
Preservation metadata: metadata voor digitale archieven op basis van het OAIS-informatiemodel¹



Timmy Leysens
Medewerker Cocosnet
Koepel van de Vlaamse Noord-
Zuidbeweging - 11.11.11
vzw Studie en Publicaties
Vlasfabriekstraat 11
1060 Brussel
Tel.: 02 536 19 30
E-mail: timmy.leysens@11.be

I N S T A P

Het langdurig bewaren van digitale informatie kreeg de voorbije jaren veel aandacht in de bibliotheek- en archiefwereld. Rond deze materie duiken allerhande problemen op, maar in vele opzichten wordt aan een oplossing gewerkt.

selectiebeleid van organisaties als bibliotheken en archieven die met digitale informatieobjecten te maken krijgen. De juridische problemen hebben betrekking op de intellectuele eigendomsrechten en de achterophollende wetgeving inzake wettelijk depot.

Ondanks de complexe problematiek zijn er een aantal lovenswaardige initiatieven die de weg naar een oplossing voorbereiden. De archiefwereld tracht greep te krijgen op de elektronische records – de digitale tegenhangers van de papieren documenten en dossiers. Het lijkt erop dat van archivariissen een grondige wijziging van de traditionele methoden wordt verwacht. Men spreekt dan ook wel eens van een paradigmaverschuiving binnen de archivalistiek. Ook de bibliotheekwereld laat zich niet onberoerd: groot-schalige projecten als NEDLIB² (Networked European Deposit Library) en CEDARS³ (Curl Exemplars in Digital Archives) brachten respectievelijk nationale en universitaire bibliotheken samen om een oplossing te vinden voor de verschillende aspecten van de problematiek.

DIGITALE PRESERVATIE: DE PROBLEMATIEK

Om meer te weten te komen over de problematiek van digitale preservatie, vormt de website PADI - Preserving Access to Digital Information (<http://www.nla.gov.au/padi/>) een goede gids. Er zijn technologische, organisatorische en juridische problemen. Technologische problemen betreffen de beperkte levensduur van de media waarop digitale informatie wordt opgeslagen – zo zal een diskette na verloop van tijd onleesbaar worden – en de snelle evolutie van de technologie – er verschijnt om de zoveel maanden wel een nieuwe versie van je tekstverwerker; ook besturingssystemen en hardware veranderen razendsnel. Twee veelbesproken strategieën om die technologische veroudering tegen te gaan zijn migratie en emulatie. Migratie is het periodiek converteren van bestanden. Bij emulatie is de filosofie: behoud de originele bestanden, maar laat het nieuwe systeem het oude systeem nabootsen. Organisatorische problemen betreffen kosten, expertise en

In de rubriek 'Instap' worden technieken die relevant zijn voor het vakgebied van archivariissen, bibliothecarissen en documentalisten op een introducerende manier voorgesteld.

De voorliggende Instap gaat in op de preservation metadata voor digitale archieven op basis van het OAIS-informatiemodel.

HET OAIS-MODEL

Een belangrijk probleem binnen het wetenschappelijk discours over het langdurig bewaren van digitale informatie is de terminologie. Eén van de verdiensten van het *Reference Model for an Open Archival Informa-*

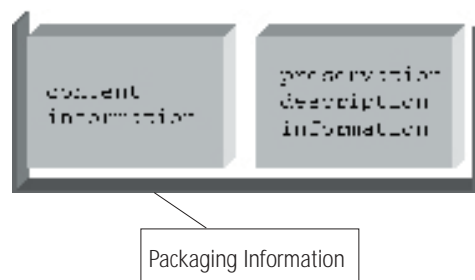
tion System (OAIS-RM) is dat er een nieuwe en goed doordachte terminologie wordt aangereikt.

Het OAIS-RM is een blauwdruk voor een digitaal archief. Zo een digitaal archief wordt nu evenwel een *Open Archival Information System* of OAIS genoemd. Het model is een creatie van het *Consultative Committee for Space Data Systems* (CCSDS), een internationale organisatie van ruimtevaartagentschappen; NASA en ESA zijn wellicht de bekendste leden. Het OAIS-model bestaat in feite uit twee modellen. Het functioneel model onderscheidt en analyseert zes belangrijke functionele entiteiten binnen een OAIS⁴. Het informatiemodel stelt de bewaarde informatieobjecten centraal. Het beschrijft wat er precies wordt bewaard in het OAIS en welke informatie het OAIS te allen tijde over die bewaarde digitale objecten moet worden bijgehouden om ze *langdurig* te kunnen bewaren. Dit laatste aspect maakt dat het informatiemodel een goede basis lijkt voor de beschrijving van die informatieobjecten (de metadata dus). In enkele projecten van eminente instellingen zoals de universiteiten van Oxford en Cambridge, de *Bibliothèque Nationale de France* en de *Library of Congress* is daarom het OAIS-informatiemodel als uitgangspunt genomen voor preservatie metadata-schema's.

Sterk vereenvoudigd bestaan informatieobjecten volgens het OAIS-informatiemodel uit een *data object* en *representation information*. Het *data object* is wat het archief wil bewaren. *Representation information* is de bijkomende informatie die nodig is om het *data object* te kunnen 'begrijpen'. Er kan enerzijds informatie nodig zijn om het object begrijpelijk te maken voor computers

(bv. kennis van het bestandsformaat) en anderzijds kan er informatie nodig zijn om het object begrijpelijk te maken voor mensen (bv. kennis van de taal van een tekst). Het OAIS-model spreekt hiervoor respectievelijk van *structure information* en *semantic information*. Er zijn nu – volgens CCSDS – vier belangrijke soorten of klassen informatieobjecten binnen een OAIS:

- *content information*: wat men wil bewaren, de inhoud;
- *preservation description information* (PDI): allerlei informatie die onontbeerlijk is om de content informatie langdurig te bewaren. Het OAIS-RM onderscheidt vier soorten PDI: *provenance information* (informatie over de herkomst van een object); *reference information* (informatie die een object vindbaar maakt), *context information* (informatie over de relaties tussen het object en zijn omgeving) en *fixity information* (informatie die de authenticiteit van het object garandeert);
- *packaging information*: de informatie die de vorige twee soorten met elkaar in verband brengt;
- *descriptive information*: beschrijvende informatie die nodig is om een bewaard object te kunnen terugvinden. Deze beschrijvende informatie wordt afgeleid van de content informatie en de preservation description information.



Een Archival Information Package (AIP) volgens het OAIS-RM.

Om de band tussen *content information* (bestaande uit een data object en representation information) en *preservation description information* te beklemtonen spreekt het OAIS-model

van informatiepakketten. Je kan die pakketten beschouwen als een conceptuele container van het bewaarde object met alle onontbeerlijke beschrijvende informatie. Er zijn drie soorten informatiepakketten:

- *Submission Information Packages* (SIPs): de informatiepakketten zoals ze worden ontvangen door het OAIS;
- *Archival Information Packages* (AIPs): de informatiepakketten zoals ze worden bewaard in het OAIS;
- *Dissemination Information Packages* (DIPs): de informatiepakketten zoals ze worden verspreid door het OAIS.

Het OAIS-informatiemodel is sterk schatplichtig aan het eindverslag van de *Task Force on Archiving of Digital Information* – een werkgroep die tussen 1994 en 1996 onderzocht welke problemen de archivering van digitale informatie stelde. In een hoofdstukje dat de verschillende dimensies van informatieobjecten in het digitale landschap beschrijft wordt er hierin immers gesproken van inhoud, referentie, context, herkomst en onveranderlijkheid (*content, reference, context, provenance en fixity*). Deze vier dimensies zijn volgens het document belangrijk om de integriteit van digitale informatieobjecten te bewaren. Een sterke analyse van digitale informatieobjecten is dit evenwel niet (dat was ook niet het doel van de tekst). CCSDS nam deze 'dimensies' over als basis voor haar taxonomie van informatieobjectklassen. *Packaging information* en *descriptive information* zijn nieuw. *Preservation description information* (PDI) neemt de begrippen *reference, context, provenance* en *fixity* over.

PRESERVATION METADATA-SHEMA'S

De bekende *Dublin Core Metadata Element Set* (www.dublincore.org) heeft als ambitie het mogelijk te maken aan de hand van een heel beperkte set kernelementen alle mogelijke informatieobjecten te beschrijven. Enkele projecten uit de bibliotheekwereld

hebben, net als *Dublin Core*, getracht op een hoog abstractieniveau informatieobjecten te beschrijven. Het doel van de metadata-elementen die deze projecten vooropstellen is evenwel niet de objecten vindbaar te maken, maar de langdurige bewaring ervan te ondersteunen. Ook hier wordt er evenwel sterk geabstraheerd. Er zijn immers verschillende soorten digitale informatieobjecten met specifieke kenmerken (een *e-book*, een gedigitaliseerde 19de-eeuwse krant, een website...). Ook de manier waarop men de informatieobjecten bewaart – de gebruikte 'strategie' – kan verschillende vormen aannemen (migratie, emulatie of een tussenvorm).

Een eerste poging om een set elementen samen te stellen die de langdurige bewaring van digitale informatieobjecten in een bibliotheek ondersteunt is '*preservation metadata for digital collections*' (1999)⁵. Dit ontwerp voor een metadata-schema van de nationale bibliotheek van Australië bevat 25 elementen en subelementen die drie niveaus kunnen beschrijven: collecties, objecten en individuele bestanden (in het geval van objecten die uit meerdere bestanden bestaan). De elementen beschrijven hardware- en software-afhankelijkheden (zoals bestandsformaat en systeemvereisten), hoe een bepaalde manifestatie van een object tot stand is gekomen (bv. door scanning, door migratie van een eerdere manifestatie), administratieve informatie zoals de reden voor bewaring, relaties met andere objecten en een element om de authenticiteit te controleren (bv. door een controlecijfer). Een implementatie op basis van dit schema is er nooit gekomen.

CEDARS – een project waaraan de universiteiten van Cambridge, Leeds en Oxford deelnamen – baseerde zijn '*metadata for digital preservation*'⁶ sterk op het eerder besproken OAIS-informatiemodel. Er wordt een onderscheid gemaakt tussen *representation information* en *preservation description information*. De elementen van het onderdeel *representation infor-*

mation zijn sterk afhankelijk van de bewaarstrategie die het CEDARS-experiment toepast – een combinatie van migratie en emulatie waarbij een informatieobject geconverteerd wordt naar een soort 'minimumversie' (CEDARS spreekt van de Onderliggende Abstracte Vorm) die toch alle belangrijke eigenschappen van het oorspronkelijke object bevat. Daarna wordt deze Onderliggende Abstracte Vorm omgezet tot één bitstream. De elementen van het onderdeel *preservation description information* zijn onderverdeeld in *reference information* (unieke identificatie en beschrijving), *provenance information* (beschrijft de herkomst van het informatieobject, documenteert wijzigingen aan het object en geeft informatie in verband met intellectuele eigendomsrechten), *context information* (relaties met andere informatieobjecten) en *fixity information* (informatie die de authenticiteit garandeert).

NEDLIB is een Europees project dat de problematiek van de wettelijke deponeering van digitale publicaties onderzoekt. In '*Metadata for the Long Term Preservation of Electronic Publications*'⁷ werd eveneens de terminologie van het OAIS-informatiemodel overgenomen. Dit schema biedt eigenlijk in hoofdzaak een gedetailleerde beschrijving van hardware- en software-afhankelijkheden. In OAIS-terminologie is dit *representation information*. *Preservation description information* is bij NEDLIB minder belangrijk dan bij het CEDARS-schema. Er wordt ook kritiek geleverd op de PDI-categorieën: met name de categorieën *context information* en *provenance information* worden onbevredigend genoemd. De PDI bevat bij NEDLIB slechts enkele elementen die unieke identificatie, controle van de authenticiteit en wijzigingen van de *representation information* documenteren. Deze beperkte invulling van de PDI is grotendeels te wijten aan het onderscheid dat – op een hoger niveau – wordt gemaakt tussen descriptieve metadata (met onder meer de bibliografische beschrijvingen van de objecten), administratieve metadata

en *preservation metadata* (bij NEDLIB dus veeleer een technische beschrijving die de langdurige bewaring moet ondersteunen). *Preservation metadata* krijgt bij het NEDLIB-schema bijgevolg een veel beperktere invulling dan bij het CEDARS-schema.

De vorige drie metadata-schema's werden verzoend in een recent document van een gezamenlijke werkgroep van het *Online Computer Library Center* (OCLC) en de *Research Libraries Group* (RLG): '*A metadata framework to support the preservation of digital objects*'⁸. De werkgroep bevatte vertegenwoordigers van onder meer de *Library of Congress*, de *British Library*, NEDLIB, CEDARS en de nationale bibliotheek van Australië. Ook nu deed men een beroep op de OAIS-terminologie. De *representation information* bestaat uit een uitgebreide technische beschrijving van de hardware- en software-omgeving van het digitaal informatieobject. De PDI bestaat opnieuw uit de vier categorieën met elementen over de unieke identificatie, de context, de herkomst en de integriteit.

Enigszins verrassend is het feit dat dit onderzoek totnogtoe weinig weerklank vindt in de archiefwereld. Eén recent artikel⁹ plaatste alvast – ongetwijfeld terecht – enkele vraagtekens bij de toepasbaarheid van de voorgestelde *preservation metadata*-schema's in archieven. Vooral wat betreft de gebruikte terminologie (archivarissen zullen enigszins verbaasd zijn over de nieuwe betekenis van *provenance* en *context* in de bovenstaande *preservation metadata*-schema's) bestaan er misverstanden.

CONCLUSIE

Bibliotheken en archieven hebben als taak het culturele erfgoed te bewaren, ook wanneer dit erfgoed een digitale vorm aanneemt. Hoe men digitale duurzaamheid kan garanderen is nog geen uitgemaakte zaak. Het '*Reference Model for an Open Archival Information System*' levert alvast een

Preservation metadata

nuttige blauwdruk voor een digitaal archief. Vele projecten nemen deze standaard dan ook als uitgangspunt.

Preservation metadata is een recente topic binnen de discussie omtrent digitale duurzaamheid. Pogingen om metadata-schema's te ontwerpen die de langdurige bewaring van digitale informatieobjecten ondersteunen, leveren uiteenlopende resultaten op. Ondanks het alom gerespecteerde OAIS-informatiemodel heeft de theorievorming hier wellicht nog een lange weg te gaan. Ook praktische implementaties laten op zich wachten.

ENKELE REFERENTIES

- *CCSDS. Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1 BLUE BOOK.* Washington: Consultative Committee for Space Data Systems, 2002.
URL: <http://www.ccsds.org/recommandreports.jsp>
- *Cedars Project en UKOLN. Metadata for Digital Preservation: the Cedars Project Outline Specification.* 2000.
URL: <http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html>
- *Cedars Project. Cedars Guide to Preservation Metadata.* 2002
URL: <http://www.leeds.ac.uk/cedars/guideto/metadata/>
- Hodge, Gail. *Metadata made simpler.* -Bethesda: NISO Press, 2001.
URL: www.niso.org
- Holdsworth, David en Sergeant, Derek M., *A blueprint for Representation Information in the OAIS model* (paper voor de Eighth NASA Goddard Conference on Mass Storage Systems and Technologies, 27-30 maart 2000, Maryland, USA). 2000
URL: <http://esdis-it.gsfc.nasa.gov/MSST/conf2000/PAPERS/DO2PA.PDF>
- Jones, Maggie en Beagrie, Neil, *Preservation Management of Digital Materials: A Handbook.* - London: The British Library, 2001.
URL: <http://www.dpconline.org/graphics/handbook/index.html>
- Lupovici, Catherine en Masanès, Julien, *Metadata for the Long Term Preservation of Electronic Publications.* - Den Haag: Koninklijke Bibliotheek, 2000. - (NEDLIB Report series; 2).
- *OCLC/RLG. Preservation Metadata for Digital Objects: A Review of the State of the Art. OCLC/RLG Working Group on Preservation Metadata.* 2001
URL: <http://www.oclc.org/research/pmwg/documents.shtm>
- *OCLC/RLG. Preservation Metadata and the OAIS Information Model. A Metadata Framework to Support the Preservation of Digital Objects. OCLC/RLG Working Group on Preservation Metadata.* 2002
URL: <http://www.oclc.org/research/pmwg/documents.shtm>
- Waters, Donald; Garret, John e.a., *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and The Research Libraries Group.* - 1996. - (CLIR report: 63).
URL: <http://www.clir.org/pubs/abstract/pub63.html>
URL: <http://www.rlg.org/ArchTF/>
- Webb, Colin, *Preservation Metadata for Digital Collections. Exposure Draft.* - National Library of Australia, 1999.
URL: <http://www.nla.gov.au/preserve/pmeta.html>

NOTEN

1. Dit artikel is gebaseerd op een eindverhandeling voor de GGS Informatie- en Bibliotheekwetenschap. Hartelijk dank aan professor Kees van der Meer voor het proeflezen en de suggesties.
2. NEDLIB onderzocht de wettelijke deponering van digitale publicaties. Resultaat van het onderzoek was een model voor een depotsysteem (DSEP - deposit system for electronic publications). Het NEDLIB-project liep van 1 januari 1998 tot 31 december 2000. De coördinatie was in handen van de Nederlandse Koninklijke Bibliotheek. De eindrapporten en meer informatie zijn te vinden op <http://www.kb.nl/coop/nedlib/>.
3. Het CEDARS-project (april 1998-maart 2002) onderzocht de langdurige bewaring van digitale informatie en werd geleid door de universiteiten van Leeds, Oxford en Cambridge. Meer informatie op <http://www.leeds.ac.uk/cedars/>.
4. Deze zes functionele entiteiten zijn *ingest* (staat in voor de ontvangst van informatiepakketten); *preservation planning* (staat in voor de langdurige bewaring); *archival storage* (opslag en retrieval van informatiepakketten); *data management* (beheer van beschrijvende en administratieve informatie); *administration* (algemene werking) en *access* (toegankelijkheid tot en verspreiding van informatie).
5. <http://www.nla.gov.au/preserve/pmeta.html>
6. <http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html>
7. Lupovici, Catherine en Masanès, Julien, *Metadata for the Long Term Preservation of Electronic Publications.* - Den Haag: Koninklijke Bibliotheek, 2000. - (NEDLIB Report series ; 2). Ook online beschikbaar op de Nedlib-website (<http://www.kb.nl/coop/nedlib/>).
8. <http://www.oclc.org/research/pmwg/documents.shtm>
9. Hofman, Hans (2002), *Review: Some Comments on Preservation Metadata and the OAIS Model*, in: *Difficult Info. A Newsletter on Digital Culture.* (October 2002) 2.

SAMENVATTING

Bibliotheeken en archieven hebben als taak het culturele erfgoed te bewaren, ook indien dit erfgoed een digitale vorm aanneemt. Hoe men digitale duurzaamheid kan garanderen is nog geen uitgemaakte zaak. Een recente ISO-standaard – het 'Reference Model for an Open Archival Information System' – levert alvast een nuttige blauwdruk voor een digitaal archief. Recente projecten als CEDARS en NEDLIB nemen deze standaard dan ook als uitgangspunt bij onderzoek naar digitale duurzaamheid, ook wat betreft de benodigde metadata bij de langdurige bewaring van digitale informatieobjecten. Hoewel er gezocht wordt naar een consensus omtrent preservation metadata, heeft de theorievorming nog een lange weg te gaan.

ABSTRACT

Libraries and archives are responsible for the preservation of the cultural heritage, also when that heritage has assumed a digital shape. How digital durability will be guaranteed still remains unclear. A recent ISO-standard - the 'Reference Model for an Open Archival Information System' - provides a first and useful blueprint for a digital archive. Recent projects, such as CEDARS and NEDLIB, take this standard as the starting point for their research into digital durability, also as concerns the metadata needed for long-term preservation of digital information objects. Although a consensus on preservation metadata is being searched for, the development of theories still has a long way to go.