

**The Subject in Question:  
The Case for Developing a UK Archival Thesaurus**

***Supporting information to accompany an application to the Heritage  
Lottery Fund under the 'Your Heritage' programme***

Peter Garrod  
University of London Computer Centre  
p.garrod@ulcc.ac.uk

## Overview

This paper sets out the rationale for funding to create a subject indexing standard for UK archives, known as the UK Archival Thesaurus. It is divided into three sections:

- **The demand** outlines the ongoing emergence - backed in part by funding from the Heritage Lottery Fund (HLF) - of a UK National Archives Network of interlinked catalogue data. It explains why subject indexes will be a vital component of the Network's gateway, when it is developed.
- **The problem** explains why a subject indexing standard for archives is necessary, if the potential for subject searching across the Network is to be fulfilled. Currently there is only limited standardisation of the indexing terms which are being used by the projects which make up Network. Unless addressed, this problem will limit the ability of users to search in a "joined up" way using the Network's gateway. Moreover, the subject standards which are being used do not contain sufficient terms reflecting the histories and experiences of ethnic minorities and the socially disadvantaged.
- **The solution** sets out our proposal to create a UK Archival Thesaurus, with funding from the HLF. This will open up the searching potential of the National Archives Network by standardising the Network's subject content. It will also ensure that this content helps to support the goals of social inclusion which have been set for the archival sector as a whole.

## The demand

Archives, and the archive services which care for them and make them accessible, are a vital national resource for all citizens of the United Kingdom. This importance has been summarised by the National Council on Archives:

Archives have a relevance to every member of the community, without exception. . . . for millions of people, regardless of formal educational attainment, the exploration of family 'roots' through archives creates a personal involvement with history which is unchallenged in its immediacy and which can both inform and be informed by further exploration of historical sites and avenues. For others, often newcomers to remote rural communities or inner cities, their immediate local environment can be given striking new value and significance when it is illuminated by an understanding through archives of how it came to be as it is. For many ethnic groups, archives have in recent years been able to unveil a legacy of interaction with the community, sometimes good, sometimes bad, but either way unarguable. For others, basic social rights and entitlement to services prove wholly dependent on access to properly preserved, objective records.<sup>1</sup>

---

<sup>1</sup> National Council on Archives, *British Archives: The Way Forward* (1999), section 3.1 <<http://nca.archives.org.uk/brarchs.htm>>.

It has been conservatively estimated that there are over 1,100 institutions in the UK which hold significant archives.<sup>2</sup> In 2000/2001, local authority archive services alone in England received over 740,000 reader visits, produced documents on over 1.6 million occasions, and held over 56,000 cubic metres of permanent collections.<sup>3</sup> The government, in its *Policy on Archives*, has identified archives as an important resource which can support its objectives in a number of areas, including promoting education and lifelong learning, economic regeneration, combating social exclusion, and supporting the modernisation of public services.<sup>4</sup>

In its 1998 report *Archives On-Line*, the National Council on Archives set out the vision of a National Archives Network, which would use the Internet to open up the archival heritage to existing and new groups of users. This would be done by digitising existing archive catalogues, by generating new catalogues for previously uncatalogued collections, and by linking the resulting catalogue data into a network searchable via a single gateway. The ability of users to locate records would no longer be hampered by the physically dispersed nature of archive collections, or by the sectoral divisions which separate archive services.<sup>5</sup> This vision of using ICT to enhance access to archives was endorsed by the government in its 1999 *Policy on Archives*, which states that:

A key priority [for the archive sector] should be to provide accurate and specific descriptions of all historical archives as quickly as possible on the emerging information networks, so that a wider public is aware of their existence . . . It is of crucial importance to the archive sector as a whole to establish a framework or infrastructure, both national and regional, to permit the development of remote, on-line access to catalogues of holdings.<sup>6</sup>

The draft Action Plan arising from the *Policy* has emphasised that archive catalogues on the Internet should aim to reach out to social groups which are currently under-represented among archive users, thereby removing barriers to social inclusion.<sup>7</sup> The main agencies concerned with archives - the Public Record Office, the Historical Manuscripts Commission, and Resource: The Council for Museums, Archives and Libraries - have embraced these goals, and have worked with individual archive services to lay the groundwork for a

---

<sup>2</sup> Public Record Office, "Introduction and Purpose" to the *Government Policy on Archives* <<http://www.pro.gov.uk/archives/archivepolicy/introduction.htm>>.

<sup>3</sup> Chartered Institute of Public Finance and Accountancy (CIPFA) statistics reproduced in London Archive Regional Council, *Out of the Past: Into the Future - Priorities for Archive Development in London* (2001) <[http://www.ling.org.uk/larc/LARC\\_Report\\_v4.pdf](http://www.ling.org.uk/larc/LARC_Report_v4.pdf)>

<sup>4</sup> *Government Policy on Archives* (Cm 4516), section 4.2 <<http://www.pro.gov.uk/archives/archivepolicy/fullpolicy2.pdf>>

<sup>5</sup> National Council on Archives, *Archives On-Line: The Establishment of a United Kingdom Archival Network* (1998) <<http://nca.archives.org.uk/ONLINE01.HTM>>.

<sup>6</sup> *Government Policy on Archives* (Cm 4516), section 8.4.1 <<http://www.pro.gov.uk/archives/archivepolicy/fullpolicy2.pdf>>

<sup>7</sup> *Government Policy on Archives: Action Plan* <<http://www.pro.gov.uk/archives/actionplan/vision.htm>>

National Archives Network. Construction of the Network is a key theme of the Regional Archive Strategies produced by the nine English Regional Archive Councils.<sup>8</sup> A number of projects are currently putting the components of the Network into place:

- Access to Archives (A2A) <<http://www.a2a.pro.gov.uk/default.asp>>, which aims to create a "virtual national archives catalogue", covering over 400,000 converted catalogue pages from about 200 archive repositories in England. A2A's central infrastructure has been funded by the government's Invest to Save budget. Many of the regional and subject-based consortia which are carrying out the work of catalogue conversion are funded by the Revenue Grants Programme of the Heritage Lottery Fund.
- Scottish Archive Network (SCAN) <<http://www.scan.org.uk/index.html>>, whose aims include Internet access to the top level catalogues of 50 Scottish archives, and to historical Scottish wills and testaments. SCAN is supported by the Heritage Lottery Fund, the Genealogical Society of Utah and the National Archives of Scotland.
- The Archives Hub <<http://www.archiveshub.ac.uk/>>, which will provide a single point of access to the descriptions of archives held in UK universities and colleges (currently over 40 institutions are included). The Hub is funded by the Joint Information Systems Committee (JISC), a strategic committee to the UK further and higher education funding bodies.
- Archives in London and the M25 Area (AIM25) <<http://www.aim25.ac.uk/index.stm>>, a project funded by the Research Support Libraries Programme (RSLP) to create a web-accessible database of descriptions of the archives and manuscript collections of over 50 higher education institutions and learned societies in the Greater London region.
- Gateway to the Archives of Scottish Higher Education (GASHE) <<http://www.gashe.ac.uk/>>, an RSLP-funded project providing electronic access to the descriptions of the archives of 10 higher education institutions in Scotland.
- CASBAH <<http://www.casbah.ac.uk/welcome.html>>, a cross-sectoral demonstrator project funded by RSLP to identify and map national research resources for Caribbean Studies and the history of Black and Asian people in Britain, and create an on-line database of these resources. The project (now completed) covered both libraries and archives.
- MUNDUS <<http://www.mundus.ac.uk/>>, another RSLP-funded project to create a web based guide to archives throughout the UK reflecting the

---

<sup>8</sup> National Council on Archives, *Archives in the Regions: An Overview of English Regional Archive Strategies*, p.12 <<http://www.resource.gov.uk/information/policy/nca.pdf>>

activities of overseas missionaries - an historically important point of interaction between Britain and the wider world.<sup>9</sup>

The next phase in the creation of a National Archives Network will involve connecting these sectoral and subject-based projects into a single integrated network, with a single point of entry for users, and the ability to search catalogue descriptions across the network. This is part of the original vision of *Archives On-Line*, and is a stated goal of the draft Action Plan for the government's *Policy on Archives*. The Action Plan calls for the strengthening of links between the component parts of the National Archives Network, and between the Network and information providers for libraries, museums and the built heritage, "so that from the user's point of view access to information becomes a seamless operation and not one which is impeded by the existence of a multiplicity of information sources".<sup>10</sup> A recent user evaluation of four of the above projects has highlighted the fact that users, rather than searching websites separately, would prefer to have a common gateway linking the various strands of the National Archives Network, and would welcome the development of such a gateway.<sup>11</sup>

Subject indexes will be a vital component in any gateway to the National Archives Network. Indexes are a necessary resource discovery tool, allowing users to identify archives in different collections which relate to the same or broadly similar topics. Whereas archival catalogues reflect the hierarchy of the body which originally created the records, and any ordering of the records established by that body, subject indexes provide a way of locating archives regardless of the context in which they were created. This point is recognised in the Historical Manuscripts Commission's report *Archives at the Millennium*, which states that:

Even material which appears to have been fully catalogued can remain intractable if there is no index or if the indexing terms omit the very issues the searcher is interested in. This applies, for example, to references in the archives to people from ethnic minorities, but it can also apply to more general uses such as genealogical and family research.<sup>12</sup>

A number of indicators suggest that subject indexes are increasingly regarded by UK archives as an essential facility for their users, and will form a key element in the gateway to the National Archives Network:

- Most of the projects outlined above have created, or are in the process of creating, searchable subject indexes of their on-line catalogues.

<sup>9</sup> The MUNDUS web site is not yet (Sept 2002) publicly available.

<sup>10</sup> *Government Policy on Archives: Action Plan*, objective 1  
<<http://www.pro.gov.uk/archives/actionplan/policy.htm>>

<sup>11</sup> National Council on Archives and National Archives Network User Research Group (NANURG), *User Evaluation: Report of Findings* (2002), sections 1.11, 10.1  
<<http://www.resource.gov.uk/information/research/nanurg.pdf>>

<sup>12</sup> Historical Manuscripts Commission, *Archives at the Millennium: The Twenty-Eighth Report of the Royal Commission on Historical Manuscripts 1991-1999* (1999), p. 23. <<http://www.hmc.gov.uk/pubs/millen/archmill.pdf>>

Substantial subject index content is already being generated by the various strands of the Network.<sup>13</sup>

- In December 2001 the High Level Thesaurus Project (HILT) conducted a survey of "stakeholders" in archives, libraries, museums, special collections, on-line information services and hubs, and government agencies and initiatives. 95% of respondents reported that their users required subject access to their resources, and a similar number indicated that subject access was required by their staff.<sup>14</sup>
- In 1996-1997 the Archives Sub-committee of the JISC's Humanities Non-Formula Funding Committee commissioned a survey of subject indexing in UK archives (hereafter referred to as the JISC: NFF survey). The consultant interviewed a small number of archive users and noted search-room archivists' perceptions of users' reactions to subject indexes. The survey found that "user demand is . . . a catalyst for the maintenance or development of subject indexes [in archives]", and that "user comments indicated the importance placed on availability of subject indexes and other self-help facilities".<sup>15</sup>

Although little data is available for individual on-line services, there is evidence that subject searching facilities, where offered, are popular with users of Internet-based archive services. The UK National Digital Archive of Datasets (NDAD), operated by ULCC on behalf of the Public Record Office, has offered its users freetext searching and subject searching facilities since its inception in 1998. Cumulative web usage statistics show that the subject index's home page is the seventh most visited page on NDAD's website (the freetext searching facility is the fifth most visited).<sup>16</sup> Data supplied by the AIM25 project suggests that searches of their subject index have accounted for about 20% of total searches, the remainder consisting of freetext searches and searches for personal, place and corporate names.<sup>17</sup> Searches of the

---

<sup>13</sup> The only project which is not known to be creating a subject index is SCAN. Searchable on-line indexes are available for A2A <<http://www.a2a.pro.gov.uk/search/focused.asp>>, AIM25 <<http://www.aim25.ac.uk/search/>>, the Archives Hub <<http://www.archiveshub.ac.uk/searchsub.shtml>> and GASHE <<http://www.gashe.ac.uk/heicolls/subjects.html>>, and will shortly be available for CASBAH. As previously noted, MUNDUS's web site is not yet (Sept 2002) publicly available.

<sup>14</sup> Dennis Nicholson et al, *High Level Thesaurus Project: Final Report to RSLP and JISC* <<http://hilt.cdlr.strath.ac.uk/Reports/Documents/Hiltfinalreport.doc>>, p. 26.

<sup>15</sup> Janet Foster, "Subject Indexing Survey and Analysis: A Report for JISC: NFF Archives Sub-committee" (April 1997) <<http://www.kcl.ac.uk/projects//srch/reports/subindex.htm>>

<sup>16</sup> NDAD: see <http://ndad.ulcc.ac.uk/>. NDAD's web usage statistics cover the period 2 March 1998 - 30 June 2002. The entry level page for the freetext searching facility (<http://ndad.ulcc.ac.uk/search/freetext.htm>) received 34635 user requests in this period, versus 16024 requests for the entry-level page for NDAD's subject index, the NDAD Thesaurus (<http://ndad.ulcc.ac.uk/search/thesaurus.htm>). Note that page requests are not equivalent to searches: some users will have visited the freetext searching page or the NDAD Thesaurus home page without proceeding to conduct a search.

<sup>17</sup> For AIM25's searching facilities, see <http://www.aim25.ac.uk/search/>. The data supplied by AIM25 for the purposes of this application covers the period July-December 2001. During that period AIM25's users conducted 2102 freetext searches, 4481 searches of AIM25's indexes of personal, place and corporate names, and 1582 searches of AIM25's subject index. Note that

Public Record Office's on-line indexes (covering persons, places, corporate bodies and subjects) increased steadily from 1,829 in April 2001 to 11,503 in January 2002, an increase of over 600 percent. Moreover, subject searches eventually accounted for over half the total number of index searches.<sup>18</sup> Elsewhere in the electronic environment, searching by subject keywords is recognised as the preferred option to meet users' needs. The Office of the E-envoy's E-Government Strategy has provided for the development of a pan-government thesaurus, to guide public searches of governmental documentation. The Government Category List (GCL), which has emerged as the top level thesaurus/taxonomy for this purpose, is searched by keyword. It will be used to add subject metadata to public sector resources which are available on the Internet, such as UK Online.<sup>19</sup>

### ***The problem***

Against this backdrop of the increasing provision and use of subject indexes in individual archive services, and by the strands of the National Archives Network, one has to set the question: will the current situation support effective subject searching via the Network's gateway, when it is created? Our contention is that this will not be the case, unless steps are taken to integrate and standardise the indexing data which is being generated.

Users of subject indexes need to be sure that a search on a subject term will return "hits" on all the records relating to that subject. However, this is not possible if an index has been created using inconsistent terminology: e.g. if different terms have been used to express the same subject. This problem can occur within an archive service when subject indexes are created in an uncontrolled way. It is even greater if one tries to search *across* the subject indexes of multiple institutions and projects, whose indexes may have been based on different (perhaps locally created) standards. Historically, this has been the situation in the archive sector. Standardisation is a relatively recent phenomenon, and subject indexes have tended to be "home grown", developed locally by institutions following in-house practice. Some examples of the type of mismatches which can result are presented in the report of the JISC: NFF survey. This included a mapping of selected subject terms used in five archive repositories employing different indexing schemes. Of the 50

---

this method of counting searches rather than page impressions probably overstates the importance of freetext searching. Many freetext searches will have returned no hits in response to the term(s) input by users. By contrast, AIM25's indexes have a browsing based interface: this means that users select the terms which they wish to search in order to conduct a search, and are therefore likely to obtain better results than freetext searching, because every indexing term will be associated with at least one catalogue description.

<sup>18</sup> See <http://catalogue.pro.gov.uk/>. Detailed breakdowns for subject searches versus searches of the PRO's indexes of persons, places and corporate bodies have been supplied for November 2001-January 2002. In these three months, subject searches accounted for 11,808 searches out of a total of 21,929 index searches, or 53.8%; the proportions within each month are comparable. Note that these figures do not include freetext searches.

<sup>19</sup> Cabinet Office, *Guide to Meta-tagging with the GCL* (Jan 2002)  
<<http://www.govtalk.gov.uk/documents/Guide%20to%20metatagging%202002-01-17.pdf>>

terms compared by the survey, exact matches across all five repositories were found in only three instances.<sup>20</sup>

Reliable searching by users of indexing data generated by more than one service or project only becomes possible when indexes are created using a common standard. This has been acknowledged by the National Council on Archives, whose *Archives On-Line* report states that "any commonality of approach which can be adopted across the [archival] profession on this matter [subject indexing] will . . . help to ensure the consistency of cross-searching between the finding aids of different repositories".<sup>21</sup> However, the problem of lack of standardisation goes beyond the archive sector. The HILT project found that a variety of subject schemes are in use in archives, libraries, museums and the electronic services community. This is likely to be preventing users from conducting effective subject searches across these sectors, though there is also a lack of hard data on user behaviour. In archives, the situation has been compounded by the absence of a strong tradition of subject indexing in many repositories (HILT found that "archives have not provided subject access until relatively recently").<sup>22</sup>

In the last four years, the second (1995) edition of the *UNESCO Thesaurus*<sup>23</sup> has been adopted by many individual archive services, and by a number of the strands of the National Archives Network, as the basis for their subject indexes. It offers several advantages over purely in-house standards:

- It has a broad subject reach, covering education; science; culture; social and human sciences; information and communication; politics, law and economics; and the names of countries and groupings of countries.
- It is easily available in electronic form, via a web site created by ULCC on behalf of UNESCO (<http://www.ulcc.ac.uk/unesco/>) and from UNESCO on CD-ROM.<sup>24</sup>
- It has a clear, coherent structure which conforms to British and international standards for thesauri, in particular BS 5723/ISO 2788, *British Standard Guide to the Establishment and Development of Monolingual Thesauri*.<sup>25</sup>

---

<sup>20</sup> Janet Foster, "Subject Indexing Survey and Analysis: A Report for JISC: NFF Archives Subcommittee" (April 1997) <<http://www.kcl.ac.uk/projects/srch/reports/subindex.htm>>

<sup>21</sup> National Council on Archives, *Archives On-Line: The Establishment of a United Kingdom Archival Network* (1998), section 4.11 <<http://www.archives.org.uk/publications/archivesonline/aol4.asp>>

<sup>22</sup> Dennis Nicholson et al, *High Level Thesaurus Project: Final Report to RSLP and JISC* (<http://hilt.cdjr.strath.ac.uk/Reports/Documents/Hiltfinalreport.doc>), pp. 4, 19, 37.

<sup>23</sup> UNESCO, *UNESCO Thesaurus: A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social and Human Science, Culture, Communication and Information* (Paris: UNESCO Publishing, 1995). ISBN 92-3-003100-3. See also <http://www.ulcc.ac.uk/unesco/>.

<sup>24</sup> CD-ROM: UNESCO, *UNESCO Databases* (Paris: UNESCO, 2001), ISBN 92-3-003810-5, ISSN 1020-0223.

<sup>25</sup> British Standards Institution, *BS 5723: 1987 (ISO 2788-1986): British Standard Guide to the Establishment and Development of Monolingual Thesauri* (London, 1987).

- UNESCO has readily granted institutions and projects permission to use the *UNESCO Thesaurus* on a non-profit making basis (and has given permission for the *Thesaurus* to be used to create a subject thesaurus for UK archives).

Six of the seven projects which make up the current components of the National Archives Network (A2A, AIM25, GASHE, Archives Hub, CASBAH and MUNDUS) have adopted the *UNESCO Thesaurus* as a standard for their subject indexing data.<sup>26</sup> Together these projects cover hundreds of individual participating archives in England, Wales and Scotland, from local and national government, higher education and the voluntary sector. The *UNESCO Thesaurus* is used for subject indexing by the Public Record Office (PRO) and by ULCC in connection with NDAD. In December 2001 an informal survey of users of the *UNESCO Thesaurus* received responses from 21 individual archives which had adopted (or intended to adopt) the *UNESCO Thesaurus*, though some had done so in connection with networking projects.<sup>27</sup> Interest in the *UNESCO Thesaurus* is reflected the high usage of the *Thesaurus*' website, which received over 15,000 visits in the first six months of 2002.<sup>28</sup>

However, despite its acknowledged strengths and widespread adoption, there is growing evidence that the *UNESCO Thesaurus*, by itself, will not satisfy the need for a subject indexing standard for the National Archives Network. It was originally developed for UNESCO's internal use and reflects the international context in which it was created. The ca.6000 terms which it contains are most effective when used to index broad, general concepts, but are less satisfactory for indexing collections at more specific levels of detail. This means that most of the archives and projects which have used *UNESCO* have found that they need to include in their indexes terms which they have either devised themselves or drawn from other subject schemes, such as *Library of Congress Subject Headings* (LCSH). Here are some examples taken from the strands of the Network:

- A2A has estimated that 10-20% of the subject terms in its database do not originate from the *UNESCO Thesaurus*.
- AIM25 had 2420 terms in its subject index in December 2001. About 75% were drawn from *UNESCO*; of the remainder, the largest single group (21% of the total) were terms devised by participating institutions or the AIM25 project team. Terms were also incorporated from other subject schemes, including LCSH and the Australian Human Services and Health Thesaurus.

---

<sup>26</sup> The Archives Hub allows participants to use either the *UNESCO Thesaurus* or *Library of Congress Subject Headings*.

<sup>27</sup> The survey was conducted through an email posted to the listservs archives-nra@jiscmail.ac.uk and unesco-thesaurus@jiscmail.ac.uk.

<sup>28</sup> 15,269 visits between 1 January and 30 June 2002. <http://www.ulcc.ac.uk/unesco/> has consistently been among the most requested pages on ULCC's website.

- GASHE allows its contributors to submit terms taken from *UNESCO* and the British Education Thesaurus (BET). In December 2001 about 7% of the subject terms in GASHE's database were drawn from the BET, with the remainder being *UNESCO* terms.
- CASBAH has supplemented the *UNESCO Thesaurus* with terms created by participating institutions or by the CASBAH project team. In May 2002, of the 289 subject terms in CASBAH's database, 46 (15.9%) were non-*UNESCO* additions, including terms such as "African diaspora", "Black peoples", "Carnival arts", "Post-war migration" and "Abolition of slavery".<sup>29</sup>

The situation is similar in individual repositories. Since 1998 NDAD has used *UNESCO* to build an on-line index (the NDAD Thesaurus: <http://ndad.ulcc.ac.uk/search/thesaurus.htm>). This also includes indexing terms devised by NDAD, where a suitable term could not be found in *UNESCO*. Almost exactly one third of the 745 "preferred" terms in the NDAD Thesaurus are non-*UNESCO*. The British Library of Political and Economic Science at the London School of Economics has used about 490 terms to index its archives at collection (fonds) level. Of these, 86% are derived from *UNESCO*, 7% created by LSE, and the remainder taken from other subject schemes and the indexes of other projects. About 68% of the subject indexing terms used by the Public Record Office in early 2002 were drawn from *UNESCO*, with the remainder being taken from other sources or created by the PRO.

The experience of CASBAH also suggests that the *UNESCO Thesaurus*, by itself, does not contain enough detailed terminology to adequately reflect the histories and experiences of ethnic minorities. CASBAH's final report on subject indexing states that the project chose the *UNESCO Thesaurus* because of its clear, coherent structure, conformance to British standards and widespread adoption in the archive community. Research by CASBAH also showed that *UNESCO* was preferable to the main alternative subject scheme, LCSH, in terms of its coverage of Caribbean, Black and Asian History. However, the project found that:

There were a number of instances where its [*UNESCO's*] existing set of terms could not adequately summarise the content of several collection descriptions incorporated into the *CASBAH* database; particularly in relation to aspects of Caribbean history and some ethnic and cultural issues featured in material on the Caribbean and South Asian diasporas. For example, using the *UNESCO Thesaurus*, records relating to the papers of 18<sup>th</sup> century sugar planters in the Caribbean were difficult to index without resorting to very general words such as 'Economics', 'Sugar' or 'Slavery', in the absence of more specific phrases, such as 'Plantation economy', 'Sugar estates', 'Slave revolts' and 'Abolition of slavery'- terminology that researchers interested in

---

<sup>29</sup> Carol Dixon, "Indexing collection descriptions for the CASBAH database" (May 2002), Appendix 1 <<http://www.casbah.ac.uk/IndexingMay2002.doc>>.

this aspect of Caribbean history might want to use as database search terms.<sup>30</sup>

CASBAH also found that the *UNESCO Thesaurus* could only offer a limited set of terms relating to the classification of ethnic groups. However, it should be acknowledged that the problem goes beyond *UNESCO*. The Black and Asian Studies Association has reported that many archives hold material relating to Black history, which remains "hidden" because it is not listed or indexed to reflect its relevance to ethnic minorities.<sup>31</sup> People from ethnic communities are more likely to live in deprived areas and suffer disproportionately from social exclusion<sup>32</sup>, so the failure of traditional archival indexes and catalogues to reflect ethnic minorities is likely to be undermining the government's goal of using archives to combat social exclusion.

As long as the components of the National Archives Network have to supplement the *UNESCO Thesaurus* with their own indexing terms or terms taken from other subject schemes, the problem of inconsistent subject terminologies will be perpetuated, albeit in a reduced form. It is possible to integrate added terms into the *UNESCO Thesaurus* structure in a coherent way within individual repositories and projects: AIM25 and CASBAH have had considerable success in this. But future users who try to search across repositories and projects via the National Network's gateway will still face the problem that different services will have amended the *UNESCO Thesaurus* in different ways. A non-*UNESCO* term used in one archive or project will not necessarily be used to index the same concept in another archive or project. As the HILT project has argued in a wider context, "if their own users need to find resources by subject in a distributed environment . . . then organisations applying controlled terminologies locally that are not consistent across the distributed environment are spending time, effort and money in what, at best, is a counterproductive way".<sup>33</sup>

### ***The solution: a UK Archival Thesaurus***

It seems clear that to realise the vision of a searchable National Archives Network, a subject indexing standard for UK archives must be developed. This UK Archival Thesaurus should build on the strengths of the *UNESCO Thesaurus*, but have an expanded set of terms reflecting areas where a need to amend *UNESCO* has been identified; including (but not limited to) the

---

<sup>30</sup> Carol Dixon, "Indexing collection descriptions for the CASBAH database" (May 2002), pp. 3-4 <<http://www.casbah.ac.uk/IndexingMay2002.doc>>.

<sup>31</sup> National Council on Archives, *Taking Part: An Audit of Social Inclusion Work in Archives* (2001), p. 26 <<http://nca.archives.org.uk/takepart.pdf>>.

<sup>32</sup> Findings of the Social Exclusion Unit, reported in National Council on Archives, *Taking Part: An Audit of Social Inclusion Work in Archives* (2001), p. 25 <<http://nca.archives.org.uk/takepart.pdf>>.

<sup>33</sup> Dennis Nicholson et al, *High Level Thesaurus Project: Final Report to RSLP and JISC* (<http://hilt.cdjr.strath.ac.uk/Reports/Documents/Hiltfinalreport.doc>), p. 6.

addition of terms reflecting the histories and cultures of ethnic minorities and the socially disadvantaged. Creating such a thesaurus would:

- (1) Reduce, and hopefully remove altogether, the need for repositories and projects participating in the Network to create their own local amendments to *UNESCO*, or incorporate terms from other schemes. This will free up time which Network participants currently spend on maintaining their local thesauri - time which could be used for other vital tasks, such as outreach to users.
- (2) Make the subject indexes of individual repositories and projects more internally consistent, and consistent with each other.
- (3) Result in effective and reliable subject searching by users of the Network's gateway.
- (4) Help to achieve the goals, set by the government and endorsed by the National Council on Archives, of promoting access to archives by all sections of the community, and eliminating the barriers to access standing in the way of the socially disadvantaged.

The University of London Computer Centre - in association with the Public Record Office and the *UNESCO* Thesaurus Working Party - is seeking funding from the HLF for a project of 15 months' duration to create the UK Archival Thesaurus. This will:

- Be based on the second (1995) edition of the *UNESCO Thesaurus* - thereby building on *UNESCO*'s strengths, and the reality of its emergence as a de facto standard used by many of the Network's components.
- Incorporate:
  - (1) Appropriate terms which have already been created by individual archives and by the projects which make up the Network;
  - (2) Appropriate terms from other schemes (such as LCSH), especially those which Network projects have already incorporated into indexes based on *UNESCO*; and
  - (3) Terms which are submitted as proposed additions during the course of the UK Archival Thesaurus project.

One of the functions of the project will be to ensure that terms are added to the UK Archival Thesaurus in a coherent way, in accordance with established British and international standards for thesaurus construction.

- Be disseminated to archives and projects without charge:
  - (1) Via a website, which will allow browsing and searching of the Thesaurus and the on-line submission of proposed amendments. The website will show the Thesaurus under construction - it will be made available "as built", thereby facilitating feedback from archivists and users of archives.
  - (2) Via a portable data file, capable of being integrated into the thesaurus facilities of the cataloguing software packages most commonly used in

UK archives (CAIRS, CALM, ADLIB).<sup>34</sup> The data file will also be distributed on request to existing and future components of the National Archives Network.

The project will employ a Project Editor and an IT specialist, and will be managed by a Management Board consisting of experienced staff from ULCC and the PRO (see the application form, Q25a for further details). High professional standards will be ensured by an Advisory Panel drawn from the UNESCO Thesaurus Working Party (UTWP), a coalition of archives and projects which are using the *UNESCO Thesaurus* for indexing, and other bodies with an interest in the archival subject standards. UTWP - which has advised on the preparation of this application - comprises 26 user members, two representatives from the Historical Manuscripts Commission, the director of the HILT project, and one member of the Society of Archivists' Professional Methodology Panel.<sup>35</sup> Following the completion of the project, ULCC will continue to maintain the Thesaurus' website and database, including the facility for the on-line submission of proposed additions. ULCC will also continue to distribute the Thesaurus data file, on request, to repositories, projects and cataloguing software suppliers. Ongoing editorial support will be provided by the UTWP, following a peer review approach (for further details, see the application form, Q25d). We also anticipate that private sector suppliers of archival cataloguing software will continue to support the dissemination of the Thesaurus via their packages (see below).

A major part of the project will be to seek out terms which reflect the histories and cultures of ethnic minorities and socially disadvantaged groups. This will be done in part by incorporating amendments which have already been made by projects such as CASBAH. Advice on the suitability of terms will be sought from user groups and other interested bodies, such as the Black and Asian Studies Association, and members of the PRO's User Advisory Group who are preparing a guide to sources for Gay and Lesbian History.

It will be important for the project to reach out to potential users: both archivists, who will use the Thesaurus to create indexing data, and members of the public, who will use indexes based on the Thesaurus to search for

<sup>34</sup> CAIRS: <http://www.cairs.co.uk/contents/contents.htm> ; CALM: <http://www.ds.co.uk/company/files/index.html>; ADLIB: <http://www.uk.adlibsoft.com/>

<sup>35</sup> Membership of the UNESCO Thesaurus Working Party (UTWP) is as follows: Sarah Aitchison, King's College, London/AIM25; Kevin Ashley, NDAD; Robert Baxter, King's College London/AIM 25; Jacky Cox, Cambridge University Library; Louise Craven, PRO, chair; Jonathan Draper, Norfolk Record Office; Sue Donnelly, BLPES; Julie Evans, CASBAH Project; Penny Feltham, Museum of Science and Industry, Manchester; Janet Foster, AIM25/Independent Consultant; Peter Garrod, NDAD; Jill Geber, Oriental and India Office Collections, British Library; John Gurney, HMC; Jenny Haynes, The Women's Library; Tim Groom, Gateway to the Past Project, Staffordshire Record Office; Joanne Ichimura, Shakespeare Institute Library; Nick Kingsley, Gloucestershire County Council; Edmund Lee, English Heritage Data Services Unit; Patricia Methven, King's College, London, HE Hub/AIM25; Pam Miller, The Data Archive, University of Essex; Dennis Nicholson, HILT, University of Strathclyde; Kate Perry, Girton College; Andrew Riley, Churchill Archive Centre; Dick Sargent, HMC; Rosemary Seton, SOAS; Charles Tattersall, PRO; Paul Sillitoe, English Waterways Trust; Bill Stocking, A2a, PRO; Ruth Vyse, Walsall Archives Service; Caroline Williams, LUCAS/Society of Archivists PMP.

archives relevant to their research. The main forum for publicising the project will be the project's website, which will include the emerging Thesaurus itself; news and reports relating to the project; an on-line tutorial which will explain the Thesaurus to archivists and to archive users; and facilities for providing feedback on the project and the Thesaurus to the project team. Other outreach work will be carried out by the Project Editor, the Management Board and the Advisory Panel. This will include:

- Posting of news relating to the project to archival listservs (e.g. archives-nra@jiscmail.ac.uk, unesco-thesaurus@jiscmail.ac.uk), to archival publications (such as the Society of Archivists' *ARC* newsletter), and to relevant websites (such as the PRO website). This will promote the involvement of as many archive services as possible in submitting terms for the Thesaurus, and commenting on proposed additions.
- Training events for archivists outlining the nature of the emerging Thesaurus and how it can be used. The Society of Archivists will be approached about integrating events into its schedule of training courses, including courses for recently qualified archivists.
- Preparation of leaflets aimed at explaining the significance of the Thesaurus and the project to archivists and archive users. This will increase the visibility of the project to those who lack Internet access.
- Talks and lectures aimed at archivists and users of archives.

To ensure that the Thesaurus meets the needs of the users of archives, a User Group will be established. We anticipate that this will be drawn from members of the PRO's User Advisory Group, and other user groups associated with repositories participating in the UTWP. The User Group will be encouraged to give feedback on the developing Thesaurus from the perspective of archive users: is its terminology reflecting the types of subjects which users want to search for?

The UK Archival Thesaurus project will collaborate closely with the second phase of the HILT project, which has been funded by the Joint Information Systems Committee for 12 months starting in May 2002. HILT will seek to establish a "pilot interactive route map" between the major subject indexing standards used by archives, libraries, museums and electronic services, including the *UNESCO Thesaurus*, LCSH, Dewey Decimal Classification, the Getty Art and Architecture Thesaurus and the Universal Decimal Classification. Data from the UK Archival Thesaurus will be submitted to any mapping service which emerges from HILT, so that non-*UNESCO* terms in the Thesaurus can be incorporated into the mapping. Where possible, assistance will be provided to HILT in the mapping of these terms into other schemes. It should be emphasised that HILT does not remove the need to develop a UK Archival Thesaurus: HILT's interactive route map will only cover extensions to subject standards which have been developed by "key UK

services", such as national repositories.<sup>36</sup> Feeding data from the UK Archival Thesaurus to the mapping service will ensure that it takes into account a broader set of subject terminology, reflecting the needs of the archives sector as a whole.

Will the UK Archival Thesaurus be accepted and used by archivists? We believe that the answer is yes, provided that it is developed in such a way as to meet archives' indexing needs. After decades of localised practice, the archive profession is now receptive to the need for standardisation. The JISC: NFF report found that all of the repositories surveyed supported the idea of a subject authority standard, and that it was feasible to create such a standard.<sup>37</sup> The widespread adoption of the *UNESCO Thesaurus* is evidence of the acknowledgement by individual archives and projects of the need for standardisation in the context of the National Archives Network. CASBAH submitted the terms which it had created to a range of bodies for "dissemination, discussion and peer review", and expressed a wish that its terms should be incorporated into "a national thesaurus for the UK archives sector".<sup>38</sup> Subject indexing is a topic of increasing interest and concern among archivists: a training day on subject indexing sponsored by the Society of Archivists in October 2001 was attended by representatives from over 30 repositories. Moreover, this desire for standardisation has been recognised by the manufacturers of the cataloguing software packages used in UK archives. ULCC - acting on behalf of UNESCO - has given the data for the *UNESCO Thesaurus* to three software companies whose products are used in UK archives, all of whom have been able to successfully incorporate the *Thesaurus*. In each case, this was subject to the condition that *UNESCO* should be supplied to customers at no extra cost. It is likely that these firms would be willing and able to incorporate a UK Archival Thesaurus based on the *UNESCO Thesaurus* into their products, if they perceived a demand for it. The potential is clearly there to use the private sector to support the continuing dissemination of the UK Archival Thesaurus, after the project's completion.

In conclusion, we urge the HLF to seize the opportunity to fund the creation of a subject indexing standard for UK archives. The time is ripe for such a standard, and the momentum for it exists. Only if such a standard is created will the vision of "joined up" searching across the National Archives Network, outlined by the National Council on Archives in *Archives On-Line*, become a reality.

---

<sup>36</sup> Dennis Nicholson et al, *High Level Thesaurus Project: Final Report to RSLP and JISC*, pp. 4-5 <<http://hilt.cdli.strath.ac.uk/Reports/Documents/Hiltfinalreport.doc>>.

<sup>37</sup> Janet Foster, "Subject Indexing Survey and Analysis: A Report for JISC: NFF Archives Subcommittee" (April 1997) <<http://www.kcl.ac.uk/projects/srch/reports/subindex.htm>>

<sup>38</sup> Carol Dixon, "Indexing collection descriptions for the CASBAH database", p. 6.