

Review of metadata standards in use by SHERPA DP repositories

Document details

Project: SHERPA DP
Work Package: 4.1
Author: Gareth Knight & Kirti Bodhmage
Version: Version 1.0
Document date: 13/02/2006
Change history:

<i>Date</i>	<i>Version</i>	<i>Author</i>
14/12/2006	First version	Gareth Knight

Contents

Summary	2
Metadata within institutional repositories	2
Repository-specific refinements	3
Subject Searching	3
Type-specific metadata	4
Recommendations.....	5
Appendix A: EPrints metadata fields.....	7
Appendix B: DSpace metadata fields	11
References	13

Summary

E-print repositories store a range of metadata in order to administer and maintain the items within their collection. This paper outlines the metadata held by repositories participating within the SHERPA DP project and outlines some recommendations for standardisation and enhancement.

Metadata within institutional repositories

Metadata is often described as 'data about data'. It is any structured data associated with a resource that describes the resource. Metadata contains a diverse range of information and may be constructed for different purposes. A metadata scheme may contain the following information:

1. **Descriptive** – Descriptive metadata is useful for the indexing, discovery and identification of a digital resource. The Dublin Core element set, in a qualified or unqualified variety, is considered the basic information necessary to describe an item.
2. **Administrative** – metadata necessary for the internal management of a digital resource. Administrative information may be further subdivided into preservation, provenance and authorisation policy metadata.
3. **Structural** – Information that is required to indicate how the items should be presented or suggest a relationship between constituent components.

Metadata in an institutional repository has a dual purpose: it is created to enable researchers to locate the research paper and allow its successful management over time. E-print repositories participating in the SHERPA DP project, and indeed the SHERPA project as a whole, implement the EPrints or DSpace repository software. These possess their own application profile¹ that is useful for describing resources.

Institutional repositories use the OAI-PMH (Open Archives Initiative for Metadata Harvesting) to maintain interoperability and enable the harvest of simple Dublin Core metadata set. Simple Dublin Core provides an interoperable framework through the provision of a core set of generic metadata elements, however it lacks the detail of metadata standards that have been developed in response to the needs of specific resource types. The absence of a controlled vocabulary and use of vague or broadly defined terms will make it difficult to locate a resource, and potentially reduce its value.

SHERPA DP repositories, similar to other e-print archives take a pragmatic approach. They attempt to balance the need to maintain interoperability through support for the minimum data recommended for Dublin Core on the oai_dc record format with the unique requirements of their institutions. The simple Dublin Core exposed by these repositories is, for most purposes, compatible with guidance issued by DC-Lib and Powell *et al.* However, there are likely to be key differences in the vocabulary of descriptive information and the type of administrative metadata held within the internal database that may be specific to the subject domain (e.g. abstract, date of accession, and date made available) and unique to the institution (e.g. originating institution). When ingesting e-print metadata into the preservation repository, the AHDS must ensure that institutional repository metadata is mapped to the appropriate Dublin Core element or that a suitable location is found within the METS framework to package the e-prints metadata.

¹ An application profile is defined as a metadata schema, which is composed of data elements drawn from one or more namespaces and optimised for a particular local use (Heery & Patel, 2000).

Repository-specific refinements

The metadata held by an institutional repository will vary according to the implemented software and the particular choices of the institution. Of the five institutions participating within the project, four of them implement the EPrints software. The base install of EPrints uses a custom schema, composed of 45 elements for each item (appendix A.) that may have particular use for an academic environment. In contrast, DSpace is marketed towards digital libraries and specifies 53 elements that comply with the DC-Lib application profile (appendix B.). The following section outlines notable difference in the areas of subject searching and type-specific metadata to show how the application profile created by SHERPA DP repositories has been refined and extended to fit their requirements.

Subject Searching

E-prints are only likely to be of use if presented in a way that allows users to retrieve them quickly and easily. Researchers are unlikely to have a clear idea of the paper they require and will wish to research the subject area by searching or browsing available resources. Repositories participating within the SHERPA DP project make extensive customization to the available metadata, ensuring that one or more of the following subject search options are available.

1. Keywords

Free text terms employed by the depositor or repository staff to describe the paper. They may identify terms or expressions used within the research paper. For example, a research paper on creating Structured PDF Files Using XML Templates may contain the keywords 'XML', 'PDF' and 'Logical Structure Insertion'.

2. Library Classification Schemes

A library classification is a system for coding and organizing library materials according to their subject. They consist of sets of controlled descriptors, organized into an enumerative, hierarchical, and/or faceted structure to allow a repository (or any library-based system) to be searched or browsed at various levels of specificity. For example, the top level will contain broad headings, which will lead to more specific headings when the researcher narrows their search. The custom DSpace installation, used by the Edinburgh Research Archive supports, but does not implement, the Dewey Decimal Classification (DDC), Library of Congress Classification Number (LCC), Library of Congress Subject Headings (LCSH), and Medical Subject Headings (MESH). These may be refined to the qualified *subject.ddc*, *subject.lcc*, *subject.lcsh* and *subject.mesh* elements.

3. Local controlled vocabulary

Finally, e-print repositories often implement a bespoke vocabulary to supplement existing qualifiers. For example, repositories participating in the SHERPA DP project classify research according to Institution (e.g. University of Nottingham), Department and Research Group. This information may be stored as three distinct fields ('institution', 'department' and 'group') or may be held in a single field. For example, '*Education: Department of Educational Studies*' in the Glasgow EPrints Service or '*The University of Leeds > Faculty of Biological Sciences (Leeds) > School of Biomedical Sciences (Leeds)*' used by the White Rose Consortium. In each case, the repository use a unambiguous marker ('>' and ':' respectively) to identify a hierarchy.

Keywords and library scheme may be mapped to the relevant Dublin Core elements, however local controlled vocabulary is likely to require new fields (i.e. institution, department, school) to preserve the unique classification method within the METS framework. Two approaches may be taken 1) SHERPA DP repositories may coordinate their efforts to produce a consistent location descriptor, and/or 2) a METS profile may be developed that mirrors the structure of existing EPrints fields. Although the project offers the opportunity for institutional repositories to compare the metadata fields they currently use, it is unlikely that they will wish to adopt a single approach that may unnecessarily restrict them at a later date. Therefore, the

Preservation Service must develop a custom METS profile that accounts for the unique fields offered by the EPrints software and institutional repositories.

Type-specific metadata

EPrint repositories may hold a wide range of research data. The default application profile offered by the EPrints software holds metadata related to particular types of resources within different elements. As a result, a metadata record held by the institutional repository may differ between different types of resources. SHERPA DP repositories typically hold three types of research paper:

1. Events

Event is a catch-all term that may be applied to any resource developed for a presentation, conference, or workshop (ERPAePRINTS). It may also include supplementary material (Nottingham ePrints) used as a teaching aid. The default installation of the EPrints software contains four elements to describe event-specific research:

event_title	Title for conference or workshop
event_location	City/town/country
event_dates	Start and end date for the event
event_type	Conference, workshop, other
Pres_type	Type of presentation

SHERPA DP repositories differ in their implementation of these sub-elements. The London LEAP and White Rose consortiums use the Event sub-elements to identify a conference in which a paper was presented, or event proceedings where it may have been published. The Event fields are a recent addition to the Eprint software, appearing in the v2.3 release. The v 2.2 release used by the Glasgow ePrints Service contains three fields applicable to conferences - 'conference' (title), 'Conference_date' and 'Conference_location' - that may be mapped to the event elements. Nottingham ePrints, White Rose and London Leap also offer an additional 'presentation_type' element that may be used to describe papers, lectures, speeches, or posters.

2. Thesis

Student theses are relatively simple to describe using eprint metadata. Unlike published e-prints that may possess complex rights issues, student theses are likely to remain the copyright of a single individual². Student theses are accepted by all repositories (Nottingham ePrints, London Leap repositories, Edinburgh Research Archive, Glasgow ERPAePRINTS and White Rose Consortium) participating within the SHERPA DP project and represent a significant percentage of their research holdings. All EPrints-based repositories support the 'thesis_type' field, which is provided in the default installation. 'Thesis_Type' may be refined to the DC:type element, as recommended by Powell et al.

3. Journals

The EPrints software offers an extensive list of journal-specific elements and, possibly as a result, is supported by every EPrints-based repository (to varying degrees). Eight elements may be identified that refer specifically to journals:

Title	Description
Publication / Journal title	The title of a journal, publication or magazine.
Number	Issue number of journal
Volume	The volume number of the journal or series in which an item appears.

² A possible exception may be applied for student theses that receive additional funding or involvement from a third-party.

Publisher	The publisher name
Page Range	The sequence of pages that the article appears within a journal, e.g. 621-630.
Pages	The total number of pages, e.g. 9. The EPrint installation used by Glasgow (v2.2) holds page range values, Earlier versions of EPrints (e.g. version 2.2 used by Glasgow repositories) hold 'Page Range' information in the 'Pages' field.
Place_of_pub	The place of publication
ISSN	A unique identifier applied to journals/periodicals (International Standard Serial Numbering).

There are minor differences in the fields supported within institutional repositories. White Rose, for example, does not support publisher, ISSN is not held by Nottingham ePrints, and Page Range is not present in the Glasgow ePrints Service profile. Similarly, 'Place of publication' is unique to UCL EPrints and is used for books, book sections, monographs, but not articles.

4. Books and Book Chapters

The Glasgow EPrints Service, Edinburgh Research Archive and London LEAP repositories also allow authors to deposit books or chapters. Three elements in the EPrint software are specific to this type:

Title	Description
Chapter	The chapter number within a book.
Series	The title of the series or set of books.
ISBN	A unique identifier applied to books

Most repositories do not hold any books in their repository or possess one or two examples submitted by authors. Given the current emphasis upon e-prints and student theses, it is unlikely that the number will increase significantly in the short-term.

Specific decisions should be made on the most effective method of managing type-specific metadata. One option is to map the three types of metadata into a refined element list. However, problems may be encountered if the institutional repository was required to map the refined list back to their existing application profile. A wiser alternative is to create suitable placeholders within the METS profile and map metadata descriptions without additional processing.

Recommendations

E-print repositories store a diverse range of institution or type-specific metadata that cannot easily be refined to the basic or qualified Dublin Core elements without loss of information. To adequately maintain and preserve e-print metadata it is recommended the AHDS (or external consultant) perform the following actions:

1. Map repository fields to a METS package

The AHDS (or external consultant) should map repository fields, currently held in EPrints and DSpace, to the qualified Dublin Core scheme and find a suitable location for repository-specific metadata within the METS framework.

2. Clarify the institutional origin of eprints

Institutional repositories should identify the institutional origin of the eprints to enable the preservation service to identify papers transferred from a particular source. Each institution should possess its own Centre Identifier. E.g. glasgow, edinburgh, leeds, kings, york. For example, the handle identifier implemented by DSpace distinguishes between institution and repository, i.e. <http://hdl.handle.net/1842/xxx>, where 1842 refers to Edinburgh, (the centre identifier) and xxx is the repository Identifier.

If the identification software cannot identify the source institution from the harvesting logs (e.g. where a consortium of institutions share a single repository) it should consult the 'institution' element within the Eprints record. When identifying eprints provided by these repositories, the statistical software implemented by the AHDS should be aware that multiple institution identifiers will be used for consortium arrangements.

3. Implement a unique identifier for e-prints within the preservation repository

The AHDS should implement an identifier system to locate eprints held within the preservation repository. The AHDS could modify their existing identifier system to apply to eprints held within the repository. An identifier should identify three type of information:

- a. **repository Identifier:** A unique alphanumeric value to identify the institution. For example, Leeds, Glasgow, Edinburgh.
- b. **Item Identifier:** The alphanumeric value assigned to the e-print by the institutional repository, e.g. e.g. "glaseprints:2003-2058", where 2003 is the year of publication and 2058 is the database record number.
- c. **Edition:** The Preservation Service may retain different versions of the item-level record for auditing purposes and should indicate this distinction in the repository identifier. For example, the first edition of a metadata record may contain bibliographic data only, the second edition may contain a research paper as a PDF and bibliographic data, while a third edition may contain an updated research paper and bibliographic record.

4. Institutional repositories should clarify the use of custom field descriptors and controlled vocabulary

Institutional repositories often implement an institution or subject-specific vocabulary to enable researchers to locate e-prints written on a similar theme or introduce diverse field descriptions for institutional structures (e.g. the different approaches to identifying schools, faculties and departments within their institution). SHERPA DP Project partners should clarify use of any bespoke vocabulary and consider standardizing its use. For example, repositories may coordinate their efforts to produce a consistent location descriptor that incorporates the following information:

- Institution – the organization responsible for the management or which owns copyright for the collection.
- Faculty – A distinct branch or division of learning within the institution
- Department – A specialized field or school of teaching within the institution or faculty.

Further enhancements to cover Groups, Centres and Virtual Centres may be made based upon further discussion with institutional repositories.

5. Institutional repositories should create checksums

Project partners should generate MD5 checksums to ensure that data and metadata is successfully transferred between the institutional repository and preservation service. Checksum values should be stored within the IR database and transferred to the Preservation Service as a separate file. The DSpace software automatically generates a checksum for each item and stores it with the item level record. However, further customisation is necessary to implement this feature within the EPrints software.

Appendix A: EPrints metadata fields

A list of elements used by institutional repositories that implement the EPrints repository software. Three types of information are identified – ‘Y’ indicates the field is used, ‘E’ indicates the field is held by the repository but not currently in use, and ‘N’ suggests the field is not implemented by the institutional repository.

Element	Institutional repository										
	Nottingham University	Glasgow ERPA ePRINTS service	Glasgow Jelit Service	Glasgow EPrints	UCL	White Rose	Royal Holloway	Birkbeck	KCL	SOAS	LSE
Creator / Author	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Title	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subjects	Y	Y	Y	Y	Y ³	N	Y	Y	Y	Y	Y
Faculties	Y	N	N	Y	N	Y	Y	Y	Y	Y	Y
Keywords	Y	Y ⁴	Y	Y	Y	Y	Y	Y	Y	Y	Y
Abstract	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Publisher	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
full_text_status	N	N	N	Y ⁵	Y	N	N	N	N	N	N
Presentation type (pres_type)	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y
Note	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Suggestions/Additional Information	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
date_sub	N	Y ⁶	Y	Y	Y	Y	Y	Y	Y	Y	Y
Date_issue	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
Year	N	Y ⁷	Y	Y	N	N	N	N	N	N	N
Month	N	Y ⁸	Y	Y	N	N	N	N	N	N	N
date_effective	N	N	N	N	Y ⁹	Y					
Book Title	N	N	N	N ¹⁰	N	N	Y	Y	Y	Y	Y
Series	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Publication/Journal Title	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Volume	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
place_of_publication	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
pagerange	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y
Pages	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Institution	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Department	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
researchgrp	N	N	N	Y	N	N	N	N	N	N	N
thesis_type	Y	Y ¹¹	N	E	Y	E ¹²	Y	Y	Y	Y	Y
Isbn	N	N	Y	Y ¹³	Y	Y	Y	Y	Y	Y	Y
Issn	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y

³ Mapped to UCL Eprint Classification.

⁴ Keywords are held for only a small number of ePrints, (approx 50 ePrints out of 1500). In many cases keywords are the copyright of the journal publisher.

⁵ Glasgow ePrints use the element ‘publicfulltext’ for full text status. This may be mapped to full_text_status without subsequent problems.

⁶ Mapped to the eprints ‘datestamp’ field.

⁷ Year is a default field in EPrints v2.2 that is subsequently replaced by data_issue in v2.3. It is part of the ‘archive’ table in the database. Year is used in all three Glasgow repositories.

⁸ Month is a default field in EPrints v2.2 that is subsequently replaced by data_issue in v2.3. It is part of the ‘archive’ table in the database. Month is used in all three Glasgow repositories.

⁹ Stores either date_sub or date_issue, whichever is used at deposit.

¹⁰ Although books & book chapters are held in the Glasgow ePrints service, the existing ‘title’ and ‘publication’ elements are used, as opposed to book_title.

¹¹ Thesis_Type is used by EPRAePRINTS service, but not by the two other ePrints services.

¹² Field is available within the database, but is not currently in use. White Rose do not currently hold any thesis material.

¹³ Element used for ISBN is called ‘dummy1’ in database table.

Id_number	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fileinfo	Y	N	N	N ¹⁴	Y	Y	Y	Y	Y	Y	Y
main	N	Y	Y	Y	N	N	N	N	N	N	N
formatdesc	N	Y	E	E	Y	N	N	N	N	N	N
Referencetext	Y	Y	Y	E ¹⁵	E	Y	Y	Y	N	Y	Y
Editors	Y	Y ¹⁶	Y	Y	Y	Y	Y	Y	Y	Y	Y
Public Domain (pubdom)	Y	Y	Y	E	N	N	N	N	N	N	N
Status	Y	Y	Y	Y ¹⁷	Y	Y	Y ¹⁸	Y ¹⁹	Y ²⁰	Y ²¹	Y ²²
Refereed	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Type	Y	Y	Y	Y	Y	Y ²³	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	N	Y	N	N	N	N	N
Month	Y	Y	Y	Y	N	Y	N	N	N	N	N
Official_URL	Y ²⁴	Y ²⁵	Y	Y	Y ²⁶	Y ²⁷	Y	Y	Y	Y	Y
Commentary (commref)	Y	Y	E	E	N	N	N	N	N	N	N
Security level	Y	Y ²⁸	Y	Y	E ²⁹	Y ³⁰	Y	Y	Y	Y	Y
event_title	N	N	N	N	Y	Y	Y	Y	Y	N	Y
event_location	N	N	N	N	Y	Y	Y	Y	Y	N	Y
event_dates	N	N	N	N	Y	Y	Y	Y	Y	N	Y
event_type	N	N	N	N	Y	Y	Y	Y	Y	N	Y
Conference date	Y	Y	Y	Y	N	Y	N	N	N	N	N
Conference location	Y	Y	Y	Y	N	Y	N	N	N	N	N
Preprint (or Associated Material) Available	N	N	N	Y ³¹	N	N	N	N	N	N	N
Preprint URL	N	Y	Y	Y ³²	N	N	N	N	N	N	N
Reportno	Y	E	E	E	N	N	N	N	N	N	N
Publisher version	N	N	N	N	N ³³	Y ³⁴	N	N	N	N	N
UCL Eprints classification	N	N	N	N	Y	N	N	N	N	N	N

¹⁴ The name of the document file uploaded into the ePrints services is held in element 'main', in the database table called 'document', e.g. 'Chaucer_article.pdf'.

¹⁵ Glasgow ePrints Service make little use of the referencetext field. It is not requested when the author deposits their document and only 10 records currently hold information in this field.

¹⁶ This element is used for books, books chapters and conference material within the three ePrints services.

¹⁷ Status is mapped to 'ispublished' field in the EPrints software.

¹⁸ Mapped to the IsPublished field.

¹⁹ Mapped to the IsPublished field.

²⁰ Mapped to the IsPublished field.

²¹ Mapped to the IsPublished field.

²² Mapped to the IsPublished field.

²³ Article, book, Monograph, Conference or Workshop Item, Patent

²⁴ Mapped as altloc

²⁵ Mapped as alternative location ('altloc' in the 'archive_altloc' table).

²⁶ URL of publisher's copy of this article (optional, unless required by publisher).

²⁷ Mapped as Associated URL/official URL, which means link to the journal homepage

²⁸ Anyone / Register users / archive staff

²⁹ validuser / staffonly

³⁰ Registered users only, white rose university only,archive staff only,users from eprints institution only

³¹ Mapped as 'preprintavail'

³² Element 'preprinturl' is held in database table 'archive_preprinturl'.

³³ An amended field that may be mapped to official_url.

³⁴ Labelled as "Published version".

Patent applicant	N	N	N	N	Y	Y	Y	Y	Y	N	Y
Patent ID No.	N	N	N	N	Y	Y	N	N	N	N	N
Monograph type	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
copyright	N	N	Y	Y	N	N ³⁵	N	N	N	N	N
prior	N	N	Y	Y	N	N	N	N	N	N	N
repro	N	N	Y	Y	N	N	N	N	N	N	N
chapter	N	Y	Y	Y	N	N	N	N	N	N	N
rights	N	N	E	E	N	E	N	N	N	N	N
RAEsubmit	N	N	N	E	N	N	N	N	N	N	N
PURL	N	N	N	E	N	N	N	N	N	N	N
subtype	N	N	E	E	N	N	N	N	N	N	N
uofa	N	E	E	E	N	N	N	N	N	N	N
corpauthor	N	Y	N	N	N	N	N	N	N	N	N
language	N	Y	N	N	N	N	N	N	N	N	N
titleofitemreviewed	N	N	Y	N	N	N	N	N	N	N	N
authorsofitemreviewed	N	N	Y	N	N	N	N	N	N	N	N
Administrative Data											
eprintid	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
userid	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
dir	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
succeeds	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
commentary	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
replacedby	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
uniqueid	Y	N	N	Y	N	Y	Y	Y	Y	Y	Y
Workflow status	Y	N	E	Y ³⁶	N	Y	Y	Y	Y	Y	Y

Field Descriptions

Creator / Author	The name of the creator or author
Title	Title of the research paper, presentation, etc.
Subjects	The Library of Congress subject classification for the document
Faculties	The faculty and department from which the document originated. Subject and faculty is declared as type "subject" within ePrints,
Keywords	Keyword to help searchers to search the papers
Abstract	Free text field for the paper abstract
Publisher	Name of the publisher
full_text_status	full text of the paper, or just the bibliographic details (e.g. title, abstract etc. only) is submitted
Presentation type (pres_type)	Identifies the type of conference paper held. Possible values may be "paper", "lecture", "speech", "poster", or "other".
Note	Additional information
Suggestions/Additional Information	Any comments to the administrators (eg regarding problems with submission). Not displayed to public
date_sub	Date of deposit/submission
Date_issue	Date on which deposit was published or produced.
Year	Year of publication
Month	Month of publication
date_effective	A system-defined date, based upon the "best" date within an e-print record or, if it is not available, the date of submission.
Book Title	The title of a book
Series	The title of the series or set of books
Publication/Journal Title	The title of a journal, publication or magazine.
Volume	volume number of the journal
Number	Issue number of the journal or series in which your item appeared.
place_of_pub	Place of publication

³⁵ Copyright held in Note field and labelled as "Copyright, Publisher and Additional Information:"

³⁶ Element holding workflow status is called 'dummy4'.

pagerange	Sequence of pages
Pages	Total number of pages
Institution	Host institution. E.g. University of Glasgow
Department	Department
Research Group	The research group from which the submission originated
thesis_type	The type of thesis held by the repository.
Isbn	A unique identifier for books intended to be sold commercially
Issn	Unique 8-digit code identifying a serial.
Id_number	A DOI identifier
Fileinfo	A refined list of filetypes and location of the files.
formatdesc	A free text field for repository staff to describe non-standard formats. E.g. "Windows Media Audio 5.3MB". The field is held internally and appears on the Abstract page for relevant e-prints.
Referencetext	Reference list for the papers
Editors	Name of the Editor/Moderator
Public Domain	If the document you are deposit is not your own but rather an old document that is now in the public domain, then please tick the following box. This will prevent your own name and address appearing with the document as the address for correspondence.(yes/no)
Status	Published/ In Press/Unpublished
Refereed	YES/NO, only refereed papers should be submitted If this version has been refereed
Eprint types	Article, Journal, Book, Book Chapter, Conference or Workshop Item, thesis, patent, Newspaper, magazine.
Year	Year in which the deposit was written / published.
Month	Month in which your deposit was published or produced
Official_URL	Alternative location of the e-print
Commentary on(commref)	If the paper is a commentary on a item that is not in this archive, then the full reference of the item is entered here
File Type (not a metadata field)	
Security level	Identifies groups with appropriate permission to access and download the paper
event_title	Title for conference or workshop
event_location	City/town/country
event_dates	Start and end date for the vent
event_type	Conference, workshop,other
Conference date	Date of conference (YYYYMMDD)
Conference	Conference title
Conference location	Location
Preprint (or Associated Material) Available	Indicates the availability of an existing preprint (Boolean)
Preprint URL	Hyperlink location to the preprint or associated metadata record
Reportno	
Publisher version	URL of publisher's copy of this article (optional, unless required by publisher).
UCL Eprints classification	UCL Department or Research Centre
Patent applicant	The name of the applicant to whom the patent was granted. May be a person or an organisation.
Patent ID No.	The patent application number of this item.
Monograph type	Monograph type. Working papers / Discussion papers/technical reports
copyright	Copyright details
prior	Details of first publication.
repro	Publisher's policy on reproducing document.
chapter	Chapter number within a book.

rights	Rights to the metadata. Category E.
RAEsubmit	The Research Assessment Exercise in which the document was included.
PURL	Persistent URL.
subtype	Journal content type, e.g. Review, Editorial
uofa	Unit of assessment (for RAE).
corpauthor	Corporate author.
language	Language of the document
titleofitemreviewed	Title of the item being reviewed.
authorsofitemreviewed	Authors of the item being reviewed.
Administrative Data	
eprintid	Record number in database
userid	User id of depositing user
dir	Holding directory on host machine
succeeds	ID of earlier version of a document
commentary	ID of document being commented on.
replacedby	ID of document replacing this one.
uniqueid	Unique identifier for the document.
Workflow status	Flag indicating Editing Phase / Subject Assignment Phase in workflow.

Appendix B: DSpace metadata fields

The DSpace repository software implements metadata that complies with the DC-Lib application profile for qualified Dublin Core. Three types of information are identified – ‘Y’ indicates the field is used, ‘E’ indicates the field is held by the repository but not currently in use, and ‘N’ suggests the field is not implemented by the institutional repository.

Metadata field Name	Description	Edinburgh Research Archive
contributor.advisor	A person, organization, or service responsible for providing unspecified advice	Y
contributor.author	A person, organization, or service responsible for the creation of the resource.	Y
contributor.editor	A person, organization, or service responsible for the editing of the resource.	Y
contributor.illustrator	A person, organization, or service responsible for the illustration of the resource.	Y
contributor.other	Catch-all for unspecified contributors	Y
coverage.spatial	Spatial characteristics of content	Y
coverage.temporal	Temporal characteristics of content	Y
date.accessioned	Date that DSpace takes possession of item	Y
date.available	Date or date range item became available to the public.	Y
date.copyright	Date of copyright	Y
date.created	Date of creation or manufacture of intellectual content if different from date.issued	Y
date.issued	Date of publication or distribution	Y
date.submitted	Recommend for theses/dissertations	Y
Date.verified	Date that the e-print was verified.	Y
Identifier		Y
identifier.citation	Human-readable, standard bibliographic citation of non-DSpace format of this item	Y
identifier.govdoc	A government document number	Y
identifier.isbn	International Standard Book Number	Y
identifier.issn	International Standard Serial Number A known identifier type common to a local collection.	Y
identifier.sici	Serial Item and Contribution Identifier	Y
identifier.ismn	International Standard Music Number	Y

identifier.other	A known identifier type common to a local collection	Y
identifier.uri	Uniform Resource Identifier'	Y
Description		Y
description.abstract	Abstract or summary	Y
description.provenance	The history of custody of the item since its creation, including any changes successive custodians made to it	Y
Description.sponsorship	Information about sponsoring agencies, individuals, or contractual arrangements for the item	Y
description.statementsofresponsibility	To preserve statement of responsibility from MARC records	Y
description.tableofcontents	A table of contents for a given item.	Y
description.uri	Uniform Resource Identifier pointing to description of	Y
Format	size or format	Y
format.extent	Duration	Y
format.medium	Physical medium.	Y
format.mimetype	Registered MIME type identifiers	Y
Language		Y
language.iso	ISO standard for language of intellectual content, including country codes (e.g. "en_US")	Y
Publisher	Entity responsible for publication, distribution, or imprint.'	Y
relation.isformatof	References additional physical form	Y
relation.ispartof	References physically or logically containing item	Y
relation.ispartofseries	Series name and number within that series, if available	Y
relation.haspart	References physically or logically contained item	Y
relation.isversionof	References earlier version	Y
relation.hasversion	References later version	Y
relation.isbasedon	References source	Y
relation.isreferencedby	Pointed to by referenced resource.	Y
relation.isreplacedby	References succeeding item	Y
relation.replaces	References proceeding item	Y
relation.uri	References Uniform Resource Identifier for related item	Y
relation.requires	Referenced resource is required to support function, delivery, or coherence of item	Y
Rights	Terms governing use and reproduction	Y
Rights.uri	References terms governing use and reproduction.	Y
source (only for harvested metadata) source.uri		Y
Subject		Y
subject.classification	Global classification systems will receive specific qualifies	Y
subject.ddc	Dewey Decimal Classification Number	E
subject.lcc	Library of Congress Classification Number	E
subject.lcsh	Library of Congress Subject Headings	E
subject.mesh	Medical Subject Headings	E
Subject.udc	Universal Decimal Classification	E
subject.other	Local controlled vocabulary; global vocabularies will receive specific qualifier.	Y
Title		Y
title.alternative	Alternative r substitute) form of title proper appearing in item, e.g. abbreviation or translation	Y
Type	Nature or genre of content	Y

References

DCMI (1998). *Dublin Core Metadata Element Set, Version 1.0: Reference Description*. Retrieved on December 9, 2004 from:

<http://dublincore.org/documents/1998/09/dces/>

Guenther, R. (2002). *DC-Library Application Profile*.

Retrieved on December 9, 2004 from:

<http://dublincore.org/documents/library-application-profile/>

Heery, R., & Patel, M. (2000). *Application profiles: mixing and matching metadata schemas*. Ariadne, 25. Retrieved on December 9, 2004 from:

<http://www.ariadne.ac.uk/issue25/app-profiles/>

Open Archives Initiative [OAI] (2003). *The Open Archives Initiative Protocol for Metadata Harvesting*.

Retrieved on December 9, 2004 from:

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

James, H. et al, (2003). *Feasibility and Requirements Study on Preservation of E-Prints*.

Retrieved on December 9, 2004 from:

http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf

Pinfield, S. Gardner, M. and MacColl, J. (2002). *Setting up an institutional e-print archive*.

Retrieved on December 9, 2004 from:

<http://www.ariadne.ac.uk/issue31/eprint-archives/>

Polfreman, M. (2004). *Metadata and Subject Searching*.

Retrieved on December 9, 2004 from:

<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-64/>