

JISC Digitisation Conference
Draft Report

Cardiff, 20 - 21 July 2007

Executive Summary

The JISC Digitisation Conference was held at the St David's Hotel and Conference Centre in Cardiff on 20/21 July 2007. It gathered together some of the leading digitisation projects, funding-bodies, publishers, archives, libraries and many of the key thinkers in the area. There was an international delegate list, drawing in representatives from the UK, France, Germany, Italy, the US, Canada and elsewhere.

The aim was to discuss the key issues affecting those engaged in digitisation and draw conclusions about how best to take these issues forward. Numerous topics were proposed, debated and argued over through the course of the two days; below are the five issues which most regularly surfaced during the conference.

1. Reintegrating the User

A persistent theme was the need to re-focus on the end user. Speakers highlighted how easy it is to lose sight of the key reason for digitisation – providing your audience with knowledge. The complexities of mass digitisation and the rush to ensure a successful process can sometimes result in a compromised final product. Too many projects have confusing or half-baked interfaces, obscuring the content they have worked so hard to digitise.

Successful technology, it was emphasised, should be invisible. End users have no need to know how a TV or a washing machine works, and the same is true of putting content online. Digitisation projects should not forget to document its processes, but behind-the-scenes concern for data capture, disclosure, aggregation, and discovery should not overshadow how the final outputs are presented to the end users.

2. Building a Mass of Content

An echoing lament throughout the conference concerned the 'silo effect' – digital content continued to be created in separate silos, unconnected to other relevant content locked away elsewhere. Users, therefore, were continually at the mercy of the idiosyncrasies of search engines and also their own patience in locating and exploiting relevant content.

New strategies are needed. The wishes of specific user communities need to be recognised and responded to but within a larger need to aggregate and present content (both new *and* existing) in particular areas that can resonate with multiple communities. Equally, development of tools and standards must continue so as to catalyse technical interoperability among digital content.

3. The Shadow of Google and YouTube

Google and YouTube have done marvellous things in making content freely available. There is much the digitisation community can learn from their expertise, as existing partnerships are already showing.

However, the public sector needs to remember the differing perspectives of the web giants. Values such as the insistence on scholarly quality, the importance of provenance and metadata, and the necessity of sharing and preserving digital data are aspects that the public sector arguably values more than the private sector. The educational communities therefore must be prepared to negotiate in the light of their own values.

4. Business Models and Sustainability

Experience has replaced early naivety - maintaining content online requires infrastructure, expertise and hard cash. The digitisation community needs to develop and deploy sophisticated business

models to enable it to support the content it has so carefully digitised. This will also have the advantage of focussing the digitisation community on what content it really values and what content it is prepared to put to one side.

No firm answers are presently available. Business models must be experimented with, both with and without collaboration with the private sector. The present climate offers an excellent window of opportunity for engagement with commercial bodies. However, it is vital that information on the success or otherwise of such partnerships is shared within the wider community.

5. Increased Collaboration

Addressing all the issues cited above requires greater collaboration between all the relevant participants - publishers, collection curators, funding bodies, user communities, vendors and standards bodies. Therefore the most urgent call from the conference was for all these stakeholders to picture how they slot into, respond to and profit from the larger sphere of digitisation. It was not a call to start from scratch - frameworks, agreements and associations already exist - but to cultivate existing partnerships and foster new ones where needs be.

Collaboration should not happen simply for its own sake – otherwise the lack of flexibility it engenders may damage final outcomes. However, when applied within the correct strategic backdrop, collaboration will help projects reach further end-users, break down silos of digitised material and provide sophisticated business models to ensure long-term access to digital content.

Further Directions: Next Steps for JISC

Any large-scale conference not only identifies common themes but signals new directions to follow; in this case the next steps that the digitisation community should be following.

For JISC such steps have guided the way to its Digitisation Strategy. The conference gave JISC the opportunity to launch a draft strategy and receive discussion from conference delegates. Now complete, the strategy illustrates JISC's aims and priorities for the second phase of its digitisation programme (www.jisc.ac.uk/digitisation), due to complete in Spring 2009. It is part of a wider and integrated e-content strategy to improve access to digital resources within the higher and further education sector. This is done in coordination with other strands of JISC work, such as e-content licensing through JISC Collections and the Digital Repositories programme.

As a result of this strategy and as part of its digitisation programme, JISC is taking these issues forward through a number of activities and initiatives that are being planned or currently underway.

1. Emphasising User Focus in Creating Digital Collections

Highlights of current JISC actions

- Creating digital resources that allow for interaction, personalisation and contextualisation
- Continuing to refine a rights framework for the flexible use of digital material within the educational community

The strategy places particular importance on emphasising user focus in creating digital collections, and the projects that form part of the programme will offer a range of functions that allow interaction with, and personalisation of, a resource. Collections will have a strong emphasis on the creation of learning packages to accompany the digitised material, thus providing a richer contextual background for a variety of users primarily within the higher and further education sector and, when possible, schools and lifelong learning.

In addition, JISC's development of an appropriate legal framework that governs the creation, delivery, use and re-use of digitised material within an educational context will ensure that educational users can freely engage with the resources while respecting Intellectual Property Rights in a manner that is neither inflexible nor too restrictive.

2. Developing a Critical Mass of Content

Highlights of current JISC actions

- Funding specific projects to bring dispersed digital content together
- Researching user needs in general and in specific areas, e.g. Islamic Studies
- Updating accepted digitisation standards to assist with continued technical interoperability

The needs of specific user communities are driving a number of initiatives that JISC will carry out in order to break down “content silos”, enhance resource discovery, and make the most of content that is already currently available on the Web but not necessarily joined together.

A number of projects within the digitisation programme are already bringing together dispersed resources within a particular theme and that are in demand by academic users (e.g. Irish studies). JISC will also fund the creation of portals in specific subject areas that will harvest metadata from relevant collections and allow cross-searching from a single access point. This will be complemented by further studies to ascertain gaps in the provision of digitised content. Further investigation into users’ needs will also consider the interest in, and potential for, digitisation of resources in particularly topical subject areas such as Islamic Studies.

The adoption of technical standards for the capture, description, and preservation of digital assets that also fosters interoperability among resources is a strong element of the JISC digitisation programme. More work is planned to update the JISC technical standards catalogue, developing it to include, where feasible, an assessment of the costs, benefits and challenges of employing such standards.

3. Creating High-quality Content for the Educational Sector

Highlights of current JISC actions

- Digitising a broad range of resources in physical formats currently ignored by other mass digitisation programmes.
- Integrating high standards and professional metadata into all digitisation projects to ensure academic quality is maintained

Most of the resources digitised through the JISC digitisation programme include special and unique collections spanning a range of formats (newspapers, images, photographs, maps, audio-visual material, population data, manuscripts and ephemera) and centuries. As discussed during the conference, such collections privilege a “quality” approach, often require a high degree of curatorial input, and perhaps demand more sophisticated business models for their sustainability than those offered by initiatives such as Google Print and YouTube.

4. Researching Viable Business Models

Highlights of current JISC actions

- Via the digitisation programme, exploring a range of possibilities for sustaining digital resources
- Undertaking international research to ascertain the feasibility of different models

Current projects within the digitisation programme are experimenting with a variety of business models which include public-private partnerships, the use of existing services and infrastructure for the delivery of content, or ensuring a smooth transaction from project to service within an institution’s own “business as usual” remit.

Projects are still testing viable models and JISC is committed to sharing the outcome of such experiments with the wider community by capturing and disseminating the lessons learnt and examples of best practice that emerge from the digitisation programme.

In coordination with the digitisation programme, the work of the Strategic Content Alliance (SCA) (<http://www.jisc.ac.uk/contentalliance>) in particular addresses key issues such as sustainability models, IPR and the need for collaboration in order to ensure seamless access to publicly funded e-content for all users. The results of such research will be invaluable for the digitisation community.

5. Undertaking National and International Collaboration

Highlights of current JISC actions

- Working with the US National Endowment for the Humanities to fund transatlantic digitisation projects
- Collaborating with the Quality Improvement Agency to contextualise digitised resources for the further education audience in the UK

The call for increased collaboration, both at national and international level, resonates in particular with JISC's current and future work. Partnerships are already a key component of the current digitisation programme – the collaboration with the Quality Improvement Agency (QIA) aims to broaden access to resources targeting in particular the further education sector. While partnership with the National Endowment for the Humanities (NEH) has resulted in a transatlantic digitisation initiative that brings together UK-US collections for the benefit of international researchers. During the two days of the conference seeds were planted and JISC will continue to pursue new and meaningful collaborations with both public and private initiatives.

Next Digitisation Conference: 2009

JISC is planning a repeat of this successful event - currently slated for summer 2009. This will once again bring together key players in digitisation from different sectors and keep the momentum going in debating, and exploring possible solutions to, critical issues within a fast changing e-content landscape.

In the mean time, we urge those in to digitisation community to continue engaging with JISC, in the hope that discussion of ideas and plans for collaboration can continue to unfold.

<http://www.jisc.ac.uk/digitisation>

List of contents

JISC Digitisation Conference, Cardiff, 20 - 21 July 2007.....	1
Executive Summary.....	2
Further Directions: Next Steps for JISC	3
Next Digitisation Conference: 2009.....	5
Reports from Plenary Sessions - 19 July 2007	8
Welcome to Wired Wales.....	8
Introduction to the Day	8
JISC Strategy and Mass Digitisation	9
Mass Digitisation and its Impact on Cultural Heritage, Education and Research Communities	10
Unlocking e-Content and Enhancing Education and Research Opportunities: An academic perspective.....	11
Roles and Priorities for Research Libraries in the UK and US in the Delivery of e- Content.....	12
Reports from Plenary Sessions - 20 July 2007	14
Cyber-infrastructure and information policy: the changing face of US digitisation developments.....	14
The JISC Digitisation Programme: an introduction to the digitisation strategy and new collections coming on stream from 2009	15
Concluding Remarks.....	16
Reports from Parallel Sessions - 19 July 2007.....	18
E-content and repositories: The challenges of managing e-content in repositories and interoperability	18
E-content collection selection and management: How do we select, prioritise and manage the creation of the digital collections of the future?.....	20
Business models and sustainability: How do we maintain and develop e-content?	21
The digitisation experience in education and research institutions: Project management advice and guidance	23
Access and identity management: Shibboleth and beyond.....	25
Reports from Parallel Sessions - 20 July 2007.....	27
<i>Mass Digitisation: Best practice and lessons learned.....</i>	<i>27</i>
<i>Online digital video: Educational developments and opportunities.....</i>	<i>29</i>
<i>Digital images: Developments in capture, conversion and workflow</i>	<i>30</i>

<i>Digital curation of digitised material: The what, why and how of digital preservation</i>	32
<i>Exploring commercial e-content developments and private/public sector partnerships</i>	33
<i>Capacity building: Investment in centres of excellence in the European Union and the US</i>	34
<i>Transforming the users' experience: How institutions can develop innovative and affordable tools to engage increasingly sophisticated audiences</i>	36
<i>Digital capture and conversion of text: Overcoming the Optical Character Recognition (OCR) challenges</i>	38
<i>The legal landscape: Copyright, IPR and licenses and mass digitisation</i>	39
<i>Developments in resource discovery portals for digitised material</i>	41

Reports from Plenary Sessions - 19 July 2007

Welcome to Wired Wales

Carwyn Jones, Counsel General and Leader of the House, National Assembly of Wales

Carwyn Jones welcomed delegates to the conference with a stimulating speech on “wired” Wales. He began by reflecting on how access to resources has changed over the years. When he attended the University of Aberystwyth, the National Library of Wales was on his doorstep. However, you needed to be a third-year student to get in, and had to request documents which in time would be fetched. You needed special permission to see rare documents, which was understandable as they were fragile. This was only 20 years ago. Now these documents are online and accessible to many more scholars. What we are experiencing is the democratisation of research.

He recalled his excitement when the 2001 census went online, providing access to the original documents. Now it is commonplace. There are numerous web sites with data and we expect it to be at our fingertips. New technical developments like broadband have been very important. He recalled what internet access was like only 10 years ago. You could click on a web site and then make a cup of tea before it appeared. Now there is broadband; access is fast and getting faster all the time.

In Wales much of the population lives in rural areas, so it was important to reach these areas with broadband as quickly as possible. If left to market forces this might have taken many years. In 2002, the Welsh Assembly Government initiated the Broadband Wales Programme to roll out broadband quickly, along with a high speed Lifelong Learning Network.

Education has been transformed. Where there used to be a blackboard and teacher, now even whiteboards are old hat. Children are comfortable with IT and expect to have instant access to information which previously might have taken hours to find in a library. We have to be ready to develop these resources further so children can learn at a faster rate than ever before. Online access to resources is particularly important for rural schools, as they may not be able to offer the same courses as those in urban centres. Online virtual learning environments ensure that *all* children can have the same opportunities.

As we use technology to shape the future, it is important to ensure that we don't create a society of “haves” and “have nots”. This can cause a divide in terms of educational opportunities unless we take steps to avoid it. Wales has made a conscious effort to ensure equality of opportunity. As with the implementation of broadband, access to PCs has been rolled out in libraries across Wales. Those who don't have access to PCs at home now have access in public centres.

Jones noted some recent achievements of the National Library of Wales, including the new custom-built bilingual IT system. The Library offers a wealth of resources, including the world's largest collection of works about Wales and other Celtic nations. He noted the importance of providing access to sounds, images and video as well as text. For example, those studying linguistics can use the sound archive and hear how accents have changed. Some Welsh dialects have died out, but we can still access these voices online. The JISC funded *Modern Welsh Journals Online* project will benefit Wales and all who study it. And as these resources are available online, scholars don't have to travel to use them.

Jones concluded by noting that Wales is developing its online resources for the education and skills of the new century and hoped to learn from the experience shared at the conference.

Introduction to the Day

David Baker (Chair), Principal, The College of St Mark and St John, Plymouth

David Baker welcomed everyone to the conference and briefly explained the aims:

- Launch the JISC Digitisation Strategy – Share it and get feedback from the community

- Showcase the collections created to date in the JISC Digitisation programmes
- Highlight future work
- Share best practice – Share exemplars, knowledge, skills and lessons learned
- Focus on the JISC role – How it is moving digitisation forward.

He encouraged delegates to use the conference blog to share their views on the issues raised during the conference. This is an opportunity to start a debate that can continue after the conference ends. He made a few suggestions about points to debate:

- Strategic digitisation of specific collections and archives versus responsive approach
- International versus UK focus
- User perspective
- Metrics/methods to determine user requirements
- Transferring best practice
- Technical techniques and standards
- Legal and licensing issues
- Business models and resource enhancement

JISC Strategy and Mass Digitisation

Malcolm Read, Executive Secretary, JISC

Mass digitisation is only one of the many activities JISC funds, so Malcolm Read explained the strategic context. JISC's mission is to "provide world-class leadership in the innovative use of Information and Communications Technology to support education and research". Mass digitisation supports JISC's strategic aim to build sustainable collections that meet the needs of life-long learning, teaching and research in the UK.

The "content triangle" shows how the three strands of JISC's e-content activities together make available these collections to the community:

- Digitisation – JISC Digitisation programmes digitise key resources
- JISC Collections – This new company licences subscription-based content like e-journals
- Repositories – JISC facilitates the development of repositories at institutional and national level for materials like e-prints, research data, and learning materials.

Read related the interesting history of how the first Digitisation programme was funded. In the UK, the case for funding must be made about 18 months in advance. At the time transatlantic network connectivity was enormously expensive and great cause of concern. Read made a case for funding the network. By the time the money arrived, the costs had come down. At the end of the funding request, he had recommended that any spare funds should be used for digitisation. As it turned out, there was £10m in spare funds!

The first Digitisation programme (2004-2007) was a great success and JISC had no difficulty in making a case for funding to support a second phase of the programme, which runs from March 2007-2009 (www.jisc.ac.uk/digitisation). These programmes are more than bulk digitisation. JISC has taken an innovative approach, choosing interesting collections and digitising sound and film material as well as text. The programmes also tackle important issues such as how best to access and manage resources and create links between them. The current, Phase Two Digitisation programme also emphasises use of the resources for learning and teaching as well as research.

Read hopes there will be a Phase Three Digitisation programme. This might be used as an opportunity to expose some of the special collections that UK higher education institutions hold and as a marketing tool to attract overseas students.

Looking at the bigger picture, JISC is one of many organisations providing e-content. Through the Strategic Content Alliance (<http://www.jisc.ac.uk/sea.html>), JISC is committed to working with other content providers like the BBC, the MLA and the NHS to give better value for publicly funded e-content. One driver is economies of scale, but another is to broaden access.

Currently when students leave school, they lose access to the resources they had there, but gain access to another set when they move on to college or university. Similarly, when they leave university and start work, the resources they can access change again. Within the NHS, many people are involved in higher or further education, and the resources they can access depend on the hat they are wearing. These barriers are confusing to users. A key aim of the Alliance is to reduce these barriers and develop a common approach to the management and availability of resources across these sectors.

The conference will be an opportunity for all to learn more about the JISC Digitisation programmes and the Strategic Content Alliance.

Mass Digitisation and its Impact on Cultural Heritage, Education and Research Communities

Chris Batt, Chief Executive, Museums, Libraries and Archives Council

Chris Batt gave a thought-provoking presentation that set the tone for the conference. It introduced themes and posed challenges that were taken up by the speakers and delegates as the conference unfolded.

The question he addressed was, “How do we make a lot of stuff useful to a lot of people?” A lot of stuff is easy, a lot of people is possible, but the “useful” bit is more tricky. The starting point is to think about what we are aiming to achieve. In his view this is a universal right to knowledge. In a knowledge society:

- Knowledge and ideas are the raw material of our futures
- Learning must be for life
- Understanding builds empowerment and cohesion
- Competency and skills-based learning replaces knowledge-based education

Knowledge empowers people to take learning journeys. It could be a simple journey to find the time of the next bus, or a complex journey to understand an aspect of genetic engineering. If this is what we aspire to, we can then think about the drivers to make it happen – the “trickiness” of making stuff useful. He outlined three building blocks for **useful** access to knowledge.

1. Technology must be invisible

The only successful technologies are the invisible ones. We don't worry about how a TV or a telephone works, we simply use it. In a digitisation project, a number of technical processes may be going on under the hood (digitisation, disclosure, aggregation, discovery) but presentation is the one that matters, because it makes the others invisible to users independently of where the content comes from.

2. Individual and community

The value chain for the MLA (Museums, Libraries and Archives) sector involves making connections between collections and users. We need to think about better ways of making these connections, and identify synergies between collections that could serve users at different levels. Every institution has its own portal, from the 24 Hour Museum to English Heritage; each is proud of its resources and making them available. However, creating lots of separate collections is not the answer. He argued that we need “massification of relevant content objects, not the multiplicity of institutions”.

The reality is that for 99% of web users, the world is Google. The paradox is that you need to know what you are looking for, and even then you may not be able to find it using Google. For example, searching Google for antique watches will locate people selling them but not information *about* them. It doesn't tell us what is in museums or the local library. Arguably Google will never be able to do that, but that is the learning journey that needs to be taken.

3. Sustainable within the public realm

Public funds go into a black box resulting in services the public want, outcomes to support public development, and trust. Batt has reservations about public-private partnerships and argued that what is needed instead is better advocacy with public funding bodies. Compared with fighting a war, the costs of mass digitisation are small, but the benefits invaluable. He feels there is a failure of advocacy. Public sector institutions need to speak with confidence and passion, and have a shared narrative about the benefits. This underlines the importance of the Strategic Content Alliance.

In conclusion he summarised what the cultural sector should aspire to:

- Knowledge fuels learning and creativity
- Lifelong learning must be universal
- Learning is a personal experience
- Content first, the institution second
- Institutions must work together
- The future is already here...

Our behaviours now determine the future. In 2040, he would like “knowledge to be a utility: as trusted, as accessible and as invisible as pure running water”.

Unlocking e-Content and Enhancing Education and Research Opportunities: An academic perspective

Matthew Steggle, Lecturer, Sheffield Hallam University

Matthew Steggle focused on the big picture and what users want. He presented five principles on digitisation projects from his perspective as an academic researcher and teacher. Each was based on real examples.

Electronic *Manipulus Florum* Project

Manipulus Florum is a 14th century collection of around 6,000 quotations in Latin from Church elders arranged under thematic headings. The manuscript was hugely popular in 15th and 16th century France, widely copied, and became a popular tool for writing sermons. In 2000, a Canadian professor (Dr Chris Nighman) got funding to create an Internet edition of the book (www.manipulusflorum.com). This was a simple digitisation – he bought an early printed copy of the book and captured the text as unformatted PDF files. Meanwhile, Steggle had been working on Renaissance literature and Thomas Nashe. Some quotations had eluded him, so he typed them into Google and found them on the Canadian site easily.

- Principle 1. Ease of access is crucial – If you can't see it, it isn't there
- Principle 2. Rich mark-up isn't always necessary – Full text search is often enough.

The Beatles Play Shakespeare

The Beatles did a Shakespeare TV special in the 1960s, now available on YouTube (<http://www.youtube.com/watch?v=D0pEZM6OEvI>). It's an excellent teaching resource, as it tells us about cultural history. From the student's perspective, it's accessible and easy to use. We can teach with YouTube, but we can't cite it in a research article. There are copyright issues, and authority issues as there are different versions of the clip.

- Principle 3. Ease of access is crucial (see rule one)
- Principle 4. It's helpful to understand how what you are viewing relates to the original.

Early English Books Online

Early English Books Online (<http://eebo.chadwyck.com>) is a comprehensive database produced by ProQuest with page images of English books published to 1700. This is another fantastic teaching resource. There's no need for the old skills of getting past the librarian and using microfilm. It is accessible to the entire university, so undergraduates can use it as well as graduates. Most

importantly, it changes what we can do. Steggle recounted how in the 1990s, he spent three years writing an article on Aristophanes and found a handful of texts that mentioned him. Now he can get 200 references in about 15 seconds.

- Principle 5. Completeness within limits is a helpful goal – Many academics prefer a complete set that lacks detail to a few perfect parts.

He encouraged digitisation projects to keep talking to users and summarised his principles:

1. Ease of access is crucial
2. Rich mark-up isn't always necessary
3. Ease of access is crucial - (see rule one)
4. A description of the process, from analogue to digital, is helpful
5. Completeness within limits is a helpful goal.

Roles and Priorities for Research Libraries in the UK and US in the Delivery of e-Content

Richard Ovenden, Keeper of Special Collections (Associate Director), Bodleian Library, University of Oxford

Michael Keller, University Librarian, Stanford University

During both the Symposium preceding the Conference and the Conference itself, Google had been mentioned several times. Ovenden and Keller decided to talk about the “elephant in the room” and share what they were allowed to share about their experiences with Google.

Richard Ovenden described Oxford's project on 19th century printed materials. In many respects it is not a complicated. Copyright isn't an issue, as all the material is in the public domain. Selection isn't an issue, as they digitise every 19th century book they can find. The project involves taking books off the shelf, digitising them and putting them back.

What makes this project different from a JISC digitisation project is the industrial scale. It is a huge logistical effort organising the move of hundreds of thousands of books from 40 buildings and getting them back on the shelves as quickly as possible. The pace is also fast. They refer to the project as “The Beast” and spend much of their time feeding The Beast. The books are already been made available on the Google Books interface. The next phase of the project is integrating the content into new services.

Ovenden commented that an unexpected lesson learned is how little they know about their collections. “We need to get back to the shelves, learn about what we have and what condition it's in”. They found that 18% of the books had never been catalogued!

Mike Keller described Stanford's project with Google. Like the Oxford project, it covers books in the public domain. The copyright situation in the US is complex, so they needed to check copyright renewal records for books published 1923-63 to find out which were actually in the public domain. A by-product of the Google project is Stanford's “copyright determinator” (<http://collections.stanford.edu/copyrightrenewals>), a searchable database of copyright renewals.

The project involves sending 1,700 documents per day to the Google lab, and in the process they have discovered about 8,000 books that need conservation treatment. Keller described Stanford's expectations for the project:

- **Indexing** – This is fundamentally an indexing project. Users can then get a copy of the book in a local library or through a bookseller. Indexing is important as it leads to increased use of collections.
- **Preservation** – Stanford will use the digital surrogates created by Google as preservation copies.

- **Searching** – They will be indexing in new ways to enable new types of searching. Taxonomic indexing based on expressions creates new linkages for more precise matching of results. Citation linking is also incredibly valuable. They are also experimenting with associative searching based on vector expressions. Very brief records can produce amazing matches.
- **Development** – They will use the digital surrogates as the testbed for new research at Stanford for literary, historical and linguistic investigations.

Keller commented that the project has been an opportunity to accelerate development that Stanford would have done anyway. Ovenden agreed – it will enable Oxford to look at areas like e-science for the humanities and new ways to add value to works like *Early English Books Online*. Effectively the project creates a more granular index to the library, allowing researchers to make better use of their time.

There was some disagreement about metadata. Ovenden noted that even Google was starting to appreciate its role in cataloguing resources. Keller argued that indexing based on metadata is limited. We should be focusing instead on more creative approaches to indexing that lead users to better search results. Over the next 10 years we should see dramatic improvements in the quantity of content that can be indexed and the quality of search results.

They outlined some of the challenges on the future landscape. These include digital-only content and the tension between open and restricted access content. The skills base will be important, so we have the expertise to use content in more challenging ways. Technical developments will provide opportunities, e.g. to improve linking and transform digital repositories. Ovenden noted the importance of the Loughborough Report (http://www.jisc.ac.uk/publications/publications/pub_digi_uk.aspx), which led to the formation of the Strategic Content Alliance. It will be important for shaping a collaborative approach to digitisation.

The discussion that followed focused largely on the Elephant and whether the public sector should work with companies like Google – does it pose risks and can we trust them?

Mike Keller supported working with Google. The Google team are clever, have big ideas and the money to put them into practice. In his view, we should take advantage of the opportunity while it is here. However, libraries need to take on the responsibility for the long-term preservation of the content that is digitised.

Others expressed reservations. **Malcolm Read** said it's not just a question of now, but would Google be around in 50 years time? **Chris Batt** said we need to think about whether working with Google is additional or a substitution. If the government perceives it as a substitution, there is a risk they will not see the need to fund digitisation because Google is doing it. This underlines the need for public advocacy and a shared narrative.

Bernie Reilly posed a scenario that Google might succeed in an area where libraries have failed – to make content free to everyone. How should the cultural sector position itself to give better value? **Richard Ovenden** noted that much of the MLA sector in the UK is owned by the state and controlled by the government. A deal with the government that made access to museums free to all has seen a dramatic increase in visits to museums. We need to make the same case in the digital world. **Chris Batt** agreed about the public advocacy issue. However, another way to give better value is to use content in as many ways as possible.

Reports from Plenary Sessions - 20 July 2007

Cyber-infrastructure and information policy: the changing face of US digitisation developments

Joyce Ray, Associate Deputy Director for Library Services, Institute of Museum and Library Services (IMLS)

Joyce Ray explained the role of IMLS, its context in the US funding landscape, and the many interesting projects they fund in the library and museum area. These range from digitisation, preservation and conservation to funding new programmes in schools for library and information science.

There are three cultural funding agencies in the US – the National Endowment for the Humanities, the National Endowment for the Arts, and the Institute of Museum and Library Services. IMLS (<http://www.ims.gov/>) is the newest of the three and started awarding grants in 1998. Its mission is to improve library and museum services throughout the US. It serves 122,000 libraries and over 17,000 museums, funding projects at state level (\$164m for 2007) and national level (\$70m). To put this in context, IMLS funding is small compared to National Science Foundation (\$6b), but roughly that of NEA and NEH combined.

The term "cyberinfrastructure" in the US is roughly equivalent to e-science in the UK and focuses on technical infrastructure and tools. Key agencies that fund cyberinfrastructure include the new Office of Cyberinfrastructure (\$100m in 2007, \$200m in 2008) and the National Science Digital Library. The Department of Education is large and funds digitisation, and the National Digital Information Infrastructure Preservation Program (NDIIPP) of the Library of Congress funds preservation (\$7-10m for 2008). The National Historical Records Commission also has a small budget for archives. There are of course numerous private foundations as well.

Ray gave a feel for how IMLS approaches funding priorities and the challenges they face. Tools are important, but so is the content. She noted that in the humanities there is not nearly enough content to support scholarly research. From her perspective, the key priorities are to build the content landscape and establish sustainable repositories to hold it. A challenge they face is that digitising content and building repositories are expensive, more so than developing tools. Other challenges are getting the right economies of scale, address copyright barriers, and eliminate silos across all areas.

She then described the programmes IMLS funds:

- **State Grants Program** – \$164m in 2007. Funds are distributed to state libraries using a formula based on population. IMLS is increasingly funding digitisation projects at state level
- **National Grant Programs** – \$70m in 2007. Competitive grants are awarded to libraries, museums, and others for many types of projects, including some digitisation and tool development.

Over the past few years IMLS has spent about \$100m on digitisation, about half through the state programme. At the national level the projects are competitive and peer reviewed. They look for interesting content and projects that can help to develop models and new tools and lead to best practice. Many involve consortia, a model that allows institutions to learn to work together. She mentioned a few projects funded by IMLS National Leadership grants:

- California Digital Library – Preservation using the OAI reference model
- Florida Center for Library Automation – DAITSS (Dark Archive in the Sunshine State), a preservation repository application which supports format migration
- Alabama Commission on Higher Education – A state-wide model based on LOCKSS
- University of Denver – Their Colorado Digitisation Project expanded to become the Collaborative Digitisation Programme, recently taken over by Bibliographic Center for Research (affiliated with OCLC)

- University of North Carolina (Chapel Hill) – A repository for film where the film is stored offline and digital copies are available online
- University of California (Santa Barbara) – Digitisation of 6,000 wax cylinder recordings. The press release stimulated enough hits to crash the server!
- Johns Hopkins – A national digital curation project involving the National Virtual Observatory and a consortium of research libraries.

New areas like cyberinfrastructure mean that professionals will need new skills. IMLS has therefore provided some funding for curriculum development for schools of library and information science. For example, the University of North Carolina (Chapel Hill) and University of Illinois (Urbana-Champaign) have developed new programmes in digital curation.

IMLS funded a study to survey the condition of collections of libraries and archives across the US, identifying 4.8b items held by these institutions. Many of these are in unknown condition or at risk. In response, IMLS launched a conservation initiative *Connecting to Collections*. Web Wise, the IMLS conference, focused on digital stewardship in 2007 (<http://www.ims.gov/news/2007/071907.shtm>) and will focus on digital tools in 2008.

Best practice guidance is also important. IMLS recently made a grant to NISO to update the *Framework of Guidance on Building Good Digital Collections* (<http://www.niso.org/framework/>). The new *Framework* will be transformed into a wiki with supporting case studies and should be available in next few months.

She underlined the importance of building trusted repositories and preservation – digitising resources originally recorded on unstable media, preserving born-digital resources and preserving the digital surrogates in case the originals are lost or damaged. Collaboration will be an important success factor – for interoperability, to achieve economies of scale, and sharing expertise on how to do it well.

During the discussion, **Alastair Dunning** asked whether IMLS provides seed or recurrent funding to projects like DAITSS. Ray indicated that IMLS was similar to JISC in that projects are typically funded for 3 years. In the case of DAITSS, they have developed their software and the project is finished. For projects over \$250K, the institution must provide matching funds and make commitment to sustainability.

David Flanders asked how projects in the UK could find out about tools IMLS has funded, start using them and collaborate across the ocean. Ray said she would be discussing this with JISC and developing a list. If a project is still running, it may be possible to arrange a collaboration.

The JISC Digitisation Programme: an introduction to the digitisation strategy and new collections coming on stream from 2009

Alastair Dunning and Paola Marchionni, Digitisation Programme Managers, JISC

Before introducing the current JISC Phase Two Digitisation Programme (www.jisc.ac.uk/digitisation), Alastair Dunning explained how what JISC funds is different from the kind of digitisation that Google does. JISC supports a variety of different projects from different institutions and its focus is on special collections.

He invited the audience to provide feedback to the JISC Digitisation Strategy through the conference blog or by email, mainly in relation to firstly, how JISC fits in with the wider UK framework to provide access to all, and secondly, on strategic subjects for digitisation on which to build a “forest of content”, not saplings. JISC is already considering, for example, Islamic Studies.

The Phase Two Digitisation Programme follows a successful Phase One which funded six large projects. The current programme constitutes an investment of £12.5m funding 16 projects that were selected after peer review and public consultation.

The programme runs from early 2007 to early 2009 and feature collections that span five centuries of social, political, economic and cultural history in the UK. Projects include a variety of formats, such as newspapers, images and photographs, cartoons, thesis, maps, cabinet papers and video and sound recordings, from a variety of institutions. Much of this material is fragile and has so far been difficult to access.

Paola Marchionni provided an overview of the 16 projects grouping them under five key themes that reflect priorities within the Digitisation Strategy. The groupings are for illustrative purposes only, and most projects could fit under almost all themes:

- **User engagement:** both during the creation of a resource as well as in its delivery, through a variety of Web 2.0 tools and functionalities ([First world war poetry archive](#); [Pre-Raphaelite resource](#); [British cartoon archive](#))
- **Protection from deterioration:** especially in the case of fragile material ([Freeze frame: historic polar images](#); [Archival sound recordings](#))
- **Contextualized resources:** a rich contextual background and teaching resources mainly for the Higher and Further Education community ([British governance: cabinet papers](#); [Voices: moving images in the public sphere](#), Historic boundaries of Britain;)
- **Delivery, access and sustainability:** a variety of sustainability models including a “free to all” approach ([Modern Welsh journals online](#)); the use of existing infrastructure ([Nineteenth century pamphlets](#)/JSTOR and [UK theses/ETHos](#)) and partnership with the commercial sector ([Electronic Ephemera](#))
- **Building a national critical mass:** including aggregating previously dispersed collections ([E-resources on Ireland](#); [British newspapers 1620-1900](#); [East London theatre archive](#); [Independent Radio News](#) Archive)

Alastair Dunning concluded outlining the programme’s expected outcomes such as the creation of a sophisticated rights framework; increased technical metadata and digitisation knowledge including technical standards; enhanced digital infrastructure and a greater awareness of digitisation needs both from users and from collection curators.

Concluding Remarks

Simon Tanner, Director, King’s Digital Consultancy Services

Simon Tanner had the difficult task of summing up the conference, but he did so in just five words – collaboration, visibility, invisibility, mass and tomorrow. His concluding remarks summarised the key themes of the conference.

Collaboration – A key theme of the conference is that collaboration is important. It can be difficult for institutions to sit down and commit to working together to achieve common goals. However, the conference has shown that it matters and we all take it seriously.

Visibility – We need to be more visible as a community, lobbying better so that everyone knows about the work the community is doing and that it matters. We also need to take away the message that users are interested in the content but not by who provides it. In this sense “content is king”, so making it visible is very important.

Invisibility – Users don’t want to worry about how to get to the content; the tools should be invisible. The real challenge is making the content (and the branding) visible but at the same time making the process as invisible as possible.

Mass – Mass is important in both senses of the word. Some projects are on a massive scale, like digitisation of newspapers. However, mass can also mean density. Some digitisation projects involve content that is unique and specialised. This can be as important and valuable even if it doesn't involve a million pages.

Tomorrow – The conference has focused on tomorrow and what will happen next. In thinking about tomorrow, we shouldn't be constrained by the present. The people who developed YouTube didn't sit down and say, "How can we revolutionise the web?" They thought it would be cool to share video with each other, then with everyone. Things grew and in the end they did revolutionise the web. Tomorrow will be different, so we shouldn't be constrained by today. We should look to the future and see tomorrow as a fresh field to take our projects into.

Reports from Parallel Sessions - 19 July 2007

E-content and repositories: The challenges of managing e-content in repositories and interoperability

Moderator: Alastair Dunning, Digitisation Programme Manager, JISC
Rachel Bruce, Programme Director, Information Environment, JISC
Balviar Notay, Programme Manager, Information Environment, JISC
Gareth Knight, Digital Preservation Officer, Arts and Humanities Data Service, King's College London

Repositories are containers for the content we digitise, and this interesting session looked at some important repository issues – strategic issues for repository development, engaging users, and preservation.

Rachel Bruce explained JISC's strategy to "create and manage a layer of scholarly academic resources readily available to all who can exploit it" and how building an interoperable network of repositories supports this strategy. Achieving this interoperability presents challenges. There is a diverse range of content to deal with, and repositories do more than simply store it. Content has a complex lifecycle involving gathering, creating, sharing, and discovery, and repositories must provide interoperable processes and workflows to support this lifecycle.

An injection of funding in 2006 was the stimulus for JISC to plan investment in repository infrastructure and commission a Digital Repositories Roadmap (http://www.jisc.ac.uk/uploaded_documents/rep-roadmap-v15.doc). The Roadmap by Andy Powell (Eduserv) and Rachel Heery (UKOLN) sets out a vision for 2010 where a wide range of scholarly content is available via open access and managed access. To achieve this vision will involve many challenges, and the organisational and policy issues may prove to be more challenging than the technical issues.

She illustrated how JISC is putting the strategy into practice, using the EThOS project (Electronic Theses Online Service) as a case study. There was a clear need for a national UK thesis service, and the role of the EThOS (<http://www.ethos.ac.uk/>) was to develop a service to meet this need. This involved several strands of work, including those related to infrastructure, business models, policy development, digitising theses, and getting take-up from the institutions. As with many projects, the devil is in the detail, and much work was done to develop practical but effective policies on IPR issues.

Balviar Notay explained how web 2.0 can help to overcome the challenge of getting users to engage with repositories. To illustrate that users are not engaging with repositories, she gave some revealing statistics from the JISC CD-LOR project (<http://www.academy.gcal.ac.uk/cd-lor/>). Their survey found that 75% of respondents use email to share content with colleagues where only 1% use repositories. In comparison, the growth and use of web 2.0 technologies is staggering – MySpace alone has 15m users daily.

The bottom line is that web 2.0 applications are easy to use, functional, and fun, where repositories are not fun. However, there is a lot we can learn from web 2.0 and apply to academic repositories. And this may be important to engage the new generation of 'digital natives'. She showed how repositories like SlideShare and Scribd use web 2.0 to enable users to easily share their slides and documents. She recommended Tom Loosemore's 15 principles for developing web 2.0 applications at the BBC (http://www.tomski.com/archive/new_archive/000063.html) and showed how they can help us to think creatively. For example, experiment, iterate, kill the failures, and build on the successes.

Web 2.0 won't do everything, so we need to think about how to merge it with the traditional digital repository. For example, web 2.0 doesn't do curation and preservation, it isn't heavy on metadata, and there are no strict policies for IPR. It's also global rather than a local phenomenon, and sometimes we need to work locally. These are issues we need to address, but the key point is that web 2.0 is part of the landscape and we can't ignore it.

Gareth Knight explained how the Arts and Humanities Data Service manages preservation. The AHDS collects, preserves and distributes high-quality digital resources for research and teaching. The content mainly comes from research projects supported by the academic funding bodies, but also libraries, archives and museums. He explained their preservation procedures and that this involves conceptualising digital objects at three levels– the bit stream, the information content (words, images), and where possible the experience of interacting with the content (speed, layout).

He outlined the process for depositors submitting data and then the workflow and ingest activities. Data to be deposited must be in agreed submission formats, and the depositor signs a licence agreement and supplies catalogue information and supporting documentation. The workflow works quite well, and there are plans to automate it to improve management, scalability and speed. However, some data requires manual intervention, and he gave case studies to illustrate the problems that arise. An interesting case involved train timetables from the 1980s. In this case they were able to find the software used to create and migrate it to a new format.

The most common problems they encounter preserving data involve rights, incomplete resources, proprietary formats, and lack of documentation. He made some recommendations about preserving data based on AHDS experience. Workflow should be scaleable and ideally automated, but sufficiently flexible to consider alternative solutions. Audit trails are also important, so you understand who did what and when and can go back and check mistakes.

The **discussion** was interesting, and the Elephant (Google) joined the discussion from time to time. Some thought web 2.0 was over-hyped and can get in the way of what we want to do in the education sector. But others noted that we don't have to choose between web 2.0 and repositories; it's more a matter of combining the best elements of both. A perceptive comment was that repositories are simply containers, perhaps the 'new database'. It is the content rather than the container that is important. If we preserve the content, we can always build a new container.

The discussion brought out a tension between the world of academic repositories and web 2.0. Trusted content and preservation are important for repositories, where fun is important for web 2.0. There is also tension about whether we should be using web 2.0 to create our own applications or working with the large commercial services like Google and Flickr and whether we can trust them. It would be useful to have a coherent plan for working with the private sector and using the tools they are creating so we can provide what our users want.

Key Themes:

- Repositories need to support the complex lifecycle of content – gather, create, share, and discover
- The Digital Repository Roadmap illustrates how JISC created a vision of the repository landscape in 2010 to work towards and identified challenges to address
- The Arts and Humanities Data Services is a useful model for digital preservation, and their recommendations about processes, procedures, and pitfalls to avoid can help other repositories on the learning curve
- Web 2.0 engages with users, allows them to share content with others easily and is fun; repositories, in contrast, are not engaging users
- There is probably a lot we can learn from web 2.0 and apply to academic repositories to engage the new generation of 'digital natives'
- There is a tension between the world of academic repositories (trusted content and preservation) and web 2.0 (fun). There is also tension about whether we should be using web 2.0 to create our own applications or working with the large commercial services, and whether we can trust them. A coherent plan for working with the private sector would be useful.

E-content collection selection and management: How do we select, prioritise and manage the creation of the digital collections of the future?

Moderator: Philip Pothen, Press and PR Manager, JISC
Ralf Goebel, Programme Director, Deutsche Forschungsgemeinschaft (German Research Foundation)
Mark Brown, University Librarian, University of Southampton

Ralf Goebel spoke about digitisation in Germany and how digital collections will evolve in the future. The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) is the central funding agency for research in Germany and provides significant funding for digitisation projects, both mass digitisation and special collections. About €15m per year is used for retrospective digitisation of historical materials, e.g. German printed books of the 16th and 17th century. A further €10m funds digitisation of 20th century material to meet the needs of the research community. Here they are working with German publishers, and peer review plays a role in the selection of content. The plan is to make this digitised content available through the DFG-funded Virtual Subject Libraries.

Goebel feels that the concept of a 'collection' as a cohesive set of resources will become less important. In the future we can expect that most content in the public domain will be available on the Internet through both government-funded initiatives and the investment of companies like Google and Microsoft. As more content becomes available, the issue of selection criteria for digitisation will become less important. What will be increasingly important are tools to navigate and manage these resources.

He concluded by commenting on JISC's digitisation strategy. He recommended the DFG's approach to digitisation where there is a balance between mass digitisation in defined subject areas (top down) and projects put forward by the community (bottom up). He also recommended more international cooperation, especially on the development of standards and tools, and a focus on national collections to minimise duplication of effort.

Mark Brown drew on the experience at Southampton in digitising niche collections. We need to look at multiple approaches to content selection. The user-led approach is important, but we should also consider factors like the rarity and uniqueness of collections, national institutional missions, thematic subject approaches, and a collection approach to get critical mass.

He emphasised that users are not interested in the mechanics of digitisation but in effective techniques for using them. Issues like the integration of collections, searching effectiveness, and value-added functionality are therefore important. He questioned whether the Google approach to digitisation is the right approach, where there is no apparent link between user and collection.

He also noted that there are various approaches to business and sustainability models. Google's commercial model is that content is free at point of use. If this approach is taken, we need to consider who is responsible for sustainability.

In the **discussion**, the issue of business models was considered further. Google's business objectives are well-aligned with the scholarly community in that content should be free. However, the Google model is based on advertising on top of free content. More content means more potential advertising. They don't therefore offer free content but leverage the content of others. It was also felt that Google is digitising the 'low-hanging fruit', leaving gaps in collections and the difficult material to be met by the public purse.

There seemed general agreement that there is room for a number of business models. However, institutions have been (and will be) around for hundreds of years and need to think long term. Commercial partners will want to recover their investment, but will also be prepared to take risks. Concerns were also raised about their long term commitment to hosting, updating, and functionality. We therefore need to understand their motivation in investing in digitisation.

Key Themes:

- The German experience suggests that a balance between mass digitisation in defined subject areas (top down) and projects put forward by the community (bottom up) is a useful approach
- In a world with an abundance of content, navigation may be a more important issue than selection; the development of tools for navigation may be a useful area for international collaboration
- We need to look at multiple criteria for content selection - not just following user demand but responding to the uniqueness and rarity of collections, national institutional missions, particular thematic subjects to get critical mass
- User demand is unlimited, so setting priorities is challenging
- Initiatives from commercial providers like Google and Microsoft should be seen as a window of opportunity, but institutions have to think in the long term
- There was concern that commercial initiatives may focus on the 'low-hanging fruit', leaving gaps in collections and the difficult material to be met by the public purse
- We need to understand the motives of commercial partners in investing in digitisation and their commitment to sustainability.

Business models and sustainability: How do we maintain and develop e-content?

Moderator: Stuart Dempster, Director, Strategic Content Alliance, JISC
Catherine Draycott, Chair, BAPLA
Peter Kaufman, Chief Executive Officer, Intelligent Television
Dan Burnstone, Director of Publishing, Chadwyck-Healey, ProQuest

Catherine Draycott focused on business models and sustainability for picture libraries. Previously, the models for licensing images have been very complex. For rights managed licences, the fee depends on the use and can vary from a nominal fee (reproduction in an academic journal) to much larger (use in an advertising campaign). Recently large players have entered the market buying up many of the stock photo libraries that deal in generic images. They have brought with them new business models, causing concern among BAPLA members. These include the 'royalty free' model which allows unlimited use, and the microstock model which prices according to size (about \$1-35). Some feel threatened by these new models, but others wonder if they could be a way to tap new audiences.

Other models coming into the market include monthly subscriptions, where you pay a monthly fee and can download images up to an agreed limit. Some clients are also starting to challenge fees - the Telegraph is dictating the fees it will pay on a take it or leave it basis.

The legal environment is also evolving. She noted the Gowers Review of copyright in the UK, concern about 'orphan images' where the owner cannot be traced, and concern among creators about piracy. The new culture of access is stimulating the development of model licences like Creative Commons and Creative Archive. About Wellcome Images (<http://images.wellcome.ac.uk/>) makes 150,000 images available online from the Wellcome library collection. Wellcome Trust uses a Creative Commons non-commercial attribution licence for the historical images, and a similar Creative Commons licence which forbids manipulation of contemporary images.

A key issue that picture libraries in the MLA sector face is how to fulfil their remit to educate in a sustainable way. The large costs of running a picture library mean it won't generate much income. She noted that even with a fully automated service, fully e-commerce enabled with images online for users download, there will still need to be some human mediation for 80% of the transactions. Picture libraries won't make a profit because of the high costs, but they can provide invaluable public exposure and huge reputational value for the organisation.

One option for covering the costs would be to distinguish between commercial and non-commercial use. Museums need to subsidise education, but not the commercial sector. The National Museums Directors' Conference has commissioned a legal firm to develop a model contract for museums defining the parameters of free access to their collections and the distinction between commercial and non-commercial use.

She also noted that exchange models are something to consider. Creators want their work to be preserved and for people to be able to access it. Perhaps there is an exchange model that could be win-win where creators allow access to their content in return for it being housed and preserved. This needs to be scoped out more.

Peter Kaufman considered how making our collections more relevant and valuable can help us in thinking about business models and sustainability. The e-Content Policy and Strategy Symposium spent some time discussing how the commercial sector thinks about value – the value of content and the value of investing in it. There are lessons we can learn. He offered three observations relevant to digitisation and business models.

Firstly, if we want to build collections of lasting value, we will need to take an imaginative leap into the future. Michael Jensen wrote a seminal article in the *Chronicle of Higher Education*, *The New Metrics of Scholarly Authority* (<http://chronicle.com/free/v53/i41/41b00601.htm>). Jensen argues we need to think about new business models that allow us to compete in “computability”. This means making our content more available for indexing, linking and tagging, and improving the metadata for identifying, categorising, contextualising, and summarising it. This is the number one thing we can do to increase the value of our content.

Secondly, the Wall Street investment banks and analysts are looking at the same future trends and how the commercial companies like Google and MySpace are helping to define them. Analysts like Gartner tell us that too much choice can be a bad thing (“consumer vertigo”), and that making content searchable may be as good if not a better investment than making more content available.

Thirdly, he noted that creative partnerships will be important and used the CBS Interactive Audience Network as an example. It is a partnership between online content providers like AOL, Microsoft and Bebo, and social network application providers. The aim is to allow social network users to include clips from the providers into their blogs, wikis, and community pages.

Finally, Kaufman made a few comments on the JISC digitisation strategy (<http://jiscdigi2007.pbwiki.com/Digitisation+strategy>). It uses terms like ‘knowledge economy’ in preference to more loaded terms like ‘knowledge commons’ or ‘knowledge society’. It doesn’t talk about ‘free’ and avoids open access, focusing instead on business models and public/private partnerships. He feels this is a good practical approach, but if it is to be followed through, it requires sustained understanding of how commercial companies are approaching digitisation.

Dan Burnstone described how ProQuest approaches business models and sustainability as a commercial publisher. His area of the business specialises in large collections in the humanities and social sciences for the academic market. They have two main business models:

- * Subscription - An annual subscription fee gives customers access and updates
- * Perpetual Archive - A purchase model which includes ownership of the content and associated metadata in perpetuity, and the right to take delivery of data and metadata for local institutional use if so desired. An annual access fee is charged for access to the product on ProQuest's servers. This covers storage and hosting, software licences, interface enhancements, technical support and other costs.
- * Other models include partial open access in which a service is free to users in a defined territory where public funding has helped to produce it but is sold elsewhere. In such public-private collaborations a royalty is returned to the public digitisation partners.

However, business models are changing, and ProQuest is seeing more public-private partnerships. ProQuest is pleased to be in partnership with JISC and the Bodleian Library on a Phase 2 Digitisation project (http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation/ephemera.aspx) and is negotiating with University of Southampton on a partnership to host the 18th Century Parliamentary Papers. A project to digitise the *Connecticut Courier* in the US involves a contributory model.

Sustainability involves issues related to preserving, maintaining, and enhancing their products. Regarding preservation, the Perpetual Archive model gives customers the right to make an archival copy of the data. Maintaining their databases is a continual process, and compliance with standards

is a cost of doing business. They have an ongoing interest in enhancing their resources to maintain long term relationships with their customers and ensure stability within their market.

Regarding the future, he speculated on some trends for the next 3-5 years – more commercial content will be partly owned/developed by users, the open web will be harnessed by commercial services to add value, and new partnerships and business models will develop.

In the **discussion**, Simon Tanner noted one model that had not been mentioned. The Royal Horticultural Society plan to have images on their web site, but members can have access to better more and better quality images. Income is therefore derived from memberships rather than the sale of photos.

Quantifying reputational value was also discussed. Putting a value on the kind of PR that results from high-profile awards is difficult, but Bernard Reilly said that quantifying the return would be increasingly important as more content becomes free. MLAs need to understand the returns and the value of the returns.

Key Themes:

- New players in the picture library market are introducing new business models for generic images. Some feel threatened by these new models, but others feel they may have potential to tap new audiences.
- There are large costs associated with running a picture library. Even with a fully automated service, fully e-commerce enabled with images online for user download, there will still need to be some human mediation for 80% of the transactions.
- Picture libraries won't make a profit because of the high costs, but they can provide invaluable public exposure and huge reputational value for the organisation. Being able to quantify this value may be important.
- Museums need to subsidise education, but not the commercial sector. The National Museums Directors' Conference has commissioned a model licence defining the parameters of free access and the distinction between commercial and non-commercial use.
- Exchange models between content creators and content providers may hold promise.
- Focusing on value, and how to make our collections more relevant and valuable, may help us in thinking about business models and sustainability.
- Michael Jensen's article 'The New Metrics of Scholarly Authority' argues we need to think about new business models that allow us to compete in "computability".
- Similarly, analysts like Gartner tell us that too much choice can be a bad thing ("consumer vertigo"), and that making content searchable may be as good if not a better investment than making more content available.

The digitisation experience in education and research institutions: Project management advice and guidance

Moderator: Paola Marchionni, Digitisation Programme Manager JISC

Stuart Snyderman, Digital Production Services Manager, Stanford University

Matthew Woollard, Head of Digital Preservation and Systems, UK Data Archive, University of Essex

Peter Findlay, Project Manager, Archival Sound Recordings, British Library

Stuart Snyderman outlined some of the challenges of managing digitisation projects in education and research institutions:

- *Defining internal digitisation programmes in context of mass digitisation initiatives* – This should be viewed as an 'opportunity' to be strategic rather than a challenge. Responding directly to user needs to help teaching and research by focusing on specialist collections and unique material can help ensure success and prevent overlap with other projects.

- *The Challenge of Scale* – Increasing automation while maintaining standards of quality and preservation can be a challenge.
- *Building the infrastructure to support streamlined digitisation efforts* – Human infrastructure is vital from project conception to ongoing sustainability. Technical infrastructure is also important, such as software tools to enable tracking, reporting, effective queue management and quality control. Currently these tools are unsatisfactory, and there is scope for institutions to work together on this.
- *Building services to take advantage of digitised content and creating content to take advantage of sophisticated online services* – There is still a long way to go in creating content that supports sophisticated discovery in a highly automated way.
- *Balancing local institutional context and the imperative to collaborate* – JISC openly supports collaboration, but local projects are easier, more efficient and cheaper. The balance needs to be struck to ensure efficiency and that effort is not repeated.
- *Balancing project-based and service-based work* – It can often be a challenge to manage services and large projects with the same set of resources.

Matthew Woollard identified success drivers for digitisation projects based on the experience of the Online Historical Population Reports project (Histpop) (<http://histpop.org>), a project in the JISC Digitisation programme. He described how flexibility, quality assurance, and champions can all contribute to the success of a project.

Flexibility is a success factor, particularly in relation to the following areas:

- **Project plan** – The original project specification and the final product will often be very different. Allow for flexibility in the project plan to account for this.
- **Stakeholders** – The project should be flexible enough to allow for events such as the strategic aims of the project funders changing.
- **Contracts** – Contracts should be watertight, but also flexible. Project managers should take advantage of the JISC model templates which provide such flexibility.

Quality Assurance is also a success factor for mass digitisation, helping a project move beyond delivery towards sustainability. It can be difficult to ensure that subcontractors adhere to the expected standards of the project team, although sampling small selections of the digitised content can be a way of checking for errors. Compromising is sometimes necessary, but this depends on the scope of the project. Histpop had a policy of zero tolerance and checked images one by one, while other projects advocate a 'good enough' approach. Discussion highlighted the need for collaborative tools to help manage the QA process, particularly with software to allow for automated checking.

Woollard also noted the importance of project 'champions', both within the project and externally. A champion within the project with an in-depth knowledge of the potential of the resource for education and research can help define what the end product should be. However, striking the right balance between the roles of project manager and the champion can be a challenge. One possible solution would be to have both a technical project manager and a curator in the project team.

Champions in other areas can also be crucial to pushing the project beyond delivery, such as champions within the host institution, and within JISC, such as programme managers or advisory boards, who will be able to recognise the importance of the project and can provide advice and guidance to the project team.

Peter Findlay identified communication as the key to the successful delivery of a project, particularly with regard to the following groups:

- **Users** – User panels are useful in helping to guide and inform a project. Online consultation and lab-based usability testing with representative users is also valuable during the development of a project.

- Vendors – Relationships need to be built both during and after contract negotiations to ensure the successful and timely delivery of the project.
- Institutions – Engagement with the host institution is key to ensuring support for the project.

Essentially project management is about interconnectivity between the project team, stakeholders, users, vendors, drivers, and other projects.

Key Themes:

- Project 'champions', with an in-depth knowledge of the content being digitised and its potential for education and research, can be important to the success of a project. It's important to strike the right balance between the role of project manager and that of the 'champion'.
- It is important to involve users in the project, for example through user panels and usability testing.
- Quality assurance is an important success factor, and there is a need for collaborative tools to help institutions manage the QA process. However, there were different views on what level of quality is acceptable. Some argued for 'zero tolerance' and others for the 'good enough' approach.
- Flexibility of approach and managing interpersonal relationships and communication are also important, with suppliers as well as internal members of the team, and the wider group of stakeholders.

Access and identity management: Shibboleth and beyond

Moderator: Emma Beer, Manager, Strategic Content Alliance, JISC
Nicole Harris, Senior Services Transition Manager, JISC
John Chapman, Applications Project Officer, Becta

Nicole Harris gave an overview of why access and identity management are important within the digitisation process. During the conference, access and presentation have been frequently mentioned as essential to the digitisation process. However, they are often not seen as priorities compared to the physical process of digitising material, and rarely considered in enough detail in the planning stage of projects.

Harris clarified that access and identity management are about *improving* access to resources and supporting the 'universal right to knowledge' and not about restricting access. They are an essential part of the digitisation arena and support the importance of IPR, licensing and copyright management. Access management is equally important for 'free' resources in terms of managing processes such as deposit, rights management and value-added services like personalisation.

In the US, some institutions have stopped managing identities. Institutions need to either become identity providers or ignore identities all together. The technology is there to manage identities – the challenge is implementing the technology and managing staff behind this process. This involves ensuring that basic records are correct – a long, painful, but necessary process. In the US there are technologies emerging to support this process.

Nicole concluded that the benefit of the federated approach was that most of the processes are 'behind the scenes' and invisible to end-users before handing over to John Chapman to describe the UK Access Management Federation in more detail.

John Chapman highlighted the importance of federated access management in reducing the burden of username and password management for both end-users and service providers. He described the role of the UK Access Management Federation in supporting technologies such as Shibboleth, and the importance of the joined-up approach across the entire education and research sector. This is supported by increasing interest from the NHS, museums, libraries and archives, and a government portal approach based on the same standards (SAML).

In terms of good practice, the importance of good identity management at the institutional level was highlighted by both presenters. It is essential to get basic identity management in place before more complex scenarios, such as managing multiple identities across multiple domains, can be tackled. Web 2.0 developments mean that identity is now a central concern. Social networking tools such as Facebook expose users to abuses of identity. Emerging technologies such as OpenID and CardSpace will be interesting to track, particularly in relation to web 2.0 technologies.

Key Themes:

- Access and identity management are about *improving* access to resources , not restricting access
- It is important to put access management strategies in place at the *start* of a digitisation project
- Web 2.0 and social networking tools like Facebook bring new identity challenges
- trusted and authentic resources can be as invisible as running water
- Institutions need to either manage identities or ignore them
- We need licences that allow authorised people to access resources *any place any time*.

Reports from Parallel Sessions - 20 July 2007

Mass Digitisation: Best practice and lessons learned

Moderator: David Dawson, Senior Policy Adviser (Digital Futures), Museums, Libraries and Archives Council

Ricky Erway, Programme Officer, OCLC Programs and Research

Stuart Dempster, Director, Strategic Content Alliance, JISC

Joyce Ray, Associate Deputy Director for Library Services, Institute of Museum and Library Services (IMLS), Washington DC

Ricky Erway offered some excellent guidance on negotiating licensing deals with the private sector based on a review of publicly available agreements. The starting point is to think through what an institution wants from the deal. This should be decided before anyone approaches the institution. Then it is important to decide what the bottom line is, what one really must have (or it isn't worth doing) and at what point one would walk away from any negotiations.

She explained what to look for in a good deal and pitfalls to avoid on a range of topics:

- Exclusivity
- Deliverables
- Functionality
- Non-disclosure
- Financial models.

Exclusivity is particularly important – a deal that appears to be non-exclusive may be effectively exclusive. From a practical point of view, the expense involved may preclude competitors from rescanning the content. A deal may also say that exclusivity applies only to the digital copies and we can do what we want with the originals. However, as the investment is in the digital copies, we should avoid strict controls on how we can use them.

Another area to consider carefully is the terms that survive the agreement. The private partner's responsibilities may terminate at the end of the agreement, but restrictions on ownership or use may continue for some years. Institutions should ensure that the right to manage/distribute the content is limited to the term of the agreement.

Demands for secrecy should be a red flag. Non-disclosure agreements are fine to protect proprietary information like technical and business secrets. However, they should not extend to what content is being scanned or even that an agreement has been signed. Openness is important to act in the public interest, so agreements shouldn't restrict discussion with the community.

In the discussion that followed, some participants who had negotiated deals with large players defended the terms they had agreed to, particularly with respect to openness. Compromises sometimes have to be made to get content digitised and publicly accessible. Sarah Porter commented that openness is important to ensure we can make informed decisions.

Stuart Dempster summarised lessons learned from JISC's first Digitisation programme. These have been used to plan the second programme. The lessons learnt fall into two categories – those pertinent to agencies and organisations planning to undertake digitisation and those pertinent to funding organisations, such as JISC as part of programme management.

He explained what those planning to start a digitisation project should take into account:

- **Consensus** – is digitisation compatible with your organisations mission and/or strategy. Do you have support of your management and institutional buy-in?
- **Content** – what makes the collection unique, valuable and usable? Do you have accurate statistical information i.e. exact number of items?

- **Copyright** – what content do you own and/or require rights clearance for? What licence will be use and why?
- **Catalogue** – what descriptive, technical and administrative metadata exists and to what standard?
- **Capacity** – what level of infrastructure does your organisation have to capture, convert and deliver content either directly or through third parties?
- **Competence** – what technical, curatorial and administrative expertise and/or experience can you call upon?
- **Context** – if content is king, then context is queen. Who is your audience, what do you hope to achieve, how will you engage them, what is your unique “selling point”?
- **Commitment** – digitisation is the start of a long term commitment to sustain, develop and curate content and not the end...this involves recurrent costs/time/resources and necessitates a business model from the start i.e. will you or a third party deliver and develop the content?
- **Convergence** – can the content be “repurposed” within virtual learning environments, digital repositories, social networking and other “Web 2.0” tools?

He went on to the outline the issues which funders should take into account when undertaking a digitisation programme as well as more practical considerations for projects. These included testing assumptions and methodologies, allowing adequate time for procurement processes, documenting and carrying out Quality Assurance processes, paying particular attention to the wealth of metadata schema that projects can adopt and validate them against current standards. The importance of a good Project Manager can never be underestimated and projects should be run according to established project management processes.

The relationship with stakeholders is also crucial. Involving users in the creation and testing of the resource is invaluable as well as putting in place a proper marketing and communication strategy. The complex process of clearing copyright is time consuming and expensive and adequate time should be allowed for clearing copyright before digitisation starts.

Joyce Ray spoke about “Building the Cyberinfrastructure in the US.” She explained that while the US does not have a Ministry of Culture, and no single agency has overall responsibility for digitization or development of the information technology infrastructure, this responsibility is distributed among many federal agencies including, among others, the US Department of Education, Institute of Museum and Library Services, Library of Congress, National Endowment for the Arts, National Endowment for the Humanities, and the National Science Foundation. Many private foundations and other organizations including the commercial sector are also investing in these activities.

Mass digitization projects have been funded primarily by private sector entities such as Google. IMLS has not funded mass digitization per se, but has funded a number of large-scale digitization projects, many based on state-wide collaborations among libraries, museums and archives, or on specialized formats such as sound recordings. More than 40 states now have state-wide digital programs, most of which have received IMLS funding (<http://lists.mdch.org/bin/listinfo/digistates>).

In all, IMLS has invested approximately \$100 million US in digitisation over the past five years. IMLS has also invested in research and development of preservation repositories, guidance and best practices for the creation and management of digital content, and the development of educational programmes in graduate schools of library and information science for digital asset management and data curation.

In conclusion, Ray noted that collaboration among funders, developers and educators is key to ensuring interoperability of the cyberinfrastructure as well as realisation of economies of scale and expertise.

Key Themes:

- Ricky Erway gave guidance on negotiating licensing deals with the private sector. Key considerations are exclusivity and being able to share information about the deal with the community.

- It is important to decide what you want from a deal at the start and the point where you would walk away from any negotiations
- Digitisation is the start of a long term commitment to sustain, develop and curate content and a number of issues have to be considered before embarking on a digitisation projects, both from a project and a funder's point of view
- Collaboration among funders, developers and educators is key to ensuring interoperability of the cyberinfrastructure as well as realisation of economies of scale and expertise.

Online digital video: Educational developments and opportunities

Moderator: Darren Long, British Film Institute (BFI)

Rick Prelinger, Board President, Internet Archive

Murray Weston, Director, British Universities Film and Video Council (BUFVC)

Peter Kaufman, Chief Executive Officer, Intelligent Television

Rick Prelinger described the Prelinger Archive, the video archive he developed to “collect, preserve, and facilitate access to films of historic significance that haven't been collected elsewhere”. The Archive was bought by the Library of Congress and is available through the Internet Archive (<http://www.archive.org>). He summarised coverage, formats supported, and workflow.

The Archive is a long-term non-profit presence in a highly commercialised world. The site is free at the point of use with no advertising. The content is available on a Creative Commons licence and can be reused. However, future sustainability is an issue in terms of who will support the Archive, particularly with the emergence of YouTube.

From his perspective, the emergence of YouTube has had implications for the Archive. It has slowed down ingest and is perhaps lowering standards. He noted that more “dodgy” material was being uploaded than before, both in terms of rights (copyright infringement) and taste and decency (e.g. extremist organisations).

The Prelinger Archive is a small operation, and this affects its development. For example, there are no editing tools and video segmentation at the moment is unsupported. Although the Archive promises to store films forever, they are only just starting to think about digital preservation. He also noted that the site doesn't have educational input or annotations to help with selecting content for educational purposes. At present, annotations tend to be the sort you would expect from fans.

Prelinger argued that in less than one year, YouTube had become the default media archive and anything anyone does to bring archives online will be measured against YouTube.

YouTube offers a sense of completeness with regard to the material available; sticks to low quality preview mode, so users feel no sense of transgression, but above all it allows users to upload almost everything, annotate with relative freedom and network with one another.

Murray Weston focused on the use of video in education from the perspective of the British Universities Film and Video Council (BUFVC), which promotes the production, study and use of moving image, sound and related media in higher education and research. He explained how moving pictures are used in education, for example as evidence or as an instructional device, but also to change hearts and minds, or for live communication. However, the medium is underused, as many of the supporting infrastructures we have for text based material are not in place for moving images, such as statutory deposit, interlibrary loan, arrangements for fair dealing, or indeed training on how best to use the material.

He explained that for teachers and students important features for the use of moving images are reliable information on provenance and context, flexible arrangements for local use and reuse, and a licence for public demonstration.

Context and provenance are key features of BUFVC collections such as Film and Sound Online, thanks to the presence of appropriate metadata. Videos are not simply posted online, but are

supported by information on who, when and why created the material as well as a description of the content of the video, which provides the sociological and cultural context. This is a key difference between BUFVC and sites like YouTube which don't provide much contextual information.

Like Prelinger, Weston touched on sustainability. BUFVC has to charge for their services as they needed revenue to fund their activities. They use a subscription model and charge institutions for membership. He also noted a few issues for the future, including authenticated audience access, exposing metadata outside the authentication layer, training staff in higher/further education and the need to engage with players like YouTube.

Peter Kaufman began with a somewhat shocking profile of the commercial marketplace. Television is the primary source of information in the world today, but we have lost control of the media. Where once the "fourth estate" fulfilled a role of checks and balances in the public interest, now it serves the interests of big business that controls the media.

He made some interesting observations about the video market:

- The demand for video is huge (more than 100 million videos are watched on YouTube daily)
- The tools that allow anyone to produce videos are proliferating
- Our moving image heritage is being rapidly digitised by many organisations, including JISC
- Personal storage and portability are increasing exponentially
- Distribution is easy, with sites like YouTube, My Space, Joost and many others.

In light of these developments, he argued that there were opportunities within the education sector to engage more with the study and production of video. He described the Open Education Video Studio, a project that Intelligent Television and Columbia University were working on with funding from the Hewlett Foundation. The project aims to

Key Themes:

- YouTube has dramatically changed the video distribution landscape and is a key player we will need to engage with
- YouTube has had an impact on the Prelinger Archive, slowing down ingest and lowering standards.
- Long term sustainability is an issue for non-profit organisations such as the Prelinger Archive and the BUFVC.
- Video is, and can potentially be, used in many ways in education, however there is a lack of supporting infrastructures for this medium and training on how to use it
- Contextual information is a key requirement for the use of video in education – this is not provided on YouTube
- Peter Kaufman argued that the education sector should engage more actively with the use and production of video

Digital images: Developments in capture, conversion and workflow

Moderator: Stuart Lee, Acting Director, Oxford University Computing Services
Nigel Goldsmith, Technical Research Officer, Technical Advisory Service for Images (TASI), University of Bristol
Richard Everett, Imaging Manager, Wellcome Trust

Nigel Goldsmith began by explaining the role of TASI, a JISC advisory service providing advice and training on images. There are over 500 pages of content on the TASI web site (<http://www.tasi.ac.uk/index.html>), so participants can find more details on image capture and conversion there. He then gave a quick tutorial on image capture technologies and summarised the pros and cons of the various file formats. As most are familiar with TIFF and JPEG, he focused on less familiar and new formats like raw files, DNG, and JPEG2000.

Cameras capture images in raw format, the equivalent of a digital negative. There are advantages in using raw images as an archival format and using software to enhance them. As software changes over time, more can be done with files saved as raw files than as TIFF or another format. Raw images are also smaller and better quality than TIFF. However, the downside is that there is no single standard for raw images, these standards are proprietary (the camera manufacturer) and conversion will always be needed to view them.

Adobe recognised the standards problem and created the new Digital Negative Format (DNG), a public archival format for raw digital camera files. The standard is freely available, but it's not open source. It is therefore a step in the right direction but not an ideal solution.

He then compared the new JPEG2000 format to JPEG. It allows better compression performance, both lossy and lossless, and multiple resolutions. The quality depends on the software that is being used, so one package may work better than another. On the downside, JPEG2000 is not widely supported on the web, so users need to download a plug-in to view images. Decompression is also slow, and this is a key reason for the slow take-up.

Richard Everett described The Wellcome Trust's processes for the production of digital images focusing on workflow. The Wellcome Trust uses photographs for many purposes, including digitisation projects and promotional work.

He began with some comments on image formats. Currently images are captured as raw files, kept as raw files throughout the workflow and then saved as TIFF in the Miro database. They may revisit this. He agreed with Goldsmith that there are advantages in saving raw files, as you can do more with them. DNG is an option, but they would need to be sure it's the right choice, as it's an irreversible step.

All of the image work is done on Macs, and AppleScript has allowed them to automate the workflow. Key stages in the workflow are:

1. Image capture
2. Download to local computer
3. Basic editing (raw files)
4. Copy files to secure server
5. Assign unique number (done by AppleScript)
6. Convert raw to TIFF for the Miro database (and also surrogate JPEGs)
7. Archive the raw and TIFF files
8. Catalogue the images (populate the headers)
9. Post to web site
10. Internal delivery, if they want an enhanced version.

The discussion focused on the difficult choice of what file formats to use for archiving and distribution purposes. TIFF is the standard that has been recommended for many years, but is not as good as raw files which provide a digital negative. However, raw files are a proprietary standard where TIFF is not.

Other factors also come into play, such as storage. As choosing the "right" format is difficult, in principle images could be saved in more than one format. However, this commits an institution to managing more images. Storage space may also be an issue. The British Library commented that they would like to archive images as raw or TIFF, but have chosen JPEG2000 as a more affordable alternative where millions of images are involved.

Key Themes:

- The pros and cons of various image file formats was discussed, e.g. TIFF, JPEG, raw files, DNG and JPEG2000, both for preservation and delivery purposes
- There is no single image format that is 'best'; digitisation projects will need to consider the pros and cons and make an informed choice
- Raw files are a more flexible archival format than TIFF, but are a proprietary standard
- There was also concern that Adobe's Digital Negative Format (DNG) is not open source.

Digital curation of digitised material: The what, why and how of digital preservation

Moderator: Alastair Dunning, Digitisation Programme Manager, JISC

Neil Beagrie, British Library

Gareth Knight, Digital Preservation Officer, Arts and Humanities Data Service, King's College London

Simon Tanner, Director, King's Digital Consultancy Services

Neil Beagrie spoke about Digital Lives, a major new research project led by the British Library and funded by the Arts and Humanities Research Council.

Personal digital collections are at the core of many special collections in research libraries, but little research has been done on how user-generated content should be preserved. There is increasing tension between the traditional model of digital preservation and Web 2.0. Web 2.0 will have a massive impact on learning, the ICT market and personal information management. There are also implications for the privacy and longevity of digitised material.

The project will explore how academics, authors, cultural figures and others create and manage their personal collections of digital information, from emails and blogs to documents and web pages, and how it might be preserved. The project will include:

- Interviews and surveys to gather in-depth insights into how different individuals are handling preservation and information management
- Research into the legal and ethical implications of digitising personal content, and the protection of both individual rights and those of the host institution
- User focus groups with representative users of digital collections to identify the types of information wanted
- Investigating and evaluating preservation tools and identifying how can they be used or transferred to special collections
- Developing relationships with long-term archiving repositories to enable personal collections to be curated professionally.

Simon Tanner demonstrated how to make a good case for funding preservation – how to justify it, build a case, and use leverage to get your way.

Cost-benefit is one justification for digital preservation. Kevin Guthrie addressed this in a recent study, *JSTOR: The Development of a Cost-Driven, Value-Based Pricing Model* where he compared JSTOR with traditional forms of storage and found that, while the costs of digital libraries are more visible, they are not necessarily greater. Risk management is another important justification. It is essential to quantify the risks and consequences associated with the loss of digitised materials. Quantifying risks can be difficult, but this is important to demonstrate that the consequences are worth the cost.

He gave some practical advice on building a case for funded repositories:

- Identify a brief timeframe during which action can be taken; don't attempt to do everything immediately
- Raise awareness of the increasing dependence on digitised materials
- Produce evidence of the various cost elements
- Be persuasive
- Emphasise the strategic fit with current institutional goals
- Show a clear understanding of the relationship between costs and benefits
- Aim to benefit the designated community and stakeholders

Preservation is an important issue and has wider implications than preserving the cultural heritage. It can be helpful to get institutions to focus on the broader issue of preservation and how this relates to the institutional mission. Preserving cultural heritage is therefore part of the broader picture.

Gareth Knight focused on the importance of documentation in managing complex resources. Digital resources are increasingly complex and difficult to manage. An archive like the Arts and Humanities Data Centre (AHDS) cannot provide supporting documentation; only the original creator who deposits a collection with the AHDS can. Documentation isn't something to do at the end of the project, but an ongoing process that is integral to creating the data. Decisions should be documented when they are made, or there will be no record when the project ends and staff leave.

A preservation and curation strategy is also key to the success of the project, and this should be established as soon as possible. Useful questions to ask when developing and refining a preservation and curation strategy are:

- What do you wish to manage and preserve?
- What are the core components of the resource?
- What are the significant properties of each component?
- How do the components of a resource relate to one another?
- How does the resource relate to others?

Knight provided some examples showing how to address these issues and suggested thinking about what is unique about the resource to ensure that the value of the resource to the user is maintained.

Key Themes:

- There is increasing tension between the traditional models for digital preservation and Web 2.0
- The British Library Digital Lives project will help to develop strategies for the preservation of user-generated content
- We need to be able to make a good case to our institutions for digital preservation and justify the costs. Cost benefit, risk consequences, and legal compliance are all important justifications.
- It can be helpful to get the institution to think about preservation at a broad institutional level, i.e. preserving cultural heritage is part of the broader picture
- It's good practice to create and maintain good documentation for projects developing complex resources and to develop a strategy for preservation and curation from the start.

Exploring commercial e-content developments and private/public sector partnerships

Moderator: Liam Earney, Collections Team Manager, JISC Collections
Dan Burnstone, Director of Publishing, Chadwyck-Healey, ProQuest

Dan Burnstone gave some useful and practical advice about private/public partnerships based on his experience at ProQuest. He started by describing some of their collaborative projects, illustrating the wide range of resources developed, the different types of relationships, and resulting innovations. One of the collaborative projects ProQuest is involved in is the JISC funded Electronic Ephemera: Digitised Selections from the John Johnson Collection with the Bodleian Library. He then gave some cardinal rules for public/private partnerships:

- **Goals** – Before entering a collaboration, both sides should be completely clear about what they want to achieve and their requirements in the short and long term. A shared understanding of these goals will define the relationship from the start.
- **Legal framework** – A robust legal framework should document the roles and responsibilities of all parties in the project.
- **Communication** – Clear, open, and frequent communication is important
- **Project planning** – Well-defined roles so everyone knows what they are doing and works to a project plan.

There can be many benefits in working with the commercial sector, as they bring many strengths and skills to the table. Marketing and promotion is a key area. Digitisation projects involve great time and expense, so it's in everyone's interest to ensure there is a real market for the new resource and to reach the intended audience. The commercial sector has an existing infrastructure and wide experience in market research, promotion, and sales. They can also offer assistance in rights clearance and IPR since they have departments set up to handle these arrangements

Another consideration is technical capabilities. The commercial sector has robust services and can develop value-added features. They are also increasingly aware of the standards used in the education sector and will have experience from past projects.

Burnstone also outlined some areas where collaboration would be useful in terms of sharing views and reaching a common understanding, and both sides can leverage activities being undertaken by the other. For example:

- **Resource lifecycle** – Publishers would like guidance from the public sector on issues like the long term embedding of resources within education and research, and on the long term archiving requirements the commercial sector should work towards.
- **Avoiding duplication** – The private sector does its best to understand what digitisation projects the public sector is initiating and will adjust its own plans accordingly. This raises the question of how much information can/should be shared while allowing both sides to pursue projects they feel are important.
- **Best practice on collaboration** – It would be useful for JISC to provide projects with best practice guidance on collaborating with the private sector.

Key Themes:

- Before entering into a public/private collaboration, both sides should be completely clear about what they want to achieve in the short and long term
- It is important to have a sound legal framework for the collaboration so both sides are clear about their roles and responsibilities.
- The private sector can bring a lot to the table, including capabilities and experience in market research, promotion, delivering robust services, developing value-added features, and legal assistance.
- There are opportunities to collaborate more broadly than creating resources, e.g. on lifecycle issues like preservation and embedding, and agreeing best practice for collaborative projects.

Capacity building: Investment in centres of excellence in the European Union and the US

Moderator: David Dawson, Senior Policy Adviser (Digital Futures), Museums, Libraries and Archives Council

Kevin Guthrie, President, Ithaka

Anne-Marie Millner, Manager, Content Management and Capacity Building, Canadian Heritage Information Network

Kevin Guthrie noted that the centre of excellence is an important concept in the academic world, e.g. an institute for advanced study with an intellectual focus on organising interdisciplinary academic activity within a university. Single institutions cannot do everything, so a push towards specialisation which promotes centres of excellence is beneficial for the community at large. However, we should think about how the term "centre of excellence" applies to digital work and how the status is awarded; it is often self-awarded rather than decided by peers.

He explained Ithaka's mission – to help the worldwide scholarly community take fullest advantage of advances in information technologies through helping to develop business discipline in the fields of strategic services, research and the shared services of IT, human resources and finance. As part of this mission Ithaka has developed "incubated entities" that could be considered centres of excellence, each with a different role:

- Aluka – An international collaboration to create a digital library of scholarly resources from and about the developing world, focusing initially on Africa
- NITLE – An initiative promoting liberal education
- Portico – An initiative for the long-term preservation of scholarly literature.

As JISC is a focal point of “best practice”, it could consider awarding the status to appropriate centres in the UK.

Anne-Marie Millner described the work of the Canadian Heritage Information Network (CHIN), an agency of the Government of Canada’s Department of Canadian Heritage. CHIN (<http://www.chin.gc.ca/>) is a national centre of excellence which “enables Canadian museums to engage worldwide audiences through the use of innovative technologies”. Their business model as a centre of excellence involves:

- Enabling collaboration within the heritage community
- Providing resources and tools to reach a worldwide audience
- Enhancing capacity building in both large and small heritage institutions.

She described how CHIN works through its programmes. The Virtual Museum of Canada is a high-profile portal for museum content and builds capacity for content creation within the museum sector through applied learning. The Knowledge Exchange builds capacity by providing resources like e-tutorials and best practice, encouraging peer-to-peer engagement through wikis and blogs, and providing access to experts through interviews and podcasts.

David Dawson noted some developments in the EU mentioned by **Pat Manson**. Competence centres are being developed through the FP7 research programme. The idea is to reduce costs by building expertise and making it available to others. He mentioned Presto Space (<http://www.prestospace.org/project/index.en.html>), a project on the digital preservation of audiovisual collections that acts as a virtual centre of excellence. This is an interesting model, and there’s more funding to come from the EU for this sort of thing.

In summing up, Dawson noted that sharing expertise will reduce costs, resulting in more digitisation of resources and competence building across universities. The starting point for a centre of excellence is existing excellence and turning this into a competence centre.

The discussion picked up on the point made by Guthrie about how the concept of centres of excellence relates to digitisation. Would an institution become a centre of excellence by doing digitisation work or by gaining expertise in outsourcing it? This would depend on the fit with institutional aims. Dawson noted a MLA North East programme to promote engagement between the cultural sector and small to medium enterprises (SMEs). This creates a virtual circle by training SMEs to raise their skills, making the cultural sector smarter customers, and putting them together to promote growth in the region.

Another topic of discussion was aggregating information about expertise. Perhaps funding agencies (like IMLS or JISC) could take a role in this area. This could be helpful for new applicants for funding and newly funded projects.

Finally, it was noted that there is no agreed definition of what “centre of excellence” means or how the status is awarded. It can be earned through experience and hard work, but it can also be conferred by virtue of funding or self-awarded. It would be useful to differentiate between these models and consider badging or branding.

Key Themes:

- Sharing knowledge and building capacity can help to reduce costs
- Centres of excellence are well established in the context of academic interdisciplinary research but perhaps need defining in the context of digitisation
- There is no agreed definition of what “centre of excellence” means or how the status is awarded. Badging or kitemarking could be useful to ensure the term is used in a meaningful way.

- The Canadian Heritage Information Network (CHIN) is an interesting example of how a centre of excellence builds capacity within the cultural sector.

Transforming the users' experience: How institutions can develop innovative and affordable tools to engage increasingly sophisticated audiences

Brian Kelly (moderator and speaker), UKOLN, University of Bath
Adrian Arthur, Head of Web Services Team, British Library
Alistair Russell, Developer, MSpace, University of Southampton

Brian Kelly opened the session using scenario planning to explore a future where Web 2.0 has won and services for delivery are in place. Google and YouTube is the environment that people use. There is enormous investment, so it's a sustainable environment. It is engaging users and getting users to engage with content. It is tapping into our core mission – learning – not just formal learning, but lifelong learning and learning as part of society. The scenario is already happening in some areas, for example in schools.

If this is the future, there will be challenges. A key challenge will be to get our audiences back. They were a captive audience when there were only VLEs. Increasingly users will expect to access resources in Web 2.0 environments. There will be benefits for us in making our content available where the users are.

But there are also opportunities if we are prepared to learn from the successes of the Web 2.0 environment. The higher education sector is good at this, as we've been doing it for years. We need to decide how to apply these technologies well, contextualise them so they fit in well with learning, teaching, research, and administrative processes. We will also need to think globally, work in a mixed economy, and be prepared to work with third-party developers. Regarding tools, institutions may not need to develop their own, as they may find they are already out there – Google, YouTube, etc.

At the CILIP Umbrella conference in June, Lynne Brindley gave an inspiring presentation about new developments at the British Library. She urged delegates to embrace the Web 2.0 world and “just do it”. He noted that there may be risks, for example a third-party provider may go out of business, but this is nothing new. There are also risks in *not* doing it.

Adrian Arthur described how the British Library is transforming and improving the user's experience and what they are learning from users. Web 2.0 is a key part of the BL's new strategy. In June 2007, Facebook, MySpace, and Bebo accounted for 15% of UK web pages viewed. It's therefore an important phenomenon we all need to take note of.

The BL is increasingly taking advantage of the web. They have migrated some traditional services to the web, like BL Direct for document supply. They have also developed new resources to serve their traditional markets in new ways, like Images Online. He noted some key themes in their approach to web development such as the importance of usability and accessibility, optimising the use of their (scarce) internal resources, and working with funders and partners to maximise what they can offer to their users.

He described two of their innovative Web 2.0 developments that relate to the user experience:

- **Turning the Pages** – This allows users to turn the pages of a book, e.g. in a gallery using a touch screen, or on the web using a mouse. It provides a rich user experience and is very compelling. The initial development was with Armadillo and they are working with Microsoft on the next version.
- **Google maps mashup** – London: A Life in Maps is an interactive map providing information about exhibitions. They used the Google Maps API and populated it with pushpins. This was done cheaply and easily inhouse.

The lessons they learned apply to Web 1.0 development as well as Web 2.0:

- Ask “Who are your users?” and “What are they trying to do?” Both are critical to ensure a web resource works for them
- Lab based usability testing – This gives a real understanding of how people interact with a resource
- Don’t assume everyone is the same – The Google Maps mashup is a great application, but about half of their users prefer using text links. A really good solution for some isn’t enough; you have to think about the rest.
- A compelling interface really pulls people in to your content – Turning the Pages is really easy to get started with.
- There is no substitute for innovation – It can come from within the organisation or from working with external partners
- Be as professional as you can within the limits of your resources.

Arthur also described Archival Sound Recordings (<http://sounds.bl.uk/>), a project in the JISC Digitisation programme to open up their sound archive to a wider audience. This includes 12,000 segmented recordings on a range of topics for learning, teaching and research.

They formed a user panel at the start of the project and worked with them to develop the user interface. In addition to usability testing, they also did accessibility testing to ensure the site worked for disabled users. He gave a demo illustrating that the user interface is quite simple – they found that users wanted to get to the content quickly and easily. There is an Amazon-like feature that tells users that “people who listen to this also listen to...”. Lessons learned from this project included keep it simple, start thinking about the web site at the start of the project, and start working with users as soon as you possibly can.

Alistair Russell described the approach MSpace and taking to develop the user interface for Newsfilm Online (<http://newsfilm.bufvc.ac.uk/>). Newsfilm Online is a project in the JISC Digitisation programme led by BUFVC. It will feature 60,000 segmented clips from the last hundred years of television news and cinema newsreels. The aim is to make it available for teaching, learning and research across a wide range of academic disciplines.

Like previous speakers, he emphasised the importance of Web 2.0 and engaging with users. Children are growing up using Facebook and are very web savvy. They know how to browse, tag, and social networking is popular. Mashups are common, and there are lightweight tools like Yahoo Pipes and Microsoft Popfly which allow non technical users to aggregate content from different sources in a user friendly way.

Sophisticated users know what they like and how to find it, but a question MSpace has been addressing is, what if users don’t know what they like? If a user likes classical music, they can build up a profile and get recommendations based on the profile. But a user who knows nothing about classical music doesn’t know where to start. The MSpace user interface is an exploratory search tool that allows users to discover, explore the data and relationships and make an informed choice. He showed a demo of how the MSpace user interface will work and noted that a beta release will be available in September 2007.

MSpace was originally based on principles of the semantic web. Russell noted that the blog on the O’Reilly site characterised Web 2.0 as instant superficial gratification where the semantic web is more of a deep and meaningful relationship with data. There may be tension between the approaches, but both have similar goals, to share and aggregate data. Web 2 has the interface and the semantic web gives power to the data so it’s usable. In developing MSpace, they have combined the best aspects of both to move towards a Web 3.0 environment.

Key Themes:

- Brian Kelly posed a “radical” future scenario where Web 2.0 has won and services for delivery are in place. There are issues related to sustainability, risk, and preservation, but little disagreement that Web 2.0 figures in our future. There are also risks in *not* embracing Web 2.0

- Increasingly our users will expect to see their services delivered in a Web 2.0 environment, and there will be benefits for us in making our content available where the users are.
- Adrian Arthur noted that Web 2.0 is a key part of the British Library strategic plan. Key themes in their use of web 2.0 are the importance of usability testing, optimising the use of their (scarce) internal resources, and working with funders and partners to maximise what they can offer to their users.
- There are already lightweight Web 2.0 tools like Yahoo Pipes and Microsoft Popfly which allow non technical users to aggregate content from different sources in a user friendly way.
- Alistair Russell described MSpace which solves the interesting problem of how to browse when you don't know what you want, combining Web 2.0 and semantic web technologies.

Digital capture and conversion of text: Overcoming the Optical Character Recognition (OCR) challenges

Moderator: Paul Ell, Director, Centre for Data Digitisation and Analysis, Queen's University Belfast
Aly Conteh, Head of Digitisation, British Library
Julian Ball, Project Manager, BOPCRIS, University of Southampton
Martin Locock, Project Manager, Welsh Journals Online, National Library of Wales

Aly Conteh spoke about the British Library experience of digitising newspapers from the 17th, 18th and 19th centuries (http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation/newspapers2.aspx) as part of the JISC Digitisation Programme. They are also digitising 25m pages of 19th century books in conjunction with Microsoft, which involves a throughput of about 1m pages a month. He outlined the challenges from their perspective.

Firstly, OCR technology tends to be tuned for modern printed materials. Historic newspapers present difficulties because of the column format and fine print. A key challenge is therefore that there is no comprehensive set of tools that allows us to work with these historic materials. There are tools like Abbyy 8 and Old English, but Abbyy is probably the only company doing research in this area. So how can the community stimulate the development of the tools needed to open up content for searching?

Another problem is the "character word accuracy" issue. It is not possible to reach a 99.9% accuracy without manual intervention, so we need to decide on the level of accuracy that is acceptable in order to enable searching. Here the unit of currency is the word, and if the accuracy rate is only 50% then we're losing a great deal of searchable content. Finally, he noted the need for methodologies and benchmarking to measure the effectiveness of OCR. To measure accuracy we have to physically count it.

Julian Ball described digitisation of the *18th Century Parliamentary Papers* (<http://www.bopcris.ac.uk/18c/>), also part of the JISC Digitisation Programme. He showed the Swiss robotic scanner used for the large folio volumes and explained the workflow to capture the images and text. When maps are scanned, OCR captures the place names and these are then used for searching.

Both OCR and re-keying are used to capture texts. Where the text will be exposed to the user, typically re-keying and triple OCR are used to ensure accuracy. OCR to lower standards is used for searchable PDFs and hit term highlighting. He gave examples to show how ABBYY 8 and Old English coped with various 18th century texts. Where the OCR output contains errors, fuzzy searching can improve search results.

Finally, he showed how digitisation can be used to improve accessibility for disabled users. For the visually impaired, maps and diagrams can be created with raised punch dots similar to Braille. The OCR output can also be used to generate Braille and output for the hearing impaired.

Martin Locock explained problems the Welsh language poses for OCR. Diagraphs and diacritics are an integral part of the Welsh language and quite common rather than incidental. They can't be ignored as they affect the meaning of a word. He described the approach they have taken for various projects.

Books from the Past (<http://www.booksfromthepast.org/>) is an online collection of books of national cultural interest which have long been out of print. The project acted as a pilot to explore the issues of digitising a range of historical texts. The intention was to create clean TEI (Text Encoding Initiative) text to accompany the scanned pages. However, OCR wasn't able to capture the diagraphs and diacritics correctly and the text had to be cleaned manually. They are looking at new solutions to improve accuracy.

Welsh Journals Online (<http://www.llgc.org.uk/index.php?id=2244>), part of JISC Digitisation Programme, covers 20th and 21st century Welsh journals. Here the OCR text is used for searching but not displayed to users. They are therefore exploring solutions to optimise search results. One option would be to separate interfaces for Welsh and English searching and apply special rules to the Welsh language content. Another would be to use silent fuzzy searching (lookup tables) or ask users "Did you mean...?".

The discussion focused on accuracy for mass digitisation projects. The OCR tools available today don't provide the accuracy we might like. Re-keying or correcting the OCR isn't practical, so we may need to compromise and say this is the best we can do at present. In the future better tools may become available. In the meantime we can explore techniques like fuzzy searching to optimise searching for users.

Key Themes:

- OCR adds value to digitisation projects by unlocking the content in scanned pages for end users. It enables searching and can be used to increase accessibility for disabled users (e.g. Braille, relief maps, and sound).
- OCR tools are optimised for modern text, so digitisation of texts with special requirements presents challenges. Speakers outlined the problems they faced for newspapers (columns of text with small fonts), historic texts (ligatures), and Welsh language materials (diagraphs and diacritics).
- Perhaps the community could drive forward the development of new OCR tools and then material would be re-digitised in the future. A stimulus is needed from the community to drive forward more research in this area.
- Deciding on an acceptable level of accuracy is an issue and benchmarking is needed to measure it. For mass digitisation projects, we may need to be pragmatic and say that this is the best we can do at the current time.
- Techniques like fuzzy searching can help to optimise searching for end users and make the most of the OCR output.
- Most projects are using Abbey 8 or Old English and have a high throughput of digitised content. They are moving away from re-keying and correcting the OCR, and are taking a "best effort" approach. Fuzzy searching can improve the search results on OCR'd texts.

The legal landscape: Copyright, IPR and licenses and mass digitisation

Moderator: Emma Beer, Manager, Strategic Content Alliance, JISC
Emanuella Giavarra, Legal Counsel, JISC
Naomi Korn, IPR Consultant, JISC

This session covered rights clearance and licensing and provided a wealth of best practice for digitisation projects from both a practical and legal perspective.

Naomi Korn focused on the practical issues relating to IPR clearance. Digitisation projects can involve a wide range of content, e.g. images, text, music, film, sound recordings and e-theses. Clearing rights can be complex, as it can involve different types of copyright, other types of intellectual property, more than one rights holder per item, and a range of legal issues. Ideally, rights should be cleared either when the content is created or when it is acquired. Human resources and costs therefore need to be considered early on. It is therefore important to develop rights clearance strategies at the start of the project. Practical advice included the following:

- Make IPR clearance an integral part of project planning
- Make it easy for rights holders to give consent
- Adhere to the terms and conditions of your funding bodies
- Make risk management a part of your strategy; this should be a policy decision within the organisation
- Be aware of outstanding issues such as orphan works (where the content owner cannot be found)
- Make sure the time/cost of clearing rights is proportionate to the value of the work.

Risk management is an important area. When rights are not cleared appropriately, this poses legal risks. It can also adversely affect relationships with rights holders, funding bodies, and project partners, and result in loss of trust. It's therefore important to think through who will be liable, understand how prepared they are to take risks, and have procedures for Due Diligence. Disclaimers may not cover you.

Open content licenses, such as the Creative Commons license and the Creative Archive license can be useful when licensing digitised content.

In conclusion, IPR needs to be embedded within the organisation at the very beginning of the project. The project team should make informed decisions regarding copyright and develop a straightforward approach that everyone can understand.

Emanuella Giavarra focused on the legal issues for rights clearance and licensing. Copyright clearance can be done in the following ways:

- An email or letter, although this has limited enforcement
- A Memorandum of Understanding, which must be subject to contract
- A contract

A contract is good practice, even where the material is out of copyright. She noted that under English law, a contract must include an offer, acceptance of the offer, an intention to contract, and a consideration. A model contract (for out of copyright material) and a license agreement (in copyright material) are available for JISC funded projects. When clearing rights for high risk material, it is good practice to have a full contract. As project teams are making decisions on behalf of their institutions, investment in professional indemnity insurance is also worth considering.

Where projects have two or more partners, they should sign a consortium agreement to specify their duties, obligations, and the liabilities of the project partners. This should cover issues like responsibility for rights clearance and ownership of intellectual property. JISC will soon release an updated model consortium agreement for JISC funded projects.

Giavarra also gave guidance about options for end user licences. For open access, any of the following may be used to allow uses permitted by copyright law, Creative Commons (for CC licences), or use you have managed to clear:

- Creative Commons licence – Suitable for low risk material, but as the liability travels with the license this may not be preferable to the end user or institution.
- Your own “click on” license – Provides flexibility
- Terms and conditions – The extent to which these are enforceable is questionable. When terms and conditions are set out in email format, it is good practice to include an “I agree”

statement which can be ticked by the person who would be otherwise signing the contract. A signature is not necessary if “I agree” is stated.

For controlled access, a own “click on” licence, a signed licence, or terms and conditions may be used for any use of material that has been cleared. She encouraged projects to consult the JISC Collections Model Licence for a list of permitted uses that should be cleared.

Key Themes:

- It is important to embed IPR within the project and develop rights clearance strategies from the start
- When clearing rights, a contract is always better than a letter
- It is important to understand the legal risks and to take the high risks very seriously
- Where risks are low, projects need to think about when they can “run a red light”. Bear in mind that the institution is liable, so find out what risks they are prepared to take
- The risks are more than legal. Trust and relationships the project has built are also at stake.

Developments in resource discovery portals for digitised material

Moderator: Balviar Notay, Programme Manager, Information Environment, JISC
Ian Dolphin, Head of e-Strategy, University of Hull

Ian Dolphin focused on resource discovery and portal development from the institutional perspective. His working definition of a portal is “a thin layer which aggregates, integrates, personalises, and presents information, transactions and applications to the user seamlessly and securely according to their role and preferences”. He noted the importance of personalisation – there’s a political imperative to deliver personalised services to meet individual students’ needs as well as evidence that users want this flexibility.

He described the work of the JISC-funded CREE project (<http://www.hull.ac.uk/cree/>) in this area. The first phase of the project proved the concepts of how resource discovery should work in the context of institutional portals. The next phase will focus on configuring portals so they really work the way users want. This will include issues like enabling advanced search features, displaying results to make them as usable as possible, and enabling request and delivery in a seamless way. It will also address broader issues like the “handover point”, the point at which it is preferable for users to leave the portal to continue their exploration in external environments. He covered a range of issues like standards, tools, and Web 2.0.

Balviar Notay focused on the broader context of portal development, within JISC and internationally. To illustrate this, she described the Go-Geo! portal (<http://www.gogeo.ac.uk/>), a geographically-oriented access point to the JISC Information Environment. Go-Geo! promotes awareness of geospatial data within the academic sector and the wider GI community, so users can make more effective use of this data and find related resources.

The aim of developing portals like Go-Geo! is not to simply build another portal. A key aim is to explore the technical design issues for creating seamless networks of portals and to make the technology invisible. She noted the importance of metadata for resource discovery across multiple services. The JISC data centre EDINA has been instrumental in developing the architectures and concepts that will work behind the scenes.

Key themes:

- Portals and resource discovery are about making resources more accessible and usable to end users while making the technology invisible
- The UK Government recently highlighted the importance of personalising services to meet individual students’ needs in higher education
- The CREE project has proved the key concepts of institutional portal design; the next phase will focus on configuring portals so they work the way users really want

- The Go-Geo! portal illustrates how JISC is developing “not just another portal” but architectures to create seamless networks of portals.