Annex 4: responses to questions 2.1, 2.2 and 2.3:

2.1 Have you recently completed, or participated in, a digital preservation project (i.e completed during 2003)?

2.2 Are you currently running, or participating in, a digital preservation project or projects?

2.3 Are you currently planning a digital preservation project?

2.1 Have you recently completed, or participated in, a digital preservation project (i.e completed during 2003)?

Arts and Humanities Data Service

JISC funded *Requirements and Feasibility Study on Preservation of E-Prints*; commissioned a review of preservation practices within the AHDS (see also JISC entry below)

The British Library

Domain.uk
Completed BL pilot project to investigate viability of selective web archiving. 2001-2. Funded from existing resources. The deliverables included a small collection of 100 web sites, captured a number of times, and some focused experience in the issues surrounding web archiving.

Sound Archive acetate disks
The transfer of ca 20,000 "acetate" direct cut audio discs to an off-line digital platform (CD-R) during the period 1995-2000. This was an internally funded BL project using external contract labour for the transfers.

Voluntary deposit of electronic publications storage (VDEP)
Yes, we are performing preservation activities to support the collection of voluntary deposit material we have collected ahead of the passage of legal deposit for electronic items. We are using the Ex-Libris product DigiTool to secure this material.

Cambridge University Library

CUL participated in the CAMiLEON project (www.si.umich.edu/CAMiLEON/), led by the Universities of Michigan and Leeds, and funded by the NSF and JISC. (The UK element finished in Dec 2002, the US element in Sept 2003.) The objective was to explore emulation as a digital preservation strategy. See the project website for a list of the expected deliverables.

Council for the Central Laboratory of the Research Councils

We have an ongoing preservation project. This year the hardware upgrade was completed, but the project continues as it has for 20 years

JISC

Information: http://www.jisc.ac.uk/index.cfm?name=programme_preservation, or phone for further detail.

Collections

Web-archiving feasibility study, by UKOLN (University of Bath), and including a study on legal issues to do with web archiving, by the Centre of IT and Law, University of Bristol. Report produced Dec 2002, available JISC website.
JISC e-journal archiving study, by Maggie Jones, final report October 2003, available JISC website.
JISC e-prints study, *Requirements and feasibility study on preservation of e-prints.* Joint study by the Arts and Heritage Data Service, Estonian Business Archives and SHERPA, University of Nottingham. Final report October 2003.
e-science curation audit: *Data Curation for e-Science in the UK: an audit to establish requirements for future curation and provision*, by the Digital Archiving Consultancy. Consultation draft published in May 2003.
Services, Advice and Support
File Format Representation and Rendering Project: *Survey and assessment of sources of information on file formats and software documentation, Final report* produced by the Representation and Rendering Project, University of Leeds, (undated)
Revision of Records Life Cycle Report: a revision of the 1999 *Study of the Records Lifecycle*. New version to be available from April 2003.
Partnership
Digital Preservation Coalition
JISC
ERMS study

Online Computer Library Center

OCLC and RLG jointly sponsored the formation of a working group, composed of international experts from a variety of institutional backgrounds, to address metadata requirements for digital preservation. The group produced a white paper which defined the scope and role of preservation metadata, and reviewed and synthesized a number of existing preservation metadata schema. The group also produced a comprehensive preservation metadata framework articulating the types of information needed to support the long-term preservation of digital materials, along with a set of "prototype" metadata elements. (Completed June 2002)

Another project produced a white paper entitled: "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making" Abstract: Economic issues are a principal component of the research agenda for digital preservation. Economics is fundamentally about incentives, so a study of the economics of digital preservation should begin with an examination of the incentives to preserve. Securing the long-term viability and accessibility of digital materials requires an appropriate allocation of incentives among key decision-makers in the digital preservation process. But the circumstances under which digital preservation takes place often lead to a misalignment of preservation objectives and incentives. Identifying circumstances where insufficient incentives to preserve are likely to prevail, and how this can be remedied, are necessary first steps in developing economically sustainable digital preservation activities. (Completed April 2003)

Oxford University Library Services

The development of the Oxford Digital Library (ODL), a digitisation project (http://www.odl.ox.ac.uk/) with its aim to develop an overarching infrastructure

with planned content creation, management, and delivery, has developed standards for ensuring long-term access to the digital content created through its programmes. Operational from July 2001, a project of continuous development. The ODL has been funded from within Oxford University, and from the Mellon Foundation.

The National Archives:

Digital Archive: see website; www.pro.gov.uk/about/preservation/digital/. The Digital Archive at the National Archives stores electronic government records and makes these records available for the first time to all who wish to consult them.

PRONOM: see website. PRONOM is an application developed by the Digital Preservation Department at the National Archives for managing information about the file formats used to store electronic records, and the software applications needed to render these formats. It will be available via the internet from October 2003.

BBC Domesday Commmunity:

Objectives: To migrate the 1986 BBC Domesday Community disc data onto a Microsoft Windows PC platform and make it accessible once more.

Funding Source: The National Archives

Partners: BBC, LongLife Data Ltd, ATSF

Deliverables: Archive-quality DigiBeta tapes storing BBC Domesday Project video and image data. BBC Domesday Community application for presenting 1986 BBC Community data to the public once more at The National Archives at Kew.

UK Data Archive

This has been an ongoing process for the Data Archive for over 20 years.

ULCC

DELOS/NSF expert study on digital preservation research agenda – Mar 2002 – autumn 2003. Joint NSF/EU funding (5th Framework).

Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation (2003)

Margaret Hedstrom, University of Michigan

Seamus Ross, HATII, University of Glasgow

Kevin Ashley, University of London Computing Centre

Birte Christensen-Dalsgaard, Statsbiblioteket (Denmark)

Wendy Duff, University of Toronto
Henry Gladney, HMG Consulting
Claude Huc, French Space Agency
Anne R. Kenney, Cornell University
Reagan Moore, San Diego Supercomputer Center
Erich Neuhold, Fraunhofer (Darmstadt)
Prepared for:

National Science Foundation's (NSF) Digital Library Initiative
&
The European Union under the Fifth Framework Programme by the Network of
Excellence for Digital Libraries (DELOS)

The project aimed to survey current research activities, identify gaps, and develop a white paper proposing future research directions in the area of digital preservation. Some of the potential areas for research include repository architectures and inter-operability among digital archives; automated tools for capture, ingest, and normalization of digital objects; and harmonization of preservation formats and metadata. Also opportunities for development of commercial products in the areas of mass storage systems, repositories and repository management systems, and data management software and tools.

ICA workbook on electronic records management (2001-2004). ICA funding + volunteer effort from members. Aims: practical handbook for archivists, primarily in government, and also for IT staff.

Wellcome

The Wellcome/JISC Web Archiving study. See: http://library.wellcome.ac.uk/projects/archiving.shtml. See also JISC entry above.

2.2 Are you currently running, or participating in, a digital preservation project or projects?

AHDS

JISC funded Hybrid Archives Project, running 1 Feb 03 to 31 Jan 05. The project aims to '…..develop and test a hybrid model for depositing institutional assets that provides a bridge between the complexity (and burdens) of full 'traditional' deposit of institutional collections and the more simplified approach embodied in harvesting methodology. The Hybrid Archives model would take elements of both traditional deposit methodology and harvesting methodology and re-work them to produce an integrated and cohesive deposit model…..'. Also described as '….. an extensible and sustainable 'partial deposit' model that would allow host institutions to retain control and dissemination of their collections, while simultaneously employing OAI and Z39.50 technologies to integrate rich, complex metadata into AHDS web-based search systems.' AHDS, working with the Courtauld Institute, the Theatre Museum and Royal Holloway.

JISC and CURL funded SHERPA project. 'SHERPA – Securing a Hybrid Environment for Research, Preservation and Access', is hosted by the University of Nottingham, a three year project for completion October 2005. The project will work 'by establishing 'e-print archives' which comply with the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI PMH) using eprints.org software. The contents of these e-print repositories (and the harvestable metadata) will be freely available to all the HE and FE community

and beyond. The creation, population and management of the repositories will be the core of the Project. Learning outcomes from these activities will also be shared with the community in a variety of ways, including the establishment of mechanisms to assist proactively other institutions in setting up and managing repositories.' The project also aims to work in conjunction with publishers, eg OUP. In conjunction with the Universities of Leeds, Sheffield and York and the British Library and Oxford University Library Services. http://www.jisc.ac.uk/index.cfm?name=project_sherpa&src=alpha

internal development of the AHDS Digital Repository

BBC

We are in year four of an archive preservation project, and some of the material being transferred to modern formats is digital (DAT audio tape, D3 videotape).

The British Library

SHERPA: (see AHDS entry on SHERPA above)

pre & post print server to capture independent BL scholars' research outputs. CURL collaborative project. 2003-6. Aim is 750 papers.

Selection for web archiving
A small pilot project within the BL which tested the feasibility of rapidly archiving British and other related websites on one topic. The case study focused on websites promulgating information and opinion about the conflict in Iraq. This involved selection of representative sites in UK, Ireland and further a field, co-ordination between English and Arabic-language selectors within the Library, liaison with the Imperial War Museum, the issuing of permission request letters, and the logging of responses from content-owners. This pilot started in 2003 and is likely to be incorporated into other web archiving activities being conducted (see below).

LOCKSS
We are participating in the beta testing stage of the LOCKSS, a project with Andrew W. Mellon foundation funding, to build production quality software to archive and preserve e-journals. http://lockss.stanford.edu/projectstatus.htm.

The Digital Object Management (DOM) programme
We are just embarking on a programme to put in place systems and processes to manage all the types of digital material received and created by the British Library - the Digital Object Management (DOM) programme. The programme will address issues of digital preservation indirectly – e.g. consideration of appropriate formats and metadata in the collection / creation of material – and directly in that preservation requirements will be assessed and functionality planned. As part of this programme we are also conducting a risk analysis of digital materials in the Library's collections to address any urgent preservation needs.

The programme will encompass several projects listed below.

a) VDEP2 (extension to voluntary deposit of electronic publications storage)
Currently planning for VDEP2, which will allow us to handle ongoing current voluntary intake. It is becoming clearer from progress of legal deposit

legislation that it may be several years before we have any compulsory replacement for the voluntary scheme.

b) Storage of digitised masters
A separate storage project has been started for digitized master images. Digitised masters will be retained due to the investment in their creation, their future use as part of the Libraries digital collection and due to the significant costs of re-digitisation at a later date. Lossless compression will be used to store the images on DVD-R with a simple tracking mechanism to record the location of each file.

c) Risk analysis of digital materials
This work will produce an assessment of risk posed to digital materials, a ranked order of predictive life expectancy of digital materials and a hierarchy of vulnerability and urgency allowing work priorities to be set.

Manuscript collections containing digital material
The Department of Manuscripts has begun the collection of private papers which include born digital material. It is assumed that many such collections acquired in the future will be partly or wholly digital. A full time Curator of Electronic Manuscripts was appointed as from 1 September 2003.

e-literature project
The Department of Manuscripts is also trailing the capture of literary manuscripts and email with the assistance of an external commercial partner. This includes examining possible conversion of email to and storage in XML formats.

PDF document collection
This involves downloading and archiving of PDF documents at the BL. This involves systematic current awareness, issuing of permissions letters, downloading, storage on a local drive, cataloguing, and making available in the Library's reading room. It is not specifically digital preservation oriented project but will be likely to feed into the DOM system (see above) and allow the material to be saved for the long term.

Sound Archive Betamax transfer
The digital to digital transfer of ca 9,000 Betamax (PCM F-1 digital audio) tapes to an off-line digital platform (CD-R) during the period since 2001. This is an internally funded BL project using external contract labour for the transfers.

1   Web Archiving Programme
A major project is the Web Archiving Programme: a programme to put in place the necessary infrastructure and procedures to ensure that the British Library, in partnership with others, can play a strategic role in extending its responsibility for the national published archive to electronic media including the web.

Web archiving project partners: British Library, JISC, National Library Scotland, National Library Wales, The National Archives, Wellcome Trust

It is the intention that the British Library should build on the pilot Domain.UK project (now called Britain on the Web) undertaken in 2001/02 and take a

leading role in collecting the UK web presence in partnership with other institutions nationally and internationally. Given the huge scale and dynamic nature of the web (currently more than 4 million UK-based websites), the Library does not consider it practicable nor affordable to aim at truly comprehensive coverage of the UK Web presence. The approach will be selective.

The programme will encompass several projects listed below.

a) PANDAS implementation project
This is a two-year project during which production web crawling is scheduled to start in 2004. It will involve at least the following partners: the British Library, the Joint Information Systems Committee (JISC), National Library of Wales, National Archives and the Wellcome Library – and will use a licensed copy of the National Library of Australia's PANDAS software to collect and manage web sites according to a defined selection policy. Rights to use individual sites will need to be cleared from rights-holders and the service will be hosted through a contracted third party. This project will deliver an agreed number of web-sites which will be made accessible to users. It is expected that 450-500 sites can be collected each year by each partner, although the provision of additional BL staffing resource could double this to approximately 1,000 sites per annum for the BL. The Collection Development Policy will form the basis of the initiative, which will test many aspects of working in partnership. The JISC will focus on its own sites, the Wellcome on sites of medical interest, the BL on sites at the core of its collection policy, including a selected subset of the 100 sites from the Domain.uk pilot, a selected set of sites associated with the Iraq War, a number of UK government sites based on the informal UK government documents archive initiative being undertaken in SPIS, and further sites of national and cultural UK importance as identified through the Collection Development Policy. (See also the National Archives web-archiving entry, below).

b) International Internet Preservation Consortium (IIPC)
This project, led by the Bibliotheque Nationale de France, aims to develop a sophisticated automated web crawler mechanism which will deliver to the partners a set of open-source automated tools which can be used to search the web at regular intervals, selecting relevant sites matching an agreed collection development policy. The partners includes several national libraries from Europe, North America and Australasia and the Internet Archive based in the USA. This encompasses all the international institutions with the most experience in this field to date. The work will include definition of a basic crawler, the minimum functionality required for larger libraries such as the BL to crawl their entire country domain, developing intelligent selective collection, and other functionality required by more than one partner. Subject to procurement requirements, the completion is scheduled for December 2005.

c) BL web site
The Library has systematically archived its own site since 2002, archiving monthly. There are some earlier archives, but these have not been systematic or in depth. The archive is held on a local drive within the Library's Web

Management department, but it is intended that there will be a different approach to this in future.

CURL digitisation task force
The BL is a member of the CURL Digitisation Task Force, first meeting mid September, which will look at digital strategy and digital preservation issues in its remit.

Cambridge University Library

The DSpace@Cambridge project (www.lib.cam.ac.uk/dspace) is a 3-year collaboration between CUL and MIT Libraries, ending in December 2005 and funded by the DTI-sourced Cambridge-MIT Institute (CMI), to establish a digital institutional repository for Cambridge University. The project's expected deliverables include the development of improved digital preservation functionality in the DSpace software platform.

CMI is funding a complementary CUL-MIT Libraries project, LEADIRS (www.lib.cam.ac.uk/UKseminars/), to promote digital institutional repository strategies in UK HE/FE. This will include coverage of preservation issues.

CUL is also participating in the LOCKSS project (http://lockss.stanford.edu/index.html) and is currently reviewing the extent of its future involvement.

Council for the Central Laboratory of the Research Councils

See 2.1. Data storage facility is managed by David Corney, for the benefit of CCLRC and some other Research councils and external organizations. Time frame is from 1983 – ongoing. Objectives are to provide a data archive facility to our users which disconnects them from technology or data storage formats – they always have access to their data. Funding is part PPARC and partly funded by users of the service.

In addition all our data centres and our Neutron spallation source (ISIS) are involved in ongoing active data preservation.

Persistant Archiving Research Group of the Global Grid Forum.

Electronic Science Core Programme

Working with JISC to establish the Digital Curation Centre

JISC

Collections
eLearning materials study, a feasibility study tendered in June 03.
CSR Digitisation Programme (sustainability)
Services, Advice and Support
Digital Curation Centre
Community Call (supporting institutional records management)
Partnership
Digital Preservation Coalition
JISC
ERMS implementation in JISC Executive

Online Computer Library Center

PREMIS: OCLC and RLG sponsored a second preservation metadata working group to build on the work of the first group. PREMIS (Preservation Metadata: Implementation Strategies) is a working group of international experts from libraries, museums, archives, government agencies, and business. The goal of PREMIS is to establish recommendations/best practices for implementing preservation metadata in digital preservation systems. PREMIS will 1) define a "core" set of preservation metadata elements, with a supporting data dictionary; and 2) identify and evaluate strategies for encoding, storing, managing, and exchanging preservation metadata. (Expected to run through June 2004)

Registry of Digital Masters: The purpose of the Registry is to provide a place for institutions that have created (or are otherwise responsible for) digitized versions of traditional printed monographs and serials to record:

- what specific items have been (or are about to be) digitized;

- where they can be accessed;

- the specifications followed in digitization.

OCLC Digital Archive: Production system providing services to state libraries, federal agencies and academic libraries for long-term access, management and preservation of important digital materials. Web archiving tools preserve assets such as HTML pages and PDFs from the Web; batch archiving service provided for multiple submissions of reformatted materials. Workflow integrated with OCLC's Connexion service for creation of bibliographic and preservation metadata. Bibliographic records integrated into FirstSearch and local catalogs.

OCLC Preservation Policy: The policy will outline OCLC's basic approach to the preservation of objects and metadata in OCLC Digital Archive. The policy will be one result of the activities of OCLC's preservation planning entity. The policy will be dynamic in nature and will be updated frequently. We will encourage comments on and questions about our preservation processes.

OCLC Digital Preservation Education and Planning: Consultative, educational, and outreach services for digital preservation. Includes workshops, grants and funding aids, and consulting services**.**

Natural History Museum

Periodic snapshots of the Museum's Website are taken.

Oxford University Library Services

In addition to the ODL developments mentioned in 2.1, the Library has also been a partner to the JISC-funded SHERPA project, led by the University of Nottingham.

The National Archives

Digital Archive (phase 1B)
Objectives: To develop a secure repository for storing electronic government records, together with substantial preservation metadata. To allow these records to be searched and viewed in the public reading rooms at The National Archives, Kew.

Web Archive

Objectives: To provide a web archiving service for selected UK central government websites (launched as *The UK Central Government Web Archive* in September 2003)
Timeframe: July 2003 – August 2004
Funding: The National Archives
Partners: The Internet archive
Deliverables: This service will comprise the development of a new crawl system, harvesting of the selected websites at agreed frequencies, delivery of the collected sites to the National Archives for storage, preservation and research, and provision of web access to the collection via the Internet Archive website. The initial contract will be for one year. (see also BL web archiving programme, above)
PRONOM 3:
Objectives: Web version of PRONOM on internet
Timeframe: July-Sep 2003
Funding: The National Archives
Partners:
Deliverables: Web version of PRONOM on internet

The National Library of Scotland

Working in collaboration with Edinburgh University Library to procure a digital object management system (DOMS) to bring all digital resources together through a single accessible interface. We would like the system to create an AIP or SIP using METS. The timeframe is 1-3 years for different aspects to be integrated within the system. We have discussed the proposals with the National Library of New Zealand and Getty Research Institute.

The National Library of Wales

NLW's own web archiving pilot project, Web Archive Wales, commenced on 1 September 2003, and will run for one year.

The Library is also a partner in the UK Web Archiving Consortium, a two year project (see BL and TNA entries, above).

UK Data Archive

Such projects are part of our core business.

ULCC

ERPANET (EU commission funded)

EVAMP (E-ten project on access to digitised photographs – digital preservation element in creation of sustainable metadata and image formats.)

Wellcome

Participation in the UK Web Archiving Consortium, which has been established to undertake a distributed Web archiving project. (see BL and TNA entries, above).

2.3 Are you currently planning a digital preservation project?

Arts and Humanities Data Service

Funding bids in progress (unspecified)

BBC

Planning for web archiving.

The British Library

The definition phase of the DOM programme will define a number of projects that will ultimately create an infrastructure to support the long term preservation of all the digital materials that we possess. For example, 99% of the ca. 30,000 hours of audio acquired p.a. by the British Library Sound Archive is born-digital. Under our general remit we are required to make medium and long term provision for the preservation of this material

Cambridge University Library

CUL has submitted an Expression of Interest in possible partnership with the proposed JISC/e-Science-funded Digital Curation Centre.

Council for the Central Laboratory of the Research Councils

We are bidding for the National Digital Curation centre and are planning data preservation for the DIAMOND facility.

JISC

Services, Advice and Support

Tender Invitation (Evaluating ERM systems: ERM Toolkit)

Tender/Community Call (evaluating ERM systems: MLE & Open Source; and Digital Preservation)

Service Development

JISC e-Journal archiving

Partnership

Digital Preservation Coalition

Web Archiving Consortium

Capture of JISC project websites

The National Archives

Collaborative UK Web Archiving Project

National electronic Library for Health

Making digital archiving a requirement of all our suppliers.

National Library of Scotland

PANDORA Digital Archiving System (PANDAS) project – we will be responsible for web archiving selected Scottish websites in conjunction with the British Library and the National Library of Wales. Unsure what specific details are.

Online Computer Library Center

We plan to work with state libraries to improve harvesting of government documents, including tools/techniques for content analysis and metadata extraction/creation. This work will also test interoperability across repositories.

Oxford University Library Services

We are prepared to be a subsidiary partner in the Digital Curation Centre, focusing on the ODL as a testbed, and working in collaboration with other Oxford stakeholders, especially Oxford eScience.

We are also planning a digital archives project focusing on eManuscripts, and looking in particular at institutional web archiving and email archiving.

UK Data Archive

Plans to participate in DCC.

ULCC

Yes, but no detail given

Wellcome

Consortium (UK web archiving) has been formed – but we wont start harvesting Web sites until Jun 04