

UK WEB ARCHIVING CONSORTIUM

EVALUATION REPORT: APRIL 2006



UK WEB ARCHIVING CONSORTIUM.....	1
EVALUATION REPORT: APRIL 2006	1
.....	3
Executive Summary.....	4
1. The UK Web Archiving Consortium (UKWAC).....	5
1.1 Aims, Objectives and Deliverables.....	5
1.1.1. Aims.....	5
1.1.2 Objectives.....	5
1.1.3 Deliverables.....	6
1.2 Budget.....	6
1.3 Governance.....	6
2. EVALUATION.....	7
2.1 Licence Procurement.....	7
2.2 Award of Contract to Provide the Common Infrastructure for the Pilot Project.....	8
2.3 Working Collaboratively on Selection	9
2.4 Working Collaboratively on Permissions.....	11
2.5 Working Collaboratively on Collection Management/Curatorial and Standards Issues.....	12
2.6 Digital Preservation.....	15
2.7 Testing and Developing PANDAS software.....	17
2.8 Providing Access to Websites collected as part of the project.....	19
3. Budget.....	20
4. Governance.....	20
5. Dissemination and usage.....	22
6. Working Collaboratively.....	23
6.1 Unexpected Issues.....	24
7. UKWAC: the future.....	24
7.1 Background developments.....	24
7.2 Original Propositions.....	25
8. UKWAC Recommendations	26
Key strategic recommendations.....	26
APPENDICES.....	29
Appendix 1: Standard permissions letter and documentation (examples from the British Library).....	29
Appendix 2: End User Survey Form for the UKWAC Web Site - Summary of Responses.....	32

Executive Summary

The UK Web Archiving Consortium (UKWAC) was set up in October 2003 with the objective of archiving and making accessible selected websites within the framework of a two-year pilot project and to share the costs and experiences of achieving that objective. The Evaluation Report shows that the Consortium achieved the aims and objectives of the project and has put in place an open and freely accessible single portal to the web archive collections of the six partner institutions.

In bringing together key players, and providing a forum for collaboration at strategic and operational levels, the Consortium has been successful in addressing, in a shared and collaborative manner, a range of legal, technical, operational, collection development and management issues relating to web archiving.

UKWAC has laid the foundations of a national web archiving strategy and a shared technical infrastructure for the United Kingdom (UK) and has prepared the ground for future development, taking into account the need to prepare for forthcoming secondary legislation associated with the Legal Deposit Libraries Act 2003 and the extension of legal deposit to non-print materials including websites.

The Report recommends an initial extension of UKWAC until September 2007, during which period the Consortium will need to focus on articulation of a vision of how web archiving should be progressed within UK, on formal evaluation of web archiving software, platform and tools for the future, and on digital preservation requirements, while at the same time continuing to build the UKWAC archive of websites at <http://www.webarchive.org.uk>

The members of the Consortium are: The British Library (lead partner), the Joint Information Systems Committee, The National Archives, the National Library of Scotland, the National Library of Wales, and the Wellcome Library.

1. The UK Web Archiving Consortium (UKWAC)

In October 2003 the British Library, the Higher Education Funding Council for England (for JISC), The Keeper of Public Records (for The National Archives), the National Library of Scotland, the National Library of Wales and the Wellcome Trust Limited signed the UK Web Archiving Consortium Agreement. The Agreement came into force on October 1 2003 and stipulated a duration of three calendar years from that date. It defined the roles and responsibilities of the six member institutions, each sharing a common business objective to archive and make accessible selected websites within the framework of a two-year pilot project and to share the costs and experiences of achieving this objective.

1.1 Aims, Objectives and Deliverables

A separate internal document outlining, in more detail, the aims and objectives of the UK Web Archiving Project was agreed between the partners in 2004 (UK Web Archiving Project: Aims and Objectives). This Evaluation Report measures progress against the aims, objectives and deliverables stated in that document. These were as follows:

1.1.1. Aims

To develop and evaluate a collaborative infrastructure for web archiving within the UK consisting of both shared costs and shared objectives, and exploring the nature of collaboration between partners.

1.1.2 Objectives

To procure a licence from the National Library of Australia to use the PANDAS (Pandora Digital Archiving System) software.

To award a contract to a contractor to provide the common infrastructure for the pilot project to include: hosting of the web archiving service; provision of public access to archived web sites; dedicated technical and developmental support for the web archiving service.

To work collaboratively in the achievement of a common searchable archive of selected web sites investigating solutions to issues such as: selection; obtaining of archive permissions; collection management and other curatorial issues and standards; digital preservation.

To evaluate the development of the collaborative infrastructure for web archiving with regards to assessing the permanence and long-term feasibility of such a collaborative enterprise.

To test and develop PANDAS software.

To provide access to web sites collected as part of the project.

1.1.3 Deliverables

The specific deliverables were agreed as follows:

A procurement exercise to identify a service provider to manage the technical infrastructure required for the project.

The subsequent awarding of a contract to the service provider identified to set up the web archiving infrastructure.

A common permissions licence for archiving web sites.

A common framework of web site selection policies.

A common approach to system developments agreed in response to either traditional curatorial issues, e.g. archival standards such as subject headings, or issues relating to digital preservation, e.g. preservation metadata.

A fully searchable/browsable online archive of sites collected and catalogued by UKWAC members.

A UKWAC website and discussion list for partners.

An evaluation report providing a set of recommendations on how best to proceed with the web archiving project.

1.2 Budget

It was agreed in the Consortium Agreement that each partner would bear its own costs in connection with the selection of web sites for the project; that the costs for the acquisition of hardware and software would be borne by the partners equally; and that the costs in connection with the procurement of the necessary IT infrastructure and technical support to establish a central web archiving repository would be shared equally by the parties, and that should the costs of a partner exceed £40,000, then that partner would be allowed to withdraw from the Agreement.

1.3 Governance

The Consortium Agreement provided for a governance structure, comprising the establishment of a Management Committee which would comprise one authorised representative from each partner. It was also agreed that the Management Committee should appoint a chair and a project manager. The roles and responsibilities of the Management Committee, the project manager, and the lead partner, the British Library, were clearly defined.

2. EVALUATION

The UK Web Archiving Consortium was set up on the basis of recommendations made in a joint Wellcome Trust and Joint Information Systems Committee report, entitled *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust* and available at (http://library.wellcome.ac.uk/doc_WTL039231.html).

At the time of the signing the agreement between the six partners, to the best of their knowledge, there was no body in the UK which was addressing the issue of the long-term preservation of websites or the risk that invaluable scholarly, cultural and scientific resources in web form were being lost for future generations. Various commentators had drawn attention to the risk, including Brewster Kahle of the Internet Archive - who in 1997 estimated a figure of 75 days for the life of an average web page. The issue was also highlighted during the parliamentary process which led to the successful extension of legal deposit to non-print material in October 2003.

The partners were also aware that progress was being made in other countries, e.g. in the United States of America, in Scandinavia, and in Australia, where the National Library of Australia was playing a leading role. That the task of archiving the web was not straightforward was also well known. There would be significant technical challenges, legal questions and issues around territoriality, (e.g. how is the UK defined in relation to web-sites?), and many more.

The purpose of this report is to demonstrate the progress made by UKWAC against the specific stated aims, objectives and deliverables agreed at the outset by the partners and to make recommendations for taking the work forward. The report will show that the UK Web Archiving Consortium has put in place a freely available website (<http://www.webarchive.org.uk>) which, at the time of writing in March 2006, provides access to over 1,100 websites representing the collection development policies of its members. The report will not seek to analyse or provide comparative assessments of developments which may have been made elsewhere.

2.1 Licence Procurement

Stated Deliverable: to procure an agreed and acceptable licence from the National Library of Australia to use the PANDAS software.

After considerable negotiation and iteration, involving legal advisers, a software licence agreement was signed between the National Library of Australia and the British Library Board, as lead partner of UKWAC. Separate back-to-back agreements were then needed between the British Library Board and the other UKWAC partners. These were signed in July 2004 for a term of up to 10 years. The licence made it possible for UKWAC to use the PANDAS software developed by the National Library of Australia to manage its PANDORA archive.

Success Factor: Agreement reached with National Library of Australia on use of licence.

Lessons to be Learned: Set realistic deadlines for the completion of licence agreements of this nature involving multiple and international partners. The British Library, through the expertise of its Central Purchasing Unit, took the lead role in this area of work but the process from beginning to end still lasted many months.

Recommendation: In the event of the continuation of UKWAC, it will be necessary to revisit the terms and conditions of use of the PANDAS licence, and the level of support and transparency of development of PANDAS that can be expected. Any future use of PANDAS should explore the open source software model, but ensure that mechanisms and safeguards are in place to provide an effective support environment.

2.2 Award of Contract to Provide the Common Infrastructure for the Pilot Project

Stated deliverable: procurement exercise/contract

A sub-group of UKWAC was set up to develop a detailed specification for the operational and technical support facilities required by the Consortium for the hosting and maintenance of a web archiving service. Requirements were summarised as follows: provision of the infrastructure to enable members to capture web sites; hosting of the web archiving service; provision of public access to the sites; provision of dedicated technical and developmental support. It was agreed that the contract would be for a two-year period with the possibility of subsequent extension for up to ten years at the discretion of the Consortium, the intervals of each extension also being at the discretion of the Consortium.

The procurement exercise was managed in line with European Union requirements by the British Library's Central Purchasing Unit and involved a subgroup from the Consortium which evaluated the submissions of six tenderers. The contract was won by Magus Research Ltd who were officially contracted to carry out the work on May 6 2004.

Success factor: Contract awarded thanks to the combined and collaborative effort of partners involved in the tender sub-group.

Lessons to be Learned: Set realistic deadlines to ensure that the necessary procurement deadlines and protocols can be met. The specification was at a very high-level and a number of the requirements were 'scenarios' for further development, which were not specified in a great level of detail. This resulted in the bids being over-simplified and generally quoted under what has subsequently been found to be realistic. Furthermore, PANDAS being the only solution under realistic conditions, the deficiencies of the software led to increased expenditure on testing, bug-fixing etc., preventing the Consortium from implementing significant additional functionality.

Recommendation: The initial contract with Magus Research Ltd terminates on May 31 2006. It will be necessary for UKWAC to ensure compliance with the terms of the contractual agreement, and in the event of UKWAC continuing, the necessary details of continuation with Magus Research Ltd will need to be considered. It is also recommended that contract/procurement negotiations need to be led by a single institution with appropriate infrastructure and experience in this area. It is also recommended that any software solution proposed for implementation beyond the two-year pilot is fully and rigorously evaluated by a technically qualified team as part of a full options appraisal assessment against agreed evaluation criteria. At the time of writing, it has been recommended by UKWAC to continue the Consortium for another year, and to extend the infrastructure and operations support provided by Magus Research Ltd until the end of September 2007 (see below section 7.2).

2.3 Working Collaboratively on Selection

Stated Deliverable: a common framework of web site selection policies

It was not the intention for the six partners to devise specific collection development policies for websites, but rather to develop their own collection guidelines in accordance with the broader collection policies of their own institutions.

In summary, the collecting aims of each partner for websites can be defined as selective, due to the limitations of the pilot project and the associated learning experience, and can be stated briefly as follows:

British Library: to collect sites from the UK web space by prioritising the archiving of sites of research value across the spectrum of knowledge; also to select sites which are representative of British cultural heritage in all its diversity and a small number of sites which demonstrate innovation.

Joint Information Systems Committee: to collect sites created as a result of funding awards made by JISC and sites relevant to the UK HE sector.

The National Archives: focus on six main clusters of government departments: defence and foreign policy; administration of justice and internal security; management of national resources; provision of services not provided by the market (health, education, culture); regulation and coordination of market provided services; machinery of government and delegated administrations.

National Library of Scotland: to collect websites that reflect the knowledge, culture and history of Scotland and the Scots.

National Library of Wales: to create a collection that provides a reflection of life in contemporary Wales whilst also serving the current and future research needs of Wales.

The Wellcome Library: to collect health and history of medicine-related sites from the UK domain.

As a result of the UKWAC pilot project, the collection of websites has been fully embedded in the collection development policies of each institution. Through the work of a collection development sub-group of UKWAC, a series of mechanisms and practical issues have been tested in areas such as site selection methodology (including raising the need for UKWAC archiving of significant specific sites and agreeing by which member), avoidance of duplication and overlap, and issues of territoriality. Some checklists of sites to be selected were necessary so that one library did not approach website owners already covered by another.

It should be noted that any decisions made as part of the web archiving work should not be seen as precedents for Regulations under the Legal Deposit Libraries Act 2003.

Success Factors: Through the collection development sub-group and through daily communication by means of the UKWAC e-mail lists, a framework of selection policies has been put in place, ensuring avoidance of duplication and overlap and shaped to fit the broad collection development policies of each partner. This common framework for selective archiving is not yet comprehensive in its remit; however, it can form a strong basis for further development.

Lessons to be Learned: These include the need to ensure that decisions on selection have long term validity, the need to ensure that all members are kept informed of sites being selected, and the importance of setting up frameworks of subject specialists forming a productive approach to site selection.

Recommendations: A full report on collection development issues should be produced, based on the UKWAC experience, and a set of pragmatic guidelines should be drawn up in the event of UKWAC continuing in its current or in an expanded form. The report and guidelines should respect the different and sometimes overlapping collection development remits of each member institution; illustrate issues which have been addressed with regard to territoriality and possible selection of non-UK sites; draw attention to gaps in coverage, either by subject or area; define mechanisms, preferably automated, to facilitate communication, to ensure all partners are kept informed of sites being selected or under consideration and thereby to avoid duplication, overlap and conflict; indicate best practice gained from events-based archiving e.g., collecting general election, tsunami and London bombing sites; identify effective means devised to ensure input of colleagues from the same or another institution in the selection process; consideration of issues around frequency and depth of capture, etc.

It also recommended that a collaborative framework is put in place within and across partner institutions for the effective selection of sites.

2.4 Working Collaboratively on Permissions

Stated Deliverable: common permissions licence

Pending secondary legislation and Regulations for UK websites under the terms of the Legal Deposit Libraries Act 2003, it was agreed as a principle by UKWAC that all archiving would need to be based on a rights-cleared basis. This was particularly important for the three national libraries and for the Wellcome. Permissions were not required for The National Archives whose remit has been to collect sites within their own domain. For JISC current conditions of award indicate a requirement for the recipient of funding to lodge a copy of any outputs with them. It was also agreed that the end product should be simple to use and administer and could not be in the form of a multi-page contract.

Work began in January 2004 on defining an agreed permissions licence which would be approved and used by all UKWAC partners. The National Library of Wales took the lead in providing the first draft which went through a number of iterations before final approval in October 2004. The draft and associated frequently asked questions were considered by the legal department of each institution and, in the case of the British Library, it was felt necessary to share the document with the Department of Culture, Media and Sport. There was some frustration among partners at the length of time required to arrive at the agreed wording, an inevitable by-product of the differing natures, size and scale of the various partners and the business and approval processes in place in each of them.

In terms of its use, the permissions form has served its purpose. In line with practice elsewhere, e.g. in Australia and the United States of America, the rate of granting permission has been low, approximately 25% of sites approached in the case of the British Library and 42% in the case of the Wellcome Trust, to quote two examples. There have been few outright rejections, approximately 1%. However, a large proportion of requests have not received replies, almost 75% in the case of the British Library. It is likely that this is due to the licence requiring the site owners to warrant they have permission from all third parties. The low response rate has generated an additional and unexpected work overhead, in particular in handling the large number of enquiries coming from those site owners approached. A further consequence for the British Library, the National Library of Scotland, the National Library of Wales and the Wellcome is a large discrepancy between the number of sites selected and the number of sites archived and made available for access.

The labour-intensive and prolonged nature of the permissions process has led to the consideration and implementation of other approaches for approval and rights clearance, in particular for events-based gathering where speed is of the essence, e.g. in connection with tsunami, general election and London bombings sites. In these cases, some partners took the decision to adapt the permissions approach by sending a revised and shorter permissions request stating that the UKWAC consortium would archive the site unless a negative reply was received within a given number of days. This approach is known as 'notice and takedown' and resulted in negative responses of only 2.3%. Some partners have expressed concern that this policy is at variance with the agreed approach.

Success Factors: A permissions statement acceptable to all partners was agreed and put to use. It has been revised slightly, through agreement between the partners, in the light of use and experience, and in response to some issues raised by website owners. (See Appendix 1 for the Consortium permissions licence and the text used in the adapted approach used for events-based collecting).

Recommendations: Lobby for an early Regulation to cover the collection, preservation of and access to publicly accessible websites within the UK, in line with the legislation passed in New Zealand.

Pending secondary legislation for websites through the Legal Deposit Libraries Act, and in light of the labour-intensive and prolonged processes of securing rights clearance through the present system, consider alternative approaches such as 'notice and take down' and take legal advice, including a full assessment of risk, alongside an audit of best practice elsewhere.

2.5 Working Collaboratively on Collection Management/Curatorial and Standards Issues

Stated Deliverables: common approach to developments such as subject headings, metadata, cataloguing; preservation metadata

All these issues have formed agenda items at the UKWAC Partners meetings, have been considered via the UKWAC discussion lists, and have been addressed as part of the operational work in delivering an accessible website resource.

Subject Access

The archive's web interface offers an easy-to-use means by which users can access information about the project and the contents of the archive. The archived material can be retrieved by use of a search facility or by browsing a hierarchical arrangement of topic headings.

The topic headings are based on the template developed by the National Library of Australia. This has been modified to reflect the UK environment, e.g. the heading 'indigenous populations' is not used. It was necessary to use the template as it was hardwired into PANDAS and as such practical and cost limits were imposed on numbers of headings and subheadings. This was deemed as an appropriate navigation method for a front end web interface at this point in time – more comprehensive subject and classification taxonomies will need to be considered in future development, with a more effective resource discovery model making use of appropriate metadata.

One issue was whether to base the headings on a traditional classification scheme such as Dewey or to create a new system, suited to the internet. A compromise was reached by combining elements from both types, i.e. Dewey and a scheme inspired by the Yahoo! Classification Scheme. Given the difference between the broad collecting spectrum of the national libraries and the more specific subject interests of the other members, the JISC, TNA and Wellcome made the decisions in the selection of subcategories for the headings 'Education & Research', 'Health' and 'Government & Politics' while the national libraries determined the remaining categories.

Descriptive Metadata

PANDAS provides limited descriptive metadata, this being restricted to basic publisher names, titles of websites, a limited number of subject headings, and ad-hoc collections. Also, in view of the need to focus attention on maintaining the basic functionality of the PANDAS software (see 2.7 below), it proved impossible to consider metadata developments, as had been originally envisaged in the *Specification* for work to be undertaken in conjunction with Magus Research Ltd.

Consequently, to integrate archived sites into the various collections, partners adopted a number of approaches.

a) The National Library of Scotland will be creating MARC21 records for all sites archived, and possibly for elements within websites, e.g. documents and serials (although this is still under discussion). Collection-level records will be created for the specific collections.

b) For the National Library of Wales the primary means of access to websites archived is currently via the UKWAC website. In line with the Library's cataloguing policy, every website archived through UKWAC will have its descriptive metadata stored in a MARC21 record on the Library's management system. Currently, technical and preservation metadata for all the Library's digital assets is held in METS and this is likely also to be true for archived websites in the future.

c) At the British Library, the catalogue records created have been MARC21 collection-level records, including DDC (Dewey) and LCSH (Library of Congress Subject Headings). They have been created for events (e.g. Tsunami, elections, London bombings) and special research areas, e.g. women's issues. The collection-level record being used has been developed from a trial of the Library of Congress Access-level record. Records link to the page for the specific collection which lists all linked sites.

d) At The National Archives, archived websites are being catalogued in accordance with TNA cataloguing standards, based on ISAD (G) and EAD. Each title is catalogued as a piece within the ZWEB series, and each archived instance is catalogued as an item within that piece. A single point of access to TNA's UKWAC collections is being provided through the UK Government Web Archive website, which allows access to be integrated with other TNA web archive collections, such as those collected through the Internet Archive.

e) As JISC does not fulfil the function of a repository in any other respect it does not have existing external cataloguing systems or standards to which UKWAC entries can be added. However, as part of the pilot project capturing JISC project websites a documentation database has been created for *internal* use. The purpose of this simple MS Access database has been to consolidate disparate information gathered in the course of the pilot project, and to perform certain functions:

- To record the success (or otherwise) of archiving actions
- To record instances of archiving in the Web Archive
- To note appearances of descriptive information held in PANDAS
- To summarise information about the original project as funded by JISC
- To add information about each site collected from the signed permissions forms

Plans are being considered to make this information publicly available.

The current JISC website also contains a summary of each project funded. This website is due for replacement during summer 2006 and plans to provide a link from this project summary to the archived instance of the website within the Web Archive are actively being pursued.

f) For the Wellcome Library the primary means of access to archived websites remains via the UKWAC website. However, every website selected and archived by the Wellcome Library will have its descriptive metadata stored in a MARC21 record on the Library's management system. Sites are catalogued from the internet and links are provided to both the archives instance of sites as well as the live instance.

The Legal Deposit Libraries are likely to be integrating descriptive metadata with a METS approach, which will also hold preservation, structural and rights metadata, in line with the Legal Deposit Libraries Metadata working group. The partner libraries will need to consider how the metadata recorded in PANDAS – and in their catalogues – relates to, duplicates, and is interoperable with, this approach.

Preservation and Rights Metadata

The lack of preservation metadata is a serious flaw in a digital archiving system, and will need to be addressed in future developments, in line with recommendations for descriptive metadata. This will also need to take into consideration metadata being produced or used elsewhere, e.g. in catalogues, storage layers etc, and its interoperability. The work of the Legal Deposit Libraries Metadata working group will have an impact in this area, and the use of METS to hold descriptive, preservation, structural and administrative metadata should be monitored.

Success Factors:

Subject Headings: The agreed categorisation of subject headings has proved adequate to enable simple retrieval of archived websites within the limits of the current archiving system. UKWAC collaborated well on the difficult area of the subject headings and this success may have been due to the creation of a subgroup which proved an effective way of working. However, there has been little feedback from users, apart from a small-scale survey commenced in summer 2005 (see section 5 below and Appendix 2), and no usability testing.

Cataloguing: Catalogue records have been produced successfully by the members, but in different ways, and not to a sufficient level to be able to make real evidence-based decisions on future methodologies and scalability.

Recommendations: A detailed review of the policies and practices used in these areas as part of the UKWAC pilot needs to be compiled as a product of the project. The review report needs to make some clear recommendations as to best practice for the future and to identify particular areas or issues where further work needs to be carried out, e.g. technical and preservation metadata, and in particular how these are affected by large scale archiving needs. This is particularly important should a decision be reached to continue and possibly expand UKWAC, and should UKWAC become the foundation of a national web archiving resource.

In the case of subject headings, and their use within PANDAS, it would be constructive to review the existing headings and subheadings and to ascertain the extent to which amendments can be made. The existing subgroup should be tasked to carry out this review and make recommendations for change. The subgroup should take into account the cost effectiveness of enhancing the PANDAS capability in this area in comparison with the facility for a new system, such as the IIPC Toolset, the IIPC Curator Tool, or other possibility, to provide a much broader and more effective categorisation system, ideally based on standard taxonomies and subject headings. Again, it is important to consider these in the context of a much larger scale operation where manual entry of metadata will be costly.

Any future system needs to be able to make use of existing metadata standards identified by the Consortium to manage descriptive, technical/preservation and rights metadata, and interoperate with other systems that use these standards.

2.6 Digital Preservation

Stated Deliverables: To preserve UK web content

Until the establishment of UKWAC, it is fair to say that there has been little concerted and coordinated attention paid to the long-term preservation of websites in the UK, or indeed internationally, despite the recognition that the Web has become, for many, the information source of first resort. Digital preservation is concerned with maintaining the accessibility of digital objects over the long-term, in authentic form. It may be considered to have two key aspects: passive preservation, which is concerned with secure storage of archived content, and active preservation, which provides for the transformation of that content, or its means of access, in order to ensure continued accessibility across technological boundaries.

Within this context, the progress of UKWAC to date can be summarised as follows:

Passive preservation: The current UKWAC infrastructure provides most of the key features required to ensure passive preservation of the collection, including backup facilities, integrity checking, and business continuity planning. UKWAC can therefore be considered to be providing a bitstream preservation service and, as such, there are no immediate concerns as to the viability of the archive.

Active preservation: UKWAC has not yet addressed the issue of active preservation, although some individual members are developing such services as part of wider digital preservation initiatives. PANDAS makes no provision for active preservation of websites, and does not allow for the concept of multiple technical manifestations of a single archived instance, such as might be created through a strategy of migration.

Active preservation is not expected to be an immediate issue for UKWAC, due to the age of the collected content. However, it is essential that the requirements for future preservation action are now considered.

Recommendations: There are legal and policy imperatives for some partners (and Magus Research Ltd) in this area, e.g. legal deposit legislation and e-Government. Future work will need to be aligned with broader digital preservation programmes in partner organisations, and be cognisant of the potential diversity of requirements which may need to be addressed.

It is recommended that, in the short-term, Magus Research Ltd should continue to host and maintain the archive. It is further recommended that UKWAC should explicitly address the issue of digital preservation in any second phase, and should make recommendations accordingly. Future requirements which are developed as a basis for either evaluating or procuring software will need to be informed by these recommendations. UKWAC should also consider how existing research in this field, both within partner organisations and beyond, may be utilised by the Consortium.

It is strongly recommended that any future infrastructure should maintain a clear separation between functions, such as collection, cataloguing, delivery, active, and passive preservation. In the preservation context, this would entail a distinction between a secure storage system, and any systems for performing active preservation. Such a separation will bring benefits in terms of maintainability, extensibility, and security, and will allow partners much greater flexibility in terms of their potential use of the infrastructure.

Archived content is currently stored in unaltered form and, as such, is readily available for transformation or export to other systems. It is essential that such portability be maintained in any future repository. However, consideration should be given to the adoption of emerging packaging formats for web archives, such as ARC or WARC, which are likely to benefit from excellent tool support from initiatives such as the IIPC. In the future, issues surrounding the physical hosting of the collection, and whether it is hosted by one organisation or shared between partners, are likely to be less significant than those of interoperability and the development of virtual collections. It is recommended that a sub-group be established to take this forward and to ensure that the content of the archive is preserved for the future.

2.7 Testing and Developing PANDAS software

Stated Deliverables: a fully searchable/browsable online archive of sites collected and catalogued by UKWAC members

From the outset, UKWAC was aware of certain deficiencies around the functionality of the PANDAS software. It was, however, recognized that the National Library of Australia had developed and pioneered a successful web archiving service based on the use of the PANDAS.

The procurement requirements were defined in the *Specification for a common web-archive infrastructure*, a document compiled by UKWAC as part of the tender documentation which led to the appointment of Magus Research Ltd. The *Specification* stated clearly that the Consortium intended to use PANDAS and that the tenderers should be able to demonstrate that they could host the software, provide appropriate access, and supply the necessary skills to be able to modify the software as required by the Consortium. Although the procurement documentation explicitly named PANDAS as the preferred solution bidders were also instructed to propose alternatives to PANDAS if they so wished. One such alternative was proposed but was not selected due to an existing relationship between one of the partners and the bidders. In terms of technical support and development, reference was made in the *Specification* to examples of some specific developments which the Consortium might be likely to request, e.g. the development of a PANDAS cataloguing module to allow metadata to be added, for instance, additional provenance, preservation and resource discovery fields; a Welsh version of the public interface, an essential requirement for the National Library of Wales; and developments in interoperability, e.g. with library systems used by the individual UKWAC members, and as a minimum requirement, the need to provide a mechanism to output records from the archive in MARC21 format.

Due to some issues with the software Magus Research Ltd struggled for some time to install and run PANDAS reliably in their environment. This caused considerable delay in starting the work. UKWAC began to use PANDAS, in conjunction with Magus Research Ltd, in September 2004, and commenced a thorough evaluation of its features and capabilities. It soon became apparent that the software presented a number of problems which had not been foreseen during the previous evaluation process and which posed greater problems to resolve in the UK consortial environment than in the home institution where the software had been developed and where local technical support and knowledge were at hand.

A small number of the problems can be attributed to teething troubles and training needs at the start of operation of PANDAS. However, many were due to the limited functionality of PANDAS and therefore needed to be addressed by Magus Research Ltd at extra cost.

As a result of these difficulties, the focus, agreed by the partners, has been on defining and maintaining a reasonable level of operational functionality. The software, therefore, became more expensive to maintain and it became impossible to budget for any of the proposed developments referred to above.

In parallel, work is being carried out within the International Internet Preservation Consortium (of which both the British Library and the National Library of Australia are members) to develop an archiving system based on the IIPC Toolset and a new system - the curator tool – to provide a non-technical user interface to the tools. This may supersede PANDAS software and may form a basis for a standard for a number of web archiving functions. UKWAC has provided input to the specification for the curator tool and the British Library has provided financial support and project input to this development, scheduled for test evaluation in July 2006.

Success Factors: The PANDAS software has made it possible for a UKWAC web archive to be put in place and made accessible to users. Some of the problems of functionality have been overcome, either through fixes or workarounds, although considerable problems remain. UKWAC has provided input to the specification for new curator tool software being developed via the International Internet Preservation Consortium, and which, it is hoped, will supersede PANDAS.

Lessons to be Learned. A more detailed evaluation of the PANDAS software should have been carried out in advance of commitment to the software for the Project. However, it should be noted that the UKWAC Management Committee made the decision to use PANDAS software in full awareness of some deficiencies and in the knowledge that it had worked successfully for the National Library of Australia.

Recommendations: It is acknowledged that PANDAS in its current form is not viable for the future development of UKWAC and that new and improved software is required. It is recommended that UKWAC should consider and assess all available options for replacement software by means of a thorough options appraisal linked to strict and agreed evaluation criteria and to an agreed timescale. The outcome of the options appraisal should inform the necessary next step for the development or procurement for any future system. In carrying out the assessment, one requirement must be the need to transfer and preserve the websites archived and made accessible as part of the pilot project. In addition it should be noted that the general procurement procedures need to be supplemented by appropriate technical specifications. Full evaluations by technically competent staff are essential. Avoidance of weighting the potential solutions prematurely should be avoided. The new system will need to be capable of meeting challenges presented by the deep web and new web technologies, otherwise the gap will continue to grow between the complexity of sites being created and those which can be captured.

In terms of the existing shared infrastructure, it must be ensured that this is not dependent on a particular contractor or the participation of a single Consortium member; that the hardware/software developed can be portable to a new contractor, if necessary, and that it is unencumbered by any contractor IPR; that the shared infrastructure should be developed in a way that is not dependent upon the particular software chosen (e.g. PANDAS), and that the content archived and data created, and any associated metadata, is portable to a new software solution as and when necessary, preserving persistent IDs intact.

2.8 Providing Access to Websites collected as part of the project

Stated Deliverable: an UKWAC website

The UK Web Archive went live on May 9 2005 and is available at <http://www.archive.org.uk>. The archive provides access to a broad range of sites selected and archived by each of the members, and where necessary, on a rights-cleared basis. As of March 20 2006, according to PANDAS administration statistics, over 3,600 separate instances or copies of individual web sites were accessible through the archive. These were distributed as follows:

Partner	Titles accessible	Instances accessible
British Library	622	1836
JISC	199	321
The National Archives	40	294
National Library of Scotland	38	389
National Library of Wales	85	312
Wellcome Library	188	489
Totals	1172	3641

It is envisaged that by the end of June 2006, in the region of 1,500 sites should be publicly available. In terms of accessible sites, this total falls well short of the estimates provided at the outset of the project, i.e. 6,000 sites at a rate of 500 sites per member per annum. The shortfall can be attributed to three main factors: a) the unexpectedly low response rate to permission requests and the associated additional work required in managing enquiries about permissions; b) the difficulties encountered in terms of PANDAS functionality; and c) external factors e.g. other priorities at particular times, levels of resource available in each institution, etc. It should be noted that progress on the response rate could be accentuated using alternative approaches to permissions (notice and takedown). However, the functionality issues are proving insurmountable within the budget available.

Success Factor: A functioning and accessible web archive, providing a single portal to the web archive collections of UKWAC partners. This is first of its kind in the UK, and one of the first non-thematic web archives to be made publicly available anywhere in the world. A successful search facility was developed and added as part of the interface.

Recommendation: It is recommended that a single portal to UK web archives should be maintained, whether by UKWAC or, in the future, by some other means.

3. Budget

As stated above in 1.2, each partner agreed to contribute £40,000 to cover the costs of the procurement, support and development of the required IT infrastructure including associated hardware, software and storage. All other costs, e.g. staff costs, selection etc have been met by the partner institutions. The British Library managed and monitored the Consortium budget through its Finance Department.

Success Factors: The budget has been managed effectively.

Lessons to be Learned. There is a need to ensure that all partners are kept fully informed of the budget situation on a regular and transparent basis.

Recommendation: In the event of the continuation of UKWAC, it should be agreed that regular budget statements with simple commentary are made available to each partner.

Building on work carried out by the Wellcome Trust, a further piece of work should be carried out to cost the various elements of the web archiving process from selection through to cataloguing and preservation. This would be significant both in terms of benchmarking and life-cycle costing.

4. Governance

As defined in the original Memorandum of Understanding (MoU), a Management Committee – with a representative from each UKWAC partner – was established. The Committee has been charged with setting the overall policy and direction of the project.

Since the MoU was signed (October 2003), the Management Committee has met regularly and has taken a number of significant decisions, such as deciding to use PANDAS software, and appointing Magus Research Ltd to host and develop the UKWAC project.

As with any Consortium, members of Committee have not always been in agreement. For example, there was much debate on whether there should (or should not) be a formal launch of the UKWAC archive. The “one member one vote” policy, however, defined in the MoU, proved to be effective in resolving such issues.

The Management Committee is also responsible for appointing the Project Manager, whose principal task was defined as having “responsibility for the day-to-day management for the project”. Though a Project Manager has been appointed, in the event, this role became less important than originally envisaged, as individual partner members have assumed more active responsibility for the project and for the sites they have archived. As a result, it is generally desired that key roles and functions should be defined and distributed throughout the partners.

Success Factors: In general, the current governance model has worked both in terms of moving the consortial work and agenda forward and in facilitating decision making.

Lessons to be Learned: The role of the project manager has not proved to fit the expected role definition. There has been some uncertainty as to representation at partners meetings.

Recommendations: In the light of experience, the Consortium has considered a number of options aimed at improving its governance and business management structure. These can be summarised as a) a recommendation for a period of extension up to September 2007 (see section 7.2 below); and b) an assessment of other business models which might be considered for a more permanent extension from a project into a sustainable business operation.

Recommendation for the extension period to September 2007

It is recommended that the existing governance and management structure should continue until September 2007, but with the following changes:

Arrangements for the Management Committee to be tightened to ensure effective use of time at the meetings and to ensure that the objectives of the extension are achieved, i.e. rotation of the venue of meetings; assurance that at least one representative from each partner institution is present at all meetings; aside from the UKWAC project coordinator (see below) and the meeting secretary, no more than two representatives from each partner institution to be present at Management Committee meetings;

Project management to be carried out by the British Library UKWAC Project Coordinator whose role should be more clearly defined, and whose functions should include assurance that the Management Committee agendas, decisions and business are carried out to schedule and that the work of any sub-groups is also carried out to agreed timescales.

Other specific roles, e.g. lead technical contact with Magus Research Ltd, to be agreed by partners and to be distributed among partners

Assessment of Other Business Methods for the Future

During the extension period consideration should be given to developing the most appropriate model to fit development from a project into a sustainable business model. A sub-group should be set up to assess potential options and to report back to the Management Committee.

5. Dissemination and usage

The work of UKWAC has generated considerable interest and representatives of all the partner institutions have given presentations on the work of the consortium and web archiving in general. Examples include presentations to the Association of Online Publishers, the Joint Committee on Legal Deposit, the Digital Library Federation, the Legal Deposit Advisory Panel, and various branches of the British Computer Society.

In terms of publicity and marketing, a small investment was made in an UKWAC brand and logo. There was considerable debate on the level of marketing, in particular at the launch of the service in May 2005. The partners were divided over a 'big bang' approach and a light touch dissemination to fellow professionals and information workers. The decision was made to opt for the light touch approach as some partners felt that the relatively small content might attract criticism.

The work of UKWAC was submitted to the Digital Preservation Coalition annual Conservation Awards for 2005 in the category of Digital Preservation and was shortlisted, thereby participating in the award ceremonies and raising the profile of web archiving.

The UKWAC site has a traffic/activity log analysis program to monitor usage. The program – which has been in place since April 2005 – provides a high level review of usage. Between June and August 2005 there was an average of 28,210 unique visits per month to the UKWAC, of which an average of 4,422 unique visits came from the UK. The equivalent averages for September – November 2005 were 16,894 and 1,618 and for December 2005 to February 2006 the figures were 9,195 and 1,345.

This relatively low usage level can be attributed to a number of factors including the relatively small scale of the archive and the lack of marketing. The substantial drops for the later periods can be partially attributed to the same issues but also to a decision made by the Consortium to exclude search engines from the archive.

In October 2005 an end user feedback form was added to the UKWAC site and generated limited interest, i.e. only 35 responses. No statistically meaningful conclusions can be drawn from such a small sample, although it can be noted that for more than 85% of the respondents, the archive met user needs. The report of the survey is attached at Appendix 2.

Success Factors: Limited awareness raising and positive feedback achieved.

Lessons Learned: Need for further UKWAC-related documentation, publicity material.

Recommendation: In the event of continuation of UKWAC, there is a need for a clearly defined communication and marketing plan and accompanying information/publicity material, including milestones for generating publicity.

6. Working Collaboratively

Any success of UKWAC in terms of delivering common frameworks, common agreements etc are due entirely to positive collaborative working. The two-year pilot project has proved to be a major learning experience in the face of a whole range of challenges: technical, legal and operational. The initial goals of the project have been fulfilled largely as a result of collaboration and for each individual and each partner institution, there have been significant benefits in addressing the challenges together. However, working in a consortium raises its own challenges. The benefits of shared learning and shared costs must be offset by the compromises required when working in a shared-decision making environment where decisions are made on a consensus basis. On occasions some partners have adopted a low profile, deferring decision making to other more active partners. As a result decision making has sometimes been less than agile.

On the positive side, specialist groups have worked well, for example, in collection development. The UKWAC mailing list has also been effective in moving operations and decisions forward, although, as stated above, contributions have not been evenly spread.

6.1 Unexpected Issues

The collaborative environment has also made it possible to consider and make decisions on a number of unforeseen issues or questions arising during the pilot. These have included:

Kitemarking: should UKWAC consider providing a kitemark of quality by archiving particular sites?

Validation of sites: issue of sites considering selection of their site to be a validation of quality and adding a statement to this effect to their site;

Concerns of website owners on archived sites being mistaken for current sites:

Issues around robots, exclusions, etc, leading to interventions in and crashing of particular sites, etc.

Site owners did not want the archived version of their sites to appear on Google

Success factors: Learning through working together; effective mailing list; success of specialist sub-groups.

Lessons to be Learned: More agile decision making needed

Recommendations: As part of the review of governance structure, consideration should be given to mechanisms which will facilitate the decision-making process, including the role and need for particular sub-groups. Those proposed to date are:

Digital Preservation
Collaborative Collection Development
Permissions/IPR/Legal issues
Infrastructure
Governance / business models.

7. UKWAC: the future

7.1 Background developments

As the above evaluation has shown, UKWAC has largely fulfilled its original objectives. However, for the reasons given it is not likely to reach the forecast numeric targets in terms of sites gathered and it will not be possible to deliver any software developments.

As the two-year pilot moves into its last phase, there have also been significant developments in legal deposit and the extension to non-print formats. In summer 2005 the Legal Deposit Advisory Panel (LDAP) was set up as a non-departmental public body to recommend to the Secretary of State Regulations required to implement the primary legislation embodied in the Legal Deposit Libraries Act 2003. LDAP has taken an interest in web archiving and has set up one of its working groups to look specifically at this area. Note has been taken of the difficulties associated with obtaining permissions. It may therefore be possible that Regulations for web archiving could be accelerated rather than considered at a later stage after other formats. It is understood that there is likely to be a year between any formulation of regulations and their implementation to allow for consultation, regulatory impact assessment, etc. This could suggest some time in 2007 as the earliest possible date. It is suggested that any decision on the future and structure of UKWAC needs to take this into account, including, also, consideration of the powers which are invested in legal deposit libraries by means of the Act and any collaborative relationships which might be considered for web archiving in the future.

It is also clear that a number of software/platform and tools developments are currently underway or in preparation and that full consideration needs to be given to these in the forthcoming year as a means of decision-making toward a new technical infrastructure for UKWAC.

It is proposed that these developments and their timing are taken into account in any decision on the timing of an extension to UKWAC.

7.2 Original Propositions

Within the original Aims and Objectives document, a number of potential project outcomes were listed for consideration at the end of the pilot:

The project is successful and the resultant common web archiving infrastructure continues with the same contractor

The project is deemed feasible in the long term but continues with a different contractor.

The project is deemed feasible in the long term but continues with a different software solution.

The project is deemed unfeasible over the long term, and is consequently terminated.

Partners may leave the Consortium.

New partners may join the Consortium.

In the light of the full evaluation above and the proposed recommendations at the end of each section, and of discussion between the partners, it is recommended that UKWAC should continue – and that continuation should be with the same contractor, subject to reaching satisfactory contractual agreement - for a further period of one year subsequent to the ending of the Consortium Agreement, i.e. one year following September 30 2006, and that a further review should be carried out at least six months in advance of that date, i.e. by the end of March 2007.

During this period of extension, focus should be on the following key areas:

To develop a UK web archiving strategy, and make recommendations for the future coordination of UK web archiving at national level, taking into consideration the requirements of legal deposit legislation

To produce recommendations for a future shared technical infrastructure, based on evaluation/appraisal of software/platform/tools and various cost/membership models

To achieve these objectives, the following supporting activities are also recommended:

Implementation of some improvements to the governance/management structure

Development of site to move towards original numeric targets

Completion of pieces of work referred to within the various recommendations, according to prioritization agreed by the partners

Preparedness for impact of legal deposit Regulations and definition of role of a UK web archiving infrastructure following the implementation of Regulations

Focus on preservation infrastructure issues and consideration of trusted depository status

No extension of membership during this extension

8. UKWAC Recommendations

The following comprises a list of key recommendations taken from throughout the Report.

Key strategic recommendations

1. That UKWAC should continue for a period of one year, subsequent to the ending of the Consortium Agreement, to September 30 2007
2. That Magus Research Ltd should be asked to continue to provide hosting and support services but ensuring that in doing so UKWAC does not become dependent on this contractor
3. That an interim review of options for UKWAC's future should be begun at least six months prior to the end of March 2007
4. That a formal evaluation/appraisal of web archiving software/platform/tools be undertaken and a recommendation made on the replacement for the current PANDAS application prior to September 30 2007
5. That UKWAC makes improvements to its current governance/management structure and uses the one-year extension period to assess options and make recommendations for a more effective and sustainable structure for the future

6. That during this extension period UKWAC gives assurances that the content of the UK Web Archive will not be placed at risk of loss and in so doing explicitly addresses issues of digital preservation
7. That UKWAC seeks to lobby for early Regulations for secondary legislation to be applied to UK websites.

Predicated on these recommendations being accepted, the following recommendations are also made:

8. During the extension period UKWAC continues to use PANDAS 2 as its core web archiving application
9. That all contract/procurement negotiations be led by the British Library
10. During the extension period the UKWAC Management Committee sets up ad hoc working groups to address specific issues and to provide recommendations to the Committee to agreed deadlines on one or more of the following
11. Write a preparedness report for the impact of legal deposit Regulations and definition of the role of a UK web archiving infrastructure following the implementation of Regulations
12. Identify and prioritise preservation infrastructure issues and recommendations for action
13. Identify and prioritise key developments required to the core archiving application to move partners towards meeting the original numeric targets
14. Identify and prioritise the requirements necessary for the preservation of material in the UKWAC archive
15. Identify and prioritise the steps necessary that would allow the consortium to work towards achieving Trusted Digital Repository status
16. Review the terms and conditions of use of the PANDAS licence agreement and of all agreements put in place as part of the pilot project
17. Assess and report on collection development issues, based on the UKWAC experience and provide recommendations for the future
18. Review the existing topic headings and subheadings within PANDAS and ascertain the extent to which cost effective amendments can be made or provide recommendations for a more effective categorisation system
19. Identify, prioritise and cost the various elements of the web archiving work flow from selection through to cataloguing and preservation
20. Produce clear recommendations as to full metadata requirements for the future and to identify particular areas or issues where further work need to be carried out, e.g. technical and preservation metadata, and in particular how these may be implemented in any application that replaces PANDAS
21. Identify and prioritise recommendations for a collaborative framework across partner institutions for the effective identification and selection of sites that avoids duplication of effort or duplication of contact with creators

22. Identify and prioritise recommendations for the future around alternative approaches to the current permission secured approach to web archiving including an assessment of risk and an audit of best practice in other web archiving initiatives
23. Identify and prioritise options, alternative approaches and recommendations for replacing the current UKWAC management and governance structure giving particular consideration to mechanisms which will facilitate an agile decision-making process, including the role and need for particular sub-groups
24. Identify and prioritise options, alternative approaches and recommendations for replacing the current UKWAC funding structure, giving particular consideration to determine what the average annual costs are, and what would be a realistic membership fee for alternative forms of membership, e.g. full, associate and temporary membership
25. Write a clearly defined communications and marketing plan and accompanying information/publicity material, including milestones for generating publicity
26. Identify and prioritise new end user services that could be built on either the existing PANDAS framework or any other future web archiving system

APPENDICES

Appendix 1: Standard permissions letter and documentation (examples from the British Library)

2.1 Example permission letter

Invitation to participate in Web preservation programme

Dear Sir/Madam

<http://www.site.co.uk/>

The British Library is a founding member of the UK Web Archiving Consortium (www.webarchive.org.uk) consisting of The British Library, JISC (Joint Information Systems Committee), the National Archives, the National Library of Scotland, the National Library of Wales and the Wellcome Library. The Consortium is undertaking a two year pilot project to determine the long-term feasibility of archiving selected web sites.

The British Library would like to invite you to participate in this pilot project by archiving your web site under the terms of the appended licence. We select sites to represent aspects of UK documentary heritage and as a result, they will remain available to researchers in the future. If the pilot is successful the archived copy of your web site will subsequently form part of our permanent collections.

There are some benefits to you as a web site owner in having your publication archived by the Consortium. If you grant us a licence, the Consortium will aim to take the necessary preservation action to keep your publication accessible as hardware and software changes over time.

If you are not the sole copyright owner please pass this request on to the other copyright owners. If you give The British Library permission to copy and archive your web site we will electronically store its contents on a server owned by the UK Web Archiving Consortium. We will also seek to take the necessary action to maintain its accessibility over time and ensure its future integrity. Permission to archive pertains only to the web site specified in this letter.

Please note that the Consortium reserves the right to take down any material from the archived site which, in its reasonable opinion either infringes copyright or any other intellectual property right or is likely to be illegal.

If you are happy for your site to be included in this Web archive please complete the attached copyright licence form and return it to the address given below. For more information about Copyright, the UK Web Archiving Consortium and how your archived web site will be made available please see the attached FAQ document.

Alternatively, if you require any additional information, please do not hesitate to contact me.

Enclosed:

1. Licence Document
2. Frequently Asked Questions

2.2 Example copyright licence

I/We the undersigned grant the British Library, on behalf of the UK Web Archiving Consortium, a licence to make any reproductions or communications of this web site as are reasonably necessary to preserve it over time and to make it available to the public:

Title of Web

site:

URL:

(Please feel free to change the title details as appropriate)

Third-Party Content: Is any content on this web site subject to copyright and/or the database right held by another party? Yes No

Has their permission to copy this content been granted? Yes No

Note that we will not be able to archive this website if you do not have the permissions of all third parties.

Licence granted by:

Name: (block letters)

Position: _____ Organisation: _____

E-mail: _____ Tel: _____

Any other information:

I confirm that I am authorised to grant this licence on behalf of all the owners of copyright in the website; I further warrant that nothing contained in this website infringes the copyright or other intellectual property rights of any third party.

Signature: _____ Date: _____

It would also be beneficial to our project if you could answer the following questions:

Do you presently archive this Web site yourself/yourselfs? Yes No

Details.....

Would you allow the archived web site to be used in any future publicity for the Web Archive? Yes No

2.3 Adapted documentation used by some partners for specific events-based collecting where immediate action is felt to be required

Dear Sir / Madam,

As part of the UK initiative to archive websites of research interest (www.webarchive.org.uk) we are collecting a representative sample of sites relating to the terrorist attacks in London on 7th July.

We would like to archive the site <http://site.co.uk> for a few months to cover the disaster. This tragic event is of major concern nationally and internationally and the British Library feels a responsibility to start archiving such sites as soon as possible. By archiving this material we would hope to preserve information which may help researchers globally in the future in their work on terrorism and other subjects related to this event.

Due to the urgency of the situation with information changing daily, we feel it is necessary to start archiving immediately. However, if you do not wish your site to be available through the web archive hosted by the UK Web Archiving Consortium (see link above), please contact the British Library at web-archivist@bl.uk at your earliest convenience.

We will be notifying you prior to making the sites live. In the meantime, please do not hesitate to get in touch if you have any concerns about the archiving of your site. Please find below the key frequently asked questions for your information.

Appendix 2: End User Survey Form for the UKWAC Web Site - Summary of Responses

Introduction

The 'prize draw' feedback form was promoted on the UKWAC web site during October 2005. The form was promoted on library, archive and digital preservation lists. After a year of UKWAC archiving activity it was probably worth carrying out a feedback exercise and it will be useful to compare these responses with feedback received in the future.

Users were asked to respond to 10 questions asking them about their experience of using the UKWAC web archive. The questions were based on successful user feedback exercises done at the Wellcome Library in the past.

This summary contains charts showing responses to each question, except for questions 6, 7, and 8, which were textual responses. It summarises and collates responses and does not constitute 'proper' statistical analysis.

The feedback form has been left on the UKWAC website and any further responses can be recorded as they arrive.

Summary/Overview of the responses

There were 35 forms completed and submitted. None of the forms was 'spoiled'. Whilst this may not constitute a statistically significant number of responses an interesting picture of how the archive is used begins to emerge. There have been some useful comments made that could form the basis of future developments of the UK Web Archive, such as the provision of an 'advanced' search feature.

From the responses it appears that the majority of those who returned a feedback form classed themselves as 'Information Professionals'. 'Researcher' and 'Just Curious' were jointly the next highest user group. This possibly accounts for the high number (29%) of users who use the archive to support reference/information research enquiries.

Four responses were received from users who are also engaged or interested in web archiving or digital preservation.

Significant numbers of users use the archive for general or leisure purposes – 19% and to support their learning – 18%. 14% used the archive for 'other' reasons. 11% use the archive to support academic research.

Forms were returned from 7 different countries, 4 forms carried no country of origin information, e.g. came from Hotmail type addresses.

62% of forms were returned from respondents in the UK
11% carried no country identification

9% were returned from New Zealand
6% were returned from the USA
3% equally from Australia, Canada, Israel and Singapore

Of the UK responses.

10 came from academic - .ac - e-mail addresses
4 came from Government - .gov - e-mail addresses
2 Came from a health - .nhs - e-mail address
1 came from a commercial law organisation e-mail address –
anonymised as it was the only response in this category
1 came from a commercial - .co.uk - e-mail address
1 came from an organisational - .org.uk - e-mail address
3 came from 'free' addresses

Of all responses,

33% claimed to have found the UKWAC site by 'Other' means
29% claim to use the UKWAC archive to support reference/information
research enquiries
89% said they found what they were looking for in the archive
86% claimed the archive met their needs
54% rated the archive at 4 out of 5
80% of respondents classed themselves as 'Information Professionals'
89% of respondents were not website owners

One respondent has been involved in discussions of web archiving with
the House of Commons Library
One respondent was working with a local archives group
One respondent is using UKWAC as a model, among other web
archives, for developing a web archive in a national library
One respondent was using the UKWAC archive to review other selective
web harvesting practices
One respondent was using the UKWAC archive to monitor digital
preservation activities for libraries and museums in the US
One respondent used the UKWAC archive as an archival record of their
organisations website activities

Questions 6, 7 and 8 asked users to write out their comments in text boxes. It is
not possible to quantify these responses.

General UKWAC web site statistics during this period

For the month of October 2005 there were a total of 8,273 unique visits to the
UKWAC web site. A total of 0.0042% of those visitors responded to the
feedback form. Unique hits on the feedback page were too few to feature in the
reports.

For the month of September there were a total of 7,370 unique visits to the site.

During October the top three visitors to the site by country were North America, Europe and Asia; visitors also came from South America, Oceania and Africa.

Conclusions

35 responses is not sufficient to draw statistically significant conclusions

There were approx a thousand more visits to the site in October than in September but this cannot be attributed to the promotion of the feedback form

The general promotion of UKWAC during this period through the use of assorted lists was a useful if low-key activity that should be continued

There are some useful comments in the answers to questions 6, 7, and 8

The promotion of the feedback form through the use of professional library, archive and digital preservation lists resulted in more 'Information Professionals' responding than other users

However, this may have succeeded in promoting the archive to a user group who use it in their daily work

UKWAC is attracting some international attention amongst other web archiving professionals

Whilst we attract international interest the local library, archives and preservation community may be our primary user group at this time