläuterungen gefüllt sind,
- ein umfangreiches Findbuch des Bundesarchivs, das in MidosaxML erstellt, in das EAD-Format exportiert, in MEX importiert und schließlich wieder im EAD-Format abgespeichert wurde, und
- ein MidosaxML-Findbuch aus Dresden, das ebenfalls im EAD-Format abgespeichert wurde,

gegen das RLG-Profil getestet.

Zu den Tests ist insgesamt anzumerken, dass sehr häufig Elemente als fehlend aufgelistet werden, obwohl diese in den EAD-Dokumenten vorhanden sind. Dies gilt insbesondere für die EAD-Feldbelegungsdatei, aber bezüglich einzelner Elemente für alle Testdateien. Der Grund für diese fehlerhafte Aufführung fehlender Elemente ist nicht ersichtlich. Insofern sind insgesamt die Testergebnisse mit großer Vorsicht zu beurteilen. Zu einigen Elementen können jedoch Aussagen getätigt werden. So werden durch die Ergebnisse wie erwartet einige Unterschiede des EAD-BArch-Profiles zum EAD-RLG-Profil deutlich. Viele Elemente und Attribute, die von RLG verlangt werden, sind aus Sicht des deutschen Archivwesens nicht gefordert oder anders gelöst. Die folgende Liste der angezeigten Fehler kann keinen Anspruch auf Vollständigkeit erheben. Es sollen lediglich die eindeutigen Unterschiede und Fehler in beiden Profilen benannt werden.

Fehlermeldungen

1. Im Element *ead/eadheader* ist das Attribut *relatedencoding* „mandatory“ (obligatorisch) und soll entweder den Wert "MARC21", "DC" oder "Dublin Core" enthalten. Je nachdem, was an dieser Stelle angegeben wird, werden anschließend in den Unterelementen *filedesc/titlestmt/titleproper*, *filedesc/publicationstmt/publisher* und *filedesc/publicationstmt/date* im Attribut *encodinganalog* die Werte „245\$a“ oder „Title“, „260\$b“ oder „Publisher“ und „260\$c“ oder „Date“ verlangt. Laut EAD Tag Library, Version 2002¹⁰ ist dieses Attribut nicht verpflichtend und wird auch ohne feste Werte, lediglich mit PCDATA¹¹, angeboten.

2. Im Element *ead/eadheader* wird vom RLG-Profil das Attribut *repositoryencoding* mit dem Wert „iso15511“ verlangt. Das Attribut ist in den Beispielfindbüchern des Bundesarchivs auch enthalten (außer in der Feldbelegungsdatei), jedoch fälschlicherweise mit dem Wert "Archiv" gefüllt. Dies ist ein Fehler im bisherigen EAD-BArch-Profil und wird bei der zukünftigen Umsetzung berücksichtigt werden.

3. Als Fehler werden auch die nicht enthaltenen Attribute *publicid*, *identifier* oder *url* im Element "ead/eadheader/eadid" von dem Testwerkzeug „EAD Report Card“ angemerkt. Die Attribute *publicid* und *identifier* fehlen bislang im EAD-BArch-Profil, während das Attribut *url* verwendet wird. Das Attribut *identifier* ist bei der zukünftigen Umsetzung des EAD-BArch-

¹⁰ <http://www.loc.gov/ead/tglib/>

¹¹ PCDATA (Parsed Character Data). Erlaubt ist ein Text, in dem alle im Zeichensatz vorkommenden Zeichen enthalten sind sowie Zeichen wie z.B. < und Unicode Entities.

Profils im Verbundfindmittel bereits mit dem Wert eines eindeutigen Identifiers berücksichtigt. Dagegen wird das Attribut *publicid* im Bundesarchiv als nicht notwendig angesehen.

4. Auch das Attribut *mainagencycode* im gleichen Element *eadid* wird mit Recht obligatorisch verlangt, ist aber in den Beispielfindbüchern des Bundesarchivs noch nicht vorhanden. Auch dieses Attribut wird zukünftig im Verbundfindmittel mit dem eindeutigen Identifier des bereitstellenden Archivs in den EAD-Findbüchern enthalten sein.

5. Im Element *archdesc/unitdate* werden die Attribute *type="bulk (oder inclusive)"* und *normal* verpflichtend verlangt. Die EAD Tag Library bietet beide Attribute zur Verwendung an, schreibt sie jedoch nicht vor. Aus Sicht der deutschen Archive ist es nach der bisherigen Erschließungspraxis nicht möglich, jede Laufzeit in normalisierter Form zu liefern, da nicht jedes Archiv bzw. jedes Archivsystem diese automatisch generieren kann. In diesem Attribut wird die im Element als Text angegebene Laufzeit nach der ISO 8601 maschinenlesbar in „JJJJ-MM-TT/JJJJ-MM-TT“ aufgelöst. Es ist durchaus fraglich, ob es Sinn macht, dieses Attribut obligatorisch vorzuschreiben, weil es auch Archiven außerhalb Deutschlands schwer fallen sollte, jede Laufzeit korrekt nach ISO 8601 anzugeben.

Im EAD-BArch-Profil ist das Attribut *type="bulk (oder inclusive)"*, das jeweils anzeigt, ob es sich bei der Laufzeit um den gesamten (inclusive) oder um den überwiegenden Zeitraum (bulk) handelt, nicht vorgesehen. Das Attribut ist notwendig, wenn das Element für die Laufzeit wiederholbar ist, um bei Bedarf zunächst den Gesamtzeitraum zu nennen, zu dem die Dokumente vorliegen, und dann im nächsten Laufzeit-Element den Hauptzeitraum. Im EAD-BArch-Profil ist jedoch das Element *unitdate* nicht wiederholbar. Darum erübrigt sich im EAD-BArch-Profil das Attribut *type*, um wiederholte Laufzeitangaben näher zu spezifizieren. Im deutschen Archivwesen wird im allgemeinen, wenn Dokumente zeitlich vom Hauptzeitraum ausreißen, das Jahr in Klammern je nachdem vor die Hauptzeit oder danach gestellt (z.B. <unitdate>(1949) 1950-1955</unitdate>).

6. Im Element *archdesc/dsc* wird das Attribut *type* mit dem Wert „combined“ obligatorisch verlangt. Laut EAD Tag Library sind jedoch an dieser Stelle mehrere Werte, u.a. auch „othertype“, erlaubt, so wie es im EAD-BArch-Profil umgesetzt ist. Während der Wert „combined“ vorschreibt, dass auf die Beschreibung einer Serie unmittelbar die Liste der Inhalte folgt, zeigt „othertype“ an, dass keine bestimmten Elemente folgen müssen. Das EAD-BArch-Profil setzt „othertype“ ein, damit unterhalb eines Bestandes mehrere Teilbestände zusammengefasst werden können. Aus diesem Grund folgt dem Element *archdesc/dsc* zunächst einmal das Element <c01 level="collection"> mit den dann folgenden näheren Erläuterungen zu den Teilbeständen. Dem Wert „combined“ im Attribut *type* des Elementes *archdesc/dsc* müsste sonst im EAD-BArch-Profil unmittelbar <c01 level="class"> folgen. Das entspricht jedoch nicht dem Ansatz im Bundesarchiv und ist daher unterschiedlich zu der Vorgehensweise zu RLG.

7. Als dringend wünschenswert (mandatory if applicable) führt das EAD-RLG-Profil auf, dass zu jedem Element *c0X/did* jeweils die Elemente *origination*, *unitid* und *languagematerial/language* aufgeführt werden. Im EAD-BArch-Profil sind die Elemente *origination* und *languagematerial* zwar vorgesehen, sollen jedoch nur verwendet werden, wenn sich für die jeweilige Verzeichnungseinheit ein Unterschied zum Gesamtbestand ergibt. Sie sollen nicht für jede Verzeichnungseinheit wiederholt werden, wenn keine neuen Informationen im Vergleich zu den Bestandsinformationen eingetragen werden können. Dies wären redundante Daten.

Das Element *unitid* wird im EAD-BArch-Profil auf der Ebene der Verzeichnungseinheiten (<c0X level="file">) verpflichtend für die Bestellnummer verwendet, nicht jedoch auf der Ebene der Gliederungsstufen (<c0X level="class">). Dort wird aufgrund der Erfahrungen in der Erschließungspraxis im deutschen Archivwesen die Nummer der Gliederungsstufe meist im Gliederungstitel mit aufgeführt. Sollte ein Archiv keine Nummer der Gliederungsstufe

verzeichnet haben, so kann bei der Generierung des Onlinefindbuchs aus dem EAD-Dokument automatisch die Nummerierung der Gliederung im MEX-Editor oder im MIDEX hinzugefügt werden.

Das Unterelement *language* des Elementes *languagematerial* war bislang im EAD-BArch-Profil nicht vorgesehen, wird aber im Rahmen der Weiterentwicklung des EAD-BArch-Profiles im Zuge des Verbundfindmittels mit aufgenommen werden.

8. Außerdem werden in jedem Element *unitid=Bestellnummer* auf der Ebene *cOX/did* die Attribute *countrycode* und *repositorycode* „gegebenenfalls obligatorisch“ (mandatory if applicable) verlangt. Diese Vorgehensweise ist auf der Ebene der Verzeichnungseinheit aus Sicht des Bundesarchivs ebenfalls redundant, da es genügt, auf der Bestandsebene den Länder- und Repositorycode einmal für den gesamten Bestand anzugeben.

Zusammenfassung

Zunächst ist festzuhalten, dass mit den aufgeführten Punkten 2. bis 4. und ein Teil aus 7. wirkliche Fehler markiert wurden. Es handelt sich dabei um die Angabe von ISO-Standards und Identifikatoren für das bereitstellende Archiv und das Findbuch selbst und die Ergänzung des Unterelementes *language* in *languagematerial*. Diese Angaben waren zu einem früheren Zeitpunkt aus Sicht des Bundesarchivs nicht notwendig, weil alle Onlinefindbücher aus ein und demselben Archiv stammten oder andere EAD-Findbücher nur für sich allein stehend in Onlinefindbücher umgewandelt wurden. Für beide Vorgehensweisen wurden keine Identifikatoren benötigt. Aus heutiger Sicht, vor der Entwicklung eines Verbundfindmittels mit Onlinefindbüchern mehrerer Archive, ist der Einsatz dieser Angaben nicht nur selbstverständlich, sondern auch für die technische Funktionsweise unerlässlich. Darum werden zukünftig alle Identifikatoren an den entsprechenden Stellen vergeben. Die Angabe der Sprachen im Findbuch wird ebenfalls erst für eine internationale Zusammenarbeit relevant, wofür das EAD-BArch-Profil jetzt entsprechend ausgebaut wird.

Die Fehlermeldungen zu Punkt 1. und 6. sind zwar aus Sicht des EAD-RLG-Profiles relevant, sind jedoch nicht von der EAD Tag Library zwingend vorgesehen, nach der das EAD-BArch-Profil gleichermaßen korrekt ist wie das EAD-RLG-Profil. Die Fehlermeldungen bilden demnach lediglich eine differierende Vorgehensweise der RLG zu der im Bundesarchiv ab. Allerdings ist durchaus zu überlegen, ob der Vorgehensweise bei der RLG in Punkt 1 mit der Zuordnung der entsprechenden Informationen zu MARC21- oder Dublin Core-Elementen eventuell gefolgt werden könnte. Entsprechende Entscheidungen würden jedoch erst zu einem späteren Zeitpunkt zum Ende des Projektes implementiert werden können.

Bei der Meldung des Testwerkzeuges im Punkt 5., bei der die Laufzeit im Element *unitdate* in der normalisierten Fassung nach ISO 8601 im Attribut *normal* obligatorisch verlangt wird, ist es sogar zweifelhaft, ob dies auf diese Weise international durchzusetzen ist. Dieses zu verlangen, dürfte für zahlreiche Archive sehr schwierig sein. Zumindest in Deutschland kann angesichts der bisherigen Entwicklungen im Archivwesen nicht vorausgesetzt werden, dass jede Laufzeit auch in normalisierter Form geliefert werden kann.

Zu den Fehlermeldungen 7. und 8. kann aus Sicht des Bundesarchivs angemerkt werden, dass diese zu einer unnötigen Vervielfachung redundanter Informationen führen, die im Bundesarchiv weitgehend vermieden werden sollen.