Digital Strategy: European Perspectives

Paul Ayris
Director of UCL Library Services and UCL Copyright Officer

Abstract
The purpose of this paper is to examine the role and content of digital strategies in European research libraries. The paper will look at current developments in UCL (University College London) in the UK in terms of strategic directions at an institutional level. It will then examine how UCL Library Services is developing its own strategy to meet these institutional objectives. The paper will look at changing digital architectures, which reflect the new emphasis on student-centred learning and research. Three types of e-content will be examined, and the success or otherwise of libraries in delivering this type of content examined. Open Access is a major new development in terms of the dissemination of academic outputs and recent developments in Europe will be studied. The final section of the paper will look at the concept of digital curation and attempt to examine two European developments which will shape future work in this area. The paper ends by suggesting some tentative conclusions on the success or otherwise of European digital strategy development and implementation.

UCL and UCL Library Services

Figure 1: UCL (University College London)

UCL (University College London) is the third oldest university in England, after Oxford and Cambridge. It was founded in 1826 as the original University of London.[1] It is one of the research-intensive universities in the UK and, as such, a member of the Russell Group of universities.[2]

A major development in European Higher Education is the arrival of league tables which rank universities' performance under a range of criteria. The league tables are important because they are highly visible and are used by students, undergraduate

---

[1] See http://www.ucl.ac.uk/.
[2] See http://www.russellgroup.ac.uk/.

and postgraduate, to influence their choice of university and course of study. European universities clearly wish to be ranked amongst the best in the world and so world rankings are important.

World University rankings

1	Harvard University US
2=	**University of Cambridge UK**
2=	**University of Oxford UK**
2=	Yale University US
5	**Imperial College, London UK**
6	Princeton University US
7=	California Institute of Technology (Caltech) US
7=	University of Chicago US
9	**UCL (University College London) UK**
10	Massachusetts Institute of Technology (MIT) US

Figure 2: Top 10 world University Rankings 2007

This listing is taken from the Times Higher Education – QS World University Rankings for 2007.[3] The Table shows the top ten universities. 6 of the top 10 are based in the US, and 4 are in Europe – actually in the UK. In the top 100, Australasian universities figure strongly. The Australian National University comes in as 16th, the University of Melbourne as 27th, the University of Sydney 31st, the University of Queensland 33rd=, Monash 43rd, the University of New South Wales 44th, the University of Auckland 50th, the University of Adelaide 62nd, and the University of Western Australia 64th.

Libraries, key support structures in a Higher Education setting, have a vital role to play in supporting and enhancing a University's performance in such league tables. Strategy development underpins such a requirement and UCL's Library Strategy for 2005-10 is the key strategic document in the Library which has been developed to deliver the Library's Mission and Vision.[4]

The Library Strategy is informed by 10 objectives:

1. Learning and Teaching Support (paras. 13-25)
2. Research Support (paras. 26-64)
3. Supporting the Student Experience (paras. 65-85)
4. E-Strategy Development (paras. 86-94)
5. Widening Access and Participation (paras. 95-99)
6. Fundraising Activities (paras. 100-104)
7. Partnerships (paras. 105-112)
8. Developing Library Services' Profile outside UCL (paras. 113-122)
9. Continuing Staff Development (paras. 123-129)
10. Planning, Resourcing and Communication (paras. 130-135)

Of the three top priorities which the Strategy identifies, one is key to the argument of this paper:

---

[3] See http://www.topuniversities.com/worlduniversityrankings/results/2007/overall_rankings/top_100_universities/.
[4] See http://www.ucl.ac.uk/Library/libstrat_may05.shtml.

➢ Identification of an E-Strategy to underpin all the Library's e-developments, particularly in relation to E-Learning

In particular, the Library's E-Strategy was identified as key in the following areas:

➢ Teaching and Learning
➢ Research
➢ Student experience
➢ Partnership working

What form does the Library's E-Strategy take and how does it underpin UCL's Mission?

UCL Library Services' E-Strategy
At the start of the work on the Library's strategy, UCL Library Services had a traditional ICT-type strategy which emphasised the need for systems development and inter-operability.
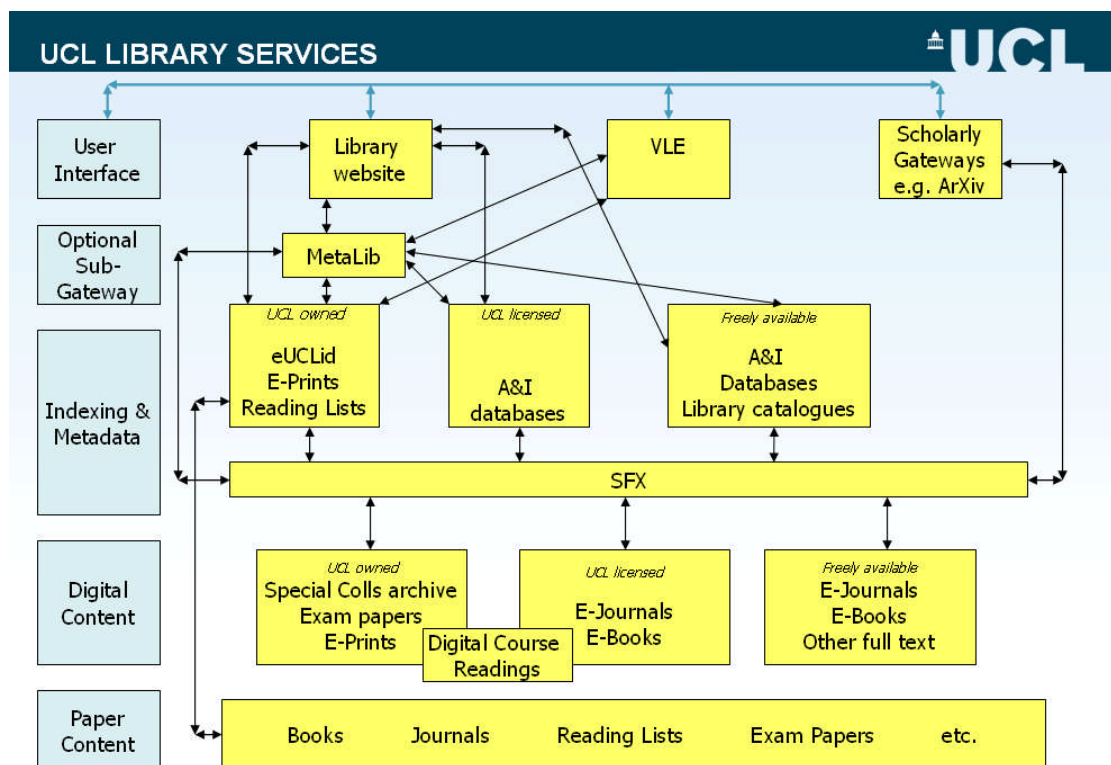

Figure 3: Interconnected digital services in UCL Library Services

The diagram shows five elements which make up the varying levels of the ICT (Information and Communications Technology) architecture for UCL Library Services; these are listed on the left-hand side of the diagram in blue. The user interface, with optional gateway, is followed by the indexing and metadata layers, which themselves are supported by digital and paper content. The Library's electronic systems and silos of data are then mapped against these elements and levels of inter-operability identified. In one way, the diagram is already quite revolutionary because the Library's paper content sits at the bottom of the structure.

During the course of the lifetime of the strategy, it became clear that the diagram needed to be revised. UCL's requirements do not fit all the modules which are available. Federated searching is not widely used by UCL academics and students. A comprehensive E-Learning platform is missing from the structure and, to support the raft of Benchmarks and identify Key Performance Indicators to measure the Library's success, library staff at all levels need to be able to generate management reports and information about given work practices. There are also new systems and services which the Library wants to introduce – for example a digital preservation and curation service. There needs to be join-up with the institution's campus-wide systems such as Student systems, Finance systems, and Human Resource systems. Also, the Library needs to take a view on how much of its information provision is made locally, and how much is derived from services which are available on the network. As a result of all these developments, the Library has begun to move to a different model for the articulation of its E-Strategy, a model which looks something like this.
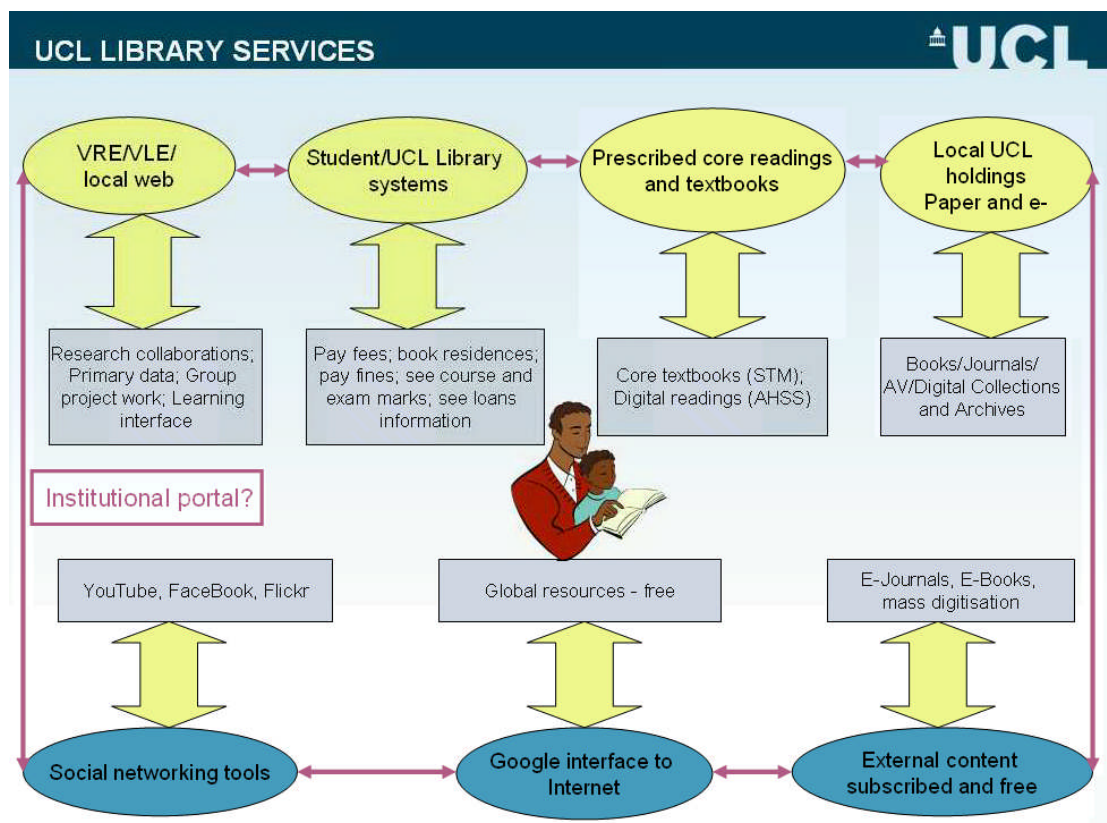

Figure 4: Emerging model for UCL Library Services' E-Strategy

Here, it is the user who stands at the centre of the strategy, not the Library or the Library's systems. The user may want to do a number of different things. He/she may be a researcher wanting to look at primary data, or to work in collaboration with others. Here he/she will want to use the Virtual Research or Learning environments or the local web. He/she may be a student wanting to pay fees or fines, and so will need to interface with a number of separate institutional systems. He/she may be an undergraduate wishing to look at digital textbooks or core readings. They may also want to look more widely at UCL's paper and digital holdings. These will all be institutional holdings or activities, and as such are represented in the top half of the diagram.

However, our user will also want to use materials that are generally available on the network. He/she will want to use social networking tools like Facebook and Flickr. There is the host of free resources available on the Internet, which our user will probably discover via Google. There will also be external content, subscribed and free, which the user will want to use – e-journals, e-books and the output of mass digitisation programmes.

In UCL, we want to try and make all these resources and services accessible to the user via an integrated front-end and so we are looking at creating an institutional portal which will be the first port of call for our user who needs to access all these services and content 24 x 7 on- and off-campus. I am pleased to be able to report that it is the Library which is chairing and leading on the portal development, in term of steering the production of a proof-of-concept system.

There is a key strategic question which deserves further consideration. The British Museum Reading Room in Great Russell Street in London presents a traditional model for library services. The Library pulls the reader into library spaces and all reading and library activities take place in a traditional library building. However, in a networked environment, the Library is just one content provider amongst an immeasurable number of such providers. In this type of environment, the Library must go to the user and push information to where the user wants to be. In part, the Library already does this. In Science, Technology and Medicine, researchers in UCL rarely set foot in a traditional library building. Digital material is pushed to them electronically at their desktop. Should this happen for teaching and learning too? Should the Library push material out to where the student is (e.g. YouTube)? And will an institutional portal be helpful in providing a one-stop shop for our user to discover and locate both local and remote content through one front-end?[5]

Figure 5: Information Behaviour of the Researcher of the Future

There is one final interesting question which needs to be posed in terms of the future development of digital strategies. The CIBER Group in UCL have recently released a report entitled *Information Behaviour of the Researcher of the Future*.[6] The Report shows that all age groups, not just young people, share the traits of what is known as the 'Google Generation'. However, when looking specifically at young people, the Report made the following conclusions. Young people

- ❑ Rely heavily on search engines
- ❑ View rather than read

---

[5] I am grateful to Lorcan Dempsey who, on a visit to UCL on 11 April 2008, helped our Operational Planning Team consider this important question.
[6] See http://www.bl.uk/news/2008/pressrelease20080116.html.
.

❑   Do not possess the critical or analytical skills needed to assess the information they find on the web

This has serious implications if materials in any university are being viewed and not read, and the quality of the materials being studied is not understood. What is happening and what is the nature of the learning experience? There needs to be further work to answer these questions. In terms of Information Literacy and Fluency programmes. It may well be libraries have a role to play in terms of the programmes that they offer in this area.

E-Journals
In one sense, there is not much to say about E-Journals. E-Journal delivery to the desktop is now standard. Our user, who appears in *Figure 4,* will have access to a very significant range of titles in digital form from research libraries. He or she will be able to access them on-campus – in the Library or in the lab – and off-site, In the UK, priorities seem to be switching to the purchase of complete publisher backfiles. It is often suggested that scientists do not read older material, but this is an exaggeration. Researchers in subjects such as mathematics, chemistry and earth science will commonly make user of older materials.

The move to electronic delivery has led to an interesting dilemma. What happens to the paper copy? Can paper copy of <u>current</u> titles simply be cancelled? Certainly in my own institution, researchers do not come into the Library to consult current titles. They are actively encouraged not to, as we prefer to push the material out to where they are. Digital preservation is far from being embedded in university libraries in Europe, and it is a risk to cancel current paper copies in this situation. There <u>are</u> some examples of sustainable digital preservation services. The E-Depot[7] in the Royal Library in The Hague, Netherlands, is a leading example of a national approach to digital curation. The E-Depot has undertaken to digitally curate commercial e-journals on behalf of Dutch libraries. There are also commercial solutions such as Portico,[8] where libraries can subscribe to the Portico digital preservation service which is offered to member libraries. Another attempt at digital preservation is the LOCKSS service,[9] where libraries enter into partnership to share the load of digital preservation using the LOCKSS software.

A particular challenge posed by the growth of e-journals is the prevalence of multiple back-runs of paper copies in university library stores. This is really looking like a nonsensical arrangement, when there are <u>significant</u> savings to be had by de-duplicating back-runs of paper holdings on a regional or national basis. This is the thinking behind the UK Research Reserve, which is a project being funded in part by HEFCE (Higher Education Funding Council for England), to provide a copy of last resort, with an electronic document delivery option.[10] Arrangements for second and third copies around the UK are being overseen by SCONUL.[11] It seems likely that, in the UK at least, the de-duplication of paper holdings across the sector is a possible outcome.

---

[7] See http://www.kb.nl/dnp/e-depot/e-depot-en.html and http://www.kb.nl/dnp/e-depot/e-depot-en.html.
[8] See http://www.portico.org/.
[9] For the UK implementation of LOCKSS, see
http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/programme_lockss.aspx.
[10] See http://www.curl.ac.uk/projects/CollaborativeStorage/Home.htm.
[11] See http://www.sconul.ac.uk.

E-Books
E-Books are a different entity altogether. In an attempt to understand the challenge of E-Books, UCL's School of Library, Archive and Information Studies is working with UCL Library Services on the SuperBook project.[12]

> ❑ 'With e-books available directly from anywhere on or off campus, and portable readers capable of holding more than 100 books, the traditional academic library will need to examine the way it manages and delivers book collections. It is the users who will drive the e-book story forward; and, unlike earlier formats, no one is watching the users of this new breed of 'super books'

The final Report of the SuperBook project will be available in the Summer of 2008, but I am very grateful to Dr Ian Rowlands of UCL SLAIS for a look at the interim findings, which he presented for me at an International Summer School for World Bank Librarians at King's College Cambridge in September 2007.

In a survey of users (n=759), the surveyors asked users what was the most popular form of content in terms of E-Books. The top three categories of content were:

❑ E-Textbooks (58.9%)
❑ Reference Books (52.4%)
❑ Research monographs (46%)

When looking at the age profile of E-Book users (n=1818), the survey found that there was a significance in the range of ages reported. Most users were aged 17-45.

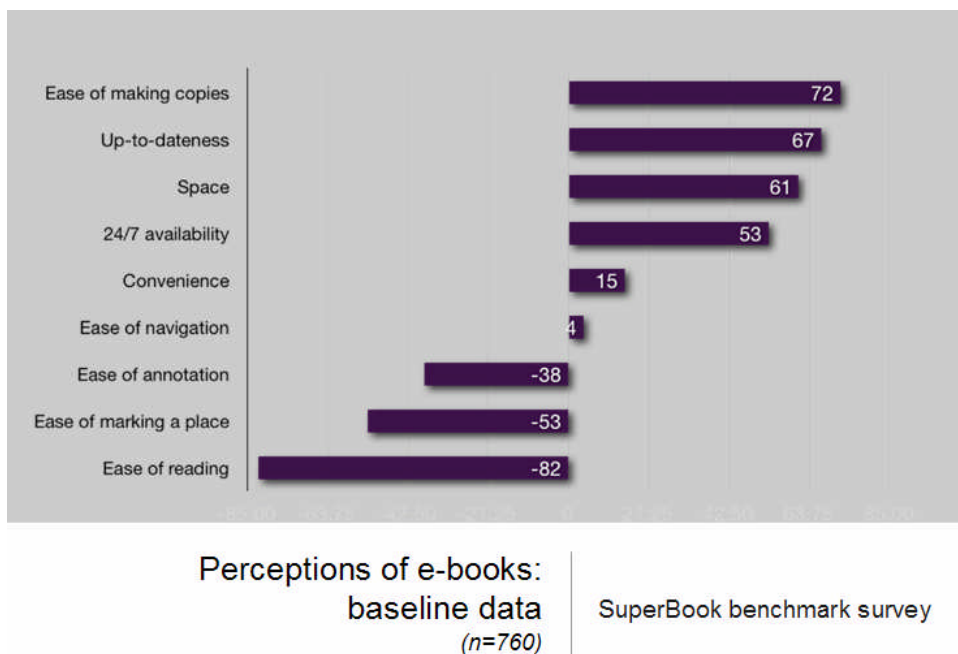A very significant finding was the value attributed to E-Books by users. This is

Perceptions of e-books: baseline data (n=760) SuperBook benchmark survey

Figure 6: Perceptions of the strengths and weaknesses of E-Books

---

[12] See http://www.ucl.ac.uk/slais/research/ciber/superbook/

recorded in *Figure 6* above. The columns to the right show levels of satisfaction with E-Books, while the columns to the left show levels of dissatisfaction. Interestingly, the most significant weakness is reported as ease of reading. If this is really the case, what is the future of mass digitisation projects such as Google Books? The findings of the SuperBook study seem to show that users do not want to read large amounts of text on a screen.

Using the technique of deep log analysis, the researchers tried to identify how users actually used the content they found. In this case, 52% of all pages viewed were pages which contained full text content. UCL Library Services also attempted to trace the routes by which users accessed content, using one of the commercial packages available in the study as a case study. The Library found that the single most important gateway for access to E-Books was actually the Library's catalogue, with 38% of users coming in this way; only 21% came in via Google.

E-Books are not yet embedded into library provision. There still remains a significant amount of work to be done to determine suitable Business models for e-Book purchase, when much of the monograph purchasing market is dedicated to sales to individuals. There are a host of technical issues. How is the mass of available content to be located and made available for discovery? Ideally, records should be de-duplicated; and the results of a search should be FRBRized (for e- and paper copy) where different editions will be retrieved and presented to the user in one search. In terms of technical developments, whose role is it to do this – vendors, third parties, libraries? Metadata standards for E-Book description also need to mature. Users will want to have access, not necessarily to whole books, but to chapters and sections of materials. Standards need to be agreed to effect this type of use.

Mass digitisation of content

In Europe, there is a growing interest in the digital delivery of library-style content. This is manifesting itself in a growth in digitisation projects on a European scale. A new focus for this activity is Europeana, a portal for the digitised output from galleries, libraries, archives and museums (GLAM) plus film and sound content.[13] The aim of Europeana is to enable access to 2,000,000 digital objects by July 2009. The prototype of the Europeana portal will be launched by Viviane Reding, European Commissioner for Information Society and Media, in November 2008. Funding for this work is being provided as part of the eContent*plus* programme,[14] which is itself a deliverable of the EU's i2010 policy.[15]

Pan-European activity requires considerable co-ordination. In the libraries sector, this will be undertaken by CENL (Committee for European National Libraries) on behalf of the national libraries[16] and LIBER (Association of European Research Libraries).[17]

The architecture which this innovative partnership will use is based on portals. Europeana will be the public front-end through which the connected user, anywhere in Europe, will access the richness and diversity of the digitised holdings of Europe's libraries. Using the OAI-PMH protocol for metadata harvesting, the Europeana portal will be fed by metadata from two supporting portals – TEL (the European Library) for

---

[13] See http://www.europeana.eu.
[14] See http://ec.europa.eu/information_society/activities/econtentplus/index_en.htm.
[15] See http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm.
[16] See http://www.nlib.ee/cenl/.
[17] See http://www.libereurope.eu.

national libraries and the LIBER portal (for Europe's research libraries). An overview of the architecture is given in *Figure 7* below.

At the centre of the vision for Europeana is the European citizen. EC-funded digitisation activity is not primarily aimed at the European researcher or student, although they will indeed benefit from the digitisation being undertaken. Rather, the aim of the work is to provide content which gives substance to the emerging identity of the European citizen. The EU is offering 50% funding for this work under the eContent*plus* call, which itself represents the first time that the EU has systematically offered money for the digitisation of content.
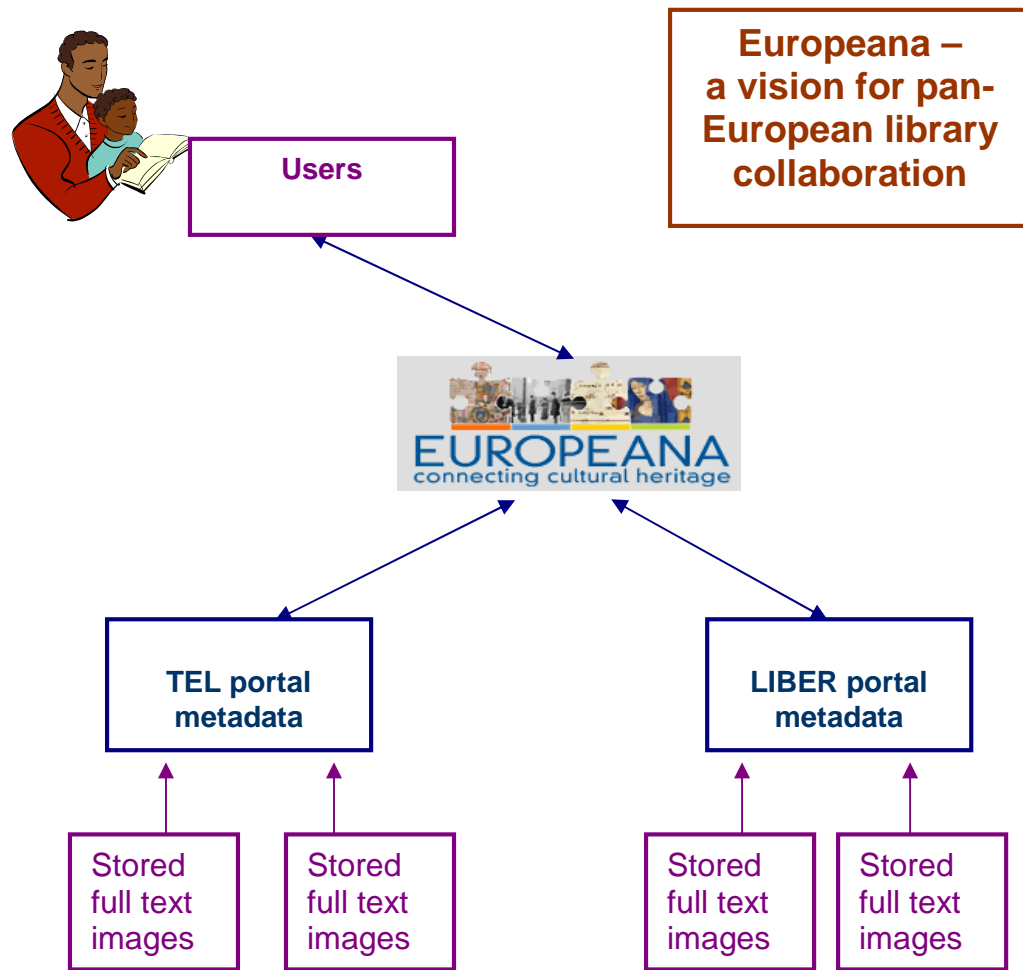

Figure 7: High level architecture for library input into Europeana for local digitisation of content

The collaboration, if it is successful, would represent the first such pan-European partnership between Europe's national and research libraries. It is an important model, not just for Europeana, but for future collaborations across the library sector in Europe. Should the project be successfully funded, the material which is made available via Europeana will help define the concept of European citizenship and identity. It is an exciting vision, and libraries are playing a part in making this a reality.

Digitisation Frameworks

As well as collaborations with Europeana, LIBER is undertaking a second line of activity via its Access Division. In October 2007, LIBER and EBLIDA held a pan-European Digitisation Workshop in Copenhagen to survey the European digitisation landscape. 95 attenders from 23 European countries formulated 25 recommendations,[18] which LIBER and EBLIDA have been asked to discuss with the Office for the EU Commissioner for Information Society and Media. The recommendations are aimed at furthering the concept of a European framework for digitisation activity.[19] The issues which the Digitisation Workshop identified can be classified as:

- ❑ Vision for European digitisation activity
    - ❑ the need for joined up thinking
- ❑ Content
    - ❑ the need for European selection criteria
- ❑ Resource discovery
    - ❑ the need for portal development
- ❑ Copyright and IPR
    - ❑ particularly around licensing and orphan works
- ❑ Standards and policies
    - ❑ the need for registries of identifiers and metadata standards
- ❑ Business Models
    - ❑ pricing and costing models; need for further EU funding
- ❑ Digital Preservation
    - ❑ definition of role and responsibilities

There is a need, across Europe, for a roadmap to guide European digitisation activity. Currently, individual countries and libraries digitise resources at will, and there is little (if any) co-ordination over the selection of content. This leads to duplication of activity and potential waste of resources around the sector. If Europeana is to become a focus for European digitisation activity, then it needs to construct a framework for future joined-up work. Central to this work is the use of selection criteria to guide the selection of materials for digitisation. Such matrixes do exist. The present author, for example, constructed such a matrix in the late 1990s,[20] but the use of such decision-making tools to guide activity is not yet widespread.

Digital preservation is a key requirement to support the long-term digital curation of digital assets. It is irresponsible to create digital resources and not to preserve them digitally for the long-term. Digital preservation is a building block of the new digital environment. In Europe, at least, there needs to be greater clarity over roles and responsibilities. There are indeed EU-funded projects such as PLANETS,[21] which are investigating the technical requirements of this activity, but there is a need in Europe

---

[18] See http://www.libereurope.eu/node/142.
[19] See http://www.libereurope.eu/node/284.
[20] See http://eprints.ucl.ac.uk/492/.
[21] See http://www.planets-project.eu/.

for a definition of digital preservation frameworks for undertaking digital preservation as a service. Developments such as the E-Depot at the Koninklijke Bibliotheek in The Hague[22] are pioneering in terms of offering the Dutch Higher Education system a national service for digital preservation. LIBER has established a Memorandum of Understanding with the E-Depot.[23] The two bodies are working in collaboration, but the actual digital preservation of digitised materials from European libraries does not currently form part of that collaboration.

Open Access
Open Access is a well-established movement in Europe. There are a number of well-founded national repository developments. SHERPA in the UK is the leading British example[24] of collaborative Open Access work. There are strong regional examples too such as the SHERPA-LEAP consortium[25] in the federal University of London, led by UCL. Here 13 institutions work together, funded in part with generous funding from the Vice-Chancellor's Development Fund, and gain from mutual support and a sharing of ideas in terms of repository development. There are also pan-European repository developments: the EU-funded DRIVER project, which currently has 13 partners[26] is attempting to construct a pan-European Open Access repository infrastructure including a portal, which will provide a one-stop search service for Europe's Open Access repositories.

The European Universities Association (EUA) has recently issued an important set of recommendations in the area of Open Access. The EUA endorsed Open Access on 26 March 2008.[27] The recommendations, which are contained in the EUA's endorsement, are aimed at a number of stakeholders.

Recommendations for University Leadership

❑ The basic approach for achieving this [Open Access] should be the creation of an institutional repository or participation in a shared repository
❑ University institutional policies should require that their researchers deposit (self archive) their scientific publications in their institutional repository upon acceptance for publication
❑ University policies should include copyright in institutional intellectual property rights (IPR) management
❑ University institutional policies should explore also how resources could be found and made available to researchers for author fees to support the emerging "author pays model" of open access

---

[22] See references in footnote 7 above.
[23] See http://www.libereurope.eu/node/151.
[24] See http://www.sherpa.ac.uk/.
[25] See http://www.sherpa-leap.ac.uk/.
[26] See http://www.driver-repository.eu.
[27] See http://www.eua.be/index.php?id=354.

Recommendations for National Rectors' Conferences

❑ All National Rectors' Conferences should work with national research funding agencies and governments in their countries to implement the requirement for self archiving of research publications in institutional repositories and other appropriate open access repositories
❑ National Rectors' Conferences should attach high priority to raising the awareness of university leadership to the importance of open access policies in terms of enhanced visibility, access and impact of their research results

Recommendations for the European University Association

❑ EUA should continue to contribute actively to the policy dialogue on Open Access at the European levels with a view to a self archiving mandate for all research results arising from EU research programme/project funding

These are seminal recommendations for European Higher Education, because they stem from a body which is itself representative of European Higher Education institutions. The guidance is simply advisory and there is no compulsion for individual universities to issue a mandate for their academics to deposit their academic outputs in an Open Access repository. The statement does, however, create a framework which can be adopted at an institutional level. The EUA also made a recommendation to itself – to encourage dialogue within Europe so that all EU-funded research can be made available in Open Access. It is a bold vision, and one which is now supported by a major body of European Vice-Chancellors.

Individual pan-European Open Access repository projects

There is a small, but growing, number of pan-European Open Access repository projects. NEEO is an EU-funded project comprising many of the economics research-intensive universities in Europe.[28] The objective of NEEO is to populate its members' institutional repositories with economics research and working papers from academics in the member institutions. Through the harvesting of metadata via OAI-PMH, NEEO is able to present the results for discovery through a central portal.

DART-Europe is a pan-European collaboration devoted to the collection and dissemination of electronic research theses.[29] The architecture is divided into three layers, or platforms, and is illustrated in *Figure 8*. At the institutional level, individual universities or, preferably, national e-theses consortia load the full text of their research dissertations into their institutional repositories. Where there is a national consortium, the most common approach would be for that consortium to harvest the metadata from local national repositories into a national portal. DART-Europe then harvests from each of these platforms into the European, or DART-Europe portal. It is via this portal that the researcher will locate and retrieve content.

---

[28] See http://www.nereus4economics.info/neeo.html.
[29] See http://www.dart-europe.eu.

At the time of writing, the DART-Europe portal (DEEP) has 34 partners from all over Europe and provides access, via harvested metadata, to 88,699 research theses. The partnership is as follows:

- *The partnership now*

❑ Selected universities from the UK, Ireland, Switzerland and Hungary
❑ National/regional consortia from the Nordic Countries, Catalonia, French-speaking Belgian universities, Germany

- *Coming soon*

❑ Switzerland, France, regional consortium from Italy
❑ DRIVER partners and countries, including the Netherlands, Flemish-speaking Belgian universities
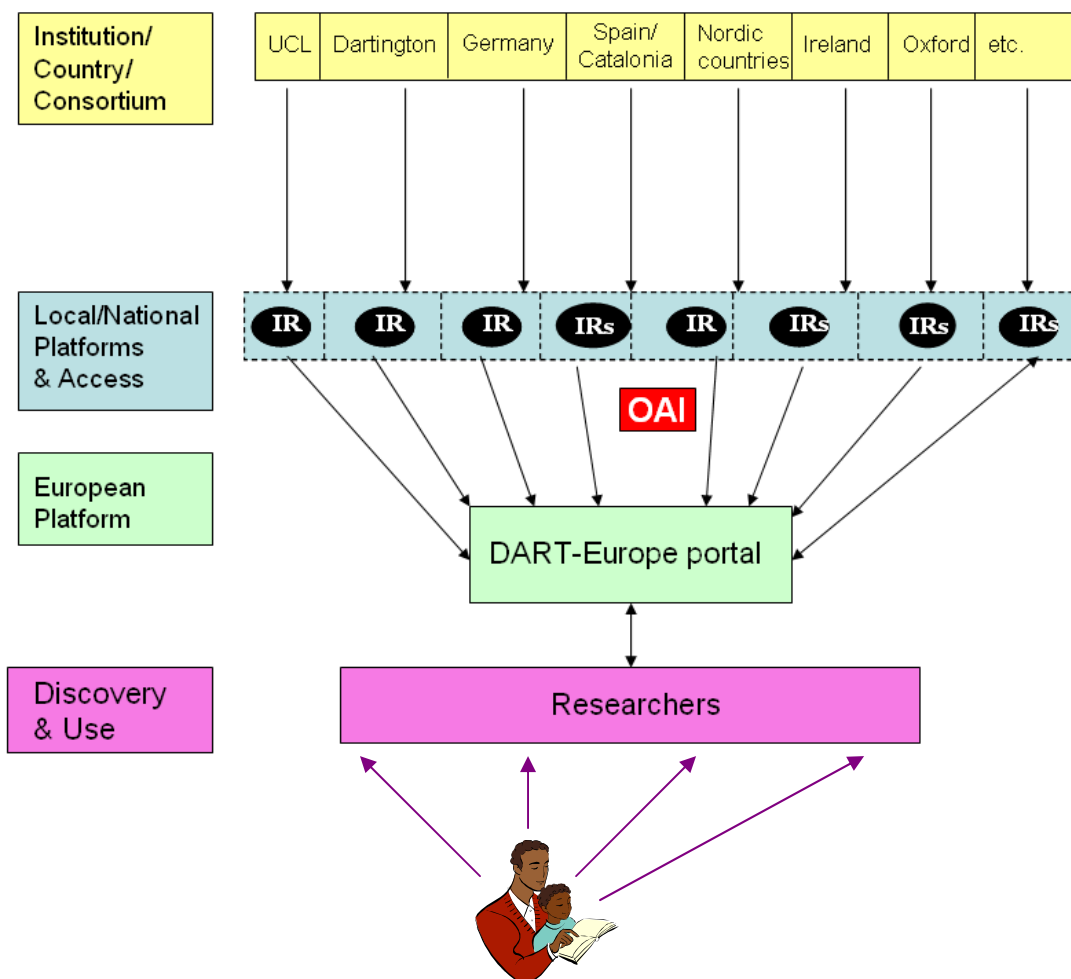❑ Austria

Figure 8: the DART-Europe architecture for e-thesis discovery

Why do research theses matter? *Figure 9* shows why as it records the top 10 downloads from UCL's Open Access E-Prints repository for a sample month. The numbers on the left-hand side indicate the ranking of a particular download.

Bibliographic details are shown in the central column and the number of downloads is given in the right-hand column.

Items 5, 7 and 8 are all Ph.D theses and they received monthly download figures of 131, 126 and 124 downloads respectively. These figures show that the electronic version of some research theses receive large numbers of consultations. The figures are far larger than the figures for consultations of the paper equivalents, which are also bibliographically hard to track down. Electronic access gives research theses great visibility. This is good for research and good for the young student, who requires visibility at the start of their career. There are also implications for the commercial publication of research theses. A good print run for a Ph.D. thesis would be 400 copies. Repository downloads are already showing many electronic consultations. Could it be that commercial publication of research theses will be replaced by dissemination via Open Access repositories?
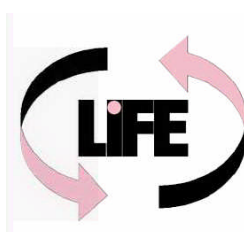
1	Cairns, S. and Sloman, L. and Newson, C. and Anable, J. and Kirkbride, A. and Goodwin, P. (2004) _Smarter choices - changing the way we travel_. Research report. Department for Transport, London, UK.	554
2	Washer, P. and Joffe, H. (2006) _The hospital 'superbug': social representations of MRSA._ Social Science and Medicine, 63 (8). pp. 2141-2152. ISSN 02779536	176
3	Adell, German (1999) _Theories and models of the peri-urban interface: a changing conceptual landscape_. Literature review. Development Planning Unit, UCL, London, UK.	174
4	Conroy Dalton, R. (2001) _Spatial navigation in immersive virtual environments_. Thesis (Doctoral.PhD), UCL (University College London).	157
5	Kalra, D. (2002) _Clinical foundations and information architecture for the implementation of a federated health record service._ Thesis (Doctoral.PhD), UCL (University College London).	131
6	Vaughan, L. and Penn, A. (2006) _Jewish immigrant settlement patterns in Manchester and Leeds 1881_. Urban Studies, 43 (3). pp. 653-671. ISSN 00420980	128
7	Miller, William L. (1995) _The picture and the letter: male and female creativity in James Joyce's Finnegans Wake_. Thesis (Doctoral.PhD), University of New South Wales.	126
8	Thomson, Catherine Claire (2003) _Danmarkshistorier: National Imagination and Novel in Late Twentieth-Century Denmark_. Thesis (Doctoral.PhD), University of Edinburgh.	124
9	Cassar, M. (2005) _Climate change and the historic environment_. Centre for Sustainable Heritage, University College London, London, UK. ISBN 0954483065	103
10	Vaughan, L. and Clark, D.C. and Sahbaz, O. and Haklay, M. (2005) _Space and exclusion: does urban morphology play a part in social deprivation?_ Area, 37 (4). pp. 402-412. ISSN 00040894	98

Figure 9: Top 10 downloads from UCL's E-Print repository, January 2007

Digital Curation

LIFE
LIBER is pursuing an important European project concerned with identifying the costs of long-term digital preservation.

The LIFE project, **L**ifecycle **I**nformation **F**or **E**-literature, is a JISC-funded collaboration between LIBER's Access and Preservation Divisions. It represents the first such

attempt that LIBER has made to gain project funding to explore development issues on behalf of the library sector in Europe.

Digital preservation is an important issue for research and national libraries, which comprise the LIBER membership. In an analogue environment, it is these libraries which are responsible for the preservation of materials which are used by the researcher. The same should be true in a digital environment, if only because few other stakeholders will undertake such preservation activity.

One of the issues which libraries have to face is that of the actual cost of digital preservation. It is this issue which the LIFE project is addressing.

LIFE is a collaboration between the British Library and UCL Library Services. Two phases of activity have currently been funded. The first has reported, and the outcomes of that work are summarised here and described much more fully in the extensive project documentation which is available from the LIFE website and in the UCL E-Prints repository.[30] Phase 2 of the project will report in the summer of 2008 and Phase 3 remains to be funded.

LIFE has taken as its starting point the need to analyse the existence of a digital object for the whole of its lifecycle from birth to death or for its permanent existence. The lifecycle can be represented in a formula and LIFE's lifecycle formula appears below as *Figure 10*.

$$L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$$

L is the complete lifecycle cost over time 0 to T. Other categories are

Aq = Acquisition,
I = Ingest,
M = Metadata,
Ac = Access,
S = Storage,
P = Preservation

Figure 10: Lifecycle Formula from LIFE

Here the lifecycle of a digital object is broken down into six stages and these are represented as separate elements in the formula as listed above. The lifecycle of a digital object can be said to comprise six elements: Acquisition of the object, Ingest into a storage/preservation system, Metadata creation, provision of Access to that object, Storage of the object and Preservation of the object.

---

[30] See http://www.life.ac.uk; also http://eprints.ucl.ac.uk/1855/ and http://eprints.ucl.ac.uk/1854/.

Lifecycle Element	Acquisition	Ingest	Metadata	Access	Storage	Preservation
Element 1	Selection (Aq1)	Quality Assurance (I1)	Characteri-sation (M1)	Reference Linking (Ac1)	Bit-stream Storage Costs (S1)	Technology Watch (P1)
Element 2	IPR (Aq2)	Deposit (I2)	Descriptive (M2)	User Support (Ac2)		Preservation Tool Cost (P2)
Element 3	Licensing (Aq3)	Holdings Update (I3)	Administrative (M3)	Access Mechanism (Ac3)		Preservation Metadata (P3)
Element 4	Ordering & Invoicing (Aq4)					Preservation Action (P4)
Element 5	Obtaining (Aq5)					Quality Assurance (P5)
Element 6	Check-in (Aq6)					

Figure 11: Detailed breakdown of each element within the lifecycle formula

Having identified these basic elements, the LIFE team went further and broke down each of these elements into sub-elements, to show the individual activities which support each element. This analysis is given in *Figure 11*. Not all elements have the same number of sub-elements. Storage has only one sub-element, that of bit-stream storage costs, whereas acquisition has six sub-elements. The identification of the sub-elements is important because it is these activities, using activity-based costing methods, which determine the costs of the whole lifecycle of a digital object, of which digital preservation is part.

To determine the costs of each of these sub-elements or activities, the LIFE team looked at the costs associated with each activity – largely hardware/software costs and staff costs pro rata (looking at the amount of time each member of staff, at their respective grade, spent on that activity). To help assess staff costs, each staff member involved in the LIFE case studies kept a diary where he/she allocated the number of hours and minutes they spent on each activity. Using their salary costs, activity-based costings could be produced; fuller details are given in the LIFE project Report.[31]

The LIFE team used three case studies as exemplars to test the LIFE lifecycle formula and approach: electronic journals acquired by UCL, the Web Archiving project run by the British Library and the British Library's VDEP (Voluntary Deposit of Electronic Publications). The results of the case study for Web Archiving are given below in *Figure 12*.

Category	Percentage of overall cost (10 year average)	Average cost per instance archived	Average cost per new title	Cost per title after 1 year	Cost per title after 5 years	Cost per title after 10 years	Cost per title after 20 years
Aq	14%	£17	£16	£108	£475	£934	£1,852
I	16%	£21	£0	£111	£557	£1,114	£2,229
M	0%	£0	£4	£4	£4	£4	£4
Ac	0%	£1	£1	£4	£15	£30	£57
S	8%	£10	£0	£54	£270	£539	£1,078
P	62%	£81	£0	£426	£2,127	£4,255	£8,509
Total	100%	£130	£21	£707	£3,449	£6,876	£13,731

The per title costs for 1, 5, 10 and 20 years are based on the average cost per title, combined with the cost of gathering a number of instances of that title. On average the Web Archiving team aims to gather just over 5 instances of each title per year. In reality titles are gathered at different frequencies depending on the nature of the title in question. These figures do not include numbers for web sites which close or remain unchanged.

Figure 12: Results of the Web Archiving case study from LIFE Phase 1

Here, the elements of the lifecycle formula are given in column 1 of the left-hand side of *Figure 12*. The associated costs for each element are given in the succeeding columns, including costs for the lifecycle for 1, 5, 10 and 20 years. So, over 20 years, the lifecycle cost for each archived website amounted to £13,731, of which 62% of the costs can be attributed to preservation. Using the lifecycle costing formula, it is now

---

[31] See references in footnote 30.

possible to identify preservation costs and to compare these with the total lifecycle costs.

LIFE was able to establish a number of headline costs for digital preservation and lifecycle management, and a selection of these is given here for objects where the lifecycle is 10 years:

❑ The lifecycle cost for a hand-held e-monograph is £48
❑ The total lifecycle cost for a hand-held e-serial is £14 per issue
❑ The total lifecycle cost for a non hand-held e-monograph is £30
❑ The total lifecycle cost for a non hand-held e-serial is £8 per issue
❑ The total lifecycle cost for a new website is £6,800

All these are predictive costs and further work needs to be done to firm up the methodology and costings. LIFE Phase 2, which is reporting in the summer of 2008, will firm up the economic modelling with advice from an academic economic consultant, widen the range of case studies to include Open Access repositories and the outputs of digitisation projects, and compare the costs of analogue and digital preservation.

UKRDS – UK Research Data Service

There is a growing recognition amongst research-intensive universities in the UK of the importance of primary data to support e-Science and e-Research. Currently, there is little national infrastructure to support the long-term preservation of such digital objects. The recent closure of the Arts and Humanities Data Service, which did have a role in the longer-term preservation of the research outputs (including primary data) for the Arts and Humanities community, underlines the fact that there is no pervasive system in place in the UK for the long-term preservation of digital data.

With this in mind, Research Libraries UK (RLUK – formerly CURL)[32] and the Russell Group IT Directors (from the research-intensive universities)[33] have issued an Invitation to Tender for a Feasibility Study to look into the development of a shared digital research data service for UK Higher Education institutions. Funding of £200,000 is being provided for the Study from HEFCE, the Higher Education Funding Council for England. At an institutional level, there is uncertainty about the costs involved in managing large data volumes and the availability of a suitably-skilled workforce to manage the new challenges posed by data curation. The Feasibility Study will address the need not just for storage capacity but for active management of the creation, selection, ingestion, storage, retrieval and preservation of research data – the data lifecycle.

There are many stakeholders in this endeavour:

❑   Research Councils and Research Councils UK (RCUK)[34]

---

[32] See http://www.rluk.ac.uk.
[33] See http://www.rugit.ac.uk.
[34] See http://www.rcuk.ac.uk.

❑   Department for Innovation, Universities & Skills (DIUS)[35]
❑   Joint Information Systems Committee (JISC)[36]
❑   Higher Education Funding Councils[37]
❑   Individual Universities

International developments in data curation will inform the Feasibility Study, and the UKRDS team are already making contacts with cognate services and developments around the world.

Conclusions

This study of digital strategies, with a particular reference to developments in Europe, has covered a range of issues and it may be possible to draw a number of conclusions.

There can be little doubt that the role of libraries is changing, and that a substantial reason for this change can be ascribed to the developments in Computer and Information Technology (C&IT). In terms of institutional architectures, libraries need to move from traditional systems-based architectures to approaches which put the user at the centre of their development work. Inter-operability is essential. The balance between local and networked delivery is changing and it is now possible to view the institutional library as just *one* source of information that users will want to use. Are users equipped for the brave new world? If the UCL SLAIS Report on the information gathering behaviour of the researcher of the future is correct, then there is cause for concern. Is this not an opportunity for libraries to re-new their efforts in offering pervasive Information Literacy and Fluency programmes to tackle a perceived need?

This paper also looked at the current perceptions of E-Books by the user community. If the perception that E-Books are not easy to read on the screen is true, does this mean that mass digitisation projects for printed book content are a waste of time? It is probably too early to draw this conclusion, but further research needs to be done. Certainly at a European level, there is growing interest in the mass digitisation of content and Europe is beginning to evolve an infrastructure to support pan-European developments in this area.

The policy statement on Open Access by the European Universities Association has delivered a powerful tool which the supporters of Open Access can use to embed Open Access delivery into institutional and subject offerings. It is too early to say how successful the EUA's initiative will be. Certainly, there are pan-European Open Access developments which are beginning to use the technology to deliver new services to their users.

In terms of digital curation, the LIFE project is establishing a generic costing model for the lifecycle curation of digital assets, which includes digital preservation. The UKRDS (UK Research Data Service) is also being scoped and costed. This would represent a major new development for Higher Education, if the Feasibility Study were to lead to a service for UK researchers.

---

[35] See http://www.dius.gov.uk/.
[36] See http://www.jisc.ac.uk.
[37] See http://www.hefce.ac.uk.

Figure 13: the Donaldson Reading Room, UCL Library Services

Digital delivery is now pervasive. *Figure 13* shows the nineteenth-century Donaldson Reading Room in UCL Library Services. It is the original nineteenth-century Library in UCL. Today in 2008, each study place is provided with power for student laptops and the whole Library is networked with radio networking. This traditional Reading Room is one of the most popular places to work in UCL Library Services, a place which both provides the monastic quiet that these students want and one which is also a highly-networked environment. Alongside the Library's provision of traditional analogue materials, and new digital formats, students can access the Virtual Learning Environment, the Web, Internet, e-mail and gain access to a global collection of digital resources and services alongside the Library's own provision – all of which UCL's students use to support their learning and research. It is this need that should drive Digital Strategy in the future – a vision where the user sits at the centre of everything he/she can see and access.