



Número 7, desembre 2001

Recursos web i metadades: informe del projecte¹

ASSUMPCIÓ ESTIVILL RIUS

Coordinadora del projecte
Facultat de Biblioteconomia i Documentació
Universitat de Barcelona
estivill@fbd.ub.es

JORDI ANDREU, JESÚS GASCÓN, JOSEP MANUEL RODRÍGUEZ GAIRÍN, CRISTÓBAL URBANO

Col·laboradors
Facultat de Biblioteconomia i Documentació
Universitat de Barcelona
andreu@fbd.ub.es; gascon@fbd.ub.es; rzgairin@fbd.ub.es; urbano@fbd.ub.es

Resum

El projecte Recursos web i metadades, que ha comptat amb una ajuda econòmica de la Divisió V per a projectes de recerca, es plantejà amb l'objectiu principal de comprovar les dificultats d'aplicació del sistema de metadades DC a la descripció de les pàgines d'un lloc web universitari i l'efectivitat d'aquestes metadades tant per a la descripció de recursos web com per a la seva ulterior cerca i recuperació. Per al desenvolupament del projecte s'han adaptat i desenvolupat una sèrie d'eines de les quals també es vol comprovar l'eficàcia. En aquest informe s'explica el desenvolupament del projecte, el seu estat el novembre de 2001 i també les eines desenvolupades.

Sumari

- 1 [El conjunt de metadades Dublin Core \(DC\) com a punt de partida del projecte.](#)
 - 2 [Objectius del projecte.](#)
 - 3 [Metodologia.](#)
 - 4 [Eines i productes desenvolupats per al projecte.](#)
 - 4.1 [Concrecions als camps del DC qualificat.](#)
 - 4.2 [Disseny de la base de dades i del sistema de recuperació.](#)
 - 4.3 [Formulari per a l'assignació de metadades.](#)
 - 4.4 [Classificació arxivística local.](#)
 - 4.5 [Guia de procediments per a l'assignació de metadades.](#)
 - 5 [Resultats provisionals i tasques pendents.](#)
 - 6 [Conclusions.](#)
 - 7 [Bibliografia.](#)
- Apèndix A. [Qualificadors del DC i qualificadors emprats en el projecte.](#)

1 El conjunt de metadades Dublin Core (DC) com a punt de partida del projecte

Des que la xarxa Internet es popularitzà a mitjan anys noranta amb la posada en funcionament del World Wide Web, el nombre de recursos que es publiquen en aquest espai ha crescut tant que l'increment es podria qualificar d'incontrolat. Els navegadors, com ara Explorer i Netscape, només permeten l'accés al recursos si se'n coneix l'adreça electrònica; els cercadors —tipus Google, Altavista, Infoseek, etc.—, si bé són d'una gran ajuda per als usuaris de la Xarxa, no precisen massa els seus resultats i sovint l'usuari recupera grans masses de recursos que no són pertinents a la seva cerca. Des de fa temps, diverses comunitats d'Internet busquen solucions a

aquests problemes, que se centren en el desenvolupament d'identificadors únics i persistents (*uniform resource names*) i de sistemes de descripció o metadades que facilitin la descoberta i recuperació dels recursos de la xarxa. El projecte Recursos web i metadades se centra en aquest darrer aspecte.

El terme metadades és relativament nou, però el concepte que hi ha al darrere ve de lluny, de les fórmules més antigues i primitives de descripció de documents per facilitar-ne l'accés. Enumerar les nombroses definicions proposades pels diferents autors i projectes ompliria moltes pàgines i obligaria a analitzar una gran quantitat de matisos que no aportarien massa elements al concepte principal. En un primer nivell, la majoria de les definicions genèriques declaren que "metadades són dades sobre dades". En un segon nivell, poden fer referència a la seva natura o les seves funcions o al tipus de recursos a què s'apliquen. A continuació es donen, a tall d'exemple, dues definicions prou autoritzades perquè provenen d'organismes professionals de prestigi, que marquen una de les diferències existents entre les diverses definicions:

"Metadata is data about data. The term refers to any data used to aid the identification, description and location of networked electronic resources. Many different metadata formats exist, some quite simple in their description, others quite complex and rich". (*Digital libraries. Metadata resources* [en línia]. IFLA, 1995–2001. <http://www.ifla.org/II/metadata.htm>).

"Metadata are structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities". (Association of Library Collections and Technical Services. Committee on Cataloging: Description and Access. Task Force on Metadata. *Final report*. June 2000. <http://www.ala.org/alcts/organization/ccs/ccda/tf-meta6.html>).

Totes dues definicions coincideixen en una part de les funcions que s'assignen a les metadades: la identificació, descripció i descoberta o recuperació de recursos. La diferència més òbvia és que la definició de l'IFLA limita l'ús del terme als recursos electrònics en xarxa. Però, de fet, aquesta és una limitació que no respon a la realitat dels sistemes de metadades desenvolupats recentment, ja que la majoria estan pensats tant per a la descripció de recursos digitals distribuïts per mitjà de xarxes com per a la de recursos tangibles.

De formats de metadades, com diu la definició de l'IFLA, n'hi ha molts. Un d'ells és el Dublin Core (DC), que —juntament amb la capçalera TEI— és dels pocs que ha estat pensat de bell antuvi per identificar, descriure i recuperar recursos digitals; el DC s'adreça, encara més concretament, a la descripció i recuperació de recursos web.

Inicialment el DC va ser desenvolupat amb la intenció que els autors i editors de recursos web poguessin assignar directament metadades a les pàgines web que creaven i les poguessin inserir a la seva capçalera de manera que als cercadors els fos més fàcil actuar sobre aquestes metadades. Així s'oferirien uns resultats més pertinents a les cerques dels usuaris del web. Tanmateix, després de sis anys de la creació del conjunt de metadades DC, els cercadors encara no saben, en la seva majoria, interpretar aquestes metadades² i, si bé hi ha molts projectes —tant en l'àmbit acadèmic, com de les biblioteques—, l'ús del DC no s'ha estès entre els creadors i editors de pàgines web —que inicialment havien de ser-ne els usuaris potencials.³

El conjunt de metadades DC té dues versions bàsiques. La versió simple consisteix en un conjunt de quinze elements força senzills que no tenen al darrere cap tipus de control. En la versió qualificada la majoria dels quinze elements poden ser qualificats amb afinaments que en precisen el contingut —per exemple, es pot assenyalar que l'element Creador és un nom personal o d'entitat—, o amb esquemes que indiquen les notacions o vocabularis controlats emprats per assignar el contingut de l'element —per exemple, es pot indicar que una matèria s'ha obtingut d'un tesaurus o d'una llista d'encapçalaments de matèries. Aquesta segona versió més complexa del conjunt d'elements DC està pensada per a aplicacions més bibliotecàries que facilitin la creació de portals temàtics de recursos gestionats per bases de dades estructurades, i també per afavorir la integració de registres DC en catàlegs i bases de dades bibliogràfiques.

Del conjunt de metadades DC destaquen una sèrie de característiques que avalen la seva tria com a punt de partida del projecte Recursos web i metadades: la seva simplicitat de creació i manteniment; una semàntica senzilla i fàcil d'entendre; el seu abast internacional i la seva extensibilitat, la qual permet que a partir dels quinze elements inicials es puguin fer extensions per a les necessitats de comunitats específiques d'usuaris o bé adreçades a recursos de disciplines concretes. Les següents són altres raons que avalen el seu ús en el projecte:

- Està pensat concretament per a la descoberta i l'accés de recursos web.
- Són molts els projectes en marxa que empen el sistema -tant d'àmbit local com nacional i internacional. No obstant això, no s'usa tant per indexar exhaustivament les pàgines d'un lloc web com per indexar els recursos de manera més global.
- Se n'està estudiant l'adopció com un estàndard ANSI/NISO.

- És un estàndard de fet quant a l'acceptació.
- És potencialment compatible amb els registres de catàleg.
- En la seva versió qualificada admet l'ús d'eines bibliotecàries.

2 Objectius del projecte

En presentar el projecte amb vista a la seva aprovació es formulaven dos objectius bàsics per a la recerca que es proposava:

- Comprovar, a partir d'un simulacre, l'efectivitat del sistema de metadades DC.
- Potenciar l'ús del DC entre la comunitat catalana d'Internet per tal de millorar l'efectivitat dels motors de cerca en la identificació i recuperació de la informació.

El simulacre havia de partir de l'assignació de metadades a una mostra de recursos i pàgines web d'un web universitari.

A mesura que el projecte ha anat avançant, s'han anat redefinint els objectius anteriors. D'una banda, en el projecte s'han aplicat una sèrie d'eines ja existents —o adaptades d'eines ja existents—, l'efectivitat de les quals s'ha hagut de provar en l'entorn del projecte: l'assignació de metadades a les pàgines d'un web universitari. D'altra banda, s'han desenvolupat eines i productes específics per al projecte de les quals també s'ha hagut de comprovar l'eficàcia.

Així doncs, els objectius del projecte Recursos web i metadades han quedat definits de la manera següent:

- Comprovar les dificultats d'aplicació del sistema de metadades DC a la descripció de les pàgines d'un lloc web universitari.
- Comprovar l'efectivitat del sistema tant per a la descripció de recursos web com per a la seva ulterior cerca i recuperació.
- Desenvolupar una sèrie d'eines que facilitin l'aplicació del DC en l'entorn del projecte.
- Comprovar l'eficàcia d'aquestes eines.

3 Metodologia

El projecte es desenvolupa en les tres fases que s'exposen a continuació. En cada fase s'ha aplicat la metodologia que s'explica amb la finalitat de desenvolupar les eines i els productes resultants i comprovar-ne l'eficàcia:

Primera fase del projecte

La primera fase del projecte s'inicià amb la revisió bibliogràfica del tema de les metadades incloent-hi diversos sistemes i esquemes que servissin per emmarcar la tria del conjunt DC. Es revisaren els estàndards següents:

- Dublin Core Metadata Initiative, *Dublin Core metadata element set*, version 1.1. <http://dublincore.org/documents/1999/07/02/dces/>.
- *Encoding Archival Description* (EAD). <http://lcweb.loc.gov/ead/>.
- Text Encoding Initiative. *TEI guidelines*, <http://www.tei-c.org/Guidelines2/index.html>.
- *ONIX product information: overview and summary list of data elements*, American Association of Publishers, <http://www.editeur.org/onixfiles2.0/ONIXProductInformationOverview2.0.pdf>.
- Visual Resources Association, *Core categories for visual resources*, <http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm>.

Així mateix, també es revisaren i analitzaren alguns dels projectes de portals temàtics, guies de recursos web i altres projectes que entren el conjunt de metadades DC.⁴ A continuació podem veure una mostra dels projectes analitzats:

Portals temàtics

- *ADAM: the gateway to art, design, architecture & media information on the Internet*, <http://www.adam.ac.uk>.
- *AHDS: Arts & Humanities Data Service*, <http://ahds.ac.uk>.
- *GEM project site: Gateway to Educational Resources*, <http://www.geminfo.org/>.
- *Internet Scout Report*, <http://scout.cs.wisc.edu/index.html>.
- *Metadata server at the State and University Library at Göttingen, Germany (SUB)*,

<http://www2.sub.uni-goettingen.de/>.

- VADS: *Visual Arts Data Service*, <http://vads.ahds.ac.uk/>.

Els següents són altres projectes revisats que també fan servir el DC:

- CORC, <http://www.oclc.org/corc/>. Projecte d'OCLC que consisteix en l'elaboració d'una base de dades de recursos web on conviuen registres MARC i registres DC.
- BIBLINK, <http://hosted.ukoln.ac.uk/biblink/>. Projecte subvencionat per la Unió Europea que és resultat de la cooperació de diverses biblioteques nacionals amb el món de l'edició electrònica. L'anàlisi de diversos sistemes de metadades era una part important del projecte.
- DESIRE, <http://www.desire.org/>. Projecte subvencionat per la Unió Europea per al desenvolupament de portals temàtics de qualitat que investigà en profunditat els aspectes relatius a la selecció, descripció i indexació dels recursos web.
- Nordic Metadata Project, <http://www.lib.helsinki.fi/meta/>. Projecte cooperatiu dels països escandinaus que fou el primer a utilitzar les metadades DC i a desenvolupar eines per a la seva assignació.

Una vegada revisada la bibliografia sobre el tema, la primera fase del projecte Recursos web i metadades continuà amb la presa de decisions i l'establiment de concrecions per a l'aplicació del conjunt de metadades DC en l'entorn del projecte i el desenvolupament de les eines que s'apunten a continuació i que s'expliquen més avall:

- Disseny d'una base de dades adaptada als elements del DC qualificat.
- Formulari per a l'entrada de metadades.
- Classificació arxivística local aplicable a un web universitari.
- Guia de procediments per a l'assignació de metadades (amb l'aplicació del DC).

Pel que fa al conjunt de metadades DC, es decidí:

- Emprar la versió qualificada d'aquest conjunt de dades.
- Emprar la sintaxi HTML, versió 4.0, perquè és la que en el moment d'iniciar el projecte reconeixen els navegadors —posteriorment, si es creu convenient, es podrà fer la conversió a la sintaxi XML.
- No inserir les metadades a la capçalera de les pàgines web, sinó a la base de dades dissenyada per al projecte —el manteniment de les metadades és més fàcil de dur a terme amb aquest sistema i, de moment, els navegadors i cercadors encara no fan un ús exhaustiu de les metaetiquetes del DC.
- Emprar, sempre que fos possible, eines de l'entorn bibliotecari català —vocabularis controlats, llistes d'autoritats, etc.— per definir el contingut dels elements del DC.

La primera fase del projecte va concloure amb la comprovació de les eines i els productes desenvolupats a partir de l'assignació de metadades a una mostra molt significativa de les pàgines web de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona.⁵ Aquesta assignació la feren dues alumnes del curs de doctorat —Silvia Argudo i Marta Somoza. També comprovaren, amb la seva pràctica, l'adequació de les eines bibliotecàries al marc del projecte i l'efectivitat de les eines i productes desenvolupats específicament per al projecte: el formulari d'entrada de dades, la base de dades, la guia de procediments per a l'assignació de metadades i la classificació arxivística local.

Segona fase del projecte

La segona fase del projecte s'inicià amb la revisió i modificació dels productes desenvolupats en la primera fase. A partir de l'experiència d'assignació de metadades es reformularen les eines i productes següents: el formulari per a l'assignació de metadades, la base de dades, la guia de procediments per a l'assignació de metadades i la classificació arxivística local. Les modificacions es van fer a partir dels comentaris de les dues alumnes esmentades i també de les observacions de l'equip de treball del projecte. Després de les modificacions oportunes, els productes anteriors han estat considerats definitius —les esmenes que s'hi poden introduir a partir d'aquest moment seran menors. La segona fase continuà amb l'assignació definitiva de metadades a les pàgines web de la Facultat de Biblioteconomia i Documentació. Està previst acabar aquesta segona fase a final de desembre de 2001.

Tercera fase del projecte

La tercera fase del projecte consistirà en la definició dels paràmetres del cercador Verity a fi que actuï sobre cada un dels camps de la base de dades de metadades i també sobre el text complet de les pàgines web de la Facultat de Biblioteconomia i Documentació. Per a la parametrització del cercador es compta amb la col·laboració d'Antonio Murillo i Miquel Llevat, experts en aquest

programa de recuperació. Verity, juntament amb una sèrie de programes complementaris, ha estat cedit gratuïtament a la Facultat per l'empresa DaVinci Consulting Tecnológico (DCT) per a fins docents i d'investigació.

Aquesta tercera fase, que s'ha de dur a terme els mesos de desembre de 2001 i gener de 2002, clourà amb l'avaluació del sistema de recuperació.

4 Eines i productes desenvolupats per al projecte

En aquest apartat s'inclou una descripció sumària de les eines i productes desenvolupats per al projecte i les seves característiques més rellevants.

4.1 Concrecions als camps del DC qualificat

Una de les primeres tasques del projecte consistí a analitzar el conjunt de metadades DC per examinar quines concrecions eren necessàries en el context del projecte Recursos web i metadades. Entre les decisions preses destaquen les següents:

- Emprar els quinze elements del conjunt de metadades DC —però es recomana no omplir el camp font, ja que la seva definició no és clara i, d'altra banda, quan hi ha una relació amb un document font, aquesta es pot expressar a partir de l'element relació;
- Utilitzar el DC qualificat i acceptar aquells afinaments i esquemes de codificació recomanats en el document *Dublin Core qualifiers*⁶ sempre que s'adiguin amb les eines emprades en el context bibliotecari català.
- L'aplicació del DC qualificat ha de prendre com a base els vocabularis de control i les notacions emprades en el context bibliotecari català.

L'apèndix A inclou una taula on es detallen els qualificadors (afinaments i esquemes de codificació) del DC i els emprats en el projecte Recursos web i metadades. De manera resumida es pot dir que s'accepten totalment els qualificadors dels elements Descripció, Tipus de recurs, Relació i Drets; en els elements Creador, Col·laborador, Editor, Matèria, Data, Format, Identificador, Font, Llengua i cobertura, els qualificadors se seleccionen dels recomanats en el document *Dublin Core qualifiers* o s'adopten qualificadors nous a partir dels vocabularis controlats i notacions emprats en el sistema bibliotecari català o desenvolupats per al projecte. En el cas de l'element Títol no s'empra el qualificador *Alternative*, ja que l'ambigüïtat de la seva definició així ho ha aconsellat.

4.2 Disseny de la base de dades i del sistema de recuperació

La base de dades per al projecte es dissenyà per acollir els camps del DC qualificat i facilitar que els quinze camps o elements puguin ser indexats i, per tant, recuperats. Les característiques de la base de dades són les següents:

- És una base de dades relacional que consta d'una taula principal o mare, una taula per a cada un dels elements del DC, una taula per als qualificadors esquemes de codificació, una taula per als qualificadors d'afinament i quinze taules per relacionar la taula principal amb els registres dels quinze elements del DC.
- La base de dades té característiques documentals; és a dir, que els camps es poden repetir i que permet el control d'autoritats de cada un dels camps del registre de manera que es poden fer canvis globals per a un valor concret d'un camp.
- La base de dades s'ha desenvolupat amb el software Microsoft Access via ODBC (*open database connectivity*) i amb comandes SQL, components que permeten que es pugui interactuar amb la base de dades via WWW.
- En el futur, la base de dades podrà migrar sense problemes al *software* Oracle.

L'esquema de la figura 1 resumeix l'estructura de la base de dades del projecte.

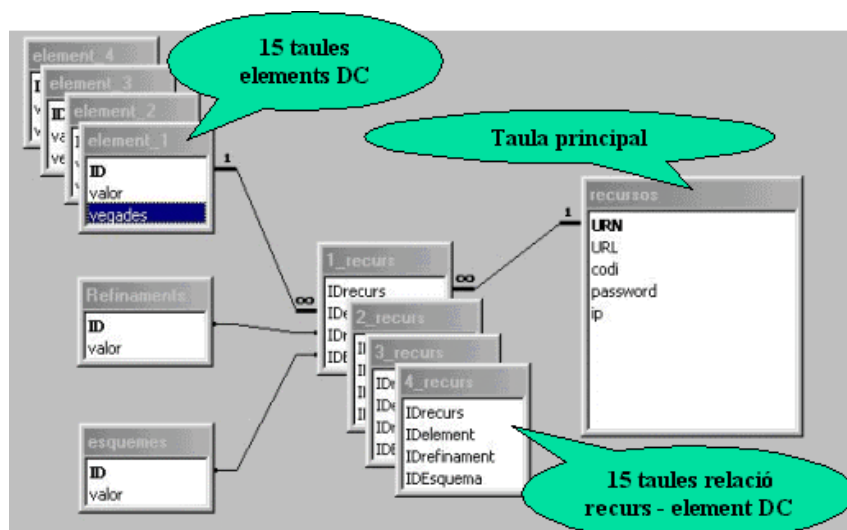
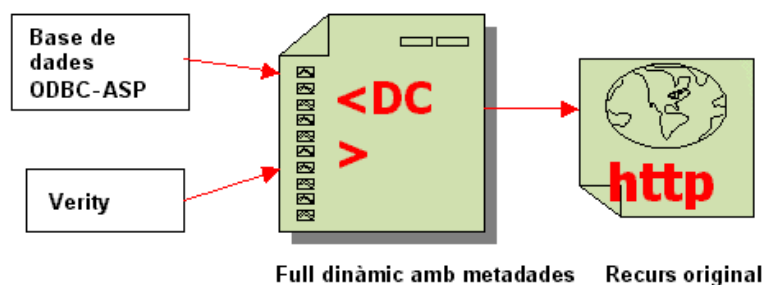


Figura 1. Esquema de l'estructura de la base de dades del projecte.

Pel que fa a la recuperació de la informació, una vegada el cercador Verity estigui parametritzat, la seva funció serà indexar i recuperar la informació a partir de les metadades DC i també directament del text complet del recurs original. L'esquema de la interacció entre l'assignació i el manteniment de metadades a partir de la base de dades, i la indexació i recuperació dels recursos mitjançant el cercador Verity, és el que es mostra a la figura 2.

Base de dades: catalogació i manteniment de metadades (sense modificar el recurs). Facilita el control d'autoritats, les modificacions globals, la gestió via WWW, etc.



Verity: indexació i recuperació. En el primer nivell Verity fa una recuperació estructurada a partir dels fulls de metadades. En el segon nivell la recuperació es fa en text lliure a partir de les mateixes pàgines web.

Figura 2. Estructura de la base de dades del projecte i del seu sistema de recuperació.

4.3 Formulari per a l'assignació de metadades

El formulari per a l'assignació de metadades recull els quinze camps del DC qualificat. El seu disseny parteix del formulari elaborat per al Nordic Metadata Project⁷ i introdueix algunes característiques noves. Se'n poden assenyalar els trets següents:

- Cada camp del formulari va precedit d'una explicació breu sobre el contingut del camp.
- Des de l'epígraf de cada element del formulari es pot accedir a la informació sobre l'element de la guia de procediments per a l'assignació de metadades.
- Tots els camps són repetibles (botó + en els camps del formulari).
- Les metadades cal introduir-les en el formulari sense codificar; en validar-les, el sistema les carrega a la base de dades. Des d'una pàgina intermèdia també es poden obtenir les metadades codificades en format HTML a punt per a ser inserides a la capçalera del recurs.
- En els casos en què el contingut d'un camp prové d'un vocabulari controlat, el terme es pot obtenir directament d'un menú desplegable i copiar en el formulari. El tipus de recurs, el seu format, la llengua i el codi de la classificació arxivística que s'aplica al recurs es poden inserir automàticament en el formulari.
- En tots els camps es poden consultar el continguts ja entrats i, si escau, copiar-los automàticament en el formulari (botó ? en els camps del formulari).

- El formulari també permet consultar metadades ja entrades i modificar-les.

La figura 3 és una imatge d'un fragment del formulari per a l'assignació de dades.

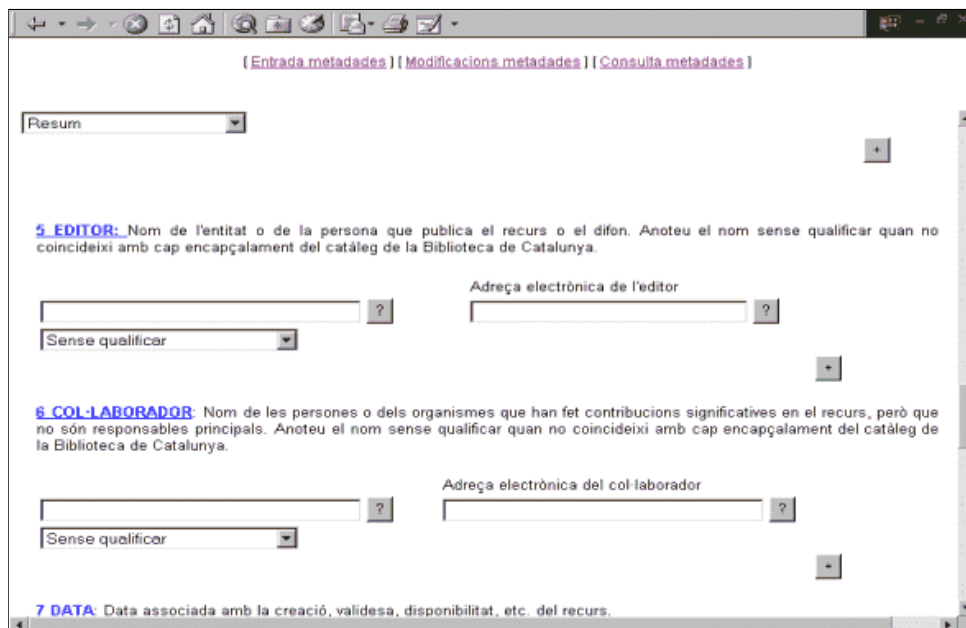


Figura 3. Fragment del formulari per a l'assignació de metadades.

4.4 Classificació arxivística local

En analitzar el web de la Facultat de Biblioteconomia i Documentació es féu palès que una part important de les pàgines podia tenir la consideració de documents de gestió i que, de fet, totes podien tenir, aparentment, un tractament arxivístic. Per aquesta raó es decidí desenvolupar, com a part del projecte, un quadre de classificació funcional adaptat a una universitat. El quadre s'organitza en les set funcions generals i les quatre d'específiques que es mostren a continuació:

Funcions generals	Funcions específiques
Organització administrativa Gestió dels recursos d'informació i de les comunicacions Relacions externes, representació i promoció Gestió dels recursos humans Gestió dels recursos econòmics Gestió dels béns mobles i dels immobles Normativa i assumptes jurídics	Gestió dels recursos acadèmics Organització de la docència Organització de la recerca Gestió dels serveis oferts a la comunitat universitària

Cada una de les classes anteriors s'organitza en diverses subclasses i així successivament fins a un tercer o quart nivell depenent del cas. Dins del conjunt de metadades del DC, els codis d'aquesta classificació s'assignen en el camp de les matèries.

El quadre de classificació s'ha desenvolupat en format de base de dades sobre Access. Cada registre inclou el codi de la classe o subclasse i les classes superiors porten una definició que facilita l'assignació de la classe o subclasse als documents en qüestió. La figura 4 mostra un exemple de registre a la base de dades de la classificació arxivística.

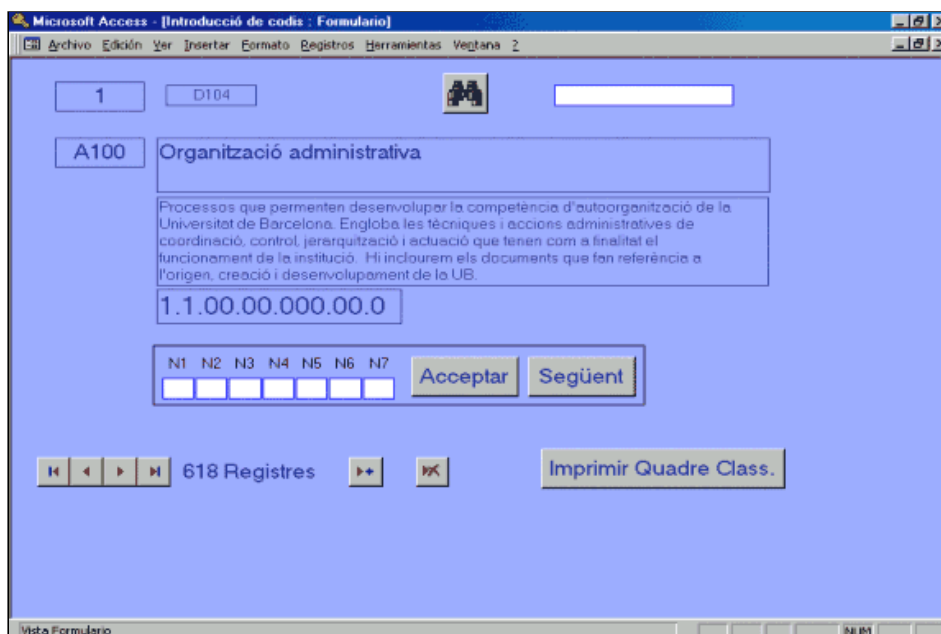


Figura 4. Registre de la base de dades amb la classificació arxivística local.

El fet que la classificació s'hagi desenvolupat com una base de dades facilita la introducció de les dades en el formulari: l'usuari només ha de fer un clic sobre la classificació per obrir-la i assignar la classe corresponent al document, sempre segons els descriptors textuais —no cal, doncs, que conegui cap codi. La figura 5 mostra com es desplega el quadre de classificació en fer un clic sobre l'element matèria del formulari quan, prèviament, s'ha escollit l'esquema de classificació arxivística.

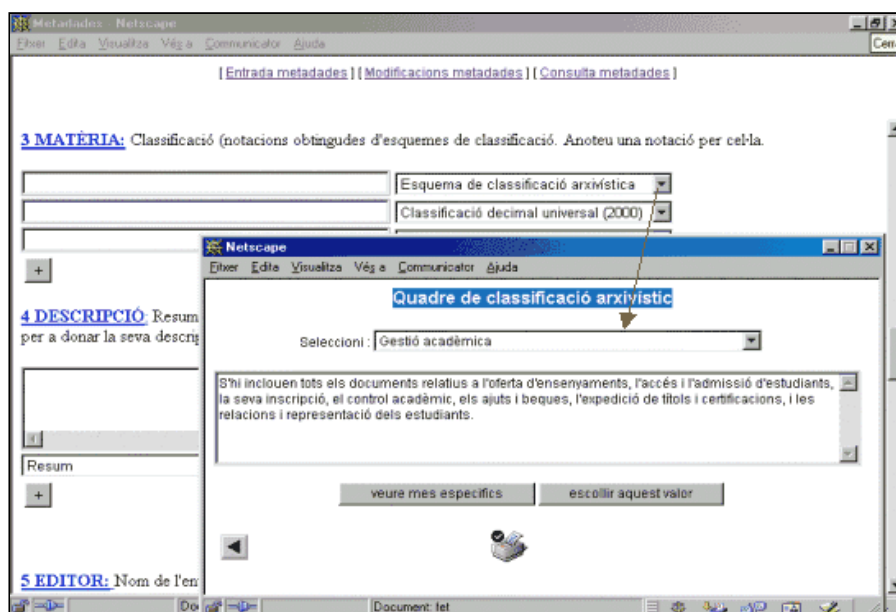


Figura 5. Consulta de l'esquema de classificació des del formulari.

El quadre de classificació desenvolupat és volgutament complet tal com correspondria a la funció arxivística que hauria de tenir per a la documentació impresa. Òbviament, a l'entorn web moltes de les classes i subclasses no seran, de moment, aplicables, ja que la documentació que s'hi fa pública és més restringida que la que es genera en suport paper.

4.5 Guia d'ús per a l'assignació de metadades

Per facilitar l'assignació de metadades s'ha desenvolupat una guia de procediments, que s'organitza al voltant dels quinze elements del DC. S'articula en un conjunt de directrius que tenen l'objectiu de facilitar una aplicació uniforme del DC. Per a la seva elaboració s'ha partit de

textos similars, com la guia recomanada per la DCMI, *Using Dublin Core* de Diane Hillman, o la desenvolupada per al Nordic Metadata Project.⁸ Tanmateix, també s'han intentat aplicar criteris propers a la normativa bibliotecària i, de fet, la versió actual de la guia està pensada perquè l'emprin persones coneixedores de la normativa catalogràfica. Per a altres aplicacions que es puguin desenvolupar en el futur, aquestes directrius s'adaptaran al context concret.

Des dels epígrafs de cada un dels camps del formulari per a l'assignació de metadades es pot consultar la secció de la guia sobre aquest element concret. A la figura 6 es veu com en fer un clic sobre l'element Editor, s'obre una finestra amb les directrius per assignar el contingut d'aquest element.

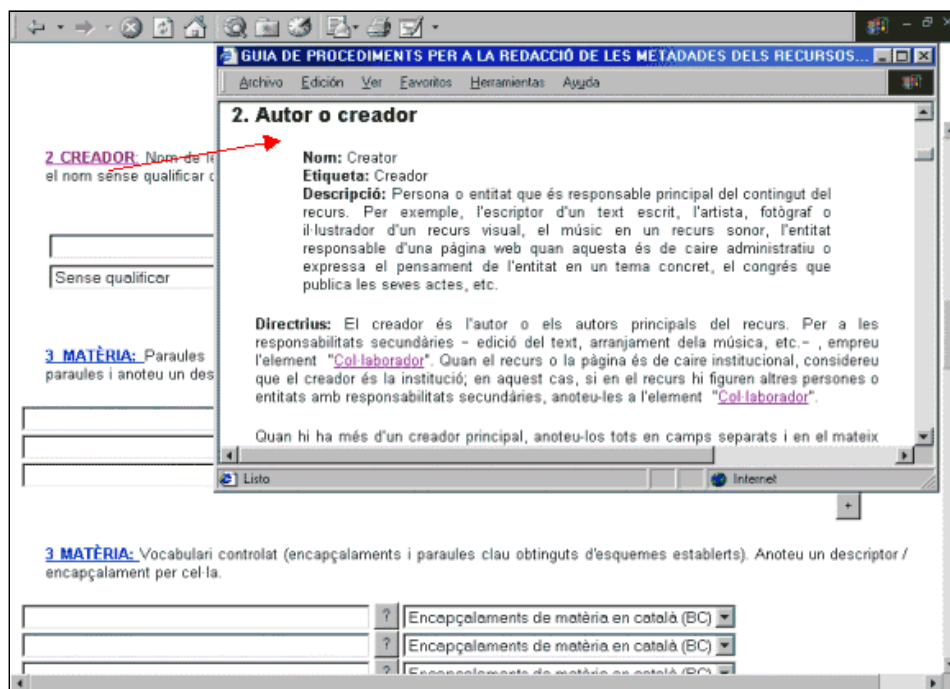


Figura 6. Consulta de la guia de procediments des del formulari.

5 Resultats provisionals i tasques pendents

En escriure aquest informe, totes les eines desenvolupades han estat comprovades amb la primera assignació provisional de metadades a les pàgines web de la Facultat i se'n disposa ja de la versió definitiva. Actualment s'està procedint a la reindexació de les pàgines web de la Facultat amb la col·laboració d'Eduard Vilagrosa, becari del projecte i alumne de la Llicenciatura.

Les aportacions del projecte que es consideren més destacables són les següents:

- Aplicació d'una sèrie d'eines —algunes de l'entorn bibliotecari català i altres de desenvolupades per al projecte.
- Indexació exhaustiva de les pàgines d'un web —la bibliografia informa sobretot de projectes que consisteixen en la indexació de recursos considerats globalment, però no ho fa pàgina per pàgina.
- Aplicació d'un esquema de classificació arxivístic a les pàgines indexades ja que s'ha considerat que el seu caràcter administratiu admet aquest tractament arxivístic.

Pel que fa a l'acabament del projecte es preveuen els terminis següents:

- Desembre 2001: assignació definitiva de metadades a les pàgines web de la Facultat.
- Desembre de 2000 – gener de 2001: establiment dels paràmetres del cercador Verity per tal que actuï sobre cada un dels camps de la base de dades de metadades i sobre el text complet de les pàgines web de la Facultat.
- Gener – febrer de 2001: avaluació final del projecte.

6 Conclusions

Les conclusions provisionals que s'esmenten a continuació s'articulen al voltant dels objectius

dels projecte:

Primer objectiu: comprovar les dificultats d'aplicació del sistema de metadades DC a la descripció de les pàgines d'un lloc web universitari.

El DC, és un conjunt de metadades simple?

- El DC és un sistema de metadades certament senzill —sobretot en la versió simple o no qualificada—, però és molt obert i, per tant, ambigu; d'aquí que la seva aplicació comporti certes dificultats sobretot perquè és difícil que es faci amb coherència.
- En la seva versió qualificada l'aplicació del DC és prou complexa perquè requereixi uns coneixements especialitzats; per això pot ser difícil que els autors i editors de pàgines web puguin emprar aquesta versió del DC si no tenen aquests coneixements.
- Malgrat el temps transcorregut des de la seva primera versió, el DC encara és un estàndard en desenvolupament, que no està fixat del tot. Els grups de treball de la DCMI estan fent propostes de modificacions que poden afectar diversos elements del DC.⁹ Aquesta manca d'estabilitat dificulta l'aplicació d'aquest conjunt de metadades.
- Using Dublin Core, de Diane Hillmann, no es publica com a recomanació fins a l'abril de 2001. Això vol dir que fins fa ben poc la comunitat d'usuaris no ha tingut unes directrius més o menys definitives que guiïn l'aplicació del conjunt de metadades DC.

Segon objectiu: comprovar l'efectivitat del sistema tant per a la descripció de recursos web com per a la seva ulterior cerca i recuperació.

El DC com a instrument de recuperació del web?

- El conjunt de metadades DC neix per indexar i recuperar recursos web al web, però al cap de sis anys de la seva creació els cercadors encara no reconeixen les metadades DC.
- Per això el seu ús més estès no és el que s'havia pensat inicialment.
- Sobretot s'empra en projectes que consisteixen en la creació de bases de dades o catàlegs de recursos web.
- Si els cercadors arriben a reconèixer les metadades DC, probablement serà el DC simple o no qualificat el que millor s'adapti als sistemes de recuperació del web per diverses raons: d'una banda, és el que poden aplicar més fàcilment els autors i editors de pàgines web; d'altra banda, el DC qualificat parteix de l'aplicació d'una sèrie d'eines bibliotecàries pensades per a entorns molt controlats que difícilment serien efectives en l'entorn de la Xarxa.

Tercer i quart objectius: desenvolupar una sèrie d'eines que facilitin l'aplicació del DC en l'entorn del projecte i comprovar-ne l'eficàcia.

Les eines bibliotecàries, serveixen per a la indexació i recuperació de recursos web?

- Són eines pensades per a entorns molt controlats -això és, entorns que inclouen xarxes de referències.
- Funcionen quan s'elaboren bases de dades de recursos web (on hi pot haver un cert control).
- Probablement el seu ús seria problemàtic en l'entorn mateix del web precisament per la dificultat d'introduir-hi el control que requereix l'ús d'aquestes eines. El DC no qualificat possiblement seria el més idoni per a la descoberta i recuperació de recursos a la Xarxa.

Quines conclusions es poden extreure de les eines concretes emprades en el projecte?

En aquest apartat es consideren els elements amb continguts que es poden derivar de les eines bibliotecàries o de les desenvolupades per al projecte.

- Agents (Creador, Col·laborador, Editor)

En absència d'una llista d'autoritats per elaborar les formes dels noms dels agents, es valora positivament l'ús dels punts d'accés bibliogràfics acceptats en el catàleg de la Biblioteca de Catalunya, ja que s'adapten bé al contingut d'aquests elements i asseguren un ús uniforme i consistent de les formes dels noms. En tot cas, l'únic problema seria discernir quina persona o entitat es pot considerar creador o editor. És per això que el *Library Application Profile* de la DCMI proposa fusionar aquests tres elements.

- Matèries (LEMC, CDU)

Aquestes eines s'han considerat adequades per indexar i classificar recursos i serveis web globalment. També són útils per indexar i classificar els recursos web monogràfics (informes, treballs, etc.), les publicacions seriades tradicionals (com ara les revistes electròniques) i els articles que contenen, els congressos i les seves ponències i comunicacions, etc.

A l'hora d'aplicar aquestes eines a totes i cadascuna de les pàgines d'un lloc web sorgeixen dificultats. La llista de matèries no té el nivell d'especificitat necessari; l'aplicació de la CDU, que sí que té aquest nivell, resultaria en notacions molt complexes que no ajudarien massa a la recuperació de la pàgina. Si es traslladés aquesta situació a un entorn més tradicional es podria fer la pregunta següent: tindria sentit indexar tots els capítols, seccions, subseccions —fins i tot pàgines— d'un document imprès?

Aquestes eines servien per indexar i classificar llocs web considerats en la seva globalitat i alguns dels documents que s'hi publiquen quan tenen la seva pròpia identitat, però seria certament forçat aplicar-les a totes les pàgines.

- Classificació arxivística local

El caràcter marcadament administratiu de les pàgines del web de la Facultat i el seu possible tractament arxivístic van aconsellar el desenvolupament d'un quadre de classificació que es pogués aplicar a la indexació d'aquestes pàgines.

L'aplicació d'aquesta classificació s'ha demostrat que és relativament senzilla, la qual cosa era bastant òbvia de bell antuvi, ja que es tracta d'una classificació *ad hoc* i, per tant, era fàcil encaixar-hi les pàgines indexades.

És una classificació adequada per codificar els documents d'una mateixa sèrie i, per tant, capaç d'agrupar-los intel·lectualment, malgrat la seva dispersió en el web.

És una classificació estable i duradora en el temps, ja que es basa en les funcions del productor i no en la seva estructura organitzativa (que pot ser molt poc estable).

Manca comprovar l'eficàcia d'aquesta classificació per a la recuperació de les pàgines i el seu emmagatzematge una vegada deixen de ser públiques.

7 Bibliografia

a) Articles, bibliografies, directrius, etc.

Abadal, Ernest; Estivill, Assumpció. *Accés als recursos web des de les biblioteques* [en línia]: *informe encarregat pel Consorci de Biblioteques Universitàries de Catalunya*. Barcelona, gener de 2000. <http://www.cbuc.es/5digital/ireweb.pdf>. [Consulta: 07/11/2001].

Ahronheim, Judith R. "Descriptive metadata: emerging standards". *Journal of academic librarianship*, vol. 24, no. 5 (Sept. 1998), p. 395-404.

ALCTS Task Force on Metadata. *Final report* [en línia]. Association for Library Collections and Technical Services, April 3, 1997. <http://www.lib.virginia.edu/alcts/>. [Consulta: 07/11/2001].

Caplan, Priscilla. "International metadata initiatives [en línia]: lessons in bibliographic control". I. Library of Congress, Cataloging Directorate, July 5, 2000. <http://lcweb.loc.gov/catdir/bibcontrol/caplan.html>. [Consulta: 15/10/2001].

Caplan, Priscilla; Guenther, Rebecca. "Metadata for Internet resources: the Dublin core metadata element set and its mapping to USMARC". *Cataloging & classification quarterly*, vol. 22, no. 3-4 (1996), p. 43-58.

Cataloging Internet resources [en línia]: *a manual and practical guide*. Nancy B. Olson, ed. 2a ed. [Dublin, Ohio]: OCLC, c1997. <http://www.oclc.org/oclc/man/9256cat/toc.htm>. [Consulta: 9/07/1998].

Chepesiuk, Ron. "Organizing the Internet: the 'core' of the challenge". *American libraries*, vol. 30, no. 1 (Jan. 1999), p. 60-63.

"CORC: new tools and possibilities for cooperative electronic resource description". *Journal of Internet cataloging*, vol. 4, no. 1/2 (2001). Tot el número està dedicat a analitzar els diversos

aspectes del projecte CORC.

Digital libraries [en línia]: *cataloguing and indexing of electronic resources*. IFLA, c1995-2001, latest rev.: 1999-06-21. <http://www.ifla.org/ll/catalog.htm>. [Consulta: 07/11/2001].

Digital libraries [en línia]: *metadata resources*. IFLA, c1995-2000, latest rev.: February 23, 2001. <http://www.ifla.org/ll/metadata.htm>. [Consulta: 07/11/2001].

Dillon, Martin. "Metadata for Web resources [en línia]: how metadata works on the Web". *Bicentennial Conference on Bibliographic Control for the New Millenium: confronting the challenges of networked resources and the Web*. Library of Congress, Cataloging Directorate, June 27, 2000. <http://lcweb.loc.gov/catdir/bibcontrol/dillon.html>. [Consulta: 15/10/2001].

Dublin Core Metadata Initiative. *Dublin Core Metadata Element set, version 1.1* [en línia]: *reference description*. DCMI, 1999-07-02. <http://dublincore.org/documents/1999/07/02/dces/>. [Consulta: 07/11/2001].

Dublin Core Metadata Initiative. *Dublin Core Metadata Initiative workshops* [en línia]. DCMI, c1995-2001. <http://dublincore.org/workshops>. [Consulta: 07/11/2001].

Dublin Core Metadata Initiative. *Dublin Core qualifiers* [en línia]. DCMI, 2000-07-11, <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>. [Consulta: 07/11/2001].

Dublin Core Metadata Initiative. *Government Working Group. DC-Gov application profile* [en línia]. DCMI, November 2001, <http://dublincore.org/groups/government/profile-200111.shtml>. [Consulta: 30/11/2001].

Dublin Core Metadata Initiative. *Libraries Working Group. Library application profile* [en línia]. DCMI, 2001-10-12. <http://www.dublincore.org/documents/2001/10/12/library-application-profile/>. [Consulta: 05/11/2001]

El-Sherbini, Magda. "Metadata and the future of cataloging". *Library review*, vol. 50, no. 1 (2001), p. 16-27. Versió en línia a l'adreça següent des dels ordinadors de les universitats catalanes: <http://www.emeraldinsight.com/0024-2535.htm>.

Estivill, Assumpció. "Organització dels recursos Internet". *Item*, núm. 18 (gen.-junt 1996), p. 42-74.

Gorman, Michael. "Metadata or cataloguing?: a false choice". *Journal of Internet cataloging*, vol. 2, no. 1 (1999), p. 5-22.

Gradmann, Stefan. "Cataloguing vs. metadata and library cataloguing: old wine in new bottles?". *International cataloguing and bibliographic control*, vol. 28, no. 4 (Oct.-Dec. 1999), p. 88-90. Versió en línia: <http://www.ifla.org/IV/ifla64/007-126e.htm>. Versió espanyola: <http://www.ifla.org/IV/ifla64/007-126s.htm>.

Heery, Rachel; Patel, Manjula. "Application profiles: mixing and matching metadata schemas [en línia]". *Ariadne*, issue 25, (September 2000). <http://www.ariadne.ac.uk/issue25/app-profiles/>. [Consulta: 30/11/2001].

Henshaw, Robin; Valauskas, Edward J. "Metadata as a catalyst: experiments with metadata and search engines in the Internet journal, First monday". *Libri*, vol. 51, no. 2 (June 2001) p. 86-101.

Hillmann, Diane I. Using Dublin Core [en línia]. DCMI, 2001-04-12. <http://dublincore.org/documents/2001/04/12/usageguide/>. [Consulta: 15/10/2001]. Hopkins, Judith. "USMARC as a metadata shell". *Journal of Internet cataloging*, vol. 2, no. 1 (1999), p. 55-68.

Introduction to metadata: pathways to digital information. Edited by Murtha Baca. [S.l.]: Getty Information Institute, 1998.

Lynch, Clifford. "The Dublin Core descriptive metadata program: strategic implications for libraries and networked information access [en línia]". *ARL: a bimonthly newsletter of research library issues and actions*, no. 196 (Feb. 1998). <http://www.arl.org/newsitr/196/dublin.html>. [Consulta: 07/11/2001].

McCallum, Sally. "Extending MARC for bibliographic control in the Web environment [en línia]: challenges and alternatives". *Bicentennial Conference on Bibliographic Control for the New Millenium: confronting the challenges of networked resources and the Web*. Library of Congress, Cataloging Directorate, June 27, 2000. <http://lcweb.loc.gov/catdir/bibcontrol/mccallum.html>. [Consulta: 15/10/2001].

Milstead, Jessica; Feldman, Susan. "Metadata: cataloging by any other name". *Online*, 23, no. 1 (Jan.- Feb. 1999), p. 25–31.

Milstead, Jessica; Feldman, Susan. "Metadata: projects and standards". *Online*, 23, no. 1 (Jan.- Feb. 1999), p. 32–40.

Nordic Metadata Project. *User guidelines for Dublin Core creation*. Nordic Metadata Project, created: 2001-04-16, last update: 1999-11-22, http://www.sics.se/~preben/DC/DC_guide.html. [Consulta: 15/10/2001].

Turner, Thomas P.; Brackbill, Lise. "Rising to the top: evaluating the use of HTML META tags to improve the retrieval of World Wide Web documents through Internet search engines". *Library resources and technical services*, vol. 42, no. 4 (Oct. 1998), p. 258–271.

Vellucci, Sherry L. "Metadata and authority control". *Library resources and technical services*, vol. 44, no. 1 (Jan. 2000), p. 33–43.

Weibel, Stuart L.; Koch, Traugott. "The Dublin Core Metadata Initiative [en línia]". *D-Lib magazine*, vol. 6, no. 12 (Dec. 2000). <http://www.dlib.org/dlib/december00/weibel/12weibel.html>. [Consulta: 15/10/2001].

Weinberg, Bella Hass. "Improved Internet access: guidance from research on indexing and classification". *Bulletin of the American Society for Information Science*, vol. 25, no. 2 (Dec.-Jan. 1999), p. 26–29. Disponible a les revistes electròniques de la Biblioteca de la UB des de l'adreça següent: <http://www.bib.ub.es/bub/bub.htm>. Weiss, Amy K.; Carstens, Timothy V. "The year's work in cataloging, 1999". *Library resources and technical services*, vol. 45, no. 1 (Jan. 2001), p. 47–58.

b) Webs d'esquemes de metadades

Dublin Core Metadata Initiative [en línia]. DCMI, c1995-2001. <http://dublincore.org>. [Consulta: 15/11/2001].

Encoded archival description [en línia]: EAD: official Web site. Network Development & MARC Standards Office, Library of Congress, 2001-08-06. <http://www.loc.gov/ead/>. [Consulta: 18/11/2001].

<Indecs> [en línia]: interoperability of data in e-commerce systems. c1999-2001. <http://www.indecs.org/>. [Consulta: 18/11/2001].

Onix product information standards [en línia]. Editeur. <http://www.editeur.org/onix.html>. [Consulta: 18/11/2001].

Text encoding initiative [en línia]. University of Illinois, Chicago, 15 August 2001. <http://www.tei-c.org/>. [Consulta: 20/11/2001].

Visual Resources Association. Information on the VRA Core [en línia]. VRA, last updated 06/02/00. <http://www.gsd.harvard.edu/~staffaw3/vra/coreinfopage.htm>. [Consulta: 20/10/2001].

c) Selecció de projectes i portals temàtics que empren o adapten el conjunt de metadades DC

ADAM [en línia]: *the gateway to art, design, architecture & media Information on the Internet*. <http://adam.ac.uk/>. [Consulta: 04/11/2001]. *AHDS* [en línia]: Arts & Humanities Data Service. *AHDS*. <http://ahds.ac.uk/>. [Consulta: 04/11/2001].

BIBLINK [en línia]: *linking publishers and national bibliographic services. Metadata*. UKOLN, 2 March, 1999. <http://www.ukoln.ac.uk/metadata/biblink/>. [Consulta: 04/11/2001].

CORC [en línia]: *Cooperative Online Resource Catalog*. OCLC, c2001. <http://www.oclc.org/corc/>. [Consulta: 25/10/2001].

DESIRE [en línia]: *Development of a European Service for Information on Research and Education*. DESIRE Consortium, c1998-2000, last updated 07 August 2000. <http://www.desire.org/>. [Consulta: 04/11/2001].

GEM project site [en línia]: *the Gateway to Educational Materials*. Syracuse: ERIC Clearinghouse on Information & Technology. <http://www.geminfo.org/>. [Consulta: 04/11/2001]. Internet Scout

Project [en línia]. c1994-2000. <http://scout.cs.wisc.edu/misc/copyright.html>.

Meta matters [en línia]. National Library of Australia, 1999-05-24. <http://www.nla.gov.au/meta/>. [Consulta: 18/10/2001].

Metadata server at the State and University Library at Göttingen, Germany (SUB) [en línia]. Göttingen: Metadata Server, c1997-2001, last update 29.11.2001. <http://www2.sub.uni-goettingen.de/>. [Consulta: 18/10/2001].

The Nordic Metadata Projects [en línia]. Helsinki, Helsinki University Library, last updated: 21 February, 2000. <http://www.lib.helsinki.fi/meta/>. [Consulta: 04/11/2001].

Pinakes [en línia]: *a subject launchpad. Designed by Dave Bond and Roddy MacLeod.* Hosted by Heriot-Watt University, Edinburgh, Scotland, last updated: 7 November 2001. <http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html>. [Consulta: 30/11/2001].

VADS [en línia]: *Visual Arts Data Service: providing, preserving & promoting digital resources for research, learning & teachin.* <http://vads.ahds.ac.uk/>. [Consulta: 04/11/2001].

d) Taules de correspondències

Day, Michael. *Metadata. Mapping between metadata formats* [en línia]. UKOLN, University of Bath, created: August 1996, last updated: 13 Feb., 2001. <http://www.ukoln.ac.uk/metadata/interoperability/>. [Consulta: 19/10/2001].

MARC 21 documentation. Marc mappings [en línia]. Library of Congress, Network Development and MARC Standards Office, October 12, 2001. <http://lcweb.loc.gov/marc/marcdoc.html>. [Consulta: 20/10/2001].

Apèndix A: Taula de qualificadors del Dublin Core i qualificadors emprats en el projecte (inclosos afinaments i esquemes de qualificació)

Elements del DC	Afinaments dels elements en el DC	Esquemes de codificació dels elements en el DC	Afinaments dels elements en el projecte	Esquemes de codificació dels elements en el projecte
Title	Alternative	—	Alternative	—
Creator	—	—	PersonalName CorporateName Adress	CBC
Subject	—	LCSH, MeSH, DDCLCC, UDC	—	LEMC, CDU2000, CAL
Description	Table Of Contents Abstract	—	Table Of Contents Abstract	—
Publisher	—	—	PersonalName CorporateName	CBC
Contributor	—	—	PersonalName CorporateName	CBC
Date	Created, Valid, Available, Issued, Modified	DCMI Period, W3C-DTF	Created, Valid, Available, Issued, Modified	W3C-DTF
Type	—	DCMI Type Vocabulary	—	DCMI Type Vocabulary
Format	Extent	—		
	Medium	IMT	Medium	IMT
Identifier	—	URI	—	URI, ISBN, ISSN
Source	—	URI		

Language	—	ISO 639-2 RFC 1766	—	ISO 639-2
Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format	URI	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format	URI
Coverage	Spatial	DCMI Point, ISO 3166, DCMI Box, TGN	—	ISO 366 CATMARC 043
	Temporal	DCMI Period, W3C-DTF	—	W3C-DTF
Rights	—	—	—	—

Notes

¹ Aquest projecte s'ha dut a terme amb una ajuda a projectes de recerca de la Divisió de Ciències en la convocatòria de 1999.

² Robin Henshaw, Edward J. Valauskas, "Metadata as a catalyst: experiments with metadata and search engines in the Internet journal, First monday", *Libri*, vol. 51, no. 2 (June 2001), p. 86-101.

³ A l'adreça següent de la Dublin Core Metadata Initiative es dona accés a gairebé setanta projectes de diferents països d'arreu del món, i a alguns d'abast internacional, que parteixen del DC: <http://dublincore.org/projects/>.

⁴ Per a una recopilació de bases de dades temàtiques de recursos web d'interès acadèmic, consulteu la referència següent: *Pinakes* [en línia]: *a subject launchpad*. Designed by Dave Bond and Roddy MacLeod. Hosted by Heriot-Watt University, Edinburgh, Scotland, last updated: 7 November 2001. <http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html>. [Consulta: 30/11/2001].

⁵ S'assignaren metadades a totes les pàgines HTML de la Facultat de Biblioteconomia i Documentació; en el cas de les pàgines en format PDF se n'assignaren només a una petita mostra dels programes de les assignatures i dels horaris de les classes.

⁶ Dublin Core Metadata Initiative, *Dublin Core qualifiers* [en línia], DCMI, 2000-07-11, <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>. [Consulta: 30/11/2001].

⁷ Nordic Metadata Project, *Dublin Core metadata template* [en línia], NMP, Created: 1997-09-26, last update: 1998-03-17, <http://www.lub.lu.se/cgi-bin/nmdc.pl>. [Consulta: 30/11/2001].

⁸ Diane I. Hillmann, *Using Dublin Core*, DCMI, 2001-04-12, <http://dublincore.org/documents/2001/04/12/usageguide/>. *User guidelines for Dublin Core creation*, Nordic Metadata Project, created: 2001-04-16, last update: 1999-11-22, http://www.sics.se/~preben/DC/DC_guide.html. [Consulta: 30/11/2001].

⁹ Vegeu les propostes següents: Dublin Core Metadata Initiative. Libraries Working Group. *Library application profile*. DCMI, 2001-10-12, <http://www.dublincore.org/documents/2001/10/12/library-application-profile/>. [Consulta: 05/11/2001]. Dublin Core Metadata Initiative. Government Working Group. *DC-Gov application profile*. DCMI, November 2001, <http://dublincore.org/groups/government/profile-200111.shtml>. [Consulta: 30/11/2001].

Data recepció: 23/11/2001 Data acceptació: 03/12/2001

Facultat de Biblioteconomia i Documentació
Universitat de Barcelona
Barcelona, desembre de 2001
<http://www.ub.edu/biblio> •  [Comentaris](#)

 [Citació recomanada](#) • [Metadades](#)
[UB](#) • [Facultat](#) • [BiD](#)