



Metadata en Dublin Core :

een inleiding

Metadata for this document:

Title	Metadata en Dublin Core : een inleiding
Creator	Ursula Oberst
Subject	DARE repositories; metadata; Dublin Core
Description	Dit document is de handout van de metadata sessie gehouden door Ursula Oberst tijdens de DARE summer school.
Publisher	Stichting SURF
Date	2005-06-28
Type	Educational material
Format	Text/richtext
Language	Dutch

Inhoudsopgave

1. Inleiding.....	2
1.1 Wat zijn metdata?	2
1.2 Wat is Dublin Core?	2
1.2 Wat is het OAI-PMH protocol?.....	4
2. Syntax en basisprincipes van Dublin Core metadata in DARE verband.....	6
2.1 Doel en reikwijdte van DARE DC	6
2.2 DARE DC Simple	6
2.3 DARE DC Qualified.....	8
2.4 XML schema's voor de codering van metadata	10
2.5 Beschrijvingen van tijdschriftenartikelen.....	12
3. Meer informatie	13
4. Bijlage: Willie Wortels (werkgroep implementatie)	14

1. Inleiding

1.1 Wat zijn metdata?

Metadata betekent letterlijk: data over data (gegevens over gegevens). Metadata is gestructureerde informatie die objecten (documenten, afbeeldingen, geluidsfragmenten) beschrijft, toelicht, lokaliseert of op een andere manier eenvoudiger vindbaar, bruikbaar of beheerbaar maakt. Metadata kan ook worden gezien als informatie over informatie. In een bibliotheek kennen we dit fenomeen in de vorm van cataloguskaarten, waarin onder andere de titel, naam van de auteur, datum van uitgave en een samenvatting van de inhoud zijn opgenomen. In het Internettijdperk wordt de term “metatdata” vaak begrepen als “beschrijvende informatie over internetbronnen”.

1.2 Wat is Dublin Core?

Dublin Core (kort DC) is een door internationale experts (van IT-specialisten tot bibliothecarissen) ontwikkelde **niet sectorspecifieke metadatastandaard**.

Er bestaan twee vormen van DC: Simple (unqualified) and Qualified.

DC Simple (of vereenvoudigd Dublin Core) bestaat uit vijftien basiselementen:

Element	Omschrijving	Voorbeeld
• dc:title	(<i>titel</i>)	Palm tree justice in the Bertoua court of appeal : the witchcraft cases
• dc:subject	(<i>onderwerp en trefwoorden</i>)	Rechtspraak; heksen, heksenprocessen; Bertoua

- dc:description (*omschrijving/abstract*) This paper examines thirty witchcraft cases reviewed by the Court of Appeal of Bertoua (East Province, Cameroon) during the period 1981-1984
- dc:source (*bron*)
- dc:language (*taal*) English
- dc:relation (*relatie*)
- dc:coverage (*dekking*) Cameroon
- dc:creator (*auteur of maker*) Fisiy, C.F
- dc:publisher (*uitgever*) African Studies Centre, Leiden
- dc:contributor (*andere medewerkers*)
- dc:rights (*rechten*) Fisiy, C.F
- dc:date (*datum*) 1990
- dc:type (*bestandstype*) Research paper
- dc:format (*bestandsformaat*) application/pdf
- dc:identifier (*bestandsidentificatie*) <http://hdl.handle.net/1887/432>

Dublin Core Qualified is een uitbreiding van de 15 basiselementen (*met de 3 elementen Audience, Provenance and RightsHolder*).

Binnen DC qualified wordt gebruik gemaakt van **extensies**, die het mogelijk maken om elementen nader te specificeren. Deze extensies worden verfijningen (**element refinements**) of qualifiers genoemd. Bijvoorbeeld "created", "valid", "issued" en "modified" zijn de aanbevolen verfijningen van het element dc:date. Dus, dc.date.created, zou dan de naam zijn voor het element creatiedatum van een document in DC qualified.

Daarnaast is bij Dublin Core qualified voor een aantal elementen vastgesteld welke gestandaardiseerde **schema's** gebruikt kunnen worden voor de inhoudelijke invulling van de elementen. Zo heeft het element dc: type twaalf aanbevolen termen: Collection, dataset, event, image, interactive resource, service, software, sound, text, physical object, still image, moving image.

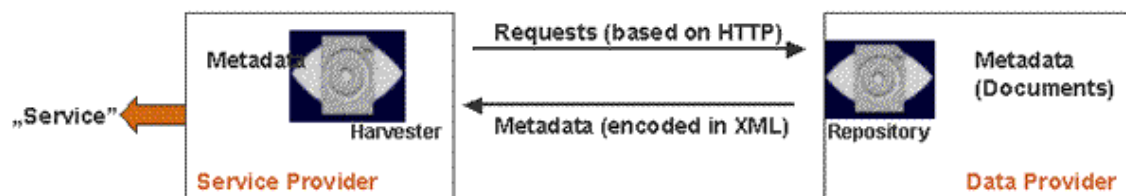
DC Simple	DC Qualified
dc.date 28 juni 2005 Er zijn geen afpraken hoe de datum zou moeten worden ingevuld, en ook niet welke datum.	dc.date. <u>created</u> 2005-06-28 * <u>element refinement</u> (door de extensie "created" te gebruiken wordt het element nader gespecificeerd: het element bevat de creatiedatum van het object) * Het W3C-DTF <u>schema</u> bepaalt het datumformaat: YYYY-MM-DD

1.2 Wat is het OAI-PMH protocol?

Het OAI-PMH protocol is een verzameling van afspraken voor het “harvesten” van metadata. Het OAI-PMH bevat o.a. afspraken over:

- het transport protocol (Hoe communiceren service and data provider met elkaar? Via http)
- **het metadata formaat** (Welk metadata formaat ondersteunen service and data provider? **Minimaal DC Simple**).
- Het response formaat (Hoe moet de metadata “verpakt” worden? In [XML](#), Extensible Markup Language)

Hoe functioneert het OAI-PMH?

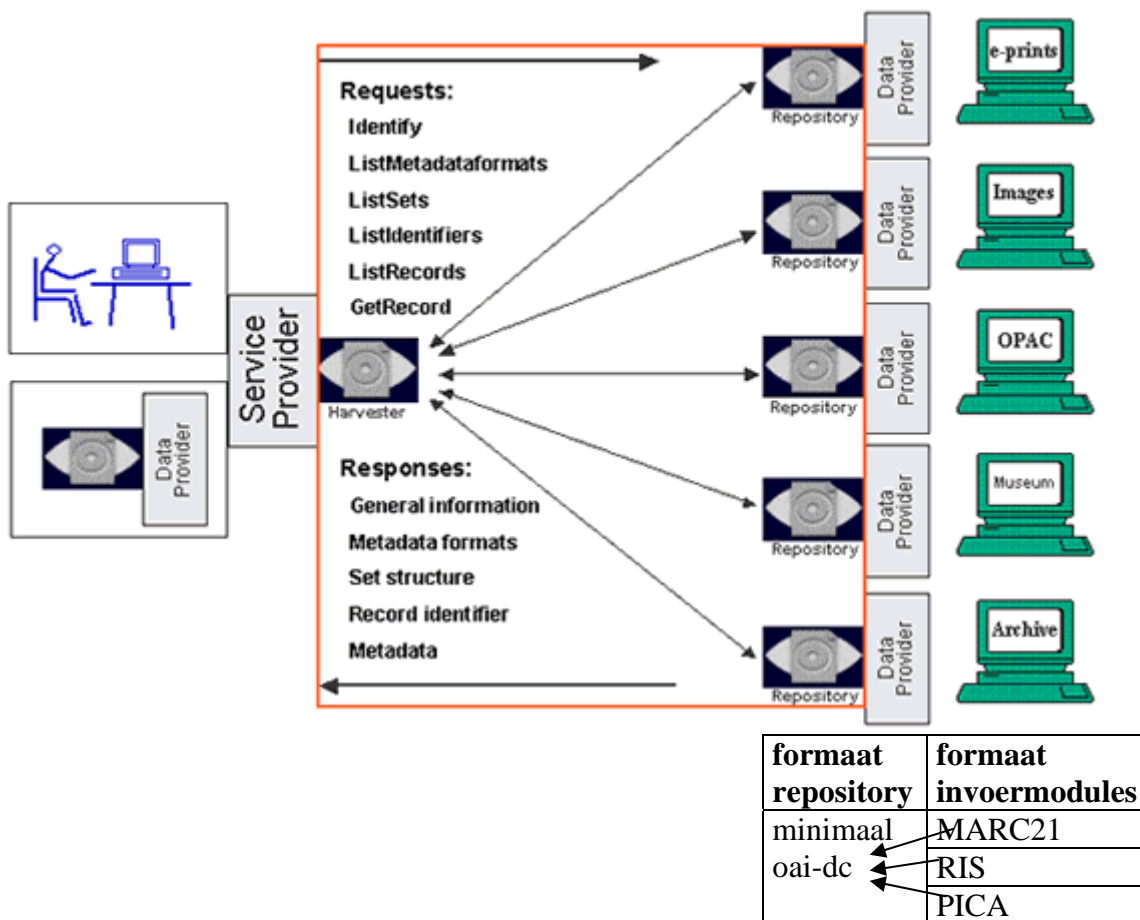


(Bron: OAI-PMH tutorial)

Een dienst (service provider) zoals b.v. Connecting-Africa stuurt “requests” (gebaseerd op HTTP) aan een repository . Mogelijk request: “Stuur mij alle metdata records die vanadag aan de repository toegevoed zijn”

(<http://archive.org?verb=ListRecords&from=2005-06-28>). Als antwoord ontvangt de dienst alle metdata records, minimaal in het formaat DC simple, verpakt in een XML schema. Data en service provider kunnen met elkaar communiceren omdat zij alle twee het OAI-PMH protocol geïmplementeerd hebben, dwz dat zij zich alle twee aan dezelfde talen houden.

Schema met meerdere data providers (en verschillende invoermodules):



(Bron: OAI-PMH tutorial)

Een repository kan op verschillende manieren gevuld worden. De auteur vervaardigt zelf een metadatabeschrijving in het invoermodule van de repository software (b.v. Dspace). De bibliotheek vervaardigt metadatabeschrijving of de repository put uit de metadatabeschrijvingen van een onderzoeksregistratiesysteem zoals METIS. Al deze verschillende systemen maken ieder van hun eigen beschrijvingsformaat gebruik (MARC21, RIS, Pica). De repository moet vervolgens de verschillende metadata formaten vertalen (“converteren”) naar minimaal oai-dc (dc simple) om de repository harvestable te maken voor de service providers.

- ⇒ Leiden en RUG hebben een GGC conversie ontwikkeld, RUG en TuE hebben een RIS conversie.
- ⇒ Voor de mapping van andere metadata formaten naar DC zie verder: de UKLON website: Mapping between metadata formats: <http://www.ukoln.ac.uk/metadata/interoperability/>

2. Syntax en basisprincipes van Dublin Core metadata in DARE verband

2.1 Doel en reikwijdte van DARE DC

In DARE verband heeft zich de werkgroep implementatie (ook de Willie Wortels genoemd) gebogen over de vraag welke metadatastandaard DARE repositories moeten gebruiken en op welke manier. Er is besloten conform het OAI-PHM protocol Dublin Core als metadatastandaard te gebruiken. De manier waarop DC binnen DARE toegepast wordt is beschreven in de DARE richtlijnen "[DARE use of Dublin Core, versie 2.0, december 2004](#)" geschreven door Maarten Feijen en Marlon Domingus. De DARE richtlijnen beschrijven in eerste instantie de mapping van een intern formaat (b.v. MARC21) naar DARE DC. Ze zijn niet bedoeld als regels voor het catalogiseren.

Versie 2.1 (geschreven door Maarten Feijen) verschijnt over ongeveer een maand en zal ook de afspraken over het beschrijven van tijdschriftartikelen bevatten.

De DARE partners hebben afgesproken twee XML schema's te implementeren: een schema for *Unqualified DC* om ook voor service providers buiten de DARE gemeenschap harvestable te zijn, een Qualified DC voor het gebruik binnen de DARE gemeenschap.

Per 1-1-2005 zijn alle DARE repositories verplicht tenminste oai-dc te ondersteunen. Harvesting binnen DARE is gebaseerd op oai-dc.

2.2 DARE DC Simple

Binnen DARE DC zijn een aantal elementen verplicht zoals b.v. dc:title. Voor een aantal elementen is vastgesteld welke gestandaardiseerde schema's gebruikt moeten worden (zoals b.v. de METIS-publicatie types in dc:type).

Quick reference sheet:

Element	Verplichte elementen	Schema's	Omschrijving en voorbeeld
Title	M(andatory)	None	A name given to the resource. <dc:title>Preliminary studies for the "Philosophical Investigations", generally known as the blue and brown books</dc:title>
Creator	M	None	An entity primarily responsible for making the content of the resource. Use inverted name, so: surname, forename, prefix <dc:creator> Janssen, J. </dc:creator> When initial and full name are both available use this formatting: <dc:creator> Janssen, J.

			<i>(John)</i> </dc:creator>
Subject	MA (andatory when applicable)	Choice of keywords and classifications is free	The topic of the resource. <dc:subject> <i>polar oceanography</i> ;</dc:subject>
Description	MA	None “Abstract” is the <u>default value</u> for dc:description	An account of the content of the resource <dc:description> <i>A number of problems in quantum state and system identification are addressed.</i> </dc:description>
Publisher	MA	None	An entity responsible for making the resource available <dc:publisher> <i>John Wiley & Sons, Inc. (US)</i> </dc:publisher>
Contributor	O (ptional)	None	An entity responsible for making contributions to the content of the resource. Examples of contributors are: a supervisor, editor, technician or data collector. <dc:contributor> <i>Evans, R. J</i> </dc:contributor>
Date	M	Date ISO 8601 W3C-DTF “Created” is the <u>default value</u> for dc:date	A date associated with an event in the life cycle of the resource. <dc:date> <i>2000-12-25</i> </dc:date>
Type	M	METIS-list with additional DCMI types .	The type of scientific output the resource is a manifestation of. <dc:type> <i>preprint</i> </dc:type>
Format	R (ecommended)	IANA list of MIME types	The physical or digital manifestation of the resource. <dc:format> <i>application/pdf</i> </dc:format>
Identifier	M	URI	An unambiguous reference to the resource within a given context. Use an URI to point to the resource (metadata).
Source	O	None	A reference to a resource from which the present resource is derived.
Language	R	ISO 639-1	A language of the intellectual content of the resource. <dc:language> <i>en</i> </dc:language>

Relation	R	none	The reference to a related resource, f.e. a previous version of an article.
Coverage	O	Period	The extent or scope of the content of the resource. Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity) <dc:coverage>NL</dc:coverage>
Rights	M	None	Information about rights held in and over the resource. <dc:rights>(c) University of Bath, 2003</dc:rights>
Audience	O	None	A class of entity for whom the resource is intended or useful. <dc:audience> Students </dc:audience>

⇒ Voor meer informatie zie de DARE richtlijnen “[DARE use of Dublin Core, versie 2.0, december 2004](#)”

2.3 DARE DC Qualified

Quick reference sheet:

Element	Element Refinement	Verplichte elementen	Schema's
Title	- Alternative	M MA	None
Creator	-	M	None
Subject	GOO, NBC, LCSH, MESH, DDC, LCC, UDC, LOCAL	MA	Choice of keywords and classifications is free. Use refinements when appropriate.
Description	- TableOfContents Abstract	MA R R	None
Publisher	-	MA	None
Contributor	-	O	None
Date	- dateAccepted	M R	Date ISO 8601 W3C-DTF

	dateCopyrighted Created Valid Available Issued Modified dateSubmitted	R R R R R R	Created is default in mapping
Type	-	M	METIS-list with additional DCMI types.
Format	- Extent Medium	R R R	IANA list of MIME types
Identifier	- Bibl. citation	M R	URI
Source	-	O	None
Language	-	R	ISO 639-1
Relation	- Isversionof Hasversion Replacedby Replaces Requiredby Requires Ispartof Haspart Isreferredby References Isformatof hasFormat Conformsto	R R R R R R R R R R R R R	none
Coverage	- Spatial	O R	Point ISO 3166 Box TGN

	Temporal	R	Period
Rights	- Access rights License Rights holder	M MA O O	None
Audience	- Mediator Education level	O O O	None

⇒ Voor meer informatie zie de DARE richtlijnen "[DARE use of Dublin Core, versie 2.0, december 2004](#)"

2.4 XML schema's voor de codering van metadata

OAI-PMH gebruikt XML schema's om record formaten te definiëren. In principe kunnen de verschillendste soorten metadata (MARC21, PICA enz.) uitgewisseld worden conform OAI-PMH als de metadata maar in XML gecodeerd zijn en er een XML schema voor gedefinieerd is.

Een in XML gecodeerd record bestaat uit de volgende delen:

- **"header"** – bevat:
 - de unieke "identificer" van een object in een repository (in bijgevoegd voorbeeld: :arXiv.org:cs/0112017).
 - een datumopgave – invoerdatum, modificatiedatum of verwijderdatum om selectief te kunnen "harvesten" (in voorbeeld: 2002-02-28).
 - nul of meerdere set specificaties
Set: gedefinieerde groep van items binnen een repository; "onderverdeling" van repository op basis van b.v. materiaalsoort of organisatorische eenheden. (Het voorbeeldrecord hoort bij twee sets *math* en *cs*)
 - facultatief: status attribuut met de waarde "deleted" om verwijderde metadatabeschrijvingen te markeren.
- **"metadata"** bevat:
 - metadata prefix: oai_dc:dc (prefix duidt aan welke metadatastandaard gebruikt wordt)
 - de XML namespace (xmlns) voor het oai_dc formaat:
xmlns:oai_dc=http://www.openarchives.org/OAI/2.0/oai_dc/
An XML namespace is a collection of names, identified by a URI reference [RFC2396], which are used in XML documents as element types and attribute names. (Definition quoted from W3C—Namespaces in XML at http://www.w3.org/TR/REC-xml-names/)
 - De XML namespace voor de DCMES (Dublin Core Metadata Element Set) elementen:
xmlns:dc="http://purl.org/dc/elements/1.1/"
XML Schemas express shared vocabularies and allow machines to carry out rules made by people. They provide a means for defining the

structure, content and semantics of XML documents. (Definition quoted from W3C Architecture Domain—XML schema at <http://www.w3.org/XML/Schema>)

- Het container schema verbonden aan de oai_dc namespace:
xsi:schemaLocation=http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd
Containers are places in OAI-PMH responses where XML complying with any external schema may be supplied. Containers are provided for extensibility and for community specific enhancements. The OAI Implementation Guidelines lists the existing optional containers and provides links to existing schemas.
- metadata
- facultatief: “about” container . Deze container bevat informatie over de herkomst van het metadata record. Het record in het voorbeeld is zelf geharvest. Het provenance statement laat zien van welke repository het record afkomstig is en wanneer het record geogst werd.

Voorbeeld:

```
<header>
  <identifier>oai:arXiv:cs/0112017</identifier>
  <datestamp>2002-02-28</datestamp>
  <setSpec>cs</setSpec>
  <setSpec>math</setSpec>
</header>
<metadata>
  <oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>Using Structural Metadata to Localize Experience of Digital
      Content</dc:title>
    <dc:creator>Dushay, Naomi</dc:creator>
    <dc:subject>Digital Libraries</dc:subject>
    <dc:description>With the increasing technical sophistication of both
      information consumers and providers, there is increasing demand for
      more meaningful experiences of digital information. We present a
      framework that separates digital object experience, or rendering,
      from digital object storage and manipulation, so the
      rendering can be tailored to particular communities of users.
    </dc:description>
    <dc:description>Comment: 23 pages including 2 appendices,
      8 figures</dc:description>
    <dc:date>2001-12-14</dc:date>
    <dc:type>e-print</dc:type>
    <dc:identifier>http://arXiv.org/abs/cs/0112017</dc:identifier>
  </oai_dc:dc>
</metadata>
<about>
  <provenance
    xmlns="http://www.openarchives.org/OAI/2.0/provenance"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/provenance
      http://www.openarchives.org/OAI/2.0/provenance.xsd">
```

```

<originDescription harvestDate="2002-02-02T14:10:02Z" altered="true">
  <baseUrl>http://the.oa.org</baseUrl>
  <identifier>oai:r2:klik001</identifier>
  <timestamp>2002-01-01</timestamp>

<metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
  </originDescription>
  </provenance>
</about>

```

(bron: [OAI-PMH protocol](#))

2.5 Beschrijvingen van tijdschriftenartikelen

Op de DARE architectuur-dag in februari 2005 is afgesproken hoe tijdschriftartikelen binnen DARE moeten worden beschreven, of precieser: in welke dc-elementen wordt welke informatie geplaatst. Het beschrijvingsmodel is gebaseerd op [de Guidelines for Encoding Bibliographic Information in DC metdatdata](#).

Schema voor DARE DC

dc: title		
dc:creator		
dc:date	default value = date issued (ipv date created)	
dc:relation	issn	
dc:identifier	tijdschrifttitel, volume (issue), eerste pagina, laatste pagina	<i>text bibliographic citation</i>
dc :identifier	open url versie 1.0 (handle, DOI)	<i>machine readable citation</i>
dc :source	-	

<dc:creator>Yu, L.</dc:creator>

<dc:creator>Apps, A.</dc:creator>

<dc:title>Studying E-Journal User Behavior Using Log Files</dc:title>

<dc:identifier>Library and Information Science Research 22 (3), 311-338</dc :identifier>

<dc:date>2000</dc:date>

<dc:relation>ISSN:0740-8188</dc:relation>

<dc :identifier>open url versie 1.0 </dc :identifier>

Schema voor DARE QDC

dc: title		
dc:creator		
dc:date	<dcterms:dateIssued>	
dc:relation	<dcterms:IsPartOf>issn	

dc:identifïer	<dcterms:bibliographicCitation>tijdschrifttitel, volume (issue), eerste pagina, laatste pagina	<i>text bibliographic citation</i>
dc :identifïer	<dcterms:dateIssued>open url versie 1.0 (handle, DOI)	<i>machine readable citation</i>
dc :source	-	

dc :identifïer

<dcterms:bibliographicCitation>Library and Information Science Research 22 (3), 311-338</dcterms:bibliographicCitation>

<dcterms:bibliographicCitation>open url versie 1.0 </dcterms:bibliographicCitation>

- ⇒ De richtlijn voor de beschrijving van tijdschriftenartikelen is nog niet alle IRs geïmplementeerd. Er is nog geen deadline voor de implementie afgesproken.
- ⇒ Aanbeveling: book chapters op vergelijkbare manier beschrijven als artikelen uit tijdschriften.

3. Meer informatie

Metadata

UKOLN's metadata website. Includes publications and presentations on metadata issues and a collection of interesting web links.

<http://www.ukoln.ac.uk/metadata/>

Dublin Core

Using Dublin Core (DC userguide)

<http://www.dublincore.org/documents/usageguide/>

Dublin Core – officiële website

<http://www.dublincore.org/>

[DCMI term declarations represented in XML schema language](http://dublincore.org/schemas/xmls/)

<http://dublincore.org/schemas/xmls/>

[Guidelines for implementing Dublin Core in XML](http://dublincore.org/documents/dc-xml-guidelines/)

<http://dublincore.org/documents/dc-xml-guidelines/>

OAI-PMH

OAI-PMH tutorial: OAI for Beginners – the Open Archive Forum online tutorial

<http://www.oaforum.org/tutorial/>

OAI-PMH, versie 2.0


















<http://www.openarchives.org/OAI/openarchivesprotocol.html>

DARE regelwerk

DARE use of Dublin Core, versie 2.0

<http://www.surf.nl/download/DARE%20use%20of%20DC%20v.%202.0.pdf>

4. Bijlage: Willie Wortels (werkgroep implementatie)

View	Instelling ▲	Naam ▼	E-mail ▼	Tel.nr. ▼	Updated ▼
	Erasmus Universiteit Rotterdam				May-25-2005
	Innervation bv/ SURF	Martin Feijen	m.feijen@innervation.nl	06 215 629 01	Mar-22-2004
	KNAW	Laurents Sesink	laurents.sesink@bureau.knaw.nl	020 462 86 85/682	Jul-15-2004
	Koninklijke Bibliotheek (KB)	Mirella van der Velde	mirella.vandervelde@kb.nl	070 314 02 16	Jun-22-2004
	KU Nijmegen	(Hans Schoonbrood)	h.schoonbrood@uci.kun.nl	024 361 79 90	Oct-25-2004
	RijksUniversiteit Groningen	Henk Druiven	H.C.G.Druiven@ub.rug.nl	050 363 80 57	Aug-20-2004
	SURFnet	Erik Saaman	erik.saaman@surfnet.nl		Mar-18-2005
	TU Delft	Arent J. Bosman	a.j.bosman@library.tudelft.nl	015 278 55 54	Dec-22-2004
	TU Eindhoven	Sjoerd Miesen	f.w.a.m.miesen@tue.nl	040 247 30 10	Mar-22-2004
	Universiteit Leiden	Marlon Domingus	domingus@library.leidenuniv.nl	071-527 2886	Nov-29-2004
	Universiteit Maastricht	Jos Odekerken	j.odekerken@ub.unimaas.nl	043 388 50 40	Nov-23-2004
	Universiteit Twente	Ben ten Tusscher	b.c.j.m.tentusscher@utwente.nl	053 4892059	Jan-20-2005
	Universiteit Utrecht	Martin Slabbertje	M.Slabbertje@library.uu.nl		Mar-22-2004
	Universiteit van Tilburg	Thomas Place (waarnemend lid)	T.W.Place@uvt.nl		Dec-23-2004
	Universiteit van Tilburg	Wilko Haast	wilko.haast@uvt.nl	013 466 03 58	Apr-25-2005
	UvA	Renze Brandsma	R.Brandsma@uva.nl	020 525 21 65	Oct-25-2004
	Wageningen Universiteit & Research	Frank Waajen	frank.waajen@wur.nl		Apr-15-2004