



Universiteit
Leiden
The Netherlands

Encoded Archival Description voor de digitale toegang tot bibliotheekcollecties

Verhaar, P.A.F.

Citation

Verhaar, P. A. F. (2009). Encoded Archival Description voor de digitale toegang tot bibliotheekcollecties. In *Handboek Informatiewetenschap voor bibliotheek en archief*. Alphen aan de Rijn: Kluwer. Retrieved from <https://hdl.handle.net/1887/16019>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/16019>

Note: To cite this publication please use the final published version (if applicable).

Encoded Archival Description voor de digitale toegang tot bibliotheekcollecties

In een tijdperk waarin meer en meer informatie via het internet te vinden is wordt ook van bibliotheken verwacht dat zij hun bezit en de informatie over dit bezit zo efficiënt mogelijk online ter beschikking stellen. Omdat bibliotheken vaak zeer omvangrijke en zeer diverse fysieke en digitale collecties beheren is het bieden van een goede toegang niet altijd een eenvoudige opgave. De keuze van geschikte metadatastandaarden is hierbij essentieel. Traditioneel werken bibliotheken met bibliografische beschrijvingen op het niveau van individuele items. Deze benadering is zeer effectief bij het beschrijven en toegankelijk maken van boeken en tijdschriften. Beschrijvingen op itemniveau zijn echter niet altijd toereikend bij het bieden van toegang tot de objecten in erfgoedcollecties. Tussen de items in dit soort collecties bestaan vaak duidelijke verwantschappen. Bovendien kan over dergelijke collecties als geheel vaak veel algemene informatie worden gegeven, over onder meer het moment en de wijze van verwerving, de omvang en de verschillende bewerkingen die hebben plaatsgevonden. Om de samenhang tussen de objecten in een collectie op een effectieve manier vast te kunnen leggen is een metadatastandaard nodig waarbij de collectie als geheel de te beschrijven eenheid vormt. Een de facto standaard voor het opstellen van collectiebeschrijvingen is *Encoded Archival Description* (EAD). Dit artikel geeft een korte introductie tot de standaard en bespreekt een aantal mogelijkheden die ontstaan wanneer de standaard wordt toegepast. Hierbij wordt speciale aandacht besteed aan de implementatie van EAD binnen de bibliotheeksector.

Ontwikkeling van EAD

Zoals duidelijk wordt aangegeven in de naam van de standaard is EAD oorspronkelijk ontwikkeld voor het beschrijven van archieven. De ontwikkeling begon in 1993 aan de University of California in Berkeley, binnen een onderzoeksteam dat werd geleid door Daniel Pitti.¹ Binnen het project werd gewerkt aan een oplossing voor het probleem dat veel archiefinventarissen alleen nog op papier beschikbaar waren en dat er dus nog geen manier bestond om deze informatie op een gestructureerde manier via computers te doorzoeken. Er bestonden wel al verschillende beschrijvingsstandaarden voor het opstellen van archiefbeschrijvingen. De belangrijkste hiervan is ISAD(G). Deze standaard is vastgesteld door de Internationale Archiefraad en bestaat uit een groot aantal richtlijnen voor het vastleggen en verklaren van de context en de inhoud van een archief.² EAD biedt een formaat waarmee gestructureerde archiefbeschrijvingen ook op een duurzame, computerleesbare en platform-onafhankelijke manier kunnen worden gecodeerd. Hierbij werd aanvankelijk gebruik gemaakt van SGML, en later van XML. EAD is alleen een datastructuur, en schrijft op zichzelf niet voor hoe een collectiebeschrijving moet worden opgesteld. De elementen en

¹ Een meer uitvoerige beschrijving van de ontstaansgeschiedenis van EAD is te vinden in *Development of the Encoded Archival Description DTD*, <<http://www.loc.gov/ead/eaddev.html>>

² Deze richtlijnen zijn te vinden op de website van de Internationale Archiefraad: <http://www.ica.org/sites/default/files/isad_g_2e.pdf>

attributen die gebruikt kunnen worden bij het coderen van een collectiebeschrijving zijn vastgelegd in de officiële EAD Document Type Definition (DTD). De zelfde structuur is inmiddels ook vastgelegd in de vorm van een XML schema. Dit is een nieuwere syntax voor het definiëren van de structuur van een XML-document.

Toepassingen binnen de bibliotheeksector

De eerste versie van EAD kwam beschikbaar in 1998. De standaard bestaat inmiddels dus ruim 10 jaar. In deze periode heeft EAD zich ontwikkeld tot een standaard die binnen de internationale archiefwereld breed wordt gedragen. In toenemende mate wordt de standaard ook binnen de bibliotheeksector toegepast. Dit is niet verwonderlijk, aangezien bij bibliotheken vaak ook verschillende archiefcollecties te vinden zijn. Bij de bibliotheek van de Universiteit van Amsterdam, bijvoorbeeld, wordt het archief van de boekhandel en uitgeverij Van Benthem & Jutting bewaard, en de Universiteitsbibliotheek in Leiden beheert onder meer het archief van A.W. Sijthoff's Uitgeversmaatschappij, en het literaire en persoonlijke archief van de Nederlandse auteur Frans Kellendonk. Bij universiteiten is het ook gangbaar dat hoogleraren op een bepaald moment hun wetenschappelijke of persoonlijke archieven overdragen aan de universiteitsbibliotheek. Dergelijke archieven bestaan meestal uit collegeaantekeningen, werkboeken of correspondentie met vakgenoten. In al deze gevallen gaat het om klassieke archivale collecties: verzamelingen van meestal unieke documenten die door een persoon of instelling zijn voortgebracht als bijproduct van de uitoefening van een bepaalde functie.

Binnen de bibliotheeksector wordt EAD zeker niet alleen toegepast voor het beschrijven van de typische archiefcollecties. De standaard kan in principe steeds worden ingezet wanneer er over een bepaalde verzameling van objecten contextuele informatie kan worden gegeven. Dit is bijvoorbeeld het geval wanneer een verzamelaar besloten heeft om zijn verzameling met bijzondere boeken in zijn geheel of deels aan een bibliotheek na te laten. Met behulp van EAD kan in deze situatie biografische informatie over de verzamelaar en informatie over de wijze van verwerving worden vastgelegd. EAD kan echter ook worden gebruikt voor het beschrijven van collecties van objecten die oorspronkelijk helemaal geen herkomst delen en die door de beherende instelling zelf zijn samengesteld op basis van een bepaalde thematische verwantschap. Zo is bij de Universiteit van Amsterdam een beschrijving gemaakt van alle kookboeken in de collectie, en zijn bij de Universiteitsbibliotheek van Leiden collectiebeschrijvingen gemaakt van alle Arabische handschriften, van alle incunabelen en van alle handgetekende kaarten. EAD stelt bibliotheken dus in staat om bepaalde objecten die van nature geen fysieke eenheid vormen toch gezamenlijk in een betekenisvol kader te presenteren. Terwijl de naam van de standaard suggereert dat deze alleen kan worden gebruikt voor het opstellen van gecodeerde archiefbeschrijvingen wordt EAD in bibliotheken dus duidelijk voor een breder doel ingezet. Het kan hierbij gaan om archieven en verzamelingen die organisch zijn gegroeid maar ook om kunstmatig samengestelde aggregaties van objecten.

Wanneer EAD binnen de bibliotheeksector wordt toegepast betekent dit een manier van beschrijven wordt geïntroduceerd die fundamenteel verschilt van de werkwijze die wordt gehanteerd bij het opstellen

van itembeschrijvingen. Een belangrijk verschil is dat collecties in EAD hiërarchisch worden beschreven. Een stuk wordt hierdoor altijd binnen een bepaalde context geplaatst. Bij collecties van archiefstukken of andere bijzondere materialen is deze context in veel gevallen noodzakelijk om de relevantie te kunnen beoordelen. Over een afzonderlijk kasboek, bijvoorbeeld, is meestal wel enige beschrijvende informatie te geven, maar een dergelijk object krijgt meestal veel meer betekenis wanneer deze wordt gepresenteerd in samenhang met de overige financiële administratie van een bedrijf. Bij de itembeschrijvingen zoals deze van oudsher binnen de bibliotheeksector werden opgesteld is deze mogelijkheid om een object in een context te plaatsen niet voorhanden. Alle informatie die relevant is voor het object moet in principe ook in de itembeschrijving zelf worden geplaatst. Bij grote herkomstcollecties kan dit mogelijk tot duplicatie van informatie leiden.

Naast EAD bestaan er binnen de bibliotheeksector ook andere metadatastandaarden voor het vastleggen van informatie over verzamelingen van objecten. Standaarden zoals METS, MPEG-DIDL en het meer recentelijk ontwikkelde OAI-ORE kunnen worden gebruikt om de structuur van samengestelde digitale objecten vast te leggen. De drie genoemde standaarden zijn voorbeelden van structurele metadata. Ze richten zich primair op het samenbrengen en het ordenen van digitale bestanden. bovengenoemde standaarden zijn voorbeelden van structurele metadata. EAD verschilt van deze standaarden omdat deze zich niet primair richt op digitale bestanden. EAD is een beschrijvende metadatastandaard en biedt op de eerste plaats een vocabulaire voor het beschrijven van fysieke collecties, en kan dus ook zonder problemen worden toegepast wanneer geen enkel collectieonderdeel is gedigitaliseerd. De standaard die wat betreft het gebruiksdoel waarschijnlijk het meest lijkt op EAD is *USMARC for Archival and Manuscripts Control* (MARC-AMC). De Binnen de familie van MARC-standaarden is bijvoorbeeld een variant ontwikkeld met de naam *USMARC for Archival and Manuscripts Control* (MARC-AMC). De hoeveelheid informatie die kan worden vastgelegd in MARC-AMC is echter beperkter dan bij EAD. Een belangrijke tekortkoming is dat de collectierecords zich beperken tot het hoofdniveau. MARC-AMC biedt geen mogelijkheid om de hiërarchische samenstelling van de collectie op een heldere manier te beschrijven.³ Voor bibliotheken kan de meerwaarde van EAD dus onder meer liggen in het feit dat het een manier biedt om contextuele informatie over collecties op een logische en duurzame manier vast te leggen.

Voordelen

Door gebruik te maken van een vaste standaard voor het maken van collectiebeschrijvingen kunnen bibliotheken een bepaalde mate van uniformiteit doorvoeren in de manier waarop informatie over collecties wordt aangeboden. Bij veel instellingen bestaat het probleem dat dergelijke informatie op vele verschillende manieren beschikbaar wordt gesteld. Op de webpagina's over de bijzondere collecties van een bibliotheek zijn meestal enkele beschrijvingen van de belangrijkste collecties te vinden. Hiernaast bevatten analoge media, zoals gedrukte catalogi, inventarissen, of

³ Een meer uitvoerige vergelijking van EAD en MARC-AMC is te vinden in Carini en Shepherd (2004).

kaartenbakken, vaak ook veel nuttige informatie. Soms kan een instelling er voor kiezen om een gedrukte catalogus integraal in te scannen, en om deze daarna als PDF-document op het web te plaatsen. Omdat bij bibliotheken vaak zeer uiteenlopende materialen worden beheerd zijn er in veel gevallen ook uiteenlopende methoden en formaten gebruikt bij het vastleggen van informatie over deze collecties en deelcollecties. Voor eindgebruikers kan deze veelheid aan bronnen snel zeer verwarrend werken. Door gebruik te maken van EAD wordt het gemakkelijk om via één centrale webpagina toegang te bieden tot alle informatie over de verschillende collecties. Een belangrijk voordeel van de standaard is dat deze in vrij algemene termen is gedefinieerd. Ook wanneer collecties uit heterogene materiaaltypen bestaan, zoals manuscripten, prenten, tekeningen, foto's en kaarten, kunnen deze toch op een uniforme wijze worden beschreven.

Het uniformeren en standaardiseren van informatie brengt meestal met zich mee dat bestaande formaten moeten worden geconverteerd. Verschillende instellingen hebben geëxperimenteerd met het converteren van gedrukte of getypte catalogi naar EAD-bestanden. De werkwijze is hierbij over het algemeen dat de tekst via OCR uit de ingescande catalogus wordt gelicht, en dat de tekst hierna via speciaal ontwikkelde programmatuur naar het EAD-formaat wordt getransformeerd. Wanneer informatie op deze manier is gemigreerd kan deze eenvoudig via het internet worden aangeboden, en eindgebruikers hoeven in de meeste gevallen de bibliotheek dan niet meer fysiek te bezoeken om de catalogi te raadplegen. De tijdsinvestering die nodig is voor dergelijke conversietrajecten hangt sterk af van de manier waarop de oorspronkelijk inventaris is opgesteld.⁴ In het merendeel van de gevallen kan de conversie niet geheel geautomatiseerd verlopen en is nog een handmatige redactie- of correctieslag nodig.

EAD is, zoals aangegeven, gebaseerd op de syntaxis van XML. In de collectiebeschrijvingen worden een groot aantal aspecten gecodeerd door middel van *tags* die expliciet aangeven hoe de gecodeerde onderdelen moeten worden geïnterpreteerd. De opgemaakte tekst krijgt hierdoor het karakter van een database. Net zoals in een database binnen de afzonderlijke velden kan worden gezocht, kan in EAD-documenten gericht worden gezocht binnen specifieke elementen. Wanneer iemand zoekt op een jaartal kan een full-text zoekactie ook andere soorten cijfers opleveren, zoals bijvoorbeeld geldbedragen of delen van inventarisnummers. Omdat in EAD-documenten jaartallen specifiek als jaartallen zijn gemarkeerd, kunnen zoeksystemen worden ontwikkeld waarbij dit soort ambiguïteiten worden opgeheven. In EAD kunnen naast data ook persoonsnamen, namen van organisaties, geografische termen, of onderwerpen expliciet worden benoemd en gelokaliseerd. Om zulke zoeksystemen goed te laten functioneren is het wel van belang dat deze inhoudelijke onderdelen ook op een correcte en consistente manier zijn gecodeerd. Deze laatste bezigheid vergt vaak veel tijd, aangezien deze verfijnde codering meestal alleen handmatig kan worden aangebracht.

⁴ Karsch en Keuning (2007) hebben aangetoond dat voor de uitvoering van een dergelijk conversietraject rekening moet worden gehouden met een gemiddelde tijdsinvestering van 40-60 minuten per pagina van een toegang.

Het feit dat EAD op XML is gebaseerd brengt ook met zich mee dat de informatie op een zeer flexibele manier kan worden gepresenteerd. Zoals het geval is bij de meeste XML-standaarden staat de codering van de inhoud volledig los van de uiteindelijke presentatie van deze inhoud. De beschrijving wordt hiermee medium-neutraal. Om de informatie in een EAD op het web te presenteren wordt vaak gebruik gemaakt van de XSLT. Dit is een techniek voor het opstellen van stylesheets. XSLT kan bijvoorbeeld worden gebruikt om het EAD-bestand om te zetten naar een HTML-bestand dat via het web kan worden aangeboden. Via een techniek als XSL:FO is het ook mogelijk om het EAD-bestand naar een PDF-document om te zetten zodat deze gemakkelijk op papier kan worden geprint. Instellingen zijn volledig vrij om te bepalen hoe de collectiebeschrijven op het web worden vormgegeven. Het is eveneens mogelijk om webomgevingen te ontwikkelen waarin eindgebruikers zelf invloed kunnen uitoefenen op de manier waarop de informatie wordt gepresenteerd. Omdat in EAD-bestanden vaak zeer lange teksten worden opgenomen kan het voor eindgebruikers soms lastig zijn om de relevante onderdelen te vinden. Wanneer EAD wordt gehanteerd als onderliggend formaat is het mogelijk om dynamische webpagina's te maken waarop eindgebruikers de collectieonderdelen zelf op verschillende manieren kunnen sorteren. Met name bij die collecties die uit vele honderden stukken bestaan kan het praktisch zijn wanneer eindgebruikers gericht kunnen zoeken binnen de EAD. Dit systeem kan eindgebruikers dan subsets uit de EAD te laten genereren waarin niet alle stukken worden getoond, maar waarin alleen die onderdelen waarin de door hen opgegeven zoektermen voorkomen. Wanneer met EAD wordt gewerkt kunnen dit soort zoeksystemen in theorie worden ontwikkeld. Zoekresultaten kunnen worden gepresenteerd op een manier die bekend is van zoekmachines zoals Google. Op een website wordt niet de gehele tekst getoond, maar alleen die zinnen die de opgegeven zoektermen bevatten. EAD biedt deze mogelijkheid, maar in de praktijk zijn er nog maar weinig bibliotheken die dit potentieel ook echt benutten.

Er zijn ook presentatievormen denkbaar waarbij eindgebruikers niet alleen invloed kunnen uitoefenen op de presentatie van de EAD, maar ook op de inhoud daarvan. Wanneer collectiebeschrijvingen in digitale vorm beschikbaar komen kunnen op basis van deze informatie ook websites worden gebouwd waarop eindgebruikers worden uitgenodigd om hun kennis over die collecties te delen. Het grote publiek kan dan mogelijk de informatie die nog ontbreekt aanvullen of alle onjuiste informatie, eventueel gemodereerd, corrigeren. Aan de Universiteit van Michigan is in 2006 een project opgestart waarbij met deze mogelijkheid is geëxperimenteerd. Op de website van het project, *Polar Bear Expedition Digital Collections*, is het mogelijk om commentaar te geven op de beschreven stukken, en op de gedigitaliseerde objecten in deze collectie.⁵ De mogelijkheid om bepaalde onderdelen van de collectiebeschrijving te laten annoteren kan zeer zinvol zijn, met name wanneer deze wordt gebruikt door wetenschappers die onderzoek doen naar bepaalde collecties en die hierdoor gegevens kunnen aandragen waar bibliotheekmedewerkers zelf niet over beschikken. Wanneer delen van de collectie zijn gedigitaliseerd kan het mogelijk ook interessant zijn om ook meer publieksparticipatie mogelijk te maken tijdens de aanmaak van de collectiebeschrijvingen. De EAD in ontwikkeling kan bijvoorbeeld in een wiki

⁵ Meer informatie over dit project is te vinden in Yakel et al. (2007).

worden ondergebracht, en aan de hand van de scans zouden onderzoekers of andere geïnteresseerden bepaalde collectieonderdelen zelf kunnen metadateren. Wel zal een centrale redactie dan moeten waken over de consistentie van de beschrijvingen.

Standaardisering van de informatie zorgt er voor dat deze informatie eenvoudiger kan worden uitgewisseld. Externe partijen kunnen op basis van deze gestandaardiseerde informatie weer nieuwe diensten bouwen. Een goed voorbeeld van een waardevolle dienst is een website waarop de collectiebeschrijvingen van vele verschillende instellingen tegelijkertijd doorzocht kunnen worden. Onderzoekers die op zoek zijn naar relevant materiaal kunnen dan gebruik maken van deze centrale voorziening, en hoeven niet meer de websites van alle individuele instellingen langs te lopen. Dit maakt het verrichten van historisch onderzoek aanzienlijk eenvoudiger. Voorbeelden van zulke metacatalogi zijn *Access to Archives*,⁶ *Archive Grid*,⁷ *Archive Hub*⁸ en *Calames*.⁹ In Nederland biedt de website *De Geschiedenis van Zuid-Holland* de mogelijkheid om via één scherm de archieven van alle deelnemende instellingen gelijktijdig te doorzoeken.¹⁰ Het uniformeren van het formaat van de informatie is uiteraard maar een aspect van het doorvoeren van meer interoperabiliteit. Andere belangrijke obstakels bij het doorvoeren van een eenduidige toegang worden gevormd door het feit dat er verschillende talen gebruikt kunnen zijn, het feit dat er vaak verschillende termen en verschillende onderwerpscodes worden gebruikt. Bij het verbeteren van de toegang tot de collecties van verschillende instellingen moeten eigenlijk ook dit soort inhoudelijke aspecten van de collectiebeschrijving worden geuniformeerd. Juist hier ligt waarschijnlijk een van de grootste uitdagingen voor de erfgoedsector.

Structuur van een EAD-document

In de EAD Tag Library, die te vinden is op de website van de Library of Congress, zijn toegankelijke omschrijvingen te vinden van alle elementen en attributen die gebruikt kunnen worden bij het opstellen van een EAD-document.¹¹ De EAD Tag Library bespreekt in totaal zo'n 150 elementen. Slechts een paar van deze elementen zijn verplicht. De standaard is dus zeer flexibel en biedt individuele instellingen veel vrijheid. Dit laatste kan zowel een voordeel als een nadeel zijn. Het positieve aspect van deze flexibiliteit is dat de codering volledig kan worden afgestemd op het materiaal dat beschreven moet worden en op de wensen van de instelling. Hier staat tegenover dat het bij vele uiteenlopende codeerstrategieën lastig kan worden om gegevens uit te wisselen. Bovendien kan het dan voor de partij die de bestanden verzamelt ook zeer complex worden om de verschillende bestanden op een uniforme wijze te presenteren. Om de consistentie te bevorderen hebben een aantal instellingen richtlijnen rond het coderen in EAD geformuleerd. De Research Library Group en de Online Archives of California bijvoorbeeld hebben beide

⁶ <<http://www.nationalarchives.gov.uk/a2a/>>

⁷ <<http://www.archivegrid.org/web/index.jsp>>

⁸ <<http://www.archiveshub.ac.uk/>>

⁹ <<http://www.calames.abes.fr/pub/>>

¹⁰ <<http://www.geschiedenisvanzuidholland.nl/>>

¹¹ <<http://www.loc.gov/ead/tglib/>>

uitgebreide *Best Practice Guidelines* (BPG) opgesteld.¹² In deze documenten wordt aangegeven welke elementen en attributen verplicht zijn, en welke onderdelen optioneel. Instellingen kunnen er voor kiezen om de officiële EAD DTD wat restrictiever te herprogrammeren, zodat niet alleen kan worden gevalideerd tegen de regels van de DTD, maar ook tegen de aanvullende vereisten van de BPG. In Nederland heeft het National Archief uitgebreide richtlijnen opgesteld voor het coderen in EAD.

EAD-bestanden kunnen in principe direct worden aangemaakt en bewerkt in XML-editors zoals XMLSpy of Oxygen. Vaak is het zo dat de bibliotheekmedewerkers die zich bezig houden met de inhoudelijke aspecten van de collectiebeschrijving niet beschikken over een grondige kennis van XML. Om er voor te zorgen dat de collectiespecialisten de benodigde informatie toch op een gestructureerde manier kunnen aanleveren wordt er vaak gewerkt met templates en invulformulieren. Bij de UB Leiden wordt er gewerkt met formulieren die in Microsoft Word zijn aangemaakt. Wanneer het ingevulde tekstbestand als XML-bestand wordt opgeslagen komt de informatie beschikbaar in een formaat dat daarna, bijvoorbeeld via XSLT, eenvoudig kan worden omgezet naar een valide EAD-bestand. Bij de UB van de Universiteit van Amsterdam wordt gebruik gemaakt van het programma XMetal, en in deze XML-editor is het mogelijk om bovenop de XML-documenten bepaalde schermen aan te maken met invoervelden en keuzemenu's. Instellingen kunnen eventueel ook gebruik maken van de open source EAD-editor ICA Atom.¹³ In dit programma, dat door de Internationale Archiefraad is ontwikkeld, is het mogelijk om via een reeks van invulschermen een volledige collectiebeschrijving op te stellen. De onderliggende XML-code blijft hierbij verborgen.

Binnen de EAD-elementen kan een onderscheid worden gemaakt tussen enerzijds die elementen die tekst kunnen bevatten, en anderzijds de zogenaamde *wrapper elementen*. Deze laatste categorie elementen kunnen alleen andere elementen bevatten, en worden alleen gebruikt om het EAD-bestand te structureren. Het meest omvattende *wrapper element* is <ead>. Dit element bevat het gehele EAD-document. <ead> kan drie subelementen bevatten: <eadheader>, <frontmatter> en <archdesc>. <eadheader> bevat metadata over de collectiebeschrijving. In dit element wordt onder meer vermeld wie verantwoordelijk is voor de inhoud, wie het bestand heeft gecodeerd, wanneer het bestand is aangemaakt en welke wijzigingen zoal hebben plaatsgevonden. Het element <frontmatter> kan worden gezien als een erfenis vanuit de wereld van de analoge archiefinventarissen. Toegangen werden meestal voorzien van een titelpagina, en in <frontmatter> kan informatie worden vastgelegd die nodig is voor het generen van een dergelijke titelpagina. In veel gevallen kan de benodigde informatie ook worden ontleend aan andere elementen, en het gebruik van <frontmatter> is dan ook optioneel.

De daadwerkelijke collectiebeschrijving staat in het element <archdesc>. Binnen dit element wordt op de eerste plaats een beschrijving van de collectie als geheel gegeven. Deze beschrijving bevat zowel een beknopte beschrijving als een meer uitvoerige beschrijving. De beknopte beschrijving wordt gegeven in het wrapper element <did> (*Descriptive*

¹² <<http://www.oclc.org/programs/ourwork/past/ead/bpg.pdf>> en

<<http://www.cdlib.org/inside/diglib/guidelines/bpgead/>>

¹³ Het programma kan worden gedownload via <<http://ica-atom.org/>>

identification). Dit onderdeel heeft als doel om eindgebruikers snel een indruk te geven van de aard van de collectie. Ter illustratie is hieronder de <did> opgenomen uit een EAD-beschrijving van het archief van uitgeverij De Erven F. Bohn, dat wordt beheerd in de UB Leiden.

```
<did>
  <unittitle encodinganalog="title">Archief van
uitgeverij De Erven F. Bohn</unittitle>
<origination>
  <corpname>Bohn N.V., De Erven F.</corpname>
  </origination>
  <unitdate calendar="gregorian" era="ce"
label="Datering:" normal="1759/1969"
type="inclusive">1759-1969</unitdate>
<langmaterial>
  <language langcode="dut">Nederlands</language>,
<language langcode="ger">Duits</language>, <language
langcode="fre">Frans</language>, <language
langcode="eng">Engels</language>.
  </langmaterial>
<physdesc>
  <extent>ca. 35 meter</extent>
  </physdesc>
  <abstract encodinganalog="description"
langcode="dut">Archief (ca. 35 meter) van uitgeverij De
Erven F. Bohn te Haarlem; ca. 2250 gedrukte
werken.</abstract>
  <abstract encodinganalog="description"
langcode="eng">Archives (c. 35 metres) of the publishing
house De Erven F. Bohn at Haarlem; ca. 2250 printed
books.</abstract>
  <physloc>Universiteitsbibliotheek</physloc>
  <repository>Universiteit Leiden</repository>
</did>
```

Zoals in het fragment te zien is wordt de naam van de collectie vastgelegd in <unittitle>, en wordt de naam van de verzamelaar of vervaardiger opgenomen in <origination>. In het element <physdesc> kunnen details over de omvang of over de afmetingen worden genoteerd, en de talen die voorkomen kunnen worden vermeld in <langmaterial> (*Language of the Material*). In het element <abstract> kan een korte samenvatting van de inhoud van de collectie worden gegeven. De tekst die hierin wordt vastgelegd wordt vaak in zoeksystemen gebruikt bij de presentatie van zoekresultaten, aangezien in veel gevallen de naam van de collectie alleen niet voldoende informatie biedt.

Bij alle elementen geldt dat de betekenis en de bruikbaarheid ervan nog verder kan worden uitgebreid door middel van attributen. In het element <unitdate>, bijvoorbeeld, kan informatie worden gegeven over de datum waarop het object is vervaardigd. Hierbij wordt meestal de datum vermeld in de vorm die voorkomt op het oorspronkelijke stuk. In het “normal”-attribuut van dit element kan deze zelfde datum ook worden opgegeven in een volgens ISO 8601 genormaliseerde vorm. Deze standaardisatie zorgt er onder meer

voor dat er beter op specifieke data of op periodes kan worden gezocht. Voor veel bibliotheken is hiernaast ook het “encodinganalog”-attribuut van belang. In dit attribuut kan worden verwezen naar equivalente velden in andere metadatastandaarden zoals Dublin Core of MARC21. Deze *encodinganalogs* kunnen zeer praktisch zijn wanneer de EAD-bestanden geconverteerd moeten worden naar één van deze andere standaarden. Omdat de EAD-elementen zelf informatie met zich meedragen over hoe zij kunnen worden geconverteerd hoeven er geen afzonderlijke concordantielijsten te worden bijgehouden.

Binnen <archdesc> kunnen naast <did> ook een aantal andere elementen worden gebruikt waarmee een meer uitvoerige beschrijving van de inhoud en de context van de collectie kan worden gegeven. In de elementen <bioghist>, <arrangement> en <acqsinfo> kan informatie worden gegeven over, respectievelijk, de historische context, de samenstelling van de collectie en de verwervingsgeschiedenis. In tegenstelling tot <did> hebben deze beschrijvingen meer de vorm van korte essays. In het element <scopecontent> kan de inhoud van de collectie nader worden besproken. Binnen het element <processinfo> kan een overzicht worden gegeven van de verschillende bewerkingen die de collectie heeft ondergaan. Het kan daarbij bijvoorbeeld gaan om ontsluitingswerkzaamheden of om restauratiewerkzaamheden. Wanneer er bepaalde beperkingen van kracht zijn met betrekking tot de toegang tot de collectie of het gebruik van bepaalde stukken kan dit worden vermeld in de elementen <accessrestrict> en <userrestrict>. In <bibliography> kunnen relevante literatuurverwijzingen worden opgenomen. Wanneer er informatie moet worden vastgelegd die in geen van de beschikbare elementen past kan deze tekst worden vastgelegd in het element <odd> (*Other Descriptive Data*). Voor bibliotheken kan het van belang zijn om instructies te geven over de manier waarop het beschreven materiaal moet worden aangevraagd. Aangezien EAD is ontwikkeld voor de archiefwereld en niet voor bibliotheken biedt de standaard geen voorzieningen voor dit soort instructies. <odd> vormt in deze situatie een oplossing.

Wanneer boeken of tijdschriften worden gecatalogiseerd kennen bibliotheken meestal gecontroleerde trefwoorden toe. Bezoekers van de bibliotheek kunnen daarmee in de catalogus de bronnen over een bepaald onderwerp eenvoudig bijeen brengen. In EAD-bestanden is een vergelijkbare vorm van onderwerpsontsluiting mogelijk via het element <controlaccess>. Binnen dit element kan het type onderwerp worden aangegeven via elementen zoals <subject>, <persname>, <geogname> en <genreform>. Bij deze laatste elementen kunnen “source”-attributen worden opgenomen die duidelijk maken aan welke thesaurus de trefwoorden zijn ontleend. Veelgebruikte waarden zijn “lcnaf” of “aat” voor respectievelijk de *Library of Congress Name Authority Files* en de *Arts & Architecture Thesaurus*. In de Nederlandse context worden ook de waarden “gtt” en “nbc” gebruikt voor de *Gemeenschappelijke trefwoordenthesaurus* en de *Nederlandse basisclassificatie*. Bij deze onderwerpsthesaurering moet worden opgemerkt dat de koppeling met thesauri vaak op een vrij primitieve manier plaatsvindt. Termen worden vaak letterlijk in het XML-bestand opgenomen, en wanneer er slechts één vorm wordt overgenomen kan uiteraard geen ondersteuning worden geboden voor meertalig zoeken, of voor het samenbrengen van de verschillende spellingsvarianten van eigennamen.

Bij de ontwikkeling van de EAD DTD vormden de richtlijnen van ISAD(G) een belangrijke bron van inspiratie. De invloed van ISAD(G) op EAD

is onder meer zichtbaar in het *wrapper element* <descgrp>. ISAD(G) onderscheidt binnen archieftoegangen verschillende beschrijvingsgebieden, zoals onder meer de “Context Area”, de “Content and Structure Area”, de “Conditions of Access and Use Area”, en de “Allied Materials Area”. De EAD-elementen die hierboven zijn besproken kunnen alle worden gerelateerd aan een van deze beschrijvingsgebieden. Zo is de informatie die in <bioghist> wordt gegeven een onderdeel van de “Context Area”, en valt <scopecontent> binnen de “Content and Structure Area”. Wanneer instellingen nauw willen aansluiten bij de richtlijnen van ISAD(G) kunnen zij er voor kiezen om een aantal elementen binnen <archdesc> te clusteren door middel van het wrapper element <dscgrp>. In het “type”-attribuut van dit element kan dan worden gespecificeerd om welk beschrijvingsonderdeel van ISAD(G) het gaat.

Multi-level description

Een ander onderdeel dat van ISAD(G) is overgenomen is het principe van multi-level description. De uitgangspunten hiervan zijn dat collecties uit verschillende niveaus kunnen bestaan, dat deze niveaus hiërarchisch gerelateerd zijn, en dat de hiërarchische samenhang tussen de niveaus ook helder beschreven moet kunnen worden. In EAD-bestanden is een beschrijving op meerdere niveaus mogelijk doordat er een inventaris kan worden opgenomen die een overzicht biedt van de verschillende collectieonderdelen. Deze inventaris wordt ondergebracht in het element <dsc> (*Description of Subordinate Components*). Aan het opstellen van zo’n inventaris gaat over het algemeen een grondige analyse van de collectie vooraf. Het geheel wordt meestal opgesplitst in een aantal deelcollecties, en vervolgens wordt geïnventariseerd welke onderdelen er weer binnen deze deelcollecties onderscheiden kunnen worden. De beschrijving richt zich dus op steeds kleiner wordende onderdelen en kan eventueel worden verfijnd tot op het niveau van individuele items. Bij het opstellen van een inventaris kunnen instellingen zelf beslissen tot welk niveau van detaillering zij gaan. Ieder afzonderlijk niveau wordt vastgelegd in een <c> (*Component*). Op ieder niveau zijn exact dezelfde elementen beschikbaar. Binnen <c> is <did> een verplicht subelement. Omdat het element <c> genest kan worden in een ander <c>-element, is het mogelijk collectieonderdelen zoals series, dossiers en individuele items hiërarchisch ten opzichte van elkaar te beschrijven. In <c> kan ook het attribuut “level” worden gebruikt, waarmee kan worden aangegeven dat het beschreven niveau bijvoorbeeld een “subfonds” of een “file” representeert. Omdat de geneste structuur (voor menselijke gebruikers) snel tot verwarring kan leiden, kiezen de meeste instellingen ervoor om genummerde <c>-elementen te gebruiken. De collectieonderdelen op het hoogste niveau staan dan alle in <c01>, de onderdelen op het tweede niveau staan in <c02>, etcetera. Het is gangbaar om ieder component ook een “id” te geven, zodat vanaf andere plekken naar deze componenten kan worden verwezen. Hieronder is een voorbeeld opgenomen van een in EAD gecodeerde inventaris. Het fragment is afkomstig uit de EAD van het archief van de curatoren van de Universiteit Leiden, met stukken uit de periode 1575 tot 1815.

```
<c01 level="subfonds" id="c01">  
<did>
```

```

    <unittitle>De periode voor 1812</unittitle>
    <unitid type="EAD">I.</unitid>
  </did>
<c02 level="series" id="c01.1">
<did>
  <unittitle>Algemeene organisatie der
Universiteit</unittitle>
  <unitid type="EAD">A.</unitid>
  </did>
<c03 level="subseries" id="c01.1.1">
<did>
  <unittitle>Wetten en Statuten</unittitle>
  <unitid type="EAD">1.</unitid>
  </did>
<c04 level="item" id="c01.1.1.1">
<did>
  <unittitle>Wetten en statuten der Universiteit te
Leiden.</unittitle>
  <unitid type="EAD">1.</unitid>
  <physdesc>1 charter.</physdesc>
  <unitdate calendar="gregorian" era="ce" normal="1631"
type="inclusive">1631</unitdate>
  </did>
  </c04>
  <c04 level="item" id="c01.1.1.2">
  <did>
    <unittitle>"Wetten ende Statuten van de Universiteyt
tot Leyden".</unittitle>
    <unitid type="EAD">2.</unitid>
    <physdesc>Gedrukt, 1 deel.</physdesc>
    <unitdate calendar="gregorian" era="ce" normal="1699"
type="inclusive">1699</unitdate>
    </did>
  </c04>
  ...
</c03>
</c02>
</c01>

```

Bij klassieke archiefcollecties kan een gedetailleerde inventaris met itembeschrijvingen zeer nuttig zijn. Archiefcollecties bestaan meestal uit een grote hoeveelheid kleine objecten, en over het algemeen is het niet zinvol of niet mogelijk om van deze objecten afzonderlijke bibliografische itembeschrijvingen te vervaardigen. Wanneer een collectie uit de meer traditionele bibliotheekmaterialen bestaat, zoals boeken, atlanten, tijdschriften, brieven of prenten, zullen de items vaak ook al afzonderlijk in de bibliotheekcatalogus beschreven staan. In deze situatie is het uiteraard niet efficiënt om de itembeschrijving ook integraal in de EAD op te nemen. Om redundantie te voorkomen kan een bibliotheek dan volstaan met een inventaris waarin alleen een hoofdingeling wordt vastgelegd. Eventueel kunnen binnen de verschillende rubrieken wel externe verwijzingen naar de

itembeschrijvingen in de online catalogus worden opgenomen. Via technieken als SRU of AJAX kunnen de itembeschrijvingen daarna op het scherm wel binnen de context van de collectiebeschrijving worden gepresenteerd. Voor bibliotheken is het van belang om er voor te zorgen dat verschillende metadatastandaarden goed naast elkaar kunnen bestaan, en dat deze elkaar op een logische manier aanvullen. Wanneer er in de collectiebeschrijving links worden aangebracht naar alle items die bij de collectie horen, en wanneer omgekeerd ook in deze individuele items in de catalogus links worden opgenomen naar de collectiebeschrijving kan er een zeer zinvolle vorm van interactie tot stand worden gebracht. Bibliotheekbezoekers die in de catalogus zoeken kunnen via de links naar de EAD meer te weten komen over de bredere context van deze objecten, en doordat er in de EAD weer links terug naar de catalogus zijn kunnen ze eveneens op het spoor worden gezet van gerelateerde items. Eindgebruikers kunnen in deze situatie op verschillende manieren worden geattendeerd op relevante informatie.

Het gebruik van een inventaris is niet verplicht. Een EAD-bestand mag ook alleen een beschrijving op het hoogste niveau bevatten. Deze flexibiliteit biedt een groot voordeel. Het stelt instellingen namelijk in staat om collecties gefaseerd te beschrijven. Bibliotheken verwerven regelmatig omvangrijke collecties met vele bijzondere materialen. Het volledig beschrijven van alle items is over het algemeen een vrij arbeidsintensief proces. Bij het ontsluiten van het materiaal kan een snellere vooruitgang worden geboekt wanneer in eerste instantie alleen de hoofdkenmerken worden vastgelegd, zoals de collectievormer, de globale inhoud, de omvang en de verwervingsmethode. Een dergelijke collectiebeschrijving op hoofd niveau kan meestal met relatief weinig inspanningen beschikbaar worden gesteld, en het brede publiek kan op deze manier ook al snel kennis nemen van het bestaan van deze collectie. De inventaris kan, indien nodig, nog worden aangevuld wanneer er tijd beschikbaar komt voor deze meer verfijnde ontsluiting.

Koppelingen met externe systemen

Binnen de EAD-standaard is ook een aantal elementen gedefinieerd waarmee verschillende soorten van hyperlinks kunnen worden vastgelegd. Een voorbeeld is het element <ref>. Hiermee kunnen interne verwijzingen worden gecodeerd. Wanneer bijvoorbeeld tijdens de bespreking van de inhoud en de samenstelling van de collectie wordt verwezen naar een specifiek onderdeel van de inventaris kan via <ref> een interne link worden aangebracht naar dit component. Het target-attribuut van <ref> bevat dan de “id” van het desbetreffende collectieonderdeel. Het element <extref> kan worden gebruikt voor het tot stand brengen van koppelingen met externe bronnen, zoals relevante websites. Wanneer bepaalde onderdelen van de collectie zijn gedigitaliseerd en via een beeldbank toegankelijk zijn gemaakt kan het zinvol zijn om vanuit de EAD links aan te brengen naar deze digitale bestanden. Er kan dan een vrij vergaande vorm van digitale toegang worden gerealiseerd, aangezien er dan niet alleen informatie over de collectie wordt gegeven, maar ook objecten uit de collectie zelf via het web kunnen worden aangeboden. Een koppeling met een beeldbank kan tot stand worden gebracht met het element <dao> (*Digital Archival Object*). Het “href”-attribuut bevat de identifier of het webadres van de gedigitaliseerde versie van het beschreven object. In veel gevallen bevat een beeldbank verschillende

manifestaties van een digitaal object. Van een scan wordt meestal een archiefkopie, een presentatiekopie en een thumbnail bewaard. Wanneer in het EAD-bestand expliciet moet worden aangegeven dat er verschillende manifestaties bestaan kan er worden gewerkt met *extended links*. In deze situatie moet niet met <dao> maar met <daogrp> worden gewerkt. Binnen <daogrp> kan het herhaalbare subelement <daoloc> worden gebruikt. Het “href”-attribuut in dit element bevat de identifier of de URL van de afbeelding, en het “label”-attribuut kan worden gebruikt voor een aanduiding van het manifestatietype. Voor alle linkingmechanismen die zijn besproken geldt dat de XML-elementen alleen een manier bieden om de link vast te leggen volgens de logica van EAD. Instellingen moeten er vervolgens zelf voor zorgen dat de codes ook op de juiste manier “vertaald” kunnen worden naar goed functionerende hyperlinks. Voor dit doel kan onder meer XSLT worden gebruikt. Wanneer via het element <daogrp> gegevens zijn vastgelegd over de verschillende manifestaties van een digitaal object kan bijvoorbeeld worden ingesteld dat bij de beschrijving van de objecten steeds thumbnails moeten worden gepresenteerd, en dat de link naar de beeldbank wordt geopend wanneer op de thumbnail wordt geklikt.

Naast de verwijzingen naar een beeldbank zijn er nog andere relevante koppelingen denkbaar. Bij bibliotheken worden meestal ook vele andere informatiesystemen beheerd, zoals catalogi, *institutional repositories* met wetenschappelijke publicaties, en zoeksystemen die toegang bieden tot e-journals, databases en elektronische naslagwerken. In het geval van een wetenschappelijk archief van een hoogleraar kan het zinvol zijn om eindgebruikers ook te verwijzen naar publicaties van of over deze hoogleraar. In veel gevallen kan een bibliotheek ook toegang bieden tot de full-text versies van de bronnen die in de bibliografie bij een collectiebeschrijving worden genoemd. Bij dit soort koppelingen moeten echter worden opgemerkt dat zij mogelijk de uitwisselbaarheid van EAD-bestanden kunnen verminderen. Om dit soort links goed te laten functioneren worden via het stylesheet van de instelling vaak specifieke maatregelen getroffen om ervoor te zorgen dat de collectiebeschrijvingen goed aansluiten op de lokale bibliotheeksystemen. Wanneer de EAD wordt gepresenteerd met stylesheets van andere instellingen worden de koppelingen vaak niet goed weergegeven. De EAD-bestanden zijn dan buiten de context van de eigen instelling minder bruikbaar. Voor een deel kan de interoperabiliteit worden vergroot door gebruik te maken van open standaarden, breed ondersteunde protocollen en globale persistent identifiers.

In sommige gevallen kan het ook zinvol zijn om collectiebeschrijvingen van de eigen instelling te koppelen aan collectiebeschrijvingen van andere instellingen. Een dergelijke koppeling kan relevant zijn wanneer een verzamelaar of een instelling de collectie heeft opgedeeld en de deelcollecties heeft nagelaten aan verschillende instellingen. Een voorbeeld is de collectie van Gerard van Papenbroeck. Deze Amsterdamse koopman en verzamelaar liet het grootste deel van zijn portrettencollectie na aan de Universiteit van Amsterdam, en legateerde hiernaast een deel van zijn manuscripten en geannoteerde drukken aan de Leidse universiteitsbibliotheek. Beide instellingen hebben inmiddels een beschrijving in EAD gemaakt van het deel dat zij bezitten. Het is nu vervolgens mogelijk om deze deelbeschrijvingen te bundelen in een grotere overkoepelende collectiebeschrijving. Dergelijke collecties die een herkomst delen, maar die over verschillende instellingen verspreid zijn geraakt, kunnen virtueel weer worden verenigd. Wie

geïnteresseerd is in de volledige verzameling van deze persoon kan dan deze koepel-EAD als uitgangspunt nemen. Deze vorm van samenwerking wordt sterk vergemakkelijkt wanneer instellingen gebruik maken van dezelfde open standaarden.

Via EAD kunnen bibliotheken helder gestructureerde collectiebeschrijvingen aanbieden die ook uniform van opzet zijn. De collectiebeschrijvingen kunnen worden gecodeerd in een vorm die de on-line presentatie, de uitwisseling en de doorzoekbaarheid sterk vereenvoudigt. Bibliotheken die EAD gaan implementeren zullen zeker met een aantal vragen worden geconfronteerd: Welke programma's en hulpmiddelen kunnen worden gebruikt voor het aanmaken en bewerken van de EAD-bestanden? Hoe verhoudt de EAD-standaard zich tot de andere metadatastandaarden die al in gebruik zijn? Hoe moeten de bestanden worden gepresenteerd? Wat is er van belang bij het verzekeren van de interoperabiliteit en de uitwisselbaarheid?¹⁴ Ondanks deze uitdagingen is het inmiddels zonder meer duidelijk dat EAD vele interessante mogelijkheden biedt en dat het een zeer nuttig instrument kan vormen bij het verbeteren en het uitbouwen van de digitale toegang tot bibliotheekcollecties.

Literatuur

Carini, P. en Shepherd, K. (2004). "The MARC standard and encoded archival description". *Library Hi Tech*, 22 (1), pp. 18-27.

Karsch, M. en Keuing, J. (2007). *EAD en EAC – grensoverschrijdend en grensverleggend – Het converteren van archivistische toegangen naar XML-standaarden*. Afstudeerscriptie Informatie en Media / Archivistiek B. 2007.

Patty, W. (2008). "Metadata, Technology, and Processing a Backlog in a University Special Collections". *Journal of Archival Organization*, 6 (1/2), pp. 102-120.

Peereboom, M. en Verhaar, P. (2007). "EAD: Encoded Archival Description: De Standaard". *Informatie Professional*, 11, (2), pp. 28-31.

Pitti, D. (1999). "Encoded Archival Description: An Introduction and Overview". *D-Lib Magazine*, 5 (11),
<<http://www.dlib.org/dlib/november99/11pitti.html>>

Thurman, A. (2005). "Metadata Standards for Archival Control: An Introduction to EAD and EAC." *Cataloging & Classification Quarterly*, 40, (3/4), pp. 183-212.

Yakel, E et al. (2007). "Creating the Next Generation of Archival Finding Aids". *D-Lib Magazine*, 13 (5/6).
<<http://www.dlib.org/dlib/may07/yakel/05yakel.html>>

¹⁴ Een aantal van deze vragen worden beantwoord in "EAD voor beginners", op de website van de Taskforce Archieven, <http://www.taskforce-archieven.nl/dossiers/tienvragen/10vragen>