
DOMME NUMMERS VOOR INTERNET CONTENT : Permanente DOI vervangt URL als referentie in publicaties

Hans van THIEL
Char Technical Writing & Journalism
Passeerdersstraat 76, NL-1016 XZ Amsterdam

De ' *Digital Object Identifier* ' (DOI) is een onveranderlijk identificatienummer voor digitale inhoud dat o.m. bedoeld is om die inhoud toegankelijk te houden, ook als de locatie op het internet verandert. Behalve dat een DOI aanklikbaar is, in een applicatie die dit ondersteunt, is hij gekoppeld aan gestandaardiseerde metadata. CrossRef voor wetenschappelijke publicaties is de belangrijkste toepassing, maar DOI kan alle vormen van gedistribueerde intellectuele eigendom ondersteunen. Het is een ' barcode voor digitale netwerken '.

De tien miljoenste ' *Digital Object Identifier* ' is toegekend aan een technisch artikel in een tijdschrift van Springer Verlag. Het gelukkige nummer is DOI:10.1007/s00348-003-0647-4 en geeft, met behulp van een browser plug-in, toegang tot een website met o.m. een samenvatting. Het bijzondere is dat over vijf, vijftig of honderd jaar die DOI nog steeds naar dat artikel zal verwijzen, of naar informatie over dat artikel, wat er ook inmiddels mee is gebeurd.

Hiermee heeft de International DOI Foundation (IDF) haar eerste doel bereikt : een unieke en duurzame identifier die op een netwerk geactiveerd kan worden. Een DOI verwijst naar een inhoud en niet naar een adres, maar geeft op het moment van gebruik de dan geldige locatie.

ZONDER CATALOGUS

Het internet is wel eens getypeerd als een ' gigantische bibliotheek zonder catalogus ' en de in 1998 opgerichte IDF probeert eigenlijk een basis voor zo'n catalogus te leggen.

Ontkoppeling van digitale inhoud en locatie ondersteunt ook de exploitatie van intellectuele eigendom en bijbehorend beheer van rechten (Digital Rights Management - DRM). Als digitale tekst, audio, video of welk werk dan ook in andere handen overgaat en dus door een nieuwe eigenaar wordt beheerd, verandert de DOI niet mee.

Voor de uitgeefwereld is dan ook sterk vertegenwoordigd in de IDF, maar behalve o.m. Elsevier, Springer Verlag, de Association of American Publishers en de Nederlandse Koninklijke Bibliotheek zijn ook de ACM, Microsoft en Hewlett Packard lid van de stichting.

Ondersteuning van een DOI systeem bestaat uit

vier componenten: de DOI zelf als betekenisloos ofwel ' dom ' identificatienummer, een beschrijving van de entiteit waar de DOI naar verwijst (meta-data), een netwerktechnologie om de DOI te activeren tot levering van iets (aan de client) en een beleid om het systeem in de praktijk te laten werken.

HANDLE SYSTEM

De DOI technologie is niet van de grond af aan opgebouwd, maar vormt een implementatie van het ' *Handle System* ' (HS) dat door de (Amerikaanse) Corporation for National Research Initiatives (CNRI) is ontwikkeld. De principes zijn al in 1995 beschreven door internet pioniers KAHN en WILENSKY in hun artikel ' *A Framework for Distributed Digital Object Services* '. Met dat raamwerk wordt het mogelijk om op computernetwerken en met name het globale internet naar inhoud te verwijzen ipv naar de adressen.

Een stuk digitale inhoud ofwel object krijgt daartoe een duurzaam handvat toegewezen dat door een ' *handle system server* ' gebruikt kan worden om de client een adres of URL (Uniform Resource Locator) te retourneren.

Natuurlijk moet de relatie tussen handel en locatie dan worden bijgehouden en dat gebeurt door de ' *local handle services* ' onder verantwoordelijkheid van een ' *global handle registry* '. Handelsvolgen dezelfde syntax als DOI's en het voorvoegsel wordt centraal toegewezen en kan dan zonder omwegen naar de juiste ' *local handle server* ' verwijzen.

Een handelsstelsel is dus in zekere zin een DNS (Domain Name Service) met een extra indirectie stap.

In tegenstelling tot DNS is HS echter niet uni-

verseel geïmplementeerd op het internet. Wel onderhoudt het CNRI een ' *Public Handle Service* ' en heeft het een Java versie van een ' *local handle server* ' vrij beschikbaar gesteld. Voor Windows clienten is er een ' *handle system resolver* ' die, als web browser plug-in, handels kan verwerken. Ook voor Adobe Acrobat en Acrobat Reader is er een dergelijke plug-in.

Een DOI is, als implementatie van een handel, met deze hulpmiddelen ook direct aanklikbaar. In een standaard web browser zijn DOI's ook aanklikbaar, maar dan moeten zij deel uit maken van een URL met de op dit moment in gebruik zijnde proxy server dx.doi.org. Voor de tienmiljoenste DOI wordt dat dus :

<http://dx.doi.org/10.1007/s00348-003-0647-4>.

Het genoemde artikel van KAHN en WILENSKY heeft geen DOI maar wel een handel die op hdl.handle.net wordt beheerd. Aanklikken van hdl.handle.net/4263537/5001 (in een gewone browser zonder plug-in) produceert een html pagina die op het moment van schrijven de URL <http://www.cnri.reston.va.us/cstr/arch/k-w.html> had, maar die wanneer U dit leest best ergens anders kan staan.

METADATA

Met een duurzame identificator wordt het mogelijk om standaard gegevens aan digitale inhoud te koppelen. Het zijn deze metadata die DOI's toegevoegde waarde geven en de basis vormen voor een catalogus of, vanuit een ander gezichtspunt, een barcode voor het internet.

De IDF gebruikt hiervoor het INDECS (INteroperability of Data in E-Commerce Systems) raamwerk met een ' *data dictionary* ' die ook door MPEG-21 (Moving Pictures Expert Group) is overgenomen.

Er zijn nogal wat van dergelijke standaarden in omloop, zoals het Dublin Core Metadata Initiative en ONIX (Online Information Exchange), maar de IDF heeft diverse afbeeldingen geconstrueerd tus-sen sector specifieke formaten en INDECS.

Ieder digitaal object met een DOI heeft in ieder geval een metadata kern die o.m. een titel, een type (abstract werk, uitvoering etc.), een mode (visueel, audio, audio-visueel etc.) en een primaire agent (auteur, componist etc.) specificiert. Naast deze voor elke DOI verplichte metadata kan een digitale inhoud ook een of meerdere ap-

plicatie profielen meekrijgen. Zo kan bijvoorbeeld een DOI registratie agentschap voor spellen op mobiele telefoons een ander applicatie profiel gebruiken dan Cross-Ref, dat zich met wetenschappelijke tijdschriften, boeken en congresverslagen bezighoudt.

CrossRef is op dit moment verreweg de grootste DOI beheerder. Uitgevers deponeren tegelijkertijd met een DOI bibliografische metadata (volgens een XML schema) die onderhouden wordt in een metadata database (MDDB). Die MDDB kan door middel van zoekopdrachten ook DOI's vinden van literatuurverwijzingen in artikelen (catalogus functie). CrossRef wordt beheerd door de in 2000 opgerichte Publishers International Linking Association.

De vierde component van het DOI systeem is het beleid. De administratie van DOI's en metadata wordt bijgehouden door ' *registration agencies* ' maar de IDF wijst centraal blokken identificatienummers toe (de DOI voorvoegsels) en bepaalt de algemene voorwaarden voor het gebruik.

Verder houdt de IDF zich bezig met de ontwikkeling van het DOI systeem. De volgende stap wordt waarschijnlijk de integratie van DOI identificatie, handels en metadata met OpenURL contextafhankelijk linken. Met OpenURL kan het resultaat van een ' DOI klik ' afhankelijk worden van de betreffende aanvrager en diens rechten voor toegang tot de ' *content* '. Dit is behalve voor institutionele gebruikers, zoals bibliotheken of grote ondernemingen, ook interessant voor ' *Digital Rights Management* ' in het algemeen.

WEB:

<http://www.doi.org>

<http://www.handle.net>

<http://www.indecs.org>

<http://www.crossref.org>

<http://www.ukoln.ac.uk/distributed-systems/openurl/>

DOI VOOR EEUWIG

De enige voorwaarde waaraan een ' *Digital Object Identifier* ' moet voldoen is dat hij uniek en duurzaam is. Een DOI kan naar alles verwijzen, een abstract werk zowel als een uitvoering of een

registratie in een bepaald formaat, naar een geheel zoals een tijdschriftnummer of een gedeelte zoals een artikel in dat tijdschrift enz. Maar een DOI mag nooit naar twee zaken verwijzen en zodra een entiteit een DOI heeft gekregen kan die identificator niet worden hergebruikt, ook niet als het werk inmiddels van het internet is verwijderd. Nieuwe soortgelijke uitvoeringen of versies krijgen, indien gewenst, een nieuw DOI.

Een DOI bestaat uit een voorvoegsel en een navoegsel, beiden van willekeurige lengte, gescheiden door een '/'. Het formaat is gebaseerd op de ANSI/NISO Z39.84-2000 standaard, maar in een DOI wordt geen onderscheid gemaakt tussen kleine en hoofdletters.

10.2345/S1384107697000225
10.4567/0361-9230(1997)42:<OaEoSr>2.0.TX;2-B
10.6789/JoesPaper56

De '10' aan het begin identificeert bovenstaande regels als DOI's en het gedeelte tussen '.' en '/' is een registratienummer dat onder verantwoordelijkheid van de International DOI Foundation door een 'Registration Agency' (RA) wordt toegewezen. CrossRef is zo'n RA en heeft bijvoorbeeld 'DOI prefixes' gegeven aan o.m. de IEEE (10.1109), Academic Press (10.1006) en ook Kluwer (10.1023).

De houders van een DOI voorvoegsel kunnen voor het gedeelte na '/' elk identificatieschema gebruiken dat zij willen. Zij zijn ook volledig verantwoordelijk voor de uniekheid en duurzaamheid van hun 'DOI suffix'. Dat lijkt misschien problemen te kunnen geven, maar DOI's

worden als indirecte URL verwijzing op het internet geadmistreerd en duplicaties kunnen dus gemakkelijk worden geweerd.

De mogelijkheid om bestaande identificatieschema's te kunnen gebruiken in een DOI (als suffix) is een groot voordeel. Niet alleen officiële nummeringen als ISBN en ISSN (boeken resp. tijdschriften) maar allerlei interne administratiesystemen met willekeurige betekenissen kunnen met een geregistreerde prefix tot DOI omgevormd worden.

Dergelijke geïmpliceerde betekenissen zoals het bovenstaande 'JoesPaper56' of '(1997)' in de identificator daarboven hebben echter geen externe betekenis. Een DOI is een zgn. 'domnummer'.

Juist vanwege deze 'domheid' kan de DOI hetzelfde blijven als een entiteit verplaatst of verhandeld wordt.

Stel het onwaarschijnlijke geval dat de IEEE het tijdschrift Nature zou overnemen dan blijft het artikel:

"Cell Biology: A cat cloned by nuclear transplantation" Nature AOP, Published online: 14 February 2002, DOI:10.1038/nature723 dezelfde identificatie behouden. Dus 10.1038 wordt niet veranderd in de 10.1109 van de IEEE, ook al zou het betreffende artikel waarschijnlijk naar een IEEE website zijn verplaatst, en de url overeenkomstig zijn bijgewerkt. Alle bestaande DOI verwijzingen, zowel in druk als op het internet, blijven onveranderd geldig.

* * *