

Archives UK

Gap Analysis Study

Author:	Joe da Silva
Document ref.:	AUK Gap Analysis 10
Date issued:	2nd June 2005
Version:	Definitive 1.0

Contents

1	Management Summary	3
2	Introduction	5
2.1	Background	5
2.2	Terms of Reference.....	5
2.3	Scope	5
2.4	Approach and Methodology Followed.....	6
3	Findings and Conclusions.....	7
3.1	Key Findings.....	7
3.2	Main Conclusions	9
4	Analysis of Services.....	12
4.1	Introduction	12
4.2	Data Analysis	12
4.3	System Constraints	13
4.4	Interoperability.....	15
5	Key Issues and Risks.....	18
5.1	Specific Issues	18
5.2	Risks to be Addressed.....	18
6	Options and Recommendations.....	19
6.1	Introduction	19
6.2	Outline Options.....	19
6.3	Option Assessment	22
6.4	Key Recommendations	22
6.5	Way Forward	23
7	Appendix A Partner Responses	24
8	Appendix B Glossary	25

1 Management Summary

In light of the failure of the Heritage Lottery Fund bid for the *Linking Arms* initiative, this study was initiated to look at the gap we would need to address in order to provide access to the archives in question and enable the initiative to continue possibly under the banner of *Archives UK*. The essential remit was to be outward facing and recap on the state of the various partners' data and systems, in order to quantify and understand the extent of the interoperability problem to be solved.

The key findings and conclusions can be most succinctly summarised as follows:

- There is good compliance to ISAD (G) archival metadata definition standards, but patchy conformance to other more generic Government standards such as the e-GMS.
- There is good usage of EAD and XML for data import and export. However, in terms of full interoperability, there is some use of Z39.50 but very little evidence of OAI compliance, which is seen as the short- to medium-term way forward for effective data harvesting.

There are a number of key elements required for a complete end-to-end solution: presentation, middleware (including searching) and data. Within the context of the variable but usable nature of partners' data, the evolving Global Search engine, which currently excludes harvesting middleware facilities, the gap is, in summary, the lack of the requisite middleware to allow the Global Search engine – or some such mechanism – to gain access to the partners' data.

Although not explicitly asked of this study, there are two potential options, which are worth considering seriously: use of the Global Search engine; and adoption of the MLA led People's Network Discovery Service (PNDS). Indeed, if the Global Search project does not evolve to include data harvesting, then both options could be used together to provide a total solution.

This could be a recommendation on the proviso that:

- We continue to set the standards agenda but recognise that there may be a need to do a certain amount of data mapping and manipulation centrally as well as locally.
- There would still be a need for some back-end development work even with 'free' middleware, in terms of implementation, configuration and data cleansing. Additionally, software is rarely free, so we could be lining up additional costs certainly in terms of support and possibly also further development / enhancement.
- TNA use open standards, e.g. in developing the Global Search, so that it can either be built on or made to interoperate with external services such as the MLA software. This is particularly important in the connection mechanisms linking the search engine to data held in the Catalogue, A2A and NRA, as solutions here may be able to be applied to the large archival catalogues under consideration in this study.

There are some immediate next steps to maintain momentum, including proactively re-establishing the consortium of partners and some detailed work on costing the various options.

However, there are also some potential issues which need to be addressed, including Global Search scope and timing, proper planning for the operational running of an effective Archives UK service and the need to carry out data mapping work due to the inconsistent implementation of data standards one would expect of such varied data providers.

2 Introduction

2.1 Background

As part of the development of the technical options report for the *Linking Arms* HLF bid an informal survey was conducted of the major catalogue systems making up the National Archives Network. Such intelligence remains important as TNA looks at alternative options that can contribute to the further development and integration of the network, given the failure of the bid. However, the data from the initial survey is now over 15 months old. This study was therefore initiated to undertake a high-level systems analysis – i.e. a gap analysis, or survey – of the current and planned systems within UK archives. This document outlines the results of the study, or gap analysis.

2.2 Terms of Reference

The survey constitutes an update of the previous survey conducted as part of the Linking Arms initiative. It has looked at the following areas for current systems and planned upgrades:

- Data: standards and structures for descriptive metadata, access points and digital objects. Volume of metadata and digital objects, and predictable growth.
- Infrastructure: type and size of systems and interrelationships.
- Interoperability: mechanisms for import and export of metadata and ability to connect via protocols such as Z39.50 and OAI. Potential restrictions (e.g. proprietary systems, data protection, commercial agreements, etc.)
- e-GMS and e-GIF compatibility.

The key objectives can be summarised as:

1. Update the survey conducted over 15 months ago to ensure the conclusions are up to date, in order to inform the analysis exercise.
2. Produce a study report that summarises the gap analysis of where we are now compared to where we want to be with the National Archives Network.

2.3 Scope

The survey has included the following archives:

- AIM25
- Archives Hub
- Archives Network Wales (ANW)
- British Library (MOLCAT and IAMS)
- Commanet
- Guildhall Library (TALIS)

- Hampshire Record Office (CALM catalogue)
- National Archives of Scotland (NAS and SCAN)
- Public Record Office of Northern Ireland (PRONI)
- UK Villages (Community Heritage Store)

The survey has included the following key bodies, in order to get their strategic perspective and an appreciation of some of the key issues:

- Common Information Environment (CIE) – Enhanced Place Demonstrator
- The Museums, Libraries and Archives Council (MLA) – Knowledge Web and People's Network Discovery Service (PNDS)

The survey has not explicitly looked at the following, as these should both be covered by discussions concerning the TNA Global Search project:

- The National Archives (TNA): HMC – National Register of Archives (NRA)
- The National Archives (TNA): PRO (A2A and TNA Online Catalogue)

2.4 Approach and Methodology Followed

The project started with a review of the previous survey report and associated material, so that the information could be updated in light of discussions with the relevant archival bodies above. Most of the discussions were conducted via telephone or Email, supplemented by interviews, both internal and external as follows:

- Derek Breeden (Seamless Flow programme)
- Sonia Ranade (Global Search project)
- Dr Paul Miller (CIE)
- Chris Turner (MLA)
- Phill Purdy (MLA)
- Rachel Stockdale (British Library)

The driver for the discussions was the revised questionnaire, which was honed and concentrated on system characteristics rather than any specific requirements for archive searching.

Once all responses were collated, these were analysed and fed into the production of this study report. The study team (Bill Stocking and Joe da Silva) are grateful for the valuable input provided by all concerned.

3 Findings and Conclusions

3.1 Key Findings

In summary, the major findings from the study are as follows:

- There is widespread adherence to basic archival standards, although their flexible nature means that some data manipulation will always be necessary.
- There is widespread willingness to be involved in any Archives UK initiative.
- There are a number of potential technical options for going forward.
- The potential volume of records has increased from 4M non-TNA records to some 20M, with significant further growth forecast over the next two years.
- The range of data structures is mixed and varied, although they generally conform, or can be mapped, to an ISAD (G) hierarchy.
- The Global Search project is not planning to meet the two Archives UK related requirements as part of its first or second planned phases, namely IDX 04 (Remote content) and TEC 03 (Crawling); these are desirable requirements so may not ensure that the tool selected is extendable to meet the needs of Archives UK.
- An alternative solution may be the MLA-sponsored PNDIS – we have been offered free use of the software as it is open source; although the developed software would be ‘free’ there would still be an ongoing cost to TNA in terms of total lifecycle costs such as enhancement and support.
- Use of the CIE Place Demonstrator has not been considered as a viable option due to cost and lack of business fit. However, we can learn valuable lessons by both retaining an involvement in the CIE initiative and in the Demonstrator through A2A data provision.

The survey concentrated on questions about the current state of play with regard to the systems that Archives UK would be interacting with.

The scope of the Global Search project is as follows:

A key objective of the Public Web Search project within the Seamless Flow programme is to develop and implement a Search Engine, which meets the needs of visitors to the site in terms of usability, accessibility and accuracy, providing a unified, intelligent and fast search of all TNA online content, including born digital content from the Seamless Flow programme.

The Public Web Search project scope will include free and charged online web content and resources for the public. Where possible externally hosted systems such as the 1901 Census service, LIA collections and other UK based online archives collections will also be included. The project scope will exclude on-site resources and content exclusively for TNA and OGD staff.

The aims of the questions were as follows:

- **Data**
Establish approaches to data encoding standards, covering XML, metadata (ISAD (G)), EAD, Dublin Core and authority controls, along with data structure / indexing / classification. ISAD (G) covers data (semantics) elements; EAD covers structure (syntax) elements. This is, therefore, to do with both 'content' and 'structure'.
- **Infrastructure**
Gather key system architectural characteristics.
- **Interoperability**
Examine use of data exchange protocols such as Z39.50, SRW/SRU and OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting). The EAD / XML side of things is dealt with in the structure side of data above.
- **Updating and connecting**
Determine the frequency with which archival content, and therefore catalogues, are updated (also any potential accessibility issues) so as to plan the frequency of updates to Archives UK.
- **General**
General system or service characteristics not covered by any of the above.

Finally, a number of key observations can be made:

CIE and harvesting

- MLA is building a library view onto the CIE via its PNDS, whereas Archives UK could constitute the archive view. A wider cultural heritage view is provided by the Knowledge Web, of which PNDS forms a key part.
- The lessons learnt from the CIE Place Demonstrator could be valuable.
- As previously stated in the Linking Arms Technical Options Study, using a harvesting approach to build up metadata indexes should be the preferred architecture, as opposed to firing off Z39.50 searches – both for performance reasons and the need to manipulate data so that it can be interpreted intelligently.
- There are a number of OAI harvest repository tools around and, once implemented, a data gatherer can be written for each partner system relatively easily; this is the only realistic current solution should the Global Search engine and the PNDS not provide the facility.

MLA and the PNDS

- The PNDS development is managed by Sheffield University with the technical work being done by Knowledge Integration; Futurate are the user interface designers. The system is based on open-source components and will run on MLA managed servers.
- The PNDS is due to go live in August 2005. It will then remain supported until March 2007. Further development during this time and sustainability thereafter will be dependent on MLA resource allocation, but as it forms a key strand of MLA's Knowledge Web vision, it is thought the prospects for future MLA support are good. However, it is also clear that MLA might seek stakeholder contributions so TNA may be liable to ongoing costs.

- MLA is happy for us to use their technology 'free of charge', rather than reinvent it; they see that as a major contribution to the Archives UK initiative. However, this only means the developed open source software, so that TNA would still need to incur implementation and service running costs.
- Realistically, we would need to provide our own interface whether we implement our own instance of the middleware or supply an interface and simply plug into the planned MLA service. Both options would incur development costs.
- Commanet and Archives Hub are already signed up to the PNDS. MLA has also shown interest in the connection of A2A to the PNDS.

Global Search Engine

- Current plans are to go for a data copying approach for data provision by the target systems to be searched.
- Archives UK harvesting requirements need to be put to the Search (Resource Discovery) Board. These requirements are not currently in the project PID so, although they could potentially be part of a future phase, there is as yet no commitment.
- The current project view is that there is further research and planning needed in order to have a complete view of what the target systems look like in detail and how that can then map to the most logical technical approach to follow. The Search project may be a good first step. However, Global Search is just about searching, not data collection – harvesting should be performed by something else, which uses XML, Dublin Core, EAD, etc. to load it.

3.2 Main Conclusions

Standards are needed so that we do not gather meaningless content; these can be summarised as:

- Dublin Core (eGMS).
- ISAD (G) (General International Standard Archival Description).
- EAD (XML representation of ISAD (G))
- OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting – Dublin Core plus a bit); OAI is essentially a repository of a database available at a standard web address, giving (as a minimum) Dublin Core back each time you access it.
- Z39.50 (search protocol); used in the Library sector to define access points for a query (real time to a real database); it is old, pre web / XML and has an obscure record syntax.
- SRW (Search and Retrieve for Web), the Z39.50 replacement – sorts out Z39.50 semantics; new systems should use SRW, with some Z39.50 support for legacy systems.
- Ideal transport is XML over http (standard web), with plenty of development skills available (any Computer Science graduate).

The key point about standards is whether they are being followed *accurately*. Some interpretation will always be needed; this workload should be put at the centre, rather than on the providers.

A number of significant conclusions emerge from the analysis carried out:

- Although there is considerable non-compliance across the board, we need to consider what should be the minimum threshold for viable inclusion in the Archives UK club and the gradual scale of ever-increasing compliance:
 - The bare minimum should be based on ISAD (G), as set out in the NCA (National Council on Archives) Interoperability Protocol for the formal archive bodies; for community archives we would need mapping to some form of qualified Dublin Core
 - Ideally partners should also comply with EAD and XML, as also mandated by the NCA Interoperability Protocol
 - OAI compliance would be the perfect scenario
 - Other areas in the survey really constitute desirables
- In light of the above, there is a reasonable starting point, which suggests that it is feasible to progress with the Archives UK initiative. However, work will be required both centrally and at the source partner systems, in terms of development and data cleansing / normalisation; the latter needs to be minimised so that most of the effort happens centrally.
- None of the options discussed later are free in terms of either effort or cost.
- We need to re-think the strategy in terms of the roles of the People's Network Discovery Service (PNDS) and A2A. A2A could potentially be used as a base for building a harvesting solution onto its data loading element.
- We should seriously consider extending the scope of the Global Search project to raise the profile of the two external facing requirements so that they are incorporated into Phase 2 rather than Phase 3 or beyond.

The following conclusions can be drawn from the answers to the survey questionnaire:

- **Data**
 - General conformance to ISAD (G)
 - Widespread usage of EAD / XML, which would be a strong basis for AUK moving forward
- **Infrastructure**
 - Mix of RDBMSs (including CALM) and other off-the-shelf systems
- **Interoperability**
 - Some possibilities in terms of Z39.50 and SRW/SRU
 - No use of OAI-PMH, other than AIM25, although Commanet are currently being funded by PNDS to develop an OAI repository
- **Updating and connecting**
 - Quite regular, highlighting the potential need for some form of batch loading process

- **General**

- Not really an analysis category as such, but rather informs the above four areas. Some useful information provided, but only really for information at this stage. Would need more detailed discussions but we must be clear on the purpose of this line of analysis

In conclusion, then, the principal gap is the lack of the requisite middleware to allow the Global Search engine – or whatever search mechanism is used – to gain access to partners' data.

4 Analysis of Services

4.1 Introduction

The answers to the survey questionnaire are included at Appendix A. These are summarised below, along with an assessment of the gap in each area, by the following groupings:

- Data (both in terms of content and structure).
- Infrastructure (i.e. system related characteristics).
- Interoperability (and the extent to which we can achieve the desired inter connection).

In addition, we can conclude that various upgrades – or new implementations – are planned over the next year or so.

4.2 Data Analysis

This covers both data content, e.g. ISAD (G) usage, and data structure, e.g. EAD. It is further split into catalogue data and authority control / indexes.

Catalogue data

1. What standards are used for catalogue data?

Widespread use of ISAD (G) and some MARC (Guildhall Library and British Library).

2. How is data in your system structured? What archival levels are represented?

Various levels, with most respecting the ISAD (G) hierarchy:

Fonds / sub-fonds / series / sub-series / sub-sub-series / file / item

Community archives do not conform to ISAD (G) but data should be mapable, so it is not seen as a major issue.

3. Does your catalogue metadata include links to sets of digital surrogates or born-digital versions of the archives themselves?

Sporadic links (e.g. British Library and Hampshire Record Office who have 10,000), but generally 'no'. ANW 'intend' to. This depends on technical and rights issues and, therefore, few links have been made yet; access to digitised images would therefore be minimal to start with.

Authority control and indexes

1. Do you use authority control for catalogue elements or index terms? If so which elements and what types of index terms? If so which standards do you use?

A mixture of usage, including in-house rules. Formal approaches are mainly:

- NAMES (Personal, Corporate, Family) tend to be ISSAR (CPF) – i.e. records about Persons – AACR2 or NCA rules
- PLACES AACR2 or NCA rules
- SUBJECTS LCSH, UNESCO thesaurus or UKAT

This is an issue, as it would be difficult to index according to these parameters due to the mix of rules being adopted.

2. *If you index, at what level(s) of archival description does it take place?*

Across the board, including at all levels, at fonds / collection level or not at all. This could therefore be a problem in terms of search functions, as returns from multiple data sets would not be consistent. Desirable functionality based on 'Who, What, Where', will then be difficult to deliver.

3. *Does your system currently include Archival Authority Records? If so what standards are used? If not, do you have plans to develop these records?*

Not all archives have developed Authority Records.

Assessment:

Difference from last survey	Increased amount of digital records, mostly images Wider use of ISAAR (CPF), UKAT and UNESCO Increasing number of data levels at which indexing occurs
The gap	Still short of full ISAD (G) compliance, especially Commanet, Guildhall Library and UK Villages
Actions required	Encourage the development of compliance to ISAD (G) and NCA Interoperability Protocol

4.3 System Constraints

This covers infrastructure characteristics and helps to build up an architectural picture.

1. *Which catalogue/database system does your online resource use?*

- Relational database management system such as SQLServer, ORACLE, Access?
- Native SGML/XML – if so, which DTD or schema?
- CALM2000? Adlib? MODES? HTML pages?
- Bespoke in-house system?
- Other?

Three users of the proprietary RDBMS CALM 2000 system (Hampshire Record Office, NAS and PRONI), otherwise there is a mixture, including SQL Server and SYBASE.

Some use of EAD SGML / XML files (Archives Hub and planned by British Library IAMS).

2. *Does your system interrelate with any other systems? If so, how?*

Most of the systems are reasonably stand-alone. However, the Talis system includes all the Guildhall Library's metadata and separating that relating to the archives will be an issue; the NAS are building a new Digital Data Archive, which may in time pose problems in providing direct links to digital versions of the material described.

3. *What is the extent of your catalogue/database system?*

- Number of discrete/fonds catalogues?
- Number of file/item descriptions?
- Total number of descriptions at all levels?
- Amount of raw data/disk space?

Wide-ranging volumetrics supplied, in summary (including some guesstimates):

- | | |
|---|------------|
| • Number of discrete/fonds catalogues? | 500,000 |
| • Number of file/item descriptions? | 6,000,000 |
| • Total number of descriptions at all levels? | 12,000,000 |
| • Amount of raw data/disk space? | 8Gb |

This excludes the fact that there may be hundreds more local record offices like Hampshire at city and county level. The Hampshire volumes are:

4,600
231,000
275,000
500Mb

4. *How much growth is expected in these figures in 2 years time?*

Some significant growth forecast.

5. *Are there any plans to change catalogue infrastructure?*

Various plans, including new system implementations (British Library and PRONI) and moves to EAD 2002 (Archives Hub) or MARC21 (Guildhall Library).

6. *Does your catalogue/database system utilise specific metadata structures/schemas for resource discovery?*

- EAD?
- Dublin Core?
- METS?
- HTML <meta>?
- Other?

Although there is good EAD compliance, some do not currently support its use (ANW and British Library MOLCAT); there is further intention to comply with EAD (British Library IAMS) and some Dublin Core (Guildhall Library and PRONI).

7. *How does your electronic cataloguing system encode 'extended character sets'?*

A complete range, including UFT-8, ANSI, UNICODE. This means there may be a need to build specific transformation mechanisms for detailed harvesting (and distributed searching if that ever becomes viable).

8. *Do you hold digital records (surrogates or born-digital records) linked to the metadata? If so, are they held in the same system or separately?*

Any usage is primarily of images. Where used, the digital records are held in a separate database to the metadata. NAS are adding a digital data archive due for completion in March 2007.

9. *What is the extent of linked digital record (surrogates or born-digital records) data sets you hold?*

- Number of images?
- Amount of raw data/disk space?

Not applicable to most but the British Library has 9,000 images (920Mb) and Hampshire Record Office has 10,000 images (500Mb), as well as the planned NAS digital data archive.

Assessment:

Difference from last survey	Various upgrades to EAD 2002 New CALM implementation (PRONI) Increase in projected volumes British Library are planning a new integrated archive manuscripts system
The gap	A mixture of character coding standards
Actions required	Encourage development using open standards

4.4 Interoperability

This covers the various protocols needed for data transfer, i.e. Z39.50, SRW/SRU and OAI-PMH. There is also an overlap with the Data grouping above to address the data transfer aspects of EAD and XML.

1. *Does your catalogue/database system support EAD? (natively or import/export? Which version (1.0 or 2002)? As SGML or as XML?)*

All support EAD (with the exception of British Library MOLCAT and the Guildhall Library), either as XML or SGML, in some cases both import and export. As part of the ongoing enhancement process for CALM, the EAD Special Interest User Group have proposed

EAD import and export routines that can be customised rather than produce generic EAD content, and this has been accepted by the developers.

2. *Does your system support connection via OAI protocol?*

Some intentions (British Library and PRONI), but generally “no” (with the exception of AIM25), although CALM is reported to be planning to do this via a separate module.

3. *Does your system support connection via Z39.50 protocol?*

Some do (Archives Hub and Guildhall Library) whilst others are planning to. There is also a possibility for Hampshire Record Office but there are CALM cost and resource issues. However, in the scheme of things this is a step backwards, and effort should be concentrated on SRW web approaches.

4. *If you have digital records linked to your metadata available online, is it possible to export/harvest or otherwise access such links?*

Some doubts as to whether this can be achieved – only the British Library and Hampshire Record Office are possibles. It may therefore be some time before we can guide users to the object – however, as public money has been spent on digitising these objects, it is something we should strive for in the short- to medium-term.

5. *Are there any restrictions on the interoperability of your data (including digital records) or systems that would prevent extending the TNA global search to your systems (e.g. proprietary systems, data protection, commercial agreements, etc)?*

This could be a problem in some cases, e.g. Archives Hub and some elements of NAS. However, charging for access, as proposed by NAS, should not be an issue.

6. *To what extent do you conform to the e-GMS standard?*

The conclusion seems to be “no” to e-GMS but “yes” to Dublin Core.

7. *Which other elements of the e-GIF standard are you compatible with?*

Generally “no”, especially as there are so few that conform to e-GMS anyway. However, there is a question as to whether we would need adherence to other elements going forward, other than in principal. The general ability to support import / export as EAD, and Dublin Core in XML though, would suggest there is a level of conformance to e-GIF, in fact.

Assessment:

Difference from last survey	AIM25 is both Z39.50 and OAI compliant (the only one)
The gap	Virtually no OAI compliance

Actions required	Sell the benefits of supporting OAI, perhaps by leading by example, e.g. develop an OAI repository for A2A to test out interoperability with PNDs, etc.
------------------	---

In addition, NAS are making good progress with CALM in its use of EAD via the Scottish Interoperability Project (SIP) and this work would be applicable to other CALM users. Commanet are planning a 'pre-harvesting' approach, which includes the ability to carry out full text searching in the harvested version and the generation of special thumbnails.

5 Key Issues and Risks

5.1 Specific Issues

A number of issues arise which need to be managed:

- Timing of TNA Global Search requirements being met for Archives UK and the need to increase the scope of the project sooner rather than later.
- Development cost (out of scope of the study). Whatever option we select there will always be some development costs, which would need to fall to the centre rather than local / community archives.
- Slight potential that, even when standards are being followed, they are not being used accurately enough to enable the creation of a generic data harvesting / gathering approach or mechanism. However, ISAD (G) helps so the problem should be minimal – it will probably affect the search functionality more than the harvesting / gathering, i.e. would only enable basic searching, not 'Who, What, Where.'
- Effort needed by partners to provide data in required format.
- Potential lack of technical resources by partners to provide data as required.
- Effort needed centrally to decode, cleanse and map data / content as required.
- Content rights (IPR) – could be implemented as a licensing mechanism, but be publicly available. Creative Commons could be a suitable approach to follow.

5.2 Risks to be Addressed

There are some potential risks, which the initiative needs to address in order both to prevent and mitigate them:

- Service ownership perception and fear of loss of control by content providers; there is a potential cultural change here, i.e. making it available through the TNA web site! We need to show that this would only be just another route to the content. Although this could be a risk it is not currently an issue. All have signed up to the Linking Arms consortium; we would not be using a TNA web site, just the Global Search technology, potentially. Furthermore, the service can be branded with the consortium signing off the screens, etc. – this was never a problem with A2A as the potential issue was managed and the service suitably branded.
- Potential – perceived or actual – takeover of the archive space by the PNDs.
- Unclear message portrayed about the respective roles of Archives UK and A2A, in particular the idea that A2A could be extended to incorporate the AUK requirements.
- Difficulty in costing the total service given the extent of current knowledge.
- Any hidden agendas by the partners; we need to understand their reasons for wanting this.

6 Options and Recommendations

6.1 Introduction

Recommendations were not expressly demanded of this study. However, a number of potential options for going forward have emerged which it is of benefit to outline in order to inform any further work. Please note, though, that a full options evaluation has not been conducted, rather an indication of our current high-level view of the various options has been given. The options are clearly more complex and matrix in their relationship, in that there is an inter-relationship of pros and cons – for example, a pro of one option is a con of another, and so on. Therefore, the assessment presented below would need further analysis and refinement before more definitive and substantial recommendations can be made.

6.2 Outline Options

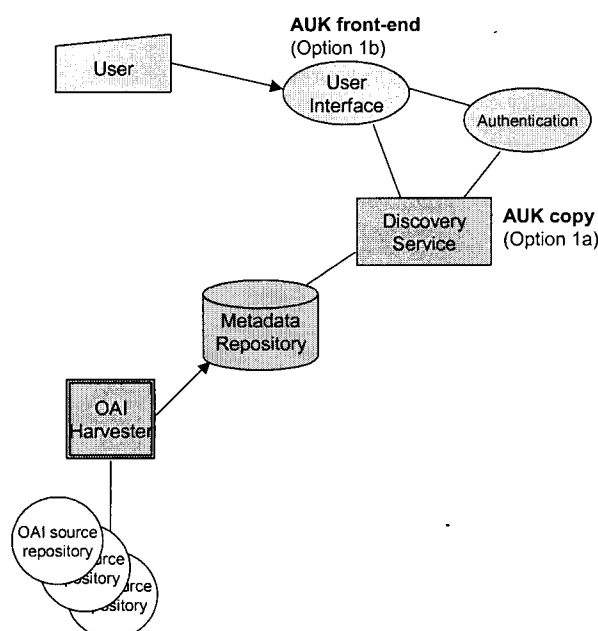
The five broad options available are as follows:

1. Use PNDS
 - 1a. As an instance
 - 1b. Plug in as another front-end
2. Extend Global Search (Phase 2+, depending on when the scope is increased to include the requisite middleware element)
3. Harvesting approach developed by TNA from scratch
 - 3a. Through updates provided by a data gatherer
 - 3b. Through physical data loading (via CD)
4. A direct searching (polling) approach
5. Extension of the A2A service (either as a Harvester and / or an OAI repository in a PNDS or other scenario)

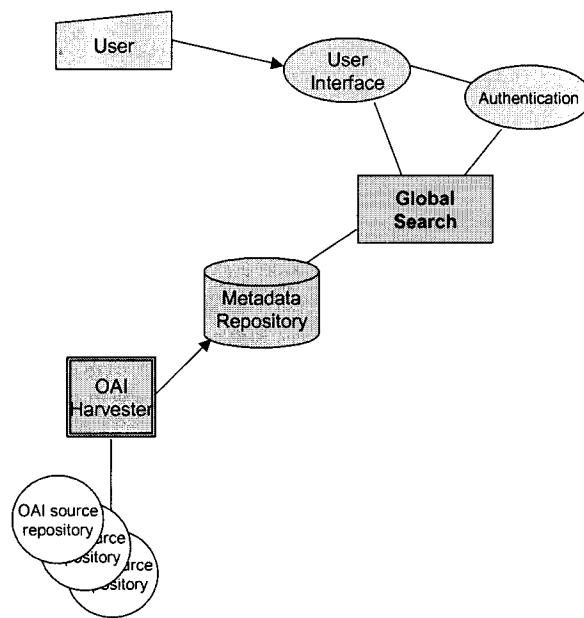
A further option was considered but discarded, namely use of the CIE Place Demonstrator.

These can be illustrated as follows:

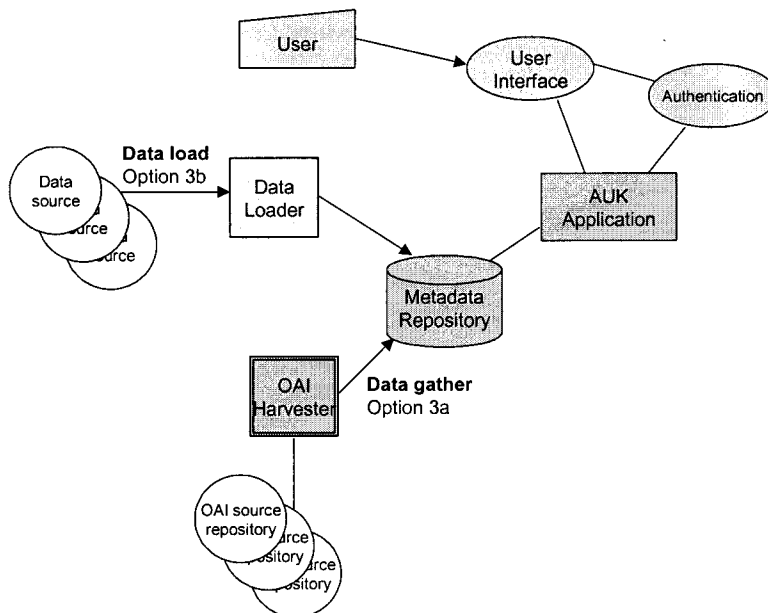
Option 1 – PNDS



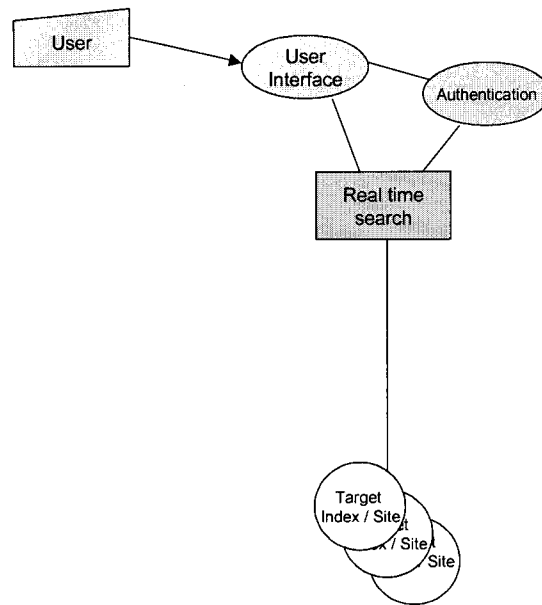
Option 2 – Global Search



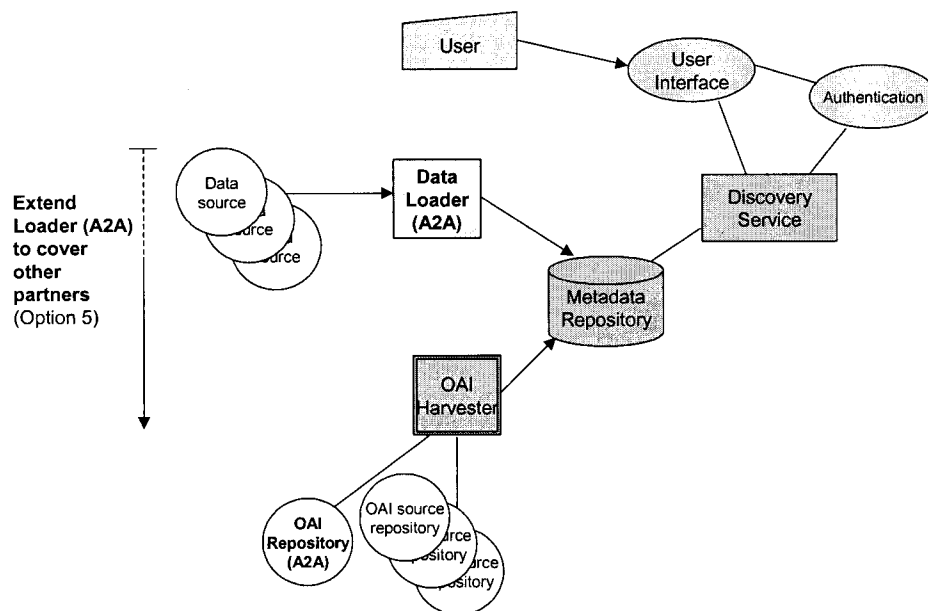
Option 3 – Bespoke Harvesting



Option 4 – Real Time Searching



Option 5 – A2A Extension



6.3 Option Assessment

The following assessment is very basic, crude, qualitative and subjective, but indicative at a high level:

Option	Major pro	Major con
1a	Instant service with some independence	Scalability unknown Sustainability
1b	Instant service without the need for hosting	Possible over dependency on the MLA web services link, needing SLAs
2	Coherence of total TNA infrastructure	Timing
3a	Own service under direct control	Data incompatibility and need for data mapping
3b	Ease of implementation	Lack of automation Only short-term quick win, not a solution – not sustainable Political (TNA are key members of CIE)
4	Up to date access to data	Performance over the internet
5	Some basic functionality to build from Political (TNA trust)	Cost of further development Political (TNA ownership)

Option 1 or Option 2 would appear at this moment to be the most viable given the funding situation – these need to be assessed in more detail before any firm recommendation can be given. Indeed, both options may be used to deliver different elements of the solution. Even with these options though, there will still be some funding issues.

Option 3 would need major funding (both the central system and development of a harvester / gatherer for each data source), Option 4 is not seen as technically viable at the moment (but may become increasingly so) and Option 5 may not be seen as a great fit to start with, so could need extensive redevelopment before it is in a position to provide the framework for an ideal solution.

6.4 Key Recommendations

Key recommendations made by this study:

- Determine strategy for Archives UK, considering also the role of A2A and whether / how it can be extended / incorporated into AUK.
- Talk to MLA about the use of PNDS. We would suggest keeping an eye on developments and working with it in some way to assess its full potential to Archives UK. Specifically, this could include the creation of an OAI harvester for A2A.
- Get the scope of the Global Search project extended to raise the profile of the Archives UK requirements. At the very minimum, ensure that it uses open standards with whichever solution is selected and implemented.

- Investigate funding for a harvesting system, i.e. extension of the Global Search engine to index and make available partners' data.

6.5 Way Forward

Regardless of the strategy and selected option for moving forward, there are a number of short-term actions, which should be carried out in order to retain initiative momentum:

- Continue and develop the consortium through a formal Memorandum of Understanding, detailing the key Roles & Responsibilities, in light of the HLF bid result. We then need to retain momentum in the initiative.
- Continue to strive to fill the conformance gap, including promotion of the NCA Interoperability Protocol and practical standards development by A2A.
- Cost the development options and seek funding.
- Re-consider the two key options / alternatives.
- Follow up with partners, including answering the outstanding questions as appropriate – either where further detailed clarification is needed or the questions were not answered in the first place.
- Determine the extension of the consortium to the likes of UK Villages and a more extensive list of local record offices.
- Continue discussions on AUK between TNA and MLA at senior levels.
- Even with good use of standards, some interpretation / manipulation / mapping of data will always be needed; the load should be put on the centre rather than the providers.

7 Appendix A Partner Responses



Partners Survey answers.obr

8 Appendix B Glossary

AACR2: Anglo-American Cataloguing Rules (AACR) are designed for use in the construction of catalogues and other lists in general libraries of all sizes. The rules cover the description of, and the provision of access points for, all library materials commonly collected at the present time.

AIM25: a JISC funded project, now completed, to create and mount for searching collection level descriptions of archives held in higher education institutions within the M25 area. The descriptions are held in a MySQL database, which has the ability to export records as EAD. The system has been developed for use by ANW. There is a proposal to extend the project by means of the London Signpost project.

ANW: Archives Network Wales. A project led by Archives Council Wales (ACW) and funded by the HLF to create 4500 collection level descriptions of the major archive collections in Wales and mount them on the Internet for searching by December 2005. The descriptions will be held in a MySQL database, based on the system used by the AIM25 project.

Archival authority records: A record holding the authority controlled name and other contextual details of the people, families and corporate bodies who create the archives or whom content show to be otherwise significant. ISAAR (CPF) is the standard for these records. Such records are linked to the NRA and are found in PROCAT. They may be organized nationally as the NNDP.

Archival levels: Archive catalogues have a structure reflecting archival arrangement and the principles of provenance and original order. These principles state that it is necessary to describe the context of records and the relationships between them as well as the records themselves. Only in this way will the evidential value inherent in archives be preserved. The three broad levels are given standard names by ISAD (G): fonds (and its parts); series (and its parts); and files and items.

Archives Hub: a JISC funded project to create and mount for searching collection level descriptions of archives held in higher education institutions in the UK. Some multi-level catalogues are also now included. Archives Hub was a partner in the development of ANG. While the programme is hosted at MIMAS, software development takes place at the University of Liverpool and the system is based on Z39.50 and EAD.

Authority controlled names: The names of people, families and corporate bodies that are archive creators or otherwise significant in a group of archives in the form authorised by the NCA Rules for the construction of personal, place and corporate names (1997). These names may be used as access points, with other index terms, to archival catalogues (as in the NRA) and will form the authorised name in archival authority records.

CALM: Archives management systems marketed by DS Ltd. This system is now used by over 100 holders of archives in the UK, especially in the local government sector.

CIE: Common Infrastructure Environment, public sector body promoting cross-Government interoperability.

Dublin Core: A generic metadata standard for cultural resources.

e-GIF: e- Government Interoperability Framework.

e-GMS: e-Government Metadata Standard, a component of e-GIF.

EAC: Encoded Archival Context. A developing data format standard for the automation of archival authority records. It is intended to be fully compatible with the 2nd edition of ISAAR (CPF).

EAD: Encoded Archival Description. A data format standard for the automation of archive catalogues, whether collection or multi-level in nature. It is fully compatible with ISAD (G). EAD is an SGML DTD and compatible with XML and related technologies; an official schema is in development. The Society of American Archivists and the Library of Congress jointly maintain it. EAD is extensively used in A2A and the Archives Hub and is a common import/export format in archive catalogue databases, such as CALM and The Catalogue.

HLF: Heritage Lottery Fund.

IAMS: Integrated Archive and Manuscripts System, the planned new British Library catalogue.

Index terms: Terms used as access points to archival catalogues relating to the content of the archives. They may include the authority controlled names of people, families, corporate bodies and places in a format authorized by the NCA Rules for the construction of personal, place and corporate names (1997). Subjects may also be included, format being based on thesauri such as UNESCO Thesaurus or Library of Congress Subject Headings (LCSH).

ISAAR (CPF): International Standard Archival Authority Record for Corporate bodies, Persons and Families. The ICA standard for Archival Authority Records. A second edition, which is radically different from the first is in the final consultation phase and is due for final publication in August 2004.

ISAD (G): General International Standard for Archival Description. The ICA standard for archival catalogue records, which is now in its second edition.

LCHS: The Library of Congress subject headings classification system.

London Signpost: A project proposed to extend the AIM25 project to repositories within London outside higher education. Currently a bid is being developed for submission to the HLF.

LSC: London Subject Classification, a Guildhall Library specialist classification.

MARC: Libraries record format.

METS: Metadata Encoding and Transmission Standard: METS is a standard for encoding descriptive, administrative, and structural metadata regarding digital objects. It is expressed as an XML schema and maintained by the Library of Congress.

MLA: The Museums, Libraries and Archives council is the national development agency working for and on behalf of museums, libraries and archives and advising government on policy and priorities for the sector.

MOLCAT: The British Library's Manuscripts OnLine CATalogue.

NAS: National Archives of Scotland. NAS are currently developing a new version of their electronic catalogue using the CALM database (NAS e-Cat), which has the ability to import and export EAD records.

NCA rules: The National Council on Archives rules for the construction of personal, place and corporate names.

NNAF: National Name Authority File: A proposed project to collect archival authority records that conform to the second edition of ISAAR (CPF) on an organizational basis and make them available in EAC.

OAI-PMH: The Open Archives Initiative Protocol for Metadata Harvesting.

PNDS: People's Network Discovery Service, an MLA-sponsored harvesting system, forming part of the Government's Knowledge Web.

PRONI: Public Record Office of Northern Ireland. A project funded by the HLF is currently retro converting archive catalogues of PRONI's holdings for loading into e-CATNI.

SCAN: Scottish Archival Network. A project by led in partnership by the NAS and funded by the HLF. Modules giving access to images of Scottish Wills and information about Scottish History are already available. There will also be about 20000 collection level descriptions of archive collections from repositories across Scotland.

SGML: Standard Generalized Markup Language

SRW/SRU: The web successor to Z39.50.

TALIS: The library management system used by the Guildhall Library.

UKAT: UK Archival Thesaurus, a subject thesaurus that has been created for the archive sector in the United Kingdom. It is a controlled vocabulary, which archives can use when indexing their collections and catalogues.

UNESCO Thesaurus: A controlled vocabulary developed by the United Nations Educational, Scientific and Cultural Organisation which includes subject terms for the following areas of knowledge: education, science, culture, social and human sciences, information and communication, and politics, law and economics. It also includes the names of countries and groupings of countries: political, economic, geographic, ethnic and religious, and linguistic groupings. The UNESCO Thesaurus allows subject terms to be expressed consistently, with increasing specificity, and in relation to other subjects. It can be used to facilitate subject indexing in libraries, archives and similar institutions

XML: Extensible Markup Language.

Z39.50: A network protocol, which allows searching of remote databases.