

Open Archives Forum , Berlin 27/29 March 2003

How Real Archivists can learn to love the OAI

A review of the potential for using the Open Archives Initiative Metadata Harvesting Protocol in conventional archives

Report by George MacKenzie and Göran Kristiansson

Abstract

The OA Forum has commissioned this report from practising archivists in the UK and Sweden. It looks at the potential for using the Open Archives Initiative Protocol for Metadata Harvesting as a simple means of disseminating and exchanging archive catalogues. The world of conventional archives is interested in exchanging metadata, and has widely adopted international data structure standards produced by the International Council on Archives. It has also shown interest in a system for encoding catalogues known as Encoded Archival Description, or EAD. Archive descriptions are complex and collection based, proceeding from the general (fonds or collection level) to the particular (item level). The report briefly examines two implementations of OAI, the University of Illinois Project and the AIM25 project in the UK. It also considers a related hybrid implementation in Australia, and a planned use of the protocol in A2A, another UK project. It observes that OAI can be used for exchanging archive descriptions, but that there are problems. The first is difficulty in accurately reflecting linkages between levels of description. The second is the inconsistent application of EAD. The report also looks briefly at alternative means archivists are using for exchanging metadata, particularly the Z39.50 protocol. The report concludes that OAI will be used by conventional archives only if three conditions are fulfilled. First, archivists must be confident that compliant descriptions will respect archival principles, second, descriptions must be produced with little effort from existing systems, and third, archivists must believe that the wider OAI user base contains sufficient numbers of potential users. It is also important that any sensitive intellectual property rights in the catalogue descriptions are protected. The report suggests possible strategies in which archives would produce OAI compliant records for parts of their descriptions only.

George MacKenzie is
Keeper of the Records of Scotland,
head of the National Archives of Scotland, Edinburgh, UK.

Göran Kristiansson is
senior archivist at the regional archives,
Lund, Sweden

1. Introduction:

1.1 The Open Archives Initiative is a misleading title; it is an initiative certainly, but it's not necessarily open and it has little to do with archives. Most archivists react with surprise to the use of the word archives in the OAI, and some react with irritation, because it is quite different from their own definition of the term. Some also confuse it with the OAIIS, the Open Archival Information Systems model for long-term preservation of digital materials. Peter Hirtle gives a good explanation of the problem and the confusion in an editorial in D-Lib Magazine.¹ In fact, the OAI developed out of a desire by those involved in scholarly electronic prints, to share metadata about these prints widely. It deliberately avoids the model of distributed searching across many servers, such as would be done by the Z39.50 protocol, in favour of servers provide metadata in bulk which can then be harvested and re-used. It aims to be a simple interface, with very low entry barriers, and to shift the responsibility for implementation away from the holders of metadata (known as data providers) and on to those agencies who wish to offer aggregated services (known as service providers). The result is a metadata harvesting protocol which is now attracting increasing interest in the digital library world. It uses the Dublin Core standard as a mandatory, lowest common denominator, but this does not prevent data providers offering other metadata schemes, or service providers requesting in other schemes. It is entirely neutral about the content to be described, or how information is to flow, and so can be applied to almost any type of metadata provided it can be encoded in the eXtensible Mark-up Language, XML.²

1.2 The adoption of the term archives in the OAI, initially by the scholarly e-print community and now more widely, is not itself the problem, but it is important to understand what archivists mean by it. A generally agreed definition would be "documents created or accumulated by a person or organisation in the course of their business and preserved because of their continuing value". Archives are not, therefore, simply objects that have been stored, they are unique documents with intrinsic value. Archives have value not only for the information they contain, but also as evidence, and this underlies the descriptive systems and standards that archivists use.

1.3 Archivists are relative latecomers to standardised description. They have been much slower to develop tools and processes than the library world. This is largely because archives are by definition unique and arranging and describing them is therefore much more complex than for books or serials. Until fairly recently, archives described their holdings in individual ways, using locally determined rules. This did not matter, since users of archives had to make physical visits to see the records, and could have the house catalogues explained by the archivists. However, as electronic communications have spread, archivists have seen the desirability of exchanging and disseminating data, and have recognised the need to have standardised tools to do so. For users consulting catalogues over the Internet, there is no archivist available to mediate and explain the local rules. In the United States, where archives are often

¹ D-Lib Magazine, April 2001 <http://www.dlib.org/dlib/april01/04editorial.html>

² See also Clifford Lynch "Metadata Harvesting and the Open Archives Initiative" ARL Bimonthly Report 217 (August 2001) <http://www.arl.org/newsltr/217/mhp.html>

associated with library services, archivists using computers for descriptions began to adopt library type standards and adapt them. The USMARC-AMC standard was adopted by the Society of American Archivists in the mid 1980s and is now used fairly extensively in the United States, mainly to describe collections of archives, rather than individual items³.

1.4 The International Council on Archives (ICA) has developed two data structure standards for archive descriptions⁴: the International Standard for Archival Description (General) or ISAD (G), which is a general standard and the International Standard for Archival Authority Records (Corporate, Personal and Family) or ISAAR (CPF) which is a standard for describing the organisations and individuals who create records. Both are essentially sets of data elements, and give little or no guidance on the content. They have been widely accepted and are now in common but not comprehensive use in Europe and elsewhere in the world. The European Archive Network (EUAN) project, funded by the European Commission, examined interoperability problems for archives across national borders and identified geographic, cultural and language barriers. It concluded that the two ICA standards were fundamental to international description, and were an essential step in any systems for international data exchange⁵.

1.5 In a parallel move, a document type definition (DTD) for encoding archival descriptions has been developed by a group of interested scholars led by Daniel Pitti. Encoded Archival Description (EAD) matches the ISAD data elements and provides archivists with a means of encoding their finding aids. EAD is in fairly wide use in the archive world, especially in North America. It is now maintained by the Network Development and MARC Standards Office of the Library of Congress in partnership with the Society of American Archivists. In December 2002, the release of a new edition of EAD was announced⁶. The EAD website currently lists 56 implementers, mainly universities in the United States, and 26 further co-operative sites.⁷ Many archive institutions in the United States have begun to aggregate their finding aids using EAD; the Online Archive of California, Texas Archival Resources Online, and in the UK, the Access to Archives (A2A) project.

1.6 The development of aggregated services of this kind indicates the considerable interest in the archive world in interoperability. The OAI may have attractions in this respect, particularly since there is considerable archive material held outside archive institutions, in libraries and museums for example. These institutions do not normally use full archival systems for description, and instead modify their own object-based systems, with varying degrees of success. A metadata protocol that is used across the whole sector offers the possibility of exchanging metadata on archives wherever they are preserved.

2. Context: How archives describe their holdings

³ Victoria Irons Walch, Standards for Archival Description (SAA, 1994) on-line version at <http://www.archivists.org/catalog/stds99/chapter3.html>

⁴ See <http://www.ica.org/biblio.php?pbodycode=CDS&ppubtype=pub&plangue=eng>

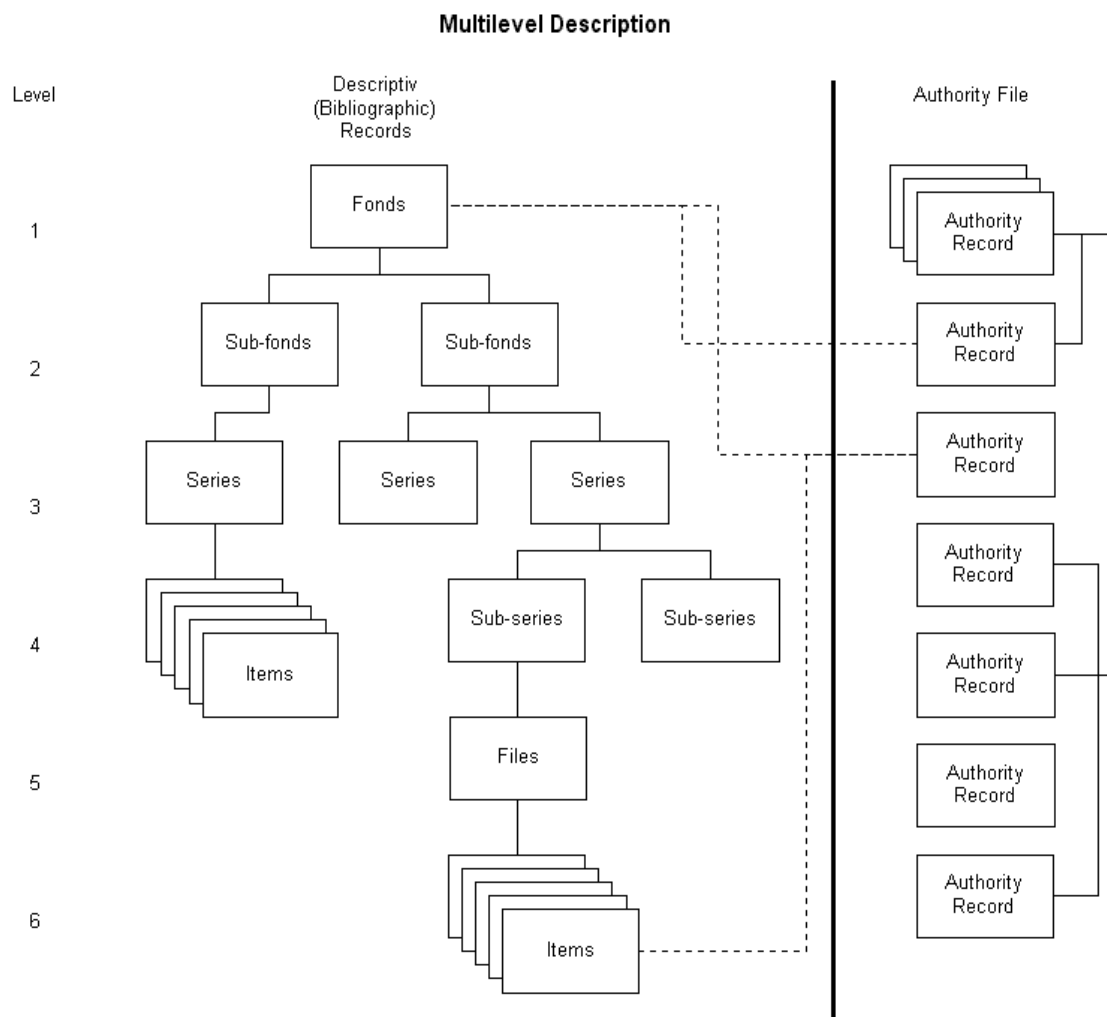
⁵ See <http://www.euan.org/>

⁶ See <http://xml.coverpages.org/ni2002-12-20-a.html>

⁷ See <http://www.loc.gov/ead/>

2.1 Archival Description Systems

The aim of all archival description is simple. It is to identify and to explain the archive materials and make them accessible to the user. All the complex rules developed by archivists, collectively and individually, are in the end based on this aim. The essential idea underlying the way archivists arrange and describe their holdings is context. What this means is that you cannot fully understand the content of a document unless you know about the context of its creation. Unlike a book, or a serial, or even a museum object, an archive document cannot simply be described on its own, because its value and its significance, and our understanding of it, increase when it is associated with similar documents. Archivists regard the provenance of documents as of fundamental importance and seek to group all the documents relating to an individual or an organisation together. They also seek to respect and retain the original order of the documents in order to explain further how they were created, stored and used. This concept, often known as the principle of provenance and original order, underlies all archival theory and practice. From this, it follows that archive descriptions involve a separation of content information and context information. Information about what is in a record is separated from information about how the record was created, by whom, when and why. This principle is clearly implemented in the ISAD and ISAAR standards, and is illustrated in the figure given below.



Source: H. Stibbe *Archivaria* 34, p. 122

2.2 The second important thing about archival description is that it proceeds from the general to the specific, from the collective to the individual, from the collection or fonds to the item. This has three important implications, which are incorporated in rules for using the ISAD(G) standard. The first rule is that descriptive information must be relevant to the level of description. The common attributes of a fonds (collection), or of a sub- are therefore given only at that level, while the attributes of a series or file are given at those levels. The second rule is that it is vital to link each level of description with the next highest level. This will both make the position of the unit being described clear in the hierarchy, and provide the necessary link to contextual information. The third rule is that information should not be repeated at different levels; rather it must be given at the highest relevant level and the links between levels will help the user to see the wider picture. The figure below illustrates the different hierarchies in description, and the links with authority files or contextual descriptions. There is further illustration of archival descriptive systems in the examples of searching in the UIUC catalogue, given in Annex 2.

2.3 To illustrate context in description further, consider the following example, which is from one of the collections in the National Archives of Scotland relating to events in the Island of Skye in the late 19 century. Using the public access catalogue to search on the place name Braes, the following reference appears as one of many hits.

GD1/36/1/1	Correspondence of Sheriff Ivory with Peter Speirs, Sheriff Substitute, Joshua MacLennan, Procurator Fiscal of Portree and others, concerning the rent strike at Braes on Skye and related intimidation of tenants.	1881- Jan 1882
------------	---	----------------------

This is helpful, but it does not explain what these papers are, or where they came from. By further clicking on the left hand side of the screen, the following additional contextual information is made available, which shows which collection the papers are from and who created it.

Miscellaneous small collections of family, business and other papers	
	Papers of Sheriff William Ivory (1825-1915)
	Correspondence of Sheriff William Ivory (1881-1896)
CountryCode	GB
RepCode	234
Repository	National Archives of Scotland
Reference	GD1/36/1/1
Title	Correspondence of Sheriff Ivory with Peter Speirs, Sheriff Substitute, Joshua MacLennan, Procurator Fiscal of Portree and others, concerning the rent strike at Braes on Skye and related intimidation of tenants.
Date	1881-Jan 1882
Extent	4 items

2.4 The archival approach to description can be further illuminated by looking at another typical implementation, in the Swedish archives. The new Arkis 2 system being implemented in the Swedish National Archives, has a relational database using SQL Server and uses EAD as an exchange format. Once XML is more widely used and supported by web browsers, EAD will be used as an output format for displaying Arkis 2 searches on the Riksarkivet website.

The Riksarkivet set up their National Archive Database (NAD) in 1990, covering the holdings of the national archives and the provincial archives in Sweden. In 1993 it received a boost, with the provision of 1,000 young unemployed people to work on it, under a government scheme. An early principle was not to re-invent things, and accordingly the MARC-AMC standard was adopted, not to create records in a MARC system but to use it as an exchange format.

2.5 The data model for the Arkis 2 system, which is given in Annex 1, shows the relations of the different parts to each other, distinguishing the intellectual or knowledge level from the physical or operational level. Arkis 2, unlike its predecessor Arkis 1, allows true multi-level descriptions. It has been designed to cover all archival activities, including, for example, locations of documents. Arkis 2 has, from the outset, been designed as an Internet available service. During the development phase, EAD emerged and its value was quickly recognised. It is used in the same way as MARC-AMC was used, as a means of tagging data elements in the system to allow the export of archival information. In their application, some parts of the descriptions are tagged using a simple XML editor, both to give formatting instructions (boldface, italic, etc.) and to indicate certain descriptive elements (names within scope and content etc.) Arkis 2 can display multiple levels of information, including the automatic construction of an organisation chart, based on the descriptive levels. The display follows the format used in Windows file trees. This allows users to see how the levels of description are derived from organisational levels, in a graphical way.

2.6 Context is an essential part of archival description. It is important to understanding the content of a document or series of documents when they are described in a paper based system. It is crucial to understanding when we move to electronic descriptive systems because the relationships which were explicit in the paper system become implicit and less visible. When searching moves to other domains, and information from archives is combined with that from other institutions, context becomes critical.

2.7 The Dublin Core standard which forms the mandatory, lowest common denominator for the OAI protocol is not used for archive cataloguing because it is not capable of reproducing the richness of archive descriptions and in particular to show the necessary levels and relationships. While no archives use it as a native description format, the EUAN project identified its potential for indexing and describing web pages.

3. Who is Using OAI in Archives?

3.1 Very few archive organisations know or use OAI. The list of OAI repositories contains only two archive implementations. These are briefly surveyed below, together with two further examples. It is important to stress that these sites (and archives generally) provide access to catalogue information only and not to digital materials.

3.1.1 The University of Illinois at Urbana-Champaign (UIUC). This is the biggest implementation, which is one of 7 institutions funded by the Mellon Foundation to carry out research into interoperability and the OAI. UIUC is an OAI service provider, taking descriptive metadata from around 40 institutions, mainly American university and research libraries. Its contributing institutions are also, in some cases, aggregating descriptive data from other agencies, for example the Online Archives of California, which brings together over 60 archival and other bodies holding archives. Another contributor to UIUC is the UK based AIM25 project, which is separately described below. Examples of searches in both UIUC and AIM25 are given in Annex 2. The metadata contributed ranges in size from a few tens of records to nearly a million, and the materials described range from museum artefacts to archival documents.

3.1.2 UIUC is particularly investigating conversion from EAD encoded descriptions to Dublin Core for exposure to OAI harvesters. Their conclusion is that there are difficult, but not insurmountable barriers to doing so. The barriers lie mainly in the inconsistency with which archivists have employed EAD, which in turn are a result of its relatively relaxed rules and permissive model. UIUC researchers found that many of the descriptions encoded in EAD were not well enough structured to allow them to be searched properly in an OAI environment. The project points out that "It is not surprising that the EAD records analysed for the Illinois OAI-PMH project reveal many differences in tag structures and encoding patterns between institutions."⁸ Based on their research, project archivist Christopher Prom challenges archivists whether the long term benefits in interoperability are worth their efforts in doing the encoding into EAD. However, he goes on to suggest, paradoxically, that OAI may provide the incentive to overcome the problems of inconsistency. If it is used as a front end to EAD metadata, he argues, OAI records could possibly mitigate the encoding differences found between institutions and between the finding aids of different cataloguers⁹. Prom and colleagues also found barriers when attempting to transfer the different hierarchies in archival descriptions from EAD to OAI. To do so they had to create many OAI records from one EAD one, with links between them to indicate the hierarchies. This resulted in ballooning file sizes and many empty records. Their solution is to eliminate the listing of most levels of the EAD hierarchy in the OAI

⁸ The Illinois Open Archives Initiative Metadata Harvesting Experience, Timothy W. Cole, Joanne Kaczmarek, Paul F. Marty, Christopher J. Prom, Beth Sandore, and Sarah Shreeves, Museums and the Web 2002 <http://www.archimuse.com/mw2002/papers/cole/cole.html>

⁹ Christopher Prom "Does EAD Play well with other metadata Standards? : Searching and Retrieving EAD using the OAI protocols" draft unpublished article consulted on UIUC website.

record, which means the searcher will need to rely on a link to the full finding aid in order to view the all important context¹⁰.

3.1.3 AIM25 This UK based project covering a range of archive repositories in London has provided OAI compliant descriptions to UIUC. These have been produced in a simpler fashion than in most of the UIUC examples. AIM25 holds its descriptions in a database and can export them in EAD or other format as required. They did not find OAI compliance difficult to achieve, but found it difficult to reproduce the linkages between archive material from the same collection, or between material created by the same person or organisation held at different repositories.

3.2 In Australia there is interest in the library world in using OAI, and one archive related project, Bright Sparcs, which provides biographical and name authority information on Australian scientists, is involved in work with the National Library of Australia (NLA). Bright Sparcs brings together information on over 4,000 people involved in the development of science, technology and medicine in Australia, from archives and libraries. It is therefore not a typical archive site, though it does illustrate better the potential of using OAI for name authorities. Bright Sparcs provides Dublin Core compliant descriptions of its pages, and these are harvested by the NLA and mounted on their site.

3.3 Another UK archive network is planning to use OAI. The Access to Archives (A2A) project brings together catalogue descriptions at collection and item level from a range of participating archive organisations in England, with central editorial control provided by a team based at the Public Record Office. They have around 4 million records captured or planned for capture. They intend to produce OAI compliant descriptions at collection level only, and use these almost as a virtual “signpost” to the fuller, item level description, which will be available through the A2A site. They looked at the UIUC method of producing many OAI records from one EAD record to represent different levels, and linking them together, but felt this would be impractical in their application. Instead they intend to convert their EAD/XML descriptions to OAI via XSL. They expect this to require minimum effort, which is an important consideration in view of the size of their application. They recognise the importance of judging whether users find their approach of having harvested metadata as a signpost to fuller descriptions is helpful¹¹.

4. What Other Ways are Archives approaching Interoperability?

4.1 Archives across the world are increasingly interested in international interoperability, but as with the adoption of standards for description, they are somewhat behind libraries in understanding the potential and developing practical implementations. Once again, the main reason is the uniqueness and diversity of the materials which archives hold. The differences multiply across regional and national boundaries. A number of archives have adopted common portals or network systems, often on a geographic, thematic or institutional basis. These are all examples of an aggregated service which adds value by bringing diverse descriptions together. They

¹⁰ Christopher Prom “Re-engineering Archival Access through the OAI Protocols” draft unpublished article consulted on UIUC website.

¹¹ Information supplied by Bill Stocking, A2A project team, Public Record Office, London, 8 November 2002.

are broadly similar in principle, though not in application, to an OAI service provider. Various US archive networks have already been mentioned above, including the Online Archives of California.

4.2 An example of an institutional based network is the Higher Education Archives Hub in the UK (<http://www.archiveshub.ac.uk/>), which is using Z39.50 in a sophisticated retrieval system known as Cheshire 2. Starting as a pilot project to test concepts in the summer of 1999, it is now moving to a full service stage and by June 2003 expects to have over 20,000 collections available for searching¹². It is currently available from a central server in Manchester, but is being developed as a distributed service, in which contributors will host their own data. The aim is to give the contributors more control over the creation and editing of their descriptions. The Hub does not at present intend to provide OAI compliance, though it could do so, because it offers a fully functioning Z39.50 target, which can provide the full multi-level descriptions archives need¹³. However, they recognise the future potential of metadata harvesting protocols such as the OAI.¹⁴

4.3 One example of a national approach is the Scottish Archive Network (SCAN) (<http://www.scan.org.uk/>), which is bringing together collection level descriptions from 51 Scottish archives and will provide a fully searchable database linked back to item level descriptions in individual archives' sites. This is being achieved, not by automatic harvesting, but by manual collection and editing by archivists. The contents of the central SCAN database will not be visible to search engines, but by providing links and advertising the existence of the site, the aim is to reach users with an interest in Scottish history or archives and encourage them to search. Metadata will also in time be supplied to other UK wide initiatives. Another example is CAIN, the Canadian Archival Information Network (http://www.cain-rcia.ca/cain/e_home.html). This is a network of networks. Individual archive institutions in Canada supply descriptions to their provincial networks, which in turn produce records for the national network database.

4.4 A broadly similar approach was taken in Sweden. In the late 1980s a state committee on archival matters in Sweden concluded that the maintenance of archival information seemed to be very expensive, especially if researchers were not making use of the documents. The key problem was how to locate relevant information. Therefore it was suggested that a National Archival Database (NAD) be constructed to guide researchers towards the information and show where it was stored. The National Archives of Sweden received the necessary funding and the development of the NAD began in the autumn of 1990. The first edition of NAD on CD was produced 1993 and the present one was compiled 1998.

4.4.1 The present edition of NAD on CD-ROM contains:

- General information on about 170 000 fonds (or collections)
- Inventories (or container lists) for around 38 000 fonds.
- Lists of microfiche for loan or sale, particularly parish records
- A historical topographical database

¹² Amanda Hill "Bringing Archives Online through the Archives Hub" *Journal of the Society of Archivists* 23 2 (October 2002).

¹³ Information supplied by project manager, November 2002.

¹⁴ Hill op. cit. p.245.

- A history of regional and local administration

4.4.2 The Internet edition of NAD was produced in 2000 and is available on <http://www.nad.ra.se/>. The Riksarkivet are still in the process of converting old data into the database and expect to have information from the last two institutions, out of a total of 10, in the first half of 2003. The database will then contain more than 6.5 million descriptions.

4.4.3 The NAD consists of descriptions of both public and private archives kept in archives, libraries and museums all over Sweden. It is built up of data imported from different archival databases in central, regional, and local institutions, including the National Register of Private Archives. The descriptions are based on a national implementation of ISAD(G) and ISAAR(CPF) and the data is imported as text files encoded in a MARC format from different hard- and software environments such as UNIX, PC and Macintosh. In the future the XML application EAD and EAC will be used instead as they are better suited for hierarchical archival descriptions and are both based on ISAD(G) and ISAAR(CPF). NAD has been developed to handle both very small fonds with few levels, for example the Swedish General Consul in Calcutta, up to very complex descriptions with thousands of items such as the Swedish Foreign Office. The Archival Information System (ARKIS II) behind NAD can handle archival descriptions of all the different objects that are stored within Archival institutions, such as photographs, digital images, maps and electronic records. Digital images can be linked to descriptions of the original documents and electronic records that have been transferred to the National Archives will be presented on the web in connection with the description. Authority records can be linked to descriptions on all levels for example as creators of a fond, author of a letter, creator of or as a motif in a photograph, in accordance with multi-level description.

4.4.4 In connection with ARKIS II there has been cooperation between the Swedish National Archives and the Royal Library trying to establish a foundation for a national authority database. The aim is to have authority records on corporate bodies, persons and families that can be used by archives, libraries and museums when they describe their holdings. The National Archives is also involved in a European Union project, called Linking and Exploring Authority Files (LEAF), which has the same aim on the European level.

4.5 The EUAN project examined both centralised and distributed means of exchanging data from archive catalogues and concluded that using fonds level descriptions as the basis for advertising archive holdings to users and potential users is a valid one. The project built a simple working model based on harvesting XML descriptions at fonds level and presenting them on a single server. This showed how a system might work as a means of resource discovery. The project did not, ultimately, manage to test the model with a significant volume of international content, nor has it subsequently extended, as envisaged in the business plan, to include a critical mass of partners. However, it did find, in a limited survey of Italian and other European archives, sufficient similarity in the way archive institutions approached description to make a common approach possible, particularly at the fonds level¹⁵.

¹⁵ See final report of project, <http://www.euan.org/>

4.6 The LEAF project started in March 2001, co-funded by the European Commission Information Society Technologies Programme and is developing a model architecture for a distributed search system harvesting existing name authority information aiming at automatically establishing a user needs based common name authority file. While several national name authority files exist in the libraries sector and attempts are currently made to link those national data files with each other, no national or international name authority file is presently used in libraries, archives or museums, and there is no standardized European name record format available.

The LEAF project has three major objectives:

- provide shared access to authority information for all involved (cataloguer, reference librarians, end users etc.);
- improve the quality of existing authorities;
- improve search and retrieval functionalities of a variety of applications.

The methods/steps chosen to reach these objectives are:

- upload distributed authorities to a central system;
- link authorities which refer to the same entity;
- annotate authorities to improve content and provide additional information;
- support external services;
- save search results in a pan-European "Central Name Authority File".

When the user finds an authorized name in the LEAF-database and wants to know more about the information connected to the authority record he or she follows a link to the repository where the information can be seen in context. An important part of the LEAF-project is the XML application Encoded Archival Context (EAC) for "describing the circumstances under which records have been created and used. This includes the identification and characteristics of the persons, organizations, and families who have been the creators, users, or subjects of records, as well as relationships amongst them."¹⁶

4.7 There is not much research available about how people search archives, particularly on the Internet. An evaluation by the government agency Resource examined four networking projects in the UK during 2001 and reported in March 2002¹⁷ It concluded that sites needed to improve their publicity to increase user profiles, that they needed to focus on improving searching and navigation and becoming more user friendly. It also observed that "several users who were positive about the potential of the sites were hindered by their limited understanding of the function of archives, catalogue descriptions and the language used". In one sense this is not surprising, because archives are difficult to explain. But in another sense it represents a challenge to archivists, and it needs to be borne in mind when considering exposing archive descriptions to a wider and by definition less knowledgeable user base. Another conclusion from the study was a strong desire on the part of the evaluators to have a single gateway to the different services.

4.8 There are no agreed business models for how archives will make their holdings available in future, but there is growing realisation that remote access to finding aids, coupled with some form of access to images of the originals, offers a significant way forward. Limited services already exist in the UK aimed at the genealogy market. Typically these offer access to index (finding aids arranged in

¹⁶ More information can be found on <http://www.crxnet.com/leaf/> and <http://www.library.yale.edu/eac/>

¹⁷ See <http://www.resource.gov.uk/information/research/respubs2002.asp - nanurg>

name order) information, from which the customer can then select images to download, on a paying basis. In Scotland, services are provided by the General Register Office¹⁸, and the National Archives of Scotland is a major partner in the SCAN project, providing a range of index and catalogue information and downloads of images of certain documents, including wills¹⁹. The business model for the SCAN project specifically saw catalogues of archives as a means of attracting customers, who would then buy downloads of images. In England, the Public Record Office experienced unprecedented demand for access to images of the 1901 census in January 2002, which forced them to close their website for re-engineering²⁰. As a further example, the national archives of Australia offers remote users the opportunity to browse in the catalogue and select records which are digitised on demand within about two weeks, free of charge.²¹ Archives which are developing digitisation services of this kind are likely to be interested in reaching new markets (whether or not they charge for access to images), and the OAI may be one way of doing so.

4.9 It is less clear how far interoperability with other domains, including archives and museums, can be fitted into new business models. There do not appear to be any examples of this operating in Europe, though in the UK the SCRAN project (quite distinct from SCAN mentioned elsewhere) does provide single portal access to a range of resources from different institutions. It selects and brings together cultural images from historic sites, museums, libraries and archives and makes them available, on a subscription basis, to a largely educational market across the UK²². The success of SCRAN suggests that there is a potential market for images, and for the catalogue data to find them, from across the different domains. This may provide a further future stimulus to use OAI.

4.10 The intellectual property rights in archive catalogues, indexes and other finding aids, will almost always belong to the archive institution because their staff will have created them. Rights in the materials described will in some cases belong to the archive, but in many others will belong to others, either to private owners of collections of documents, or to the copyright owners where ownership of document and intellectual property rights have been split. In the case of archives of public bodies held in public archive institutions, there may be little or no concern about rights in the finding aids. For example the SCAN project positively welcomes the use and re-use of data from its index of Scottish wills. In the case of archives holding documents owned by others, there may be a need to maintain tight control over catalogue information, as failure to do so may impact on the rights of the private owners. Some archivists may also be concerned to keep control over their own catalogue descriptions, and be reluctant to see them re-cycled by third parties, as would happen if they were harvested by service providers. The UK Higher Education Archives Hub considers it important to have controls to protect the rights of archive repositories that contribute data to the system²³. It is likely that the concerns of contributing archives will be highest in relation to the most detailed descriptions, at

¹⁸ See <http://www.scotlandsppeople.gov.uk/>

¹⁹ See <http://www.scottishdocuments.com/>

²⁰ The service is now partially restored. See <http://www.pro.gov.uk/>

²¹ See http://www.naa.gov.au/the_collection/recordsearch.html

²² See <http://www.scran.ac.uk/>

²³ Discussion with Paul Watry, software manager, February 2003.

series and item level, because it is in these that they will generally have invested the most intellectual effort. They may, therefore, be less concerned about OAI compliant metadata at the fonds level.

5. Conclusions

5.1 Few conventional archives yet know about OAI and even fewer are using it. The research for this report has necessarily been limited, but by November 2002, it has revealed only two actual implementations in conventional archives, one planned implementation and one hybrid that bridges archives and libraries.

5.2 Archive repositories typically make catalogue information available on line and on the Internet, but do not normally provide access in digital form to more than a tiny fraction, if any, of the materials described. The archives that have implemented OAI are in this category.

5.3 OAI can be made to work for archives, but there are problems. The first and general one is how to represent adequately the complex relationships and hierarchies in archive descriptions, so that vital contextual information is not lost, or is at least available to the researcher. A second series of problems have been encountered in trying to convert from EAD encoded descriptions to OAI, which relate to the different ways in which EAD has been implemented. Greater prescription in the use of EAD would solve some of these problems.

5.4 The most realistic strategy to us is to use OAI for fonds or collection level descriptions only. There are a number of reasons for this. First, as the EUAN project found²⁴, there is likely to be more similarity of structure and approach at this level of description, making the results from different archives more consistent and usable. Second, and crucially, fonds level descriptions are non hierarchical thus avoiding the difficulty of linking levels, though they will contain less richness of content. These would be seen as a signpost to a fuller description, with a link straight into the database of the holding institution. This is broadly what the A2A project in the UK intend to do. Third, this approach may also reduce concerns about possible loss of control of intellectual property rights in the catalogue data.

5.5 Another related strategy would be to use OAI for name authorities, which are also non hierarchical. This would depend on there being recognised national or international authorities in existence, with links to archive resources. This is the case in Sweden, but not in all European countries. The name authorities will lead a user on to archive or other material. It may be worth asking the LEAF project to look into this.

5.6 What is the potential for OAI in archives? Archivists in our view are only likely to use the protocol if:

1. they are confident that OAI compliant descriptions will respect their multi-level descriptions;
2. they can export data in an OAI compliant way with little or no additional work;

²⁴ See above, paragraph 4.5.

3. they believe that OAI will let them reach new audiences that will be interested in their holdings.

5.6.1 On the first point, the strategies of using OAI for fonds (collection) level descriptions or for name authority files on record creators provide a way forward. The alternatives are to link directly from the derived OAI metadata record to a fuller record, probably a collection level record, as currently happens with the AIM25 descriptions in the UIUC system. The key question then is whether the OAI metadata is sufficiently rich to attract interested users. Research into this is required.

5.6.2 On the second point, the best means of moving from archive descriptions to OAI compliant descriptions seems to be via the EAD DTD. Prom and colleagues at UIUC have investigated this for descriptions in native EAD format and have proved the concept, though identifying problems. Further work to test results from automatically generated EAD is needed, and the institutions of the authors will be working on this in the coming months. If in future an XML schema for archives is developed out of EAD, conversion to OAI will become easier.

5.6.3 The third point may in the end be the main determinant and views on this will probably vary between archive institutions. In higher education institutions and research libraries, the customer base is scholarly and already used to searching in multi-resource databases. In this environment, archives are more likely to use OAI to reach users. In public archives, on the other hand, the customer is most likely to be an amateur or leisure historian, and less used to searching book and e-print resources. In this environment, archives may be less likely to use OAI. On the other hand, the attractions of a low cost standard for metadata dissemination across institutions are considerable. In addition, the potential for reaching new markets may be increasingly interesting to those institutions which are developing business models based on access to catalogues or other finding aids, linked in to image delivery services, as discussed above.

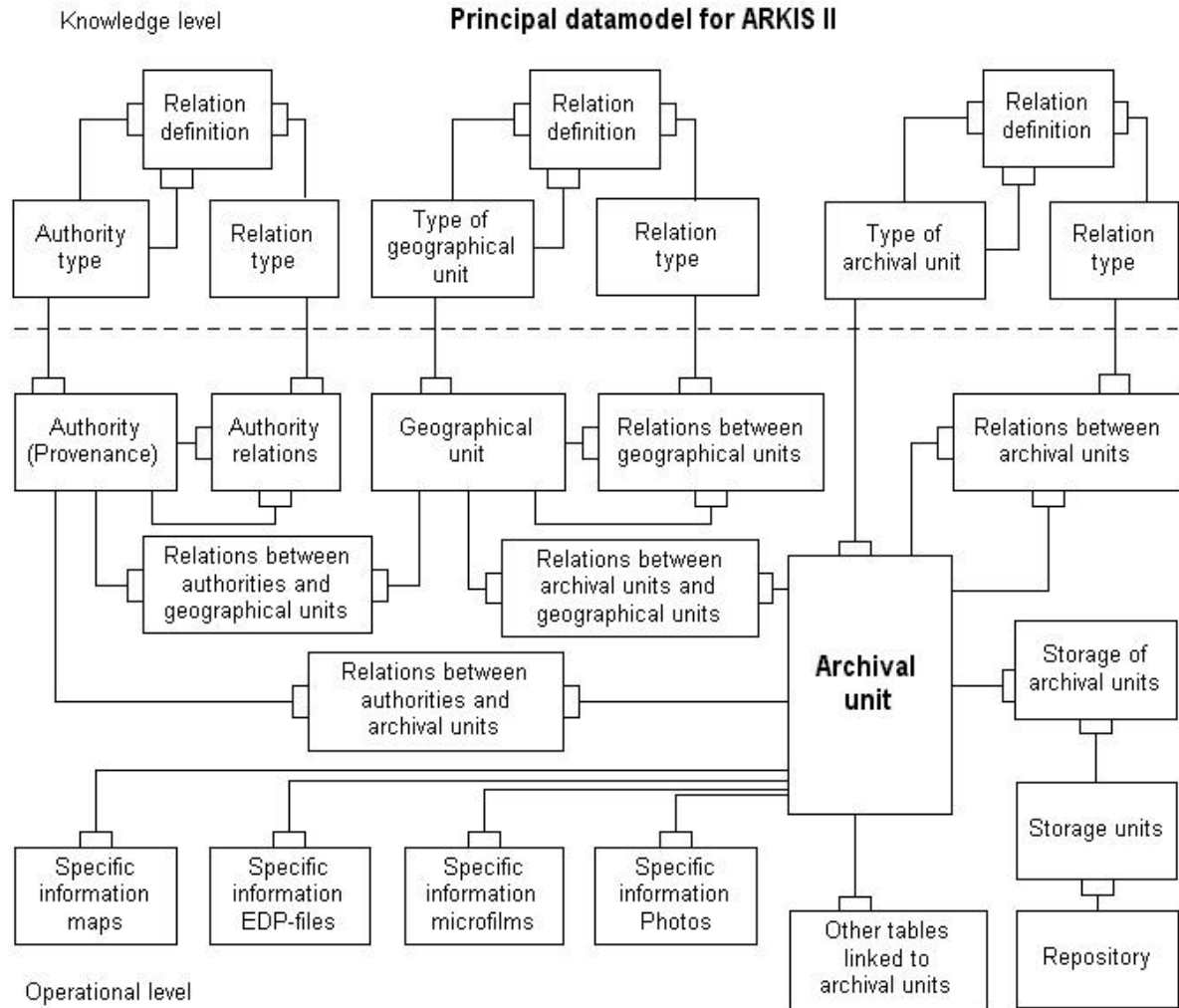
5.7 In the longer term, if interest in OAI continues to grow and it is widely used by varied organisations and domains, there will be immense quantities of information available across different fields. In these circumstances smart systems to let the user navigate and evaluate resources will be needed. The challenge will be whether OAI can provide sufficient functionality to enable users to make informed choices and find the assets they seek, and whether archives can distinguish their materials sufficiently clearly .

George P MacKenzie
National Archives of Scotland
Edinburgh
EH1 3YY
United Kingdom/ Royaume uni
<mailto:george.mackenzie@nas.gov.uk>

Göran Kristiansson
Landsarkivet i Lund
Box 2016

220 02 Lund
Sverige/ Sweden
<mailto:goran.kristiansson@landsarkivet-lund.ra.se>

Revised 6 March 2003



Example of Searching in the UIUC Catalogue

Entering <Edinburgh> in the UIUC search dialogue screen, produces over 1200 hits. These are displayed in pages, alphabetically by name of the holding repository, so AIM25 London comes first, University archives and libraries come at the end. The results displayed vary according to the institutions. One of the more informative is the Online Archive of California. The first display is limited:

Title: Final Drawings
Online access: [View hit in context of finding aid](#)
Collection: [Online Archive of California](#)
[full record](#) | [add to bookbag](#)

This is not in itself helpful, as it gives no indication of what the drawings are of, who they are by, or even how they relate to the search term. Clicking on the [full record](#) button brings a little more information, that the drawings are of the Museum of Edinburgh, 1991.

Clicking on the [View hit in context of finding aid](#) provides more contextual data:

Series 1N. Museum of Edinburgh, Scotland, 1991 See Series 5, for transparencies related to this material. **Competition Phase Design projects: drawings and documents** folder: folder 2 **Museum of Edinburgh, working drawings, ca. spring 1991** Original; plans, axonometrics, elevations; pencil on vellum: 9 drawings, 38 x 50 cm. and smaller; n.s.; German. [tears and masking tape: one badly torn] Roll 31** **4 blueprints, ca. spring 1991** Reproduction; plans, elevations, axonometrics, sections; drawings on vellum with ink notations and photocopied paper adhered; 92 x 112 cm. and smaller; n.s. [pieced together in long strips with clear tape] **Final Drawings** : Flat file folder 12** **2 blueprints, ca. spring 1991** Reproduction; final drawings; plans, elevations, sections, axonometrics; blueprint on paper; 60 x 84 cm. n.s. [pieced together in long strips with clear tape]

and by going to the frame at the left of the screen, additional contextual information is available:

PUBLIC "-//Getty Research Institute for the History of Art and the Humanities::Special Collections and Visual Resources//TEXT (US::CMAIG::920061::Daniel Libeskind Papers, 1968-1992)//EN" "libeskin.sgm" INVENTORY OF THE DANIEL LIBESKIND PAPERS, 1968-1992 Finding aid written by Hellen Lee. The Getty Research Institute for the History of Art and the Humanities Special Collections and Visual Resources 1200 Getty Center Drive, Suite 1100 Los Angeles, California 90049-1688 Phone: (310) 440-7390 Fax: (310) 440-7780 Email: reference@getty.edu URL: <http://www.getty.edu/research/library/> ©1999 Getty Research Institute, Research Library. All rights reserved. Machine-readable finding aid derived from SoftQuad Author/Editor 3.5; encoding by Hellen Lee, Jocelyn Gibbs. Description is in English.

Note: the formatting in the displays from OAC is clearer than above.

This example, from a research library rather than an archive, reflects different descriptive standards. However, the importance of contextual information in understanding the object is still clear.

In another example of a hit on <Edinburgh>, the description follows current archival practice closely. The basic display is as follows:

Title: James Clerk Maxwell Foundation
Author/Artist: James Clerk Maxwell Foundation
Subject/Description: Text of address given by Professor Cyril Domb at the unveiling of a plaque to commemorate James Cler...
[Online Access Available](#)
Collection: [ALM25 Archives in London](#)
[full record](#) | [add to bookbag](#)

Clicking on the Online Access Available button takes the user to the fonds level description, in which considerable contextual information is given.

James Clerk Maxwell Foundation

IDENTITY STATEMENT

Reference code(s): GB 0100 KCLCA K/PP45

Held at: King's College London College Archives

Title: James Clerk Maxwell Foundation

Date(s): 1992-1996

Level of description: Collection (fonds)

Extent: 1 file

Name of creator(s): James Clerk Maxwell Foundation

CONTEXT

Administrative/Biographical history: The James Clerk Maxwell Foundation was launched in 1977 to promote research and education in science and technology. In 1992 the Foundation bought the birthplace of James Clerk Maxwell in Edinburgh, sharing the cost with ICMS (International Centre for Mathematical Sciences), formed by a consortium of Scottish Universities.

CONTENT

Scope and content/abstract: Text of address given by Professor Cyril Domb at the unveiling of a plaque to commemorate James Clerk Maxwell at King College London, 24 Jun 1996; printed leaflet relating to the James Clerk Maxwell Birthplace in Edinburgh, [1992]; photocopy of Examination Papers written by Clerk Maxwell for Smith's Prizes at Cambridge University in Jan 1879, [1992-1996]; photocopies of articles relating to Clerk Maxwell, notably 'The origins of the Clerk (Maxwell) genius' by D O Forfar in the *Bulletin* of the Institute of Mathematics and its Applications, 'James Clerk Maxwell: maker of waves', based on a talk given by Forfar at a conference held at the Royal Society of Scotland on Scotland's mathematical heritage, Jul 1995, and 'Ordering the foundations', by David S Ritchie in *Natural Science*.

The full display includes additional metadata elements which have been omitted above. This is a fonds or collection level description giving the highest level metadata, which explains the context of the item or object found.

These two examples demonstrate firstly, that an OAI compliant basic entry can be used to link to a more detailed description, which can follow conventional archival descriptive rules (as does the AIM25 one) or other, library oriented rules (as does the Gety/OAC one). They also demonstrate, however, that the basic OAI metadata provided may be very thin, and this may not be particularly helpful to a user, especially if a large number of hits are returned. The examples also underline the point that on line access in the case of archive catalogues, will generally be to the catalogue itself, to the metadata, rather than to the archival object or collection. It will still, in the majority of cases, be necessary to travel to the archive or request a copy of a document from the archive.