Citation for published version:
Day, M 2002, Metadata in support of subject gateway services and digital preservation. in M Guerrini (ed.), *Le risorse electtroniche: definizione, selezione e catalogazione = Electronic resources: definition, selection and cataloguing.* Editrice Bibliografica, Milan, pp. 561-573.

Publication date:
2002

Link to publication

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Metadata in support of subject gateway services and digital preservation

Michael Day[1]
UKOLN
University of Bath
Bath BA2 7AY, United Kingdom
http://www.ukoln.ac.uk/
m.day@ukoln.ac.uk

Abstract: This paper provides an introduction to two of the metadata-related projects in which UKOLN has been a partner. It first describes the development of services known as quality controlled subject gateways and looks in more detail at the Resource Discovery Network and the EU Renardus project. It then provides an outline of recent preservation metadata initiatives and describes the way the OAIS model has been used in the Cedars project.

Introduction

While the first use of 'metadata' originated in contexts related to digital information (chiefly with regard to databases), the general understanding of the term has since expanded to include any kind of standardised descriptive information about both digital and non-digital resources. Thus, for example, the content of library catalogues and the descriptive data that lies behind abstracting and indexing services, archival finding aids and museum documentation, might all be understood to be metadata. The advantages of this are twofold. Firstly, it acts as a spur for librarians, archivists and museum documentation specialists to co-operate across professional boundaries. Secondly, it enables the cultural heritage professions to communicate more effectively with other domains that also have an interest in metadata: e.g., software developers, publishers, the recording industry, television companies, the producers of digital educational content and those concerned with geographical and satellite-based information. Metadata is currently, therefore, an important field of research and development [1].

Over the past decade, UKOLN (formerly the UK Office for Library and Information Networking) at the University of Bath has been involved in a number of metadata-based projects and initiatives. There are accounts of some former projects in overviews by Dempsey [2] and Day [3]. This paper will describe progress in some more recent projects and initiatives in which UKOLN and its partners have been involved. It will focus in particular on two strands of metadata research and development. The first of these will be

---

[1] This is a draft version of a paper on UKOLN metadata projects presented at the international conference "Electronic Resources: Definition, Selection and Cataloguing," held at the Università degli Studi di Roma «La Sapienza» and the Biblioteca nazionale centrale di Roma, Rome, Italy, 26th - 28th November 2001. A revised version has been published in both Italian and English in: Mauro Guerrini, (ed.), *Le risorse electtroniche: definizione, selezione e catalogazione = Electronic resources: definition, selection and cataloguing*, Milan: Editrice Bibliografica, 2002, pp. 193-206 (Italian), pp. 561-573 (English). ISBN 88-7075-586-X.

the continuing development of 'subject gateway' services and will focus on the progress of the European Renardus project and the UK's Resource Discovery Network (RDN). The second theme will be the ongoing development of metadata schemas to support digital preservation. The main project in this area in which UKOLN has had some involvement is the Cedars (CURL Exemplars in Digital Archives) project, but some other recent initiatives will also be described.

With regard to the title of this paper, it is important to stress that these are not UKOLN metadata projects, but metadata-related projects in which UKOLN is involved. Almost all of the metadata projects in which UKOLN is involved have multiple partners and most project outcomes are produced collaboratively. The projects described here are no exception. So while the main focus of the paper will be on work undertaken by staff at UKOLN, it is important to recognise the vital roles of our partner organisations and the project leaders.

Quality-controlled subject gateways

The Internet, and the World Wide Web in particular, is the dissemination medium of choice for a wide (and growing) range of information. One consequence of this has been the development of new types of information retrieval tools to support Internet resource discovery. Examples of these tools are the popular robot-based Web index services *Google* [4] and *AltaVista* [5]. These are based on harvesting software that periodically crawl through Web pages, following links and indexing all (or some) of the text associated with them. An alternative, but complementary, approach has been developed by Web directory services like *Yahoo!* [6] that provide links to Internet resources that are presented for users to browse in hierarchical lists. Selected resources are described by humans and assigned relevant subject terms, the user interface presenting some basic metadata in the form of a title, a brief description and a URL. A similar kind of service - based on the selection and cataloguing efforts of volunteers - is provided by the Open Directory Project [7].

Subject gateways are services that give access to selected third-party Web resources, usually with a specific focus on high-quality information suitable for use in research or higher education. On a superficial level, subject gateways share some of the same characteristics as Web directory services. However, there tends to be a much greater emphasis on the selection of resources according to some pre-defined quality criteria combined with the production of rich descriptive metadata which can enhance the potential of both searching and browsing facilities. This can be seen as analogous to the value-added selection and cataloguing services that libraries and other information professionals have traditionally provided for print formats [8]. Traugott Koch notes that the most sophisticated gateways would have policies for collection development and management, would create rich metadata complying with particular standards, and would use subject vocabularies to create a browse structure and to enhance searching [9]. He refers to these services as 'quality-controlled subject gateways.' Examples of this kind of service would include the Finnish Virtual Library [10] and the gateways that make up the UK's Resource Discovery Network (RDN).

UKOLN has previously been involved in a number of collaborative projects that have concerned the development and support of subject gateways. These included the ROADS (Resource Organisation And Discovery in Subject-based services) initiative funded by the Joint Information Systems Committee (JISC) and the two phases of the EU-funded DESIRE (Development of a European Service for Information on Research and Education) project [11]. ROADS developed software tools which would help support the establishment and maintenance of subject gateways, but also investigated associated issues like cross-searching multiple gateways, interoperability and query routing [12]. DESIRE looked at some of the same issues in a wider European context and in its second phase produced an *Information Gateways Handbook* to help support the creation and management of gateway services [13]. UKOLN's current work on subject gateways is undertaken as part of its role in the RDN and as partners in the Renardus and IMesh Toolkit projects.

The Resource Discovery Network (RDN)

The broad aim of the Resource Discovery Network (RDN) is to give the UK learning and research communities access to high-quality Internet resources across the whole range of subjects required at further education and higher education levels [14]. It is funded by the JISC, while some of its constituent services have additional support from the Economic and Social Research Council (ESRC) and the Arts and Humanities Research Board (AHRB). It is an attempt to build on the experiences of the subject gateways that previously had been funded by JISC as part of the Electronic Libraries Programme (eLib) and the ROADS project [15].

Organisationally, the RDN consists of a central organisation, the Resource Discovery Network Centre (RDNC) and a number of independent service providers called 'hubs'. The RDNC is run jointly by staff based at King's College London and UKOLN. The Centre is responsible for setting service standards, creating collection development policies and exploring strategic partnerships. Hubs provide one or more subject gateway services for their target user community, but can also offer other services. Existing hub services cover the health and life sciences (BIOME), engineering, computing and maths (EEVL), the humanities (Humbul Humanities Hub), the physical sciences (PSIgate), and social sciences, business and law (SOSIG). Three new hubs to cover hospitality, leisure, sport and tourism (Altis), geography and the environment, and the arts and creative industries are currently under development.

In order to help users search for resources described in all constituent RDN gateways, an all-RDN cross-search service known as *ResourceFinder* has been developed. Since its launch, this feature has been based on a variety of different technologies. Early versions of the feature were based on cross-searching using either the WHOIS++ protocol (as implemented by the ROADS subject gateway toolkit) or Z39.50. However, Andy Powell noted problems with these approaches related to response times and the difficulty of developing flexible browse interfaces [16]. The RDN, therefore, began to investigate the possibility of basing *ResourceFinder* and other services on a record-sharing model rather than through cross-searching. The tool chosen to test this new model was an implementation of the protocol for metadata harvesting developed by the Open Archives Initiative (OAI) [17]. Each RDN service converts records and moves them periodically to

the RDN's OAI repository, where they are stored as unqualified Dublin Core in XML. Searches of *ResourceFinder* currently return relatively simple result sets consisting of title, description, URL and an indication of which RDN service the record was taken from. It therefore does not at present use the full richness of the metadata that exists within the RDN services, although it is hoped that additional services based on OAI tools will be developed in the future.

The Renardus Project

Renardus is a project funded by the European Commission as part of its Information Society Technologies (IST) programme [18]. The twelve partners in Renardus include national libraries, research centres and subject gateway services from Denmark, Finland, Germany, the Netherlands, Sweden and the UK, co-ordinated by the Koninklijke Bibliotheek (National Library of the Netherlands). The project aims to develop a Web-based service to enable searching and browsing across a range of distributed European-based subject gateways [19].

The overall aim of Renardus is to establish a collaborative framework for European subject gateways that will benefit both users in terms of enhanced services, and the gateways themselves in terms of shared solutions. In order to achieve this aim, Renardus is firstly providing a pilot service for the European academic and research communities brokering access to those gateways that currently participate in the project. Secondly the project is exploring ways to establish the organisational basis for co-operative efforts like metadata sharing, joint technical solutions and agreement on standardisation. It is hoped that this exploration will feed back valuable experience to the individual participating gateways to suggest ways in which their services can be enhanced [20].

The services participating at the project stage in Renardus are:

- DAINet - German Agricultural information Network
- DNB-Theses (Germany)
- DutchESS - Dutch Electronic Subject Service (Netherlands)
- EELS - Engineering E-Library, Sweden
- FVL - Finnish Virtual Library
- SSGFI gateways (Anglo-American Culture (literature and history), GeoGuide, MathGuide) at the Goettingen State and University Library (Germany)
- NOVAGate (Denmark, Finland, Iceland, Norway and Sweden)
- RDN - Resource Discovery Network (UK)

A team based at the Technical Knowledge Centre and Library of Denmark (DTV) and NetLab (Lund University Libraries), developed the pilot Renardus broker [21]. It is a completely distributed system using the Z39.50 protocol to search across the content of all participating gateway services. Each participant has to set up a Z39.50 server with data normalised with reference to the Renardus data model.

Work on developing a common data model (metadata format) that would be used by the Renardus broker was led by the Renardus project team based at the Niedersächsische Staats- und Universitätsbibliothek Göttingen [22]. The team first undertook a detailed analysis of all metadata formats used by participating services and then proposed the minimum set of metadata elements which would need to be supported. These were, where possible, based on the Dublin Core (DC) Metadata Element Set. The minimum set for describing gateway content contained the DC elements Title, Creator, Description, Subject, Identifier, Language, Type, and the non-DC element Country. For administrative purposes, the data model also defined additional elements for identifying the location of the original metadata for a resource (Full Record URL) and a gateway identifier (SBIG ID). Some of these elements are able to be qualified using refinements defined by DC (e.g., Title.Alternative) or qualified by the use of schemes.

The Renardus pilot broker provides both simple and advanced searching facilities. The advanced search enables the user to define which particular metadata fields they want to search. It also allows users to filter results by document type, language or country, and to change the display defaults.

As well as the cross-search feature, the Renardus partners also wanted to offer some kind of subject browsing across all of the participating gateways. A working group of the project was set up to investigate this. In order to achieve some level of cross-browsing, a classification scheme was chosen to act as a central 'switching-language' within the Renardus broker. The scheme that was chosen was the Dewey Decimal Classification (DDC) and a research license negotiated with its maintainers, OCLC Forest Press [23]. Gateways participating in the Renardus system were expected to map DDC terms to the subject terms used in their own browse hierarchies. In addition, a software tool developed as part of the German CARMEN (Content Analysis, Retrieval and MetaData: Effective Networking) project was adapted to facilitate the mapping workflow.

Mappings were made from the DDC to the range of subject browse structures produced by participating gateways. The Renardus partners use a wide range of different types of subject classification, including general classification schemes (e.g., the Universal Decimal Classification, Nederlandse Basisclassificatie, etc.), subject specific schemes (e.g., Mathematics Subject Classification, Ei, etc.) and some that have been produced specifically for the gateway. Also, some gateways have modified considerably the schemes that they use. The precise amount of mappings that needed to be produced depended upon the complexity of the browse structures that had been implemented by gateways. It was not expected that all terms would have direct equivalents in the DDC. For this reason, a set of mapping relevance levels was defined. These indicate, for example, where there are narrower or broader equivalents or major and minor overlaps.

The cross-browse structures that could be created using this system are relatively limited. It was not possible, for example, to link directly to resource descriptions from multiple gateways in a single browse hierarchy without a greater level of centralisation than that offered by the Renardus pilot. Instead, the Renardus browse system links directly into the subject hierarchies of individual gateways. For example, a user looking for information on violins or violas would first browse through the DDC-based structure on the Renardus system until they find a relevant match (e.g., Stringed Instruments) with its relevance level. If a part of an individual gateway's browse structure has been mapped to this DDC

term, the gateway's name and logo becomes visible and the local subject term becomes a hyperlink to the relevant part of the local browse structure. Once the user follows this link, they leave the Renardus system and join the browse interface of the local gateway.

The classification mapping work remains experimental but is an interesting way of joining the browse structures of distributed services. More details are available in a conference paper produced by members of the working group [24]. The pilot Renardus broker was released for evaluation in the Summer of 2001.

Digital preservation metadata

Digital preservation has been defined by Margaret Hedstrom as "the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable" [25]. It is important to remember that preservation is normally linked with maintaining access, not just long term storage. The reasons why preserving digital information is difficult are technological, related to things like relatively short media lifetimes, obsolete hardware and software, and defunct Web sites [26]. Proposed solutions are partly technical and partly organisational, as Hedstrom's definition suggests. Various digital preservation strategies have been proposed; the most widely discussed being based on migration, emulation or technology preservation. Alternatives might include the relatively expensive data recovery programmes that are sometimes known as 'digital archaeology' [27]. Regardless, however, of which particular strategy is adopted, long-term preservation will depend upon the generation and maintenance of metadata that describes the digital information being preserved and enables its interpretation. The rest of this paper will attempt to review some recent initiatives that relate to preservation metadata for digital objects and in particular the metadata specification developed by the Cedars project. Other relevant initiatives have been described in more detail in a conference paper by Day [28]

In technical terms, the successful long-term preservation of digital information will be dependent upon organisations identifying and implementing suitable preservation strategies [29]. If one ignores strategies that involve converting digital information into non-digital forms (e.g. printouts or microforms), at the moment there are three main strategies: data migration, software emulation and technology preservation [30]. None of these options will be a perfect solution for all digital information types and it is assumed that more than one strategy may have to be adopted.

The technology preservation approach involves the preservation of a digital information object together with all of the software and hardware needed to interpret it. This may have an important short-term role for the recovery of data from obsolete storage media and platforms, but is unlikely become a viable long-term strategy. Feeney points out that collection managers who relied only upon this approach would soon end up with "a museum of ageing and incompatible computer hardware" [31].

Emulation strategies are based on the premise that the best way to preserve the functionality and 'look-and-feel' of digital information objects is to preserve its original software and run this on emulators which can mimic the behaviour of obsolete hardware and operating systems. Emulation strategies involve encapsulating a data object together

6

with the application software used to create or interpret it and a description of the required hardware environment - i.e., a specification for an emulator. Jeff Rothenberg says that the emulation approach requires "the development of an annotation scheme that can save ... explanations [of how to open an encapsulation] in a form that will remain human-readable, along with metadata which provide the historical, evidential and administrative context for preserving digital documents" [32].

A more proven preservation strategy is the periodic migration of digital information from one generation of computer technology to a subsequent one. This is currently the most tried-and-tested preservation strategy, and is often combined with some standardisation of formats on 'ingestion' into an archive in order to simplify future migration processes. However, as Seamus Ross points out, data migration inevitably leads to some losses in functionality, accuracy, integrity and usability [33]. For this reason, migration strategies will be dependent upon the creation and maintenance of metadata that records the successive migration processes and helps preserve its authenticity.

We can already see that metadata appears to be an important supporter of any digital preservation policy. Essentially, preservation metadata is all of the various types of data that allow the re-creation and interpretation of the structure and content of digital data after the bits themselves have been preserved [34]. Defined in this way, therefore, it is clear that preservation metadata needs to support a number of related, but distinct, functions. Clifford Lynch, for example, states that within a digital repository, "... metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information" [35]. The wide range of functions that preservation metadata aims to fulfil means that the production of metadata standards is not simple and that most of the currently published schemas are relatively complicated.

The Cedars project and the OAIS model

UKOLN participated in the development of a proposed preservation metadata specification as part the Cedars project [36]. Cedars was a Consortium of University Research Libraries (CURL) project funded by the JISC between 1998 and 2002 to investigate digital preservation issues and make recommendations on good practice [37]. A major strand in Cedars concerned metadata; the project produced a review of existing preservation metadata initiatives and, in March 2000, a draft outline metadata specification [38]. The Cedars specification had two main aims. Firstly, that it could be used within demonstrator services being developed elsewhere in the project. Secondly, that it would make contribution to the international standardisation of preservation metadata. Work on developing the Cedars metadata specification started in early 1999. An initial draft (for expert comment) was published in January 1999, and was broadly organised according to the information model provided in the influential *Reference Model for an Open Archival Information System (OAIS)* published by the Consultative Committee on Space Data Systems (CCSDS).

The development of the OAIS model resulted from a request from the International Organization for Standardization (ISO) that the CCSDS should co-ordinate the development of standards to support the long-term preservation of digital information obtained from observations of the terrestrial and space environments. The latest version

of the standard (CCSDS 650.0-R-2) is currently a CCSDS Draft Recommendation, and has been recently accepted as ISO DIS (Draft International Standard) 14721 [39].

The OAIS model aims to provide a common framework that can be used to help understand archival challenges, and especially those that relate to digital information. This is the model's real value: providing a high-level common language that can facilitate discussion across the different communities interested in digital preservation. The document defines a high-level reference model for an OAIS, which is defined as an organisation of people and systems that have "accepted the responsibility to preserve information and make it available for a Designated Community" [40]. The OAIS model has a much wider scope than metadata. It defines both a functional model and an information model. The functional model outlines the range of functions which would need to be undertaken by a repository, and defines in more detail those functions described within the OAIS specification as access, administration, archival storage, data management, ingest and preservation planning. The information model defines the broad types of information (or metadata) which would be required in order to preserve and access the information stored in a repository. However, it is important to realise that the OAIS standard is a reference model, not a detailed specification for any implementation based upon it. All of the different communities interested in digital preservation will have to apply the model (including the information model) in their own particular contexts, both organisational and technical.

The OAIS information model defines a number of different Information Objects covering the various types of information required for long term preservation. A basic assumption of the model is that all Information Objects are composed of a Data Object - typically a sequence of bits for digital data - and the Representation Information that permits the full interpretation of the Data Object into meaningful information. The OAIS model defines four distinct Information Objects.

- *Content Information* - the information that requires preservation.

- *Preservation Description Information (PDI)* - any information which allows the understanding of the Content Information over an indefinite period of time.

- *Packaging Information* - the information that binds all other components into a specific medium.

- *Descriptive Information* - information that helps users to locate and access information of potential interest. This could be based on information that is stored as part of the PDI, but is logically distinct.

The OAIS information model sub-divides the PDI into four distinct groupings, based on categories discussed in the 1996 report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access (CPA) and the Research Libraries Group (RLG). The task force wrote that "in the digital environment, the features that determine information integrity and deserve special attention for archival purposes include the following: content, fixity, reference, provenance and context" [41]. Accordingly, the OAIS taxonomy divides PDI into four: Reference Information, Context Information, Provenance Information and Fixity Information.

- *Reference Information* - any information that helps to identify and describe the Content Information. This would specifically include the unique identifiers used to identify the Content Information within the repository and, where appropriate, basic descriptive-type information which could be extracted to form part or all of the Descriptive Information.

- *Context Information* - defined as information that "documents the relationships of the Content Environment to its environment ... why the Content Information was created, and how it relates to other Content Information objects existing elsewhere" [42]. The CPA/RLG report suggested that 'context' should include information on the technical context of a digital object, but some of this information is assigned in the OAIS model to the Packaging Information.

- *Provenance Information* - information that documents the history of the Content Information. This might include information on its source or origin, any changes that may have taken place (e.g. migrations), and a record of the chain of custody. The CPA/RLG report says that the "assumption underlying the principle of provenance is that the integrity of an information object is partly embodied in tracing from where it came" [43].

- *Fixity Information* - refers to any information that documents the particular authentication mechanisms in use within a particular repository. Changes can either be deliberate or unintentional, but either type would adversely effect the integrity of Content Information.

The OAIS model also defines a conceptual structure for Information Packages. This is viewed as a container that logically encapsulates Content Information and its associated PDI within a single Data Object. Information Packages (IP) are defined for submission (SIP), archival storage (AIP) and dissemination (DIP). Of these, the AIP is the most important for digital preservation, as it contains "all of the qualities needed for permanent, or indefinite, Long Term Preservation or a designated Information Object" [44].

The Cedars project team took the existing draft of the OAIS information model and used it as a broad framework for an outline preservation metadata specification. It is an outline specification because in many cases it only defines the highest levels of the metadata scheme that would be required for any implementation. Also, elements and sub-elements are not specified as being 'mandatory' or 'optional,' but just given a significance level. In accordance with the OAIS's Information Package model, the project team envisaged that resources (Content Information) would be packaged together with its metadata (PDI). The specification focused on defining both the Representation Information which would enable the Content Information Data Object to be understood [45] and the Content Information's associated PDI. Less consideration was given to the specific Representation Information which would be required for the PDI Data Object, or to Packaging or Descriptive Information.

The Cedars project team was aware that the proposed metadata element set would not necessarily support all of the roles identified in the OAIS functional model, e.g. the administration or data management functions. Despite this, however, it was recognised that some of the information provided as part of the Provenance Information could help

support administrative functions such as rights management. In fact, the Provenance Information defined in the Cedars outline specification contains a number of elements specific to rights management which *may* go well beyond the OAIS model's assumption that provenance is primarily concerned with supporting the integrity of a given Data Object. This reflects the difficulty of defining simple metadata schemes where the same information can be used by functionally different parts of a system.

A quick look at the hierarchical structure of the Cedars specification demonstrates its basic dependence upon the OAIS information model. The first three levels of the hierarchy inherit the exact terminology and some of the definitions used in the OAIS model.

- The *Reference Information* section of the PDI has elements for a 'Resource Description' and a placeholder for any 'Existing Metadata.' The document doesn't make any specific recommendations as to which precise elements would be included in the 'Resource Description,' but notes that any implementation by the project would use an instantiation of the Dublin Core Metadata Element Set (DCMES). In a similar way, the precise way in which 'Existing Metadata' would be stored or utilised is not defined.

- *Context Information* has one sub-element, referring to 'Related Information Objects.' This is supposed to specify information objects that are judged to have a significant relationship to the object being preserved. Again, the precise nature of information that would be required (e.g. an identifier, descriptive information, etc.) is not defined

- *Provenance Information* comprises the largest part of the Cedars metadata specification. The 'History of Origin' sub-section is intended to record the reasons why the object being preserved was created, its custody history before ingest, and to document why it is being preserved. This section also records technical information about the original technical environment of the object and any prerequisites with regard to software, operating systems, etc. A separate section on 'Management History' is supposed to keep information about the process of ingest, and the policies and actions applied to objects since they were added to the repository. A final section on 'Rights Management' incorporates a detailed set of sub-elements to help record and manage the intellectual property rights held in objects, in order to help with providing access to resources preserved in the repository.

- *Fixity Information* contains a single sub-element, 'Authentication Indicator,' which is intended to record mechanisms used to ensure the digital object's authenticity, e.g. digital certificates or a checksum.

The Cedars specification tried not to make too many assumptions about the actual form of the digital objects being preserved or about the 'granularity' of specific objects. It was hoped that the specification would be applicable at any level of granularity, but the authors recognised that the specifics of implementation would be the responsibility of repositories. Also, the specification made no assumptions about which particular preservation strategy would be used, although this may have an impact on which particular elements would be required.

After publication of the outline specification, meetings were held in Birmingham and Cambridge to 'walk-through' the metadata element set with regard to specific resources. These raised many issues related to how the specification should be implemented and with regard to the organisation of metadata handling within a repository. This included questions about who would be responsible for generating this metadata and the relevant workflow.

There has been no serious development of the Cedars outline metadata specification since its publication in 2000. However, Cedars project participants are currently involved in the work of the Preservation Metadata Working Group supported by OCLC Online Computer Library Center and the RLG [46]. The membership of the working group is international, and includes key individuals who were involved in the development of the Cedars metadata specification and others, including those developed by the NEDLIB project [47] and the National Library of Australia (NLA) [48]. The key deliverable of the working group to date has been a white paper, a review of the state-of-the-art in preservation metadata [49]. This includes a summary of the OAIS model, descriptions of the element sets developed by Cedars, NEDLIB and the NLA and an attempt to map between them using the OAIS information model as a general framework. Also, the working group has more recently published a *Recommendation for Content Information*, which provides an expanded conceptual structure for a Content Information package and a set of metadata elements [50]. The recommendation includes elements based on the ones defined in the Cedars, NEDLIB and NLA specifications as well as new elements defined by the working group. Future work will include a recommendation on PDI.

Conclusion

This paper has attempted to describe current progress in a selection of the metadata projects in which UKOLN is currently involved. It has concentrated on two main areas. Firstly, subject gateway co-operation as expressed in the Resource Discovery Network and the Renardus project. Secondly, on the development of preservation metadata standards. These topics are important because subject gateways and digital preservation represent key areas where some of the traditional values of the information profession, with regard to, for example, the selection, description and preservation of resources, have begun to be applied in the context of the Internet.

Acknowledgements

References

1.	Lorcan Dempsey and Rachel Heery, "Metadata: a current view of practice and issues," *Journal of Documentation*, 54 (2), 1998, 145-172.

2.	Lorcan Dempsey, "ROADS to Desire: some UK and other European metadata and resource discovery projects," *D-Lib Magazine*, July/August 1996. http://www.dlib.org/dlib/july96/07dempsey.html

3.	Michael Day, "Resource discovery, interoperability and digital preservation: some aspects of current metadata research and development," *VINE*, 117, 2000, 35-48.

4.	Google: http://www.google.com/

5.	AltaVista: http://www.altavista.com/

6.	Yahoo!: http://www.yahoo.com/

7.	Open Directory Project: http://dmoz.org/

8.	Louis A. Pitschmann, *Building sustainable collections of free third-party Web resources*, Washington, D.C.: Digital Library Federation, Council on Library and Information Resources, June 2001, p. 6. http://www.clir.org/pubs/abstract/pub98abst.html

9.	Traugott Koch, "Quality controlled subject gateways: definitions, typologies, empirical overview," *Online Information Review*, 24 (1), 2000, 24-34.

10.	Finnish Virtual Library: http://www.jyu.fi/library/virtuaalikirjasto/

11.	Emma Worsfold, "Subject gateways: fulfilling the DESIRE for knowledge," *Computer Networks and ISDN Systems*, 30 (12-18), 1998, 1479-1489.

12.	John Kirriemuir, Dan Brickley, Susan Welsh, Jon Knight and Martin Hamilton, "Cross-searching subject gateways: the query routing and forward knowledge approach," *D-Lib Magazine*, January 1998. http://www.dlib.org/dlib/january98/01kirriemuir.html

13.	Martin Belcher, Virginia Knight and Emma Place, eds., *DESIRE Information Gateways Handbook*, DESIRE II project deliverable D3.4, Bristol: Institute for Learning and Research Technology, 1999. http://www.desire.org/html/subjectgateways/handbook/

14.	Resource Discovery Network (RDN): http://www.rdn.ac.uk/

15.	Lorcan Dempsey, "The subject gateway: experiences and issues based on the emergence of the Resource Discovery Network," *Online Information Review*, 24 (1), 2000, 8-23.

16.	Andy Powell, *An OAI approach to sharing subject gateway content*, poster presented at: WWW10: the Tenth International World Wide Web Conference, Hong Kong, 1-5 May 2001. http://www10.org/cdrom/posters/contents.html

17.	Open Archives Initiative: http://www.openarchives.org/

18.	Renardus project: http://www.renardus.org/

19	Lesly Huxley, "Follow the fox to Renardus: an academic subject gateway service for Europe," in: José Borbinha and Thomas Baker, eds., *Research and advanced technology for digital libraries: 4th European Conference, ECDL 2000, Lisbon,*

Portugal, (Lecture Notes in Computer Science, 1923), Berlin: Springer-Verlag, 2000, 395-398.

20.   Rachel Heery, Leona Carpenter and Michael Day, "Renardus project developments and the wider digital library context," *D-Lib Magazine*, 7 (4), April 2001. http://www.dlib.org/dlib/april01/heery/04heery.html

21.   Renardus broker: http://renardus.lub.lu.se/

22.   Heike Neuroth and Traugott Koch, "Metadata mapping and application profiles: approaches to providing the cross-searching of heterogeneous resources in the EU project Renardus," in: *DC-2001: proceedings of the International Conference on Dublin Core and Metadata Applications 2001, National Institute of Informatics, October 24-26, 2001, Tokyo, Japan*, Tokyo: National Institute of Informatics, 2001, 122-129. http://www.nii.ac.jp/dc2001/proceedings/

23.   OCLC Forest Press, a division of OCLC Online Computer Library Center, Inc.: http://www.oclc.org/dewey/index.htm

24.   Traugott Koch, Heike Neuroth and Michael Day, *Renardus: Cross-browsing European subject gateways via a common classification system (DDC)*, paper presented at: IFLA satellite meeting - Subject Retrieval in a Networked Environment, OCLC, Dublin, Ohio, USA, 14-16 August 2001. http://www.ukoln.ac.uk/metadata/renardus/papers/ifla-satellite/

25.   Margaret Hedstrom, "Digital preservation: a time bomb for digital libraries," *Computers and the Humanities*, 31, 1998, 189-202.

26.   Su-Shing Chen, "The paradox of digital preservation," *Computer*, 34 (3), March 2001, 24-28.

27.   Seamus Ross and Ann Gow, *Digital archaeology: rescuing neglected and damaged data resources*, London: South Bank University, Library Information Technology Centre, 1999. http://www.hatii.arts.gla.ac.uk/Projects/BrLibrary/

28.   Michael Day, "Metadata for digital preservation: a review of recent developments," in: Panos Constantopoulos and Ingeborg T. Sølvberg, eds., *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001, Darmstadt, Germany, September 4-9, 2001*, (Lecture Notes in Computer Science, 2163), Berlin: Springer-Verlag, 2001, 161-172.

29.   Neil Beagrie and Daniel Greenstein, *A strategic policy framework for creating and preserving digital collections*, London: South Bank University, Library Information Technology Centre, 1999.

30.   Seamus Ross, "Consensus, communication and collaboration: fostering multidisciplinary co-operation in electronic records," in: *Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996*, Luxembourg: Office for Official Publications of the European Communities, 1997, 330-336.

31.   Mary Feeny, *Digital culture: maximising the nation's investment*, London: National Preservation Office, 1999, p. 42.

32.   Jeff Rothenberg, *Avoiding technological quicksand: finding a viable technical foundation for digital preservation*, Washington, D.C.: Council on Library and Information Resources, 1999. http://www.clir.org/pubs/abstract/pub77.html

33.   Seamus Ross, "Consensus, communication and collaboration," p. 331.

34.   Bertram Ludäscher, Richard Marciano and Reagan Moore, "Preservation of digital data with self-validating, self-instantiating knowledge-based archives," *SIGMOD Record*, 30 (3), 2001, 54-63. http://www.acm.org/sigmod/record/issues/0109/index.html

35.   Clifford Lynch, "Canonicalization: a fundamental tool to facilitate preservation and management of digital information," *D-Lib Magazine*, 5 (9), September 1999. http://www.dlib.org/dlib/september99/09lynch.html

36.   Cedars project: http://www.leeds.ac.uk/cedars/

37.   Kelly Russell, "Digital preservation and the Cedars project experience," *New Review of Academic Librarianship*, 6, 2000, 139-154.

38.   Kelly Russell, Derek Sergeant, Andy Stone, Ellis Weinberger and Michael Day, *Metadata for digital preservation: the Cedars project outline specification*, Leeds: Cedars project, 2000. http://www.leeds.ac.uk/cedars/metadata.html

39.   Consultative Committee for Space Data Systems (CCSDS), *Reference model for an Open Archival Information System (OAIS)*, CCSDS 650.0-R-2, Red Book, Issue 2, July 2001. http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

40.   CCSDS, *Reference model for an OAIS*, p. 1-11.

41.   *Preserving digital information: report of the Task Force on Archiving of Digital Information*, Washington, D.C.: Commission on Preservation and Access, 1996, p.12. http://www.rlg.org/ArchTF/

42.   CCSDS, *Reference model for an OAIS*, p. 4-28.

43.   *Preserving digital information*, p. 16.

44.   CCSDS, *Reference model for an OAIS*, p. 4-33.

45.   David Holdsworth and Derek M. Sergeant, *A blueprint for Representation Information in the OAIS model*, paper presented at: 8th NASA Goddard Space Flight Center Conference on Mass Storage Systems and Technologies and 17th IEEE Symposium on Mass Storage Systems, College Park, Md., USA, 27-30 March 2000. http://esdis-it.gsfc.nasa.gov/MSST/conf2000/

46.   OCLC/RLG Preservation Metadata Working Group: http://www.oclc.org/research/pmwg/

47.   Catherine Lupovici and Julien Masanès, *Metadata for the long term preservation of electronic publications*, (NEDLIB report series, 2), The Hague: Koninklijke Bibliotheek, 2000. http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm

48.   National Library of Australia, *Preservation metadata for digital collections: exposure draft*. Canberra: National Library of Australia, 1999. http://www.nla.gov.au/preserve/pmeta.html

49.   OCLC/RLG Preservation Metadata Working Group, *Preservation metadata for digital objects: a review of the state of the art*, Dublin, Ohio: OCLC Online Computer Library Center, January 2001. http://www.oclc.org/research/pmwg/documents.shtm

50.   OCLC/RLG Preservation Metadata Working Group, *A recommendation for content information*, Dublin, Ohio: OCLC Online Computer Library Center, October 2001. http://www.oclc.org/research/pmwg/documents.shtm