

JISC DEVELOPMENT PROGRAMMES

VSM Demonstrator: *Final Report*

Project

Project Acronym	VSM Demonstrator	Project ID	
Project Title	Visual Sound Materials Portal – Demonstrator phase		
Start Date	Jan 2007	End Date	Jan 2008
Lead Institution	EDINA, University of Edinburgh		
Project Director	Co-directors: Peter Burnhill & Christine Rees		
Project Manager & contact details	Tim Stickland t.stickland@ed.ac.uk 0131 650 3309		
Partner Institutions	Glasgow Caledonian University, University of Hull, St Helens College		
Project Web URL	Project web site: http://edina.ac.uk/projects/vsm Demonstrator: http://vsm.edina.ac.uk/vsm/vsm.asp		
Programme Name (and number)	<i>Portals</i>		
Programme Manager	Balviar Notay		

Document

Document Title	<i>VSM Demonstrator Final Report</i>		
Author(s) & project role	Leah Halliday, Rick Loup, Christine Rees, Ben Soares, Tim Stickland		
Date	January 2008	Filename	
URL	http://edina.ac.uk/projects/vsmportal/outputsandactivities.html		
Access	■ Project and JISC internal	■ General dissemination	

Document History

Version	Date	Comments
V0a	March 2008	Draft submitted to JISC
Final	June 2008	Submitted to JISC and made available at project website

Visual and Sound Materials for UK Further and Higher Education

VSM Portal Demonstrator



Visual and Sound Materials Portal Demonstrator Project

Final Report

Author(s): Leah Halliday, Rick Loup, Christine Rees, Ben Soares, Tim Stickland

Date: May 2008

Acknowledgements

The Visual Sound Materials Portal Demonstrator project was funded by Joint Information Systems Committee under its Portals Programme.

The project partners were:

- Hull University: Ian Dolphin, Rob Sherrat, and Richard Garbutt
- St Helens College: Marianne Green
- Spoken Word Services based at Glasgow Caledonian University: David Donald, Iain Wallace, Caroline Noakes

The team would like to acknowledge the support and contributions of the Steering Group for the project:

- Angelo Conti (Swansea College)
- Balviar Notay (JISC)
- Linda Purdy (Sheffield Hallam University)
- Hetesh Morar (JISC)
- Anthony Theobald (TASI)
- Nick Townend (BUFVC)
- Julie Rowbotham (BBC)
- Rob Sherrat (Hull University)
- Murray Weston (BUFVC)

Related Documents

RELATED DOCUMENT	DETAILS
Appendices to VSM Portal Demonstrator Final Report (June 2008)	All documents available at: <i>http://edina.ac.uk/projects/vsmportal/outputsandactivities.html</i>
Technical plan (June 2007)	
Software documentation and archive (December 2007)	
Functional specification prior to implementation	
Draft portal/content provider agreement (June 2007)	
Draft collections strategy (September 2007)	
A notice and takedown policy for content subject to complaint	
A draft policy to support the contribution of user-generated content	
A notice and takedown policy for user-generated content subject to complaint	
External evaluation report (January 2008)	

The VSM Portal Demonstrator is available at: <http://vsm.edina.ac.uk/vsm/vsm.asp>

Contents

1. Executive Summary	1
2. Background	3
3. Aims and Objectives	4
4. Methodology	5
Phase 1 - scoping.....	5
Phase 2 - demonstrator	5
5. Implementation.....	6
Demonstrator development	6
Institutional embedding.....	10
Metadata.....	12
Content	12
Evaluation	13
6. Outputs and Results.....	14
Phase 1	14
Phase 2	14
7. Outcomes.....	17
Beneficiaries.....	17
Project achievements.....	17
Knowledge and experience gained.....	17
External evaluation	19
8. Conclusions	21
9. Implications.....	23
Potential developments	23
Possible roles for a future service	27

1. Executive Summary

“Sometimes I find it easier to retain information on a topic if I have a visual memory of it ... the visual representation of a topic can provoke an emotional response within an individual which in turn helps them connect with the subject and stimulate interest ... I think it should be used more frequently¹”

Moving picture and sound and still images represent a significant and, arguably, growing part of everyday life. They offer potential for education and scholarship across the disciplines in three ways:

- As a major part of the observational record and insight into our past and present;
- As appreciation of the aesthetics of rich cultural heritage;
- As means of visualisation and of other expression of scholarly endeavour.

The principal advantage of digital representation is the transformation into information objects that become assets with potential for re-use, across time and across place.

The Visual Sound Materials Portal Demonstrator project investigated the value and feasibility of a JISC-designated national portal for both time-based media and image collections, dedicated to the needs of UK Higher and Further Education (H/FE). It is part of JISC’s co-ordinating policy for multimedia service provision. JISC has supported both programmes of digitisation, with the focus on format, and the provision of multimedia resources for H/FE, mainly through the licensing of a number of commercial image collections, but also moving image collections. Time-based media have complex rights issues, as well as challenges for storage, metadata and retrieval.

The principal beneficiaries of a VSM Portal will be learners, educators and researchers in UK H/FE, and also the providers of content for the portal, who will benefit from exposure of their content to users in the educational community, a potentially new and important audience.

By being focused on the requirements of the H/FE community, the VSM Portal will be different from more general media-specific resource-discovery tools on the internet, such as Google Images, Google Video and YouTube. The external evaluation undertaken by Rightscom Ltd during the course of this project reported that academics, while sometimes finding suitable materials in these services, want a way of ensuring that materials are relevant, of high quality, from trusted sites, and copyright cleared for use in teaching. Developing a VSM Portal in the light of the experience of this project offers the prospect of meeting these requirements.

Following a review of the literature (including JISC-funded project reports) and an expert focus group, a scoping report was submitted with recommendations to a JISC Review Panel. The Panel’s direction for development of the demonstrator was that:

- Community-based content should not be collected by the demonstrator but suitable services providing access to such content should be portal-demonstrator targets;
- The demonstrator should utilise and test appropriate Web 2.0 technologies;
- The demonstrator should implement and test content-based image retrieval; and
- A portlet should be developed, and embedded and tested within the user environments.

During the demonstrator phase, progress was communicated to the Review panel using a project Blog.

¹ Quotation from student at Sheffield Hallam University. JISC Infonet. *Exploring Tangible Benefits of e-Learning: does investment yield interest?* Northumbria University, 2008. Page 30. Available at: <http://www.jiscinfonet.ac.uk/publications/publications/info/tangible-benefits-publication> [May 2008]

The project aimed to:

- Scope and design a suitable architecture and functionality taking in the broader context e.g. development of repositories of user-generated content and public sector search engines;
- Test the feasibility of serving different media types from a single platform using new technologies;
- Test the embedding of the demonstrator in different user environments within the UK HE/FE sector;
- Develop and test suitable licences and policies to support such a service.

The project met all of these aims and its highlights include:

- The development of the portal demonstrator itself, developed and tested within the 12-month funded period;
- The development of a portlet, also embedded and tested within the 12 month period;
- The demonstrator included a range of collections with good coverage of the different media types. Eleven collections are currently represented, with a further nineteen approached, of which nine have agreed to participate;
- Appropriate Web 2.0 technologies and content-based image retrieval were implemented and tested.

In addition, the project produced the following outputs:

- User requirements for a VSM Portal demonstrator;
- Draft technical and functional specification for a portal service;
- Documentation regarding the deployment of the portlet within a uPortal framework, that can be used to deploy the portlet on any portal framework;
- Draft collections policy, including a strategy to increase the range of content available, priorities and key criteria for a collections policy;
- Draft content licensing model;
- Takedown policy;
- Scoping report;
- Recommendations for future development and requirements
- A range of appropriate business and sustainability models intended to help inform decisions about future provision was an output from the External Evaluation.

The challenge was to integrate the facility for discovery and location of different media types and across all users within the sector (i.e. rather than specific subjects/disciplines); for example, how to present results in a coherent way when there are many more still images than moving images or sound, and when access issues and infrastructure requirements are far more complex for licensed films than for the other two media types.

There was no interest among content providers in distributed search technologies. Metadata were harvested to create a database which was the source of results of searches in the portal, which facilitated better functionality in the portal than would have been available using distributed searching.

Results from user testing at four institutions (two HEIs and two FEIs) informed a second iteration in the development cycle, while the external evaluation examined the potential impact on the learning and teaching communities of a VSM Portal service and the benefits it could deliver.

As indicated by the external evaluators, the development phase of this project was too short to allow for development, implementation and a useful user evaluation exercise. Nevertheless, implementation, testing and evaluation generated a substantial amount of useful information to inform development of a VSM Portal service. In addition, several possible roles for a future VSM Portal service have been derived from the work done during this project.

2. Background

The VSM project investigated, through initial scoping activity and the construction of a demonstrator, the value and feasibility of a JISC-designated national portal for both time-based media and image collections. The project had two phases: a scoping phase (October 2005 and January 2006) followed, after JISC review, by a demonstrator phase: construction of a demonstrator was carried out between January 2007 – January 2008.

Multimedia, both time-based and still image collections, pose several special demands upon information technology that are different to those that were first encountered with encoded numbers and words, and thus upon the development of the digital library. That said, there is much that may be regarded as generic to the digital with respect to user behaviour, digital library provision (including collection development, licensing and delivery of service), and software deployment across ICT infrastructure. What is important is critical appraisal of variants that challenge the general model. These include variants of metadata (for discovery, for retrieval, for play/display and for use), of services for access, and of authorised conditions of use. There are, perhaps, also variants in matters of collection development, with respect to material generated within academe as well as by external agencies.

The challenge identified and addressed during scoping was to re-think the means of integrating the user tasks represented by the MODELS verbs (discover, locate, request and access) into a portal in ways that reflect the different and, perhaps competing, requirements of diverse communities of use. The challenge during the demonstrator phase was to investigate the two main aspects of feasibility for the VSM Portal: one technology-focussed, and one user-focussed, so contributing further steps towards a user-led specification and development of a facility providing a single integrated access point for the range of still-image, moving-picture and sound collections in the JISC Information Environment (IE) and those that are freely accessible. It must also be consistent with shared assessment of the JISC IE as it maps onto longer term Internet developments.

The investigation of a VSM Portal is part of co-ordinating policy for multimedia service provision. JISC has supported both programmes of digitisation, with focus on format, and the provision of multimedia resources for FE and HE communities, mainly through the licensing of a number of commercial image collections. A report from the JISC Images Working Group (IWG)² indicated awareness that, “... *the technology [for delivery of images] is mature, demand for the right kind of image content ... is reported to be high, competition to provide such services is ... increasingly fragmented*”.

One of the main recommendations arising from a JISC study into the Future Programme Collecting Requirements and Priorities for Moving Pictures and Sound is that, “*The JISC community now has enough experience with time-based media to know that they carry with them more challenges than most other media ... their rights issues are more complex ... they present special challenges for storage, indexing, and retrieval. An exploration of their potential will need novel and sustained research and development.*”

In response to this challenge, the present project is seen as contributing further steps towards a user-led specification and development of a facility providing a single integrated access point for the range of still-image, moving-picture and sound collections in the JISC IE and beyond.

² JISC Images Working Group Vision Document (May 2005)

3. Aims and Objectives

The ITT³ expressed the aims and objectives of the project as:

- *Scope an appropriate architecture and functional specification for a visual and sound materials portal taking into account the JISC funded PIXUS image portal project solution;*
- *Take account of the wider context, for example, approaches being developed by search engine providers, the emerging network of interoperable digital repositories, other service providers and software solutions;*
- *Test the feasibility of serving both time-based media and images from a single portal platform;*
- *Assess the usefulness of new technology that has become available and how far it might meet user needs; and where possible to test and implement new technology based on web services and standards such as JSR168;*
- *Test the embedding of the demonstrator portal within different user environments, for example, university and college portals or web environments;*
- *Consider the range of relevant community (further and higher education and public sector) originated collections alongside formally published material that could form the basis of a collections strategy for the portal;*
- *Conduct user evaluation;*
- *Propose a content licensing and IPR structure suitable for such a portal;*
- *Develop business and sustainability models suitable for future services.*

In particular, the aims of each of the two phases were defined as:

Phase one: Scoping Study. *A detailed scoping of the functional, technical and content requirements of a visual and sound materials portal with a view to a potential long term service for the post-16 education sector. The intended outcome is to provide a body of evidence and to identify a way forward. In particular it should specify the scope of the proposed phase two demonstrator.*

Phase two: Portal Demonstrator. *Implement further and test out the findings of the scoping study by producing a robust and scalable portal demonstrator, paying particular attention to testing the feasibility of the portal in general, looking at functionality, usability and user requirements.*

The objectives for the scoping study phase were to:

- Gather and collate information on previous user requirements studies
- Review available and accessible collections
- Review portal software
- Review relevant metadata schemas and standards
- Highlight implications of legal issues on portal development
- Synthesise the outcomes from Phase 1 and propose a plan for phase 2
- Gather feedback from expert users on the specification in the scoping study report through a small expert focus group.

The objectives for the demonstrator phase were to:

- Build a demonstrable portal, with supporting policies
- Use the demonstrator in at least two FE/HE institutions
- Ensure that the demonstrator captures user requirements
- Provide formative evaluation of the Visual and Sound Materials Portal Demonstrator
- Provide a business and sustainability plan for a future portal service

³ http://www.jisc.ac.uk/fundingopportunities/funding_calls/2005/07/funding_visualandsoundportal.aspx

4. Methodology

Phase 1 - scoping

User requirements: These were derived from the literature review conducted as part of the scoping study. The project team also ran an ‘experts’ focus group to test and update the requirements.

Terms of use: Issues related to rights, licensing and responsibility for institutions, individuals, vendors and the JISC were explored with input from a Legal Issues Project Advisory Group constituted by the project.

Portal Functionality and Architecture: The scope, design and component parts of the VSM Portal were explored during Phase 1. The initial specification of the demonstrator portal was derived with reference to these findings, the user requirements and the project team’s understanding of portals.

The outcomes of Phase 1, including the specification for a portal demonstrator, were reviewed by the JISC to determine the desirable scope of the proposed second phase demonstrator and considered whether it was feasible to build the demonstrator.

The JISC review directed that:

- The VSM Portal should be a ‘thin’ portal: it should not provide functionality and facility to collect and give access to content created within the community. However, community content already held in repositories was in scope.
- EDINA was to consider the recommendations of the JISC Information Visualisation Foundation Study during development of the user interface (although noting the dependence on Windows desktop applications for some of its recommendations which will not be possible within a web portal).
- Web 2.0 functionality and approaches would have relevance for a portal. In particular, implementation of annotation was to be taken forward.
- Investigation and implementation of Content Based Image Retrieval (CBIR) should form part of the demonstrator.
- Embedding activity should implement a portlet offering search functionality based on the JSR-168 standard.

Phase 2 - demonstrator

The twelve month demonstrator phase necessitated that a number of parallel activities be carried out, the outcomes of some feeding into others, in order to implement a working demonstrator. These strands were:

Stakeholder input: A project Steering Group was set up to provide stakeholder input during the demonstrator phase. Input from the JISC Review panel during the demonstrator phase was to be facilitated by the project team.

Steering Group: The Steering Group met twice and provided input to and feedback on, for example, the range of content that should be included in the portal, the draft collections policy, and the portal demonstrator interface.

JISC Review Panel: During the JISC Review it was agreed that the project team would set up a blog⁴ as a means of continuing communication between the JISC Review Panel and project. The project team made postings in the key areas of activity.

Development of Demonstrator: An iterative development approach was taken with two iterations of the development cycle being achieved in the 12 month period. The first iteration produced a demonstrator based on the user requirements identified in the scoping phase. The second iteration of the demonstrator built on initial usability testing and evaluation, in order to further refine and add functionality and content to the

⁴ <http://vsm.edina.ac.uk/blog/blog> - requires a login to access

demonstrator. The activity also included a review of software for development and a Metadata, technology and standards watch.

Usability: was informed by initial usability testing using paper prototyping. The initial demonstrator was tested by EDINA using the Nielsen Norman group ‘walk and talk’ method of usability testing. The group of users involved in testing included teachers, librarians, staff and students. The usability testing also involved the associate partner organisations. The user testing was undertaken in parallel with the external evaluation prior to implementation of the second version of the demonstrator.

External evaluation: was commissioned to review the demonstrator, and to provide input into further scoping for any subsequent development of the portal and into the business and sustainability planning.

Embedding: Services provided by the VSM Portal should be embedded in various user environments if the VSM Portal is to be considered successful and useful. Embedding a URL to a resource is a simple form of embedding; beyond that the approach was to be based around common specifications and standards, with development of a JSR-168 portlet being determined by the JISC Review. Institutional embedding involved the project associates – University of Hull, Glasgow Caledonian, St Helens FE College.

Content: Potential content was identified during the Scoping Phase, and this work was continued in the Demonstrator Phase to ensure there was a range of suitable content.

Legal issues: The influence of rights issues on both the opportunities for producers of media to release material into the FE and HE community, and on the opportunities for students, tutors and lecturers to use such material was considered and provided input into policies needed to support the portal.

5. Implementation

Demonstrator development

Technical plan

A technical plan⁵ was produced prior to implementation. The technical documentation of the portal development in Appendix 3 gives details of the architecture at the end of the project.

Comparison of technical plan and final implementation

The technical plan begins with a review of the proposed architecture, and the final implementation is very close to this. A few aspects were clarified during development:

- The metadata of all collections were harvested, and indexed in a local database. No content providers expressed any significant interest in distributed search technologies (Z39.50 or SRW), and many ruled it out completely.
- It is not clear why harvesting should be preferred over the distributed search model. It is possible that harvesting is seen as less likely to produce high server loads, but the impression we have received is that content providers have not really considered distributed searching. It could just be that the technical staff are familiar with repository software; the concept of harvesting is entrenched in that particular community.
- From our perspective this is welcome. Since content providers all appear to favour harvesting, we can make a working assumption that we will hold metadata locally. This allows us to exploit the full functionality of our preferred database system (Zebra) and provide features such as fuzzy searching, sorting and relevance ranking.
- OAI/PMH was the most common harvesting method; downloading XML files by HTTP was also common. Some RSS feeds were used, and some files were copied by scp⁶.

⁵ <http://edina.ac.uk/projects/vsmportal/outputsandactivities.html>

- Some files were supplied as a delimited text format rather than XML. This presented no significant difficulty.
- End users were provided with access to images, films, and sound files on the content providers' web sites. In nearly all cases it was possible to implement a "deep" link, *i.e.* a link to a page that displays the item immediately.
- As planned, Content Based Image Recognition (CBIR) was implemented and added to the Portal Demonstrator as a "find more images like this" function, where the user provides a "seed" image and the CBIR function identifies other images with similar colours, shapes and textures. We encountered performance issues, and further development would be required for a quality service.

The development tasks listed in the technical plan were carried out with a few alterations:

- The "My VSM" (saved record metadata) and CBIR functionality implementation was brought forward to Version 1; this involved some rescheduling but it enabled us to include these features in the user evaluation exercise.
- The database supports SRW/SRU as well as Z39.50.
- Harvesting and conversion of metadata has not yet been automated. All the necessary code has been written and structured in such a way that this may be done, and only a little extra development would be involved (essentially configuration so that existing scripts are executed on an appropriate schedule, probably using the unix *cron* daemon).
- The "My VSM" features lack an abstract API⁷. This was due only to time constraints and not technical problems, and it has not affected the functionality offered to end users. It would still be desirable to implement an abstract API to make the functions offered by this part of the user interface more flexible and extensible in any future versions. An abstract API would also make it straightforward to add a SOAP interface, supporting a service oriented architecture (SOA), if required in future.

Functional specification

A functional specification was drafted prior to implementation⁸, with development split into two phases. The technical documentation of the Portal development in Appendix 3 includes a specification that represents the functionality of the VSM Portal at the end of the project (January 2008).

Comparison of functional specification and final implementation

The majority of the functions specified were successfully implemented in the portal demonstrator. We deviated from the specification in the following ways:

- The standard search does not limit boolean operators to simple "all of these words" or "any of these words" options. The database implementation supports CQL⁹; this is a standard query language which provides powerful searches for expert users, but also works intuitively for naive users. Queries expressed in natural language, and using conventions familiar to most users from many other web interfaces (*e.g.* using quotes around terms to be treated as an exact phrase), tend to produce results that match users expectations as well as any other scheme. We decided to permit queries expressed as CQL, rather than limit users' options.
- An advanced search was not implemented as a separate feature. Some of the planned features became available in the standard search through the use of CQL. Searching by file type is not provided because the metadata rarely provide this information. The requirement for searching by specific content providers was added as a feature of the standard search, but with an option to display or hide the inputs (hidden by default, to reduce clutter and avoid an intimidatingly complex first impression for new users).

⁶ Secure copy protocol: file copying via an encrypted network connection

⁷ Application Programming Interface

⁸ <http://edina.ac.uk/projects/vsmportal/outputsandactivities.html>

⁹ Contextual Query Language <http://www.loc.gov/standards/sru/specs/cql.html>

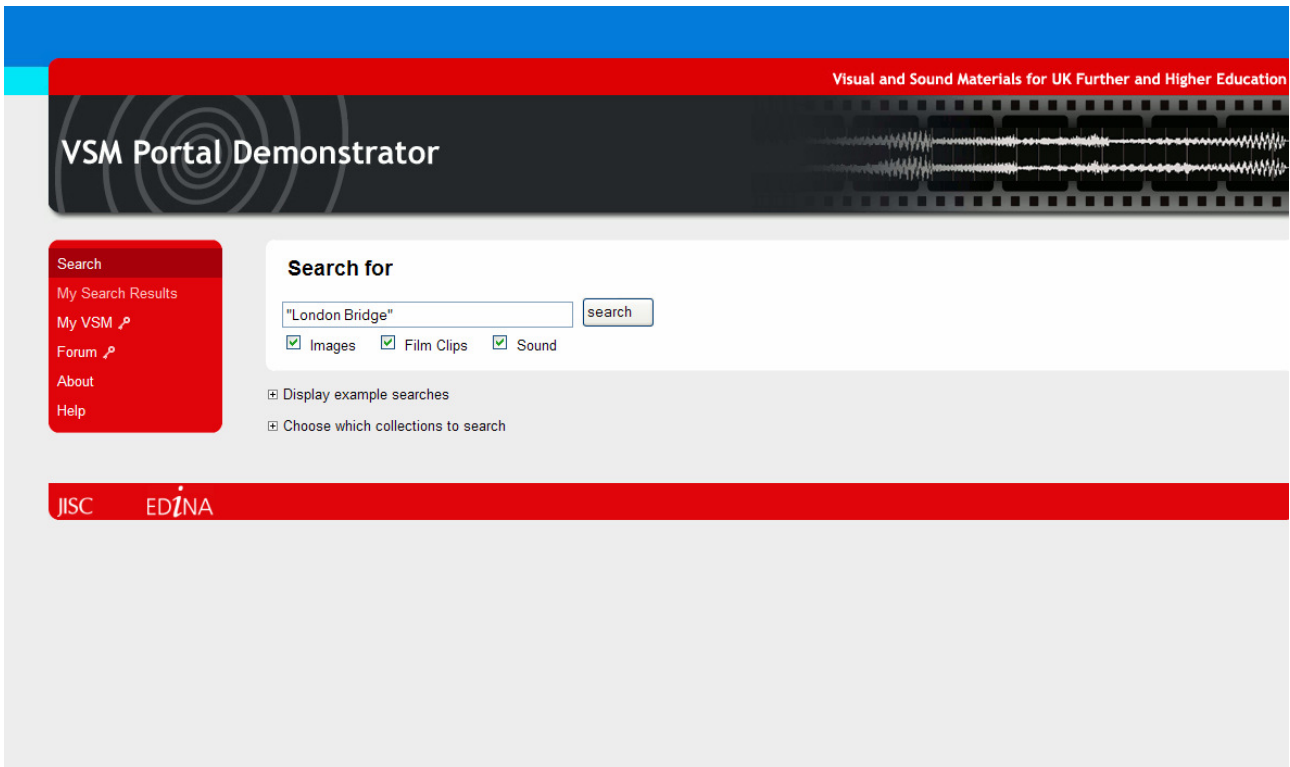
- An additional layer was added to the presentation of search results, on the advice of the Steering Group. This is the first page the end user sees after submitting a search, and presents the results grouped by content provider. The aim of this is to make the user aware of the existence of the separate collections; this was thought to be necessary partly to give due credit to content providers, but also because smaller collections that produced fewer (but feasibly the best) hits could be obscured by sheer weight of numbers if the results from all collections were merged.

It was accepted that the benefits may not be apparent to end users, and indeed the extra step required in accessing results may be unwelcome (as feedback received during user testing proved it to be), but that this did not necessarily mean it was wrong to include this feature. The issues raised are discussed later in this report.

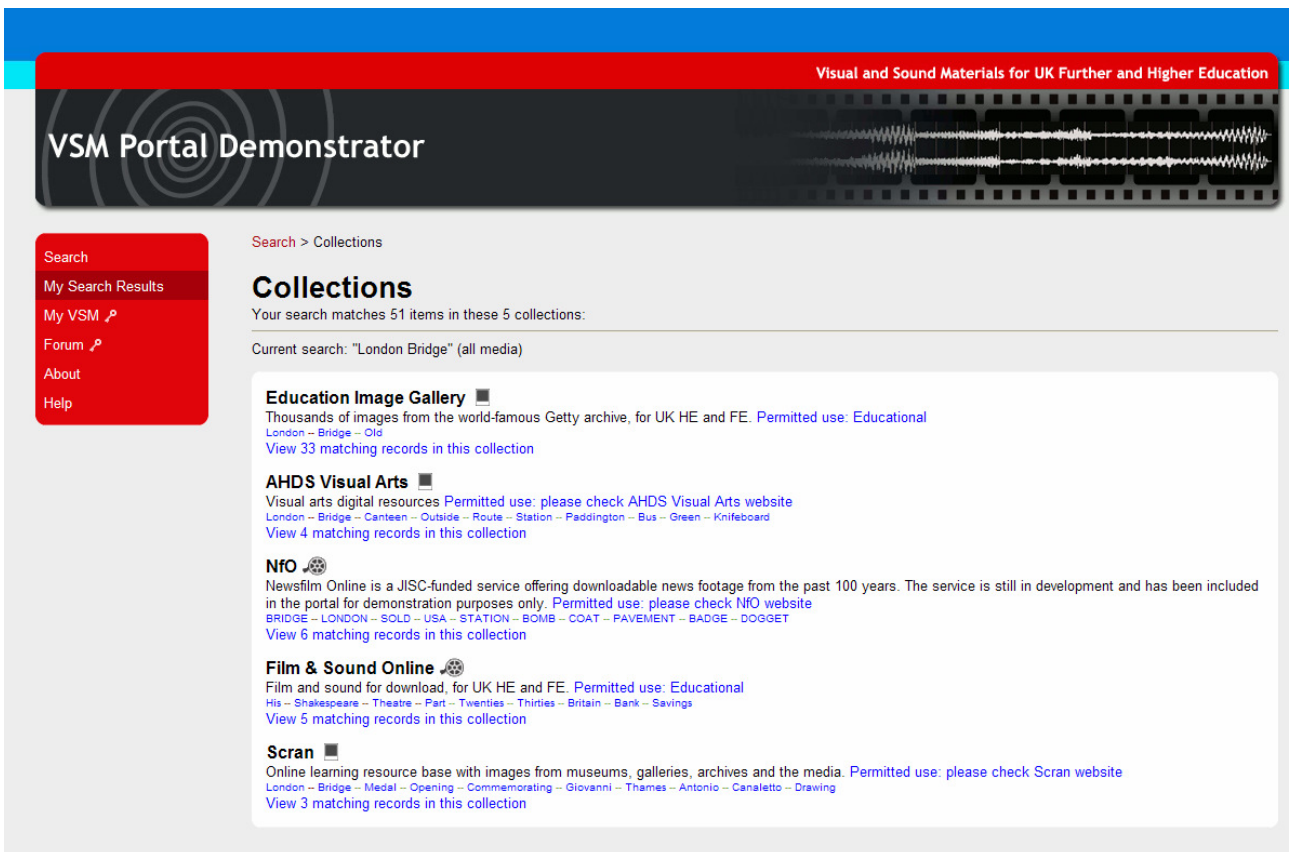
The portal exploits this extra page by displaying information about each collection, including a summary of terms and conditions. An experimental feature was added, where text from a sample of the search results is analyzed for frequently occurring terms, which are displayed to attempt to convey a quick summary of the subject area covered by the results in each collection.

- Record metadata may be emailed to users, but not in multiples. The portal design had a simple and clean cut feel, which deliberately allowed users to choose which records to save or email with a single click. Enabling saving or emailing multiple records is a two stage process (first the user must select record(s) of interest, and secondly they must save/email them), which is slightly more complex in the user interface and can produce problems for less experienced users. However, users may prefer the slightly more complex option that works with multiple records, and this may be implemented in future if there is a demand.
- Record metadata may not be saved to a local file. This was due to time constraints on development, and was not due to technical issues.
- No feedback function was implemented in the portal demonstrator. This was due to time constraints on development, and was not due to technical issues.
- When saving a record (the “My VSM” feature) some restrictions were imposed in order to complete development within the project timetable. The folder to which a record is saved does not default to the last folder used; a record cannot be copied between folders (so that a record may be found in two or more folders), and for this reason folders and their contents may not be duplicated, though records may be moved between folders; a folder may not be created without saving a record into it, and finally folders may not be renamed.
- CBIR was successfully implemented, though performance can be slow and if this feature is to be used in future extra development will be required. To keep performance within acceptable limits, we were forced to restrict the size of the database, so not all collections have been indexed.
- A function for comparing metadata between records was not implemented. It was not clear if the relatively simple descriptive metadata would make this function useful. If an end user is presented with several versions of an object (or similar objects) that differ due to terms and conditions or technical characteristics, then a comparison between such records might be useful; at present though such detailed metadata are rarely available. Other use cases or scenarios that would justify the feature were not apparent.

Two screenshots from the current portal demonstrator show the search and results pages. (See page 25 for a screenshot of a planned design for a long-term portal service.)



Screen shot showing the search page of the VSM demonstrator



Screenshot showing the results page of the VSM demonstrator

Institutional embedding

Delivery of the portal search functionality as a portlet was investigated, using the JSR 168 framework. This was achieved using the JAFER Z39.50 client¹⁰ to connect directly to the Zebra database that provides the search engine for the text metadata. Testing during development was within an instance of the uPortal¹¹ version 2.4/.2.5 framework. The portlet uses a simple Model View Controller (MVC) architecture, with the Model being a wrapper for the JAFER client, and the View making use of an XSLT engine to simply transform the XML records returned by the VSM database, and with a JavaServer Page (JSP)¹² when no database interaction is involved.

One major problem encountered was the inability of the XSLT within the uPortal framework to properly deal with (valid) XML returned by the database where the elements contain namespaces. This is unfortunate since the properly formed VSM XML records draw upon namespaced definitions from Dublin Core (DC)¹³, and also its own VSM namespaced fields. To work around this problem, a completely separate instance of the Zebra database was installed, using modified records with the namespaces removed. This temporary solution is wasteful in machine resources and also in duplicating database maintenance operations. However, a more serious potential problem is a clash of element names within the VSM and DC namespaces, though fortunately this did not happen. This situation must be rectified if further development is required.

Another problem with this approach to the portlet is that all available collections are searched simultaneously. Whilst this is actually a quick search due to the efficiency of Zebra, the division of the results into a collection level is not possible. Though records from any given collection may in principle be separately extracted from a result set, it is necessary to know which records belong to each collection prior to retrieving records, but this information is not available from the Zebra server. The VSM web portal avoids this problem by performing concurrent searches of each collection level database, however this would be a significant amount of work to adapt to a JSR 168 portlet, and time did not allow this development.

The CBIR function was also added to the portlet, utilising the SOAP interface that the CBIR component offers.

The portlet was deployed and tested at two external institutions: the University of Hull (uPortal v2.5/ingres) and Glasgow Caledonian University (uPortal v2.4/hsq). Problems were experienced by both institutions, though these were overcome by updating the uPortal framework¹⁴ and resolving a conflict between a component of the distributed portlet and host portal software¹⁵. St Helens College started but did not have the resource to complete installation within a Microsoft Sharepoint environment.

The functionality of the portlet is less than that of the actual VSM web portal, and given the inherent constraints of a portlet, is not as aesthetically pleasing. Two screen shots from the portlet are shown.

Documentation¹⁶ was produced relating to implementation on uPortal, and can be used to deploy the portlet on any portal framework.

¹⁰ <http://www.jafer.org/>

¹¹ <http://www.uportal.org>

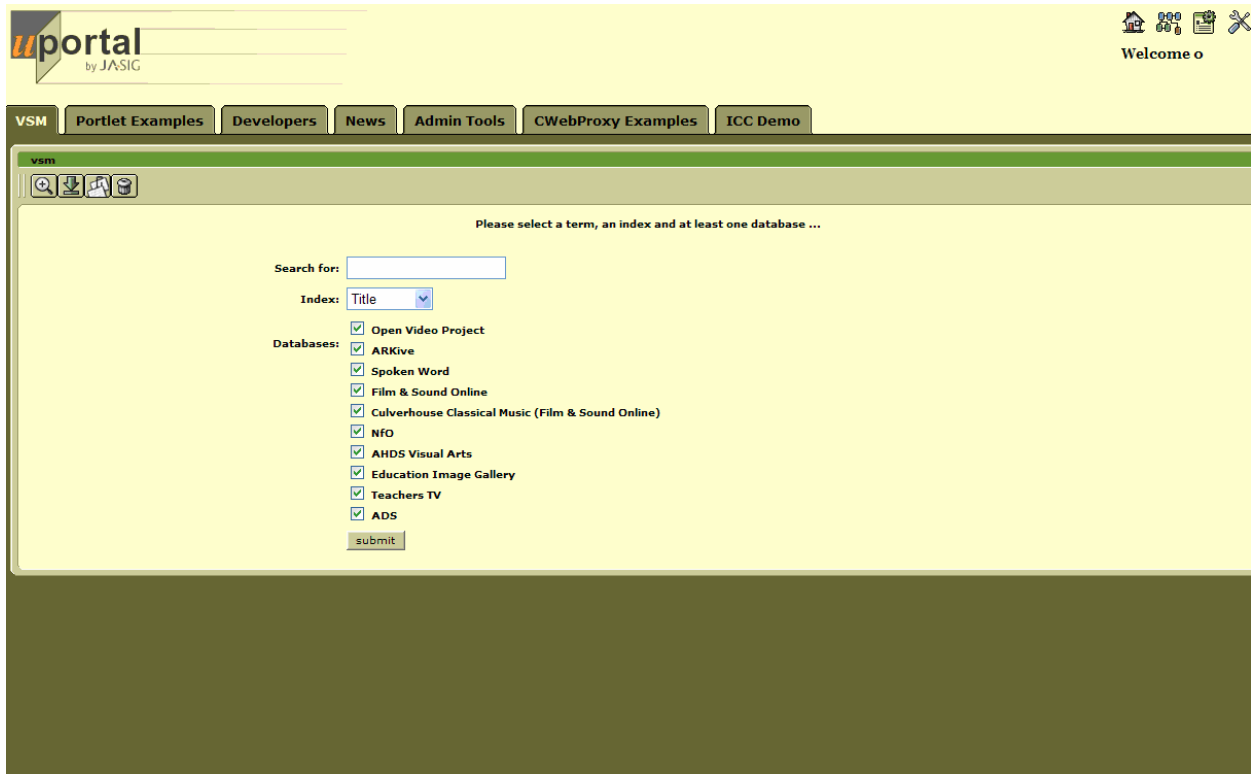
¹² <http://java.sun.com/products/jsp/>

¹³ <http://dublincore.org/>

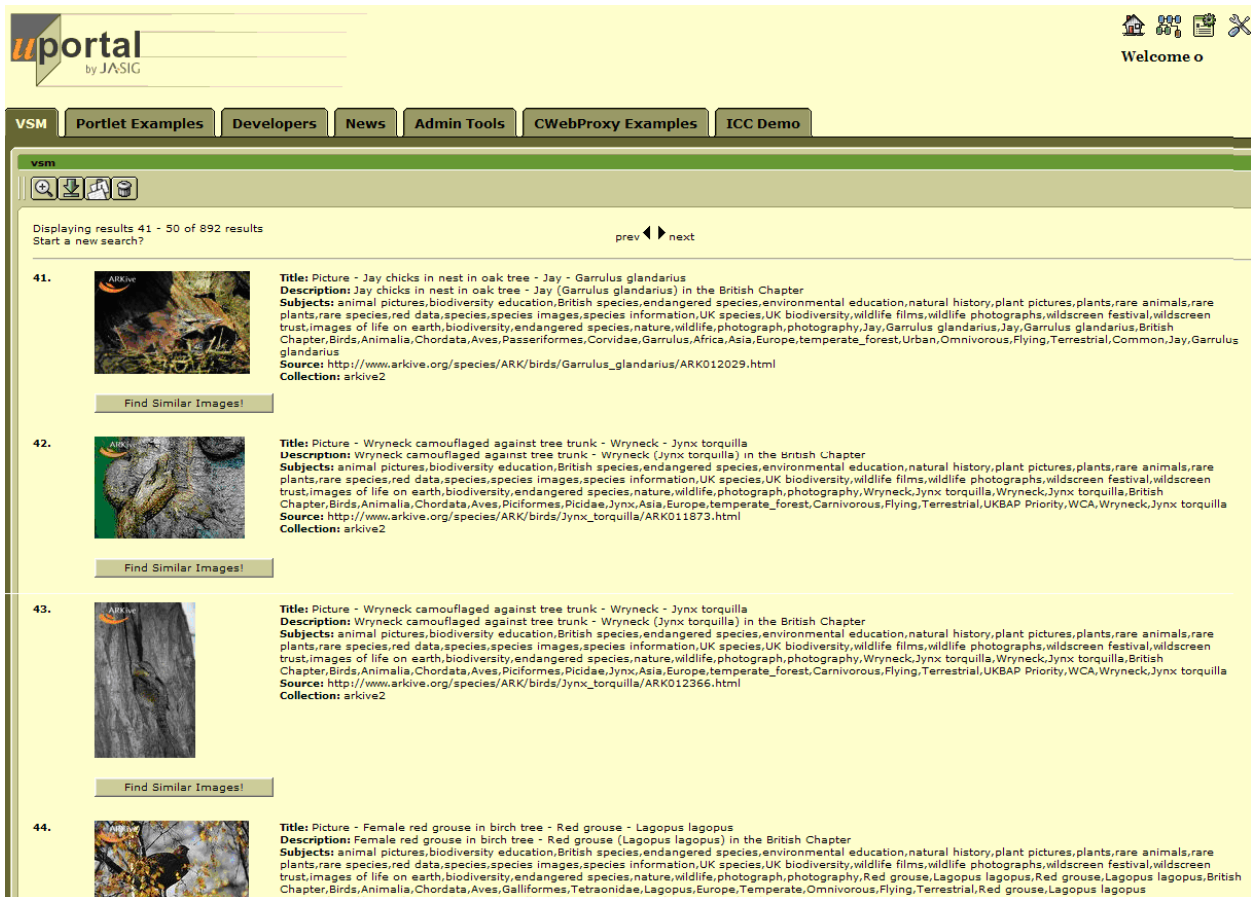
¹⁴ Installation was attempted successfully with uPortal 2.5.3.1, after failed attempts with version 2.5.2 and 2.5.3

¹⁵ The portlet was distributed as a web application archive (war) file containing a java archive (jar) file *portletapi*, which Hull reported was already being run by the Tomcat host server. After removing portletapi from the war file, the portlet was deployed successfully.

¹⁶ <http://bodach.ucs.ed.ac.uk:8180/VSMDownload/>



Screenshot showing the search page of the portlet



Screenshot showing the results page of the portlet

Metadata

The ability to adapt and/or convert existing content metadata to allow searching was considered. The project needed to take into account the varied schemas in use, in order to streamline the user experience as far as possible when navigating a heterogeneous range of collections.

Content provider metadata formats

Content providers' metadata was provided in a variety of formats. Some delimited text ("comma separated variable") was provided, but nearly all was provided as XML. Some of the XML metadata used the provider's own schema, but in the majority of cases Dublin Core (DC) was used. This is possibly because DC is mandated by the OAI/PMH¹⁷ harvesting standard.

Metadata in the content providers' own schemas were easy to interpret and use. Although XML has benefits of extensibility and may be more resistant to errors (since fields are identifiable by name), in practice delimited text presented no problems. DC metadata presented a greater challenge.

Interpretations of DC vary to a large degree, and no useful meaning may be given to "standard" terms without specific knowledge of each implementation. Inventive or even eccentric¹⁸ interpretation of DC does not necessarily prevent these metadata being useful, providing documentation is available; but the requirement for documentation explaining the meaning of every element does largely defeat the purpose of using a standard.

In more serious cases, a complete lack of clarity results from content providers coercing metadata into DC format, using terms that are inappropriate but which must be used because they are the best available fit. We have anecdotal evidence that useful information (*e.g.* the URL of a thumbnail image) has been excluded from the metadata, because there is no suitable term available in DC.

In practice non-standard schemes have been more easily and more productively interpreted for harvesting than DC metadata. This is presumably because non-standard schemes have been designed for the needs of the collection, making them more intuitive and providing a better structure.

VSM Metadata Format

The VSM database stores metadata in a scheme based on Dublin Core (DC), with the necessary addition of some elements specific to VSM. The metadata are rendered as XML for communication between software components. An attempt has been made to comply with the DC XML guidelines¹⁹ when rendering the metadata. These guidelines are vague regarding qualifiers, but in practice this was not a problem for us.

Full details and an example of a metadata record rendered as XML are available in Appendix 4.

Content

The ITT indicated:

"It is envisaged that a portal in this area should point users not only to JISC licensed material but also to free community collections from publicly funded programmes such as the BioScience Image Bank and public sector resources such as those funded under the NOF-digitise programme and the JISC Digitisation Programme.

"There is significant activity in the creation of digital repositories for different types of materials to support learning and research and it will be important for a portal to be capable of interacting with

¹⁷ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

¹⁸ We encountered a novel use of the dc:format element. The DCMI defines this element as *The file format, physical medium, or dimensions of the resource*, and notes that *Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME]*. When dealing with an image collection, one might hope that dc:format would specify the MIME type of the images, *e.g.* image/jpeg. Alternatively a descriptive string such as "JPEG image" could be useful, or at worst something like "picture" or "photo". We encountered a case of an image collection, which contained a photograph of a shoe as a JPEG image; the dc:format element for this record was given the value "nailed".

¹⁹ <http://dublincore.org/documents/dc-xml-guidelines/>

digital repositories. This cross-disciplinary discovery across a mixed economy of collections would significantly enhance the use of this material in learning, teaching and research. ”

Content providers

The project team approached 32 content providers during the course of the project seeking to interest the provider in participating in the portal demonstrator. The initial contact was made by email explaining what the VSM demonstrator aimed to achieve and outlining some of the perceived benefits a content provider could expect from participation. It could require a number of follow-up emails or phone calls to elicit a response, and involvement of technical staff to determine the feasibility for metadata transfer before successfully engaging a content provider. At the same time as contact was being made with potential content providers, a formal Content Provider agreement was being drafted. Feedback from these potential content providers was also incorporated into the Content Provider agreement.

The project steering group also made recommendations about collections that might be relevant to include in a VSM Portal for H/FE.

Evaluation

Usability testing and institutional embedding work was carried out by EDINA at four different sites, two HEIs and two FEIs, during the course of the project, and an external evaluation was also undertaken.

The usability test was designed to test the interface and functionality of the VSM Portal Demonstrator and how effectively users could complete a real-world task or project. To that end, each user was given a series of linked tasks to determine whether they could navigate and fulfil their requirements from start to finish i.e. how well the different functions operate in sequence.

External evaluation was commissioned through seeking competitive tenders from four organisations who had been approached and expressed interest in bidding for the work. The successful bidder was Rightscom Ltd. The aims and objectives of the external evaluation with regard to user needs analysis were to:

- Carry out consultation to receive feedback on the demonstrator, including:
 - Working with two institutions from the Higher and Further Education sectors to assess user experience in different user environments.
- Use the above consultation outcomes as the basis of an analysis of the user community's needs for a full portal service, including:
 - An assessment of the ways in which visual and sound materials might be used within learning, teaching and research in HE and FE;
 - An evaluation of search and browse techniques relevant to the discovery of visual and time-based media;
 - An evaluation of the use of personalisation features, such as "lightbox" and user-created metadata, such as tagging in Flickr.
- Feed back into the second iteration of the demonstrator development.

The aims and objectives with regard to strategic business analysis and planning were to:

- Review Version 2 of the demonstrator and highlight any additional scoping requirements for subsequent development of the portal;
- Focus on the viability of a full VSM Portal service and on business issues related to such a service;
- Provide an analysis of the business context indicating benefits, gaps, strengths, weaknesses, risks and opportunities for a full portal service.

The tasks of the evaluator were to:

- Evaluate Version 1 of the portal demonstrator and report to project team;
- Liaise with the usability testing and institutional embedding workpackages;
- Produce a final report, including user needs and business analysis.

6. Outputs and Results

“I like the concept and would certainly use the service”²⁰

Phase 1

The significant output of Phase 1 was a report to the JISC²¹ outlining the scope, scale and nature of work to be undertaken during Phase 2. It included options for the development of the demonstrator portal, portal software, and approaches to achieve the identified functional and user requirements. A further document was prepared for the JISC Review Panel²², providing supplementary information to aid decision making during the review. The project also produced an internal Project Investigation Report, submitted to JISC, which recorded the full details and findings of the scoping activity.

Phase 2

Phase 2 tested out the findings of the scoping study in terms of user requirements, interaction with a cross-section of potential collections and technical functionality, and highlighted the issues associated with the implementation and delivery of a portal service in this area. The main deliverables were:

- A portal demonstrator interacting with a cross-section of collections
- Delivery of the demonstrator as a portlet
- A set of supporting policies
- A final report with recommendations, including potential developments and scope of a future portal service

VSM Portal Demonstrator

The main output of the project was the VSM Portal Demonstrator²³. This enables the user to search across a number of collections of multimedia materials. The user can select between images, film clips and sound, or any of them in combination. The default option presented is to search on all of them. As well as providing advanced search options, the demonstrator also provides personalisation functionality and a forum for discussion.

Institutional embedding

Delivery of the portal search functionality as a portlet was investigated, to enable the portal to be searched directly from institutional websites. Testing during development was within the uPortal framework, and was undertaken by the University of Hull and Glasgow Caledonian University.

Content Development

Eleven collections are included in the portal demonstrator:

- AHDS – Archaeology²⁴
- AHDS Visual Arts²⁵ (Pixus²⁶)
- ARKive²⁷
- British Geological Survey (BGS)²⁸ (Pixus)
- Creative Archive - Teachers TV²⁹

²⁰ Comment from an interviewee in HE, VSM Portal Demonstrator External Evaluation report, Rightscom Ltd. Feb2008.

²¹ <http://www.edina.ac.uk/projects/vsmportal/20070529Phase1ReportFINAL4Web.pdf>

²² <http://www.edina.ac.uk/projects/vsmportal/20070528ScopingSupplementaryReport4Web.pdf>

²³ <http://vsm.edina.ac.uk/vsm/vsm.asp>

²⁴ http://ads.ahds.ac.uk/learning/image_bank/

²⁵ <http://vads.ahds.ac.uk/>

²⁶ Collections denoted with (Pixus) were those included in the JISC funded Pixus Image Portal Demonstrator project which ran from Jan-Oct 2003. See http://www.jisc.ac.uk/whatwedo/programmes/programme_portals/project_image_portal.aspx

²⁷ <http://www.arkive.org>

²⁸ From local server at EDINA, from JIDI project

²⁹ <http://www.teachers.tv>

- Culverhouse Classical Music³⁰
- Education Image Gallery³¹
- Film & Sound Online³²
- Newsfilm Online³³
- Open Video Project³⁴
- Scran³⁵ (Pixus)
- Spoken Word project³⁶

The Culverhouse Classical Music Collection from Film & Sound Online is included as a separate collection so that a discrete sound collection is present in the demonstrator.

A further nineteen collection owners were approached, of which nine agreed in principle to participate in the project.

Part of the remit of this project was to seek to include among our targets Institutional Repositories of multimedia objects. To help us explore the issues related to repositories of multimedia objects and with a view to including one among our targets, we worked with the MIDESS³⁷ team based at Leeds University.

Further details are available in Appendix 2.

Policies to support a VSM Portal

The project team was tasked with developing and testing suitable licences and policies to support a VSM Portal service. The following paragraphs describe these licences and policies, all of which are available at the project website³⁸.

Draft Collections Strategy

The draft Collections Strategy covers the aspects of scope, collection description, business model, metadata, quality, acquisition procedures and collection management. It also includes background information and experience gained during the project that will help inform the evolving strategy in future.

Content-Provider Agreement

The VSM Portal demonstrator harvests metadata and creates a database which is then searched through the user interface. It may be argued that the legal responsibility of the VSM Portal for the metadata delivered through the portal is greater than it would be if we had chosen to search databases remotely; having created a database (which attracts database rights), we are publishing these metadata to service users rather than simply facilitating their access. This is contentious; members of the VSM legal advisory group held different views regarding the need for warranties and indemnities from content providers whose services are searched remotely. The need for those warranties and indemnities for harvested content is clearer. In drafting a licence, the team sought to keep it as brief as possible whilst, at the same time, including the legal safeguards required to protect the portal service provider from legal challenge.

A number of content providers have signed the licence without disagreement although this does not imply that signature was secured quickly. It took a lot of time and repeated prompts from the VSM team to persuade content providers to invest the time required to sign the licence. Some content providers having agreed, in principle, to sign the licence, have progressed with the technical work required to expose their

³⁰ Film & Sound Online, <http://www.filmandsound.ac.uk/collections/culverhouse.shtml>

³¹ <http://edina.ac.uk/eig>

³² <http://www.filmandsound.ac.uk>

³³ From local server at EDINA pending launch of service

³⁴ <http://www.open-video.org>

³⁵ <http://www.scran.ac.uk>

³⁶ <http://www.spokenword.ac.uk>

³⁷ <http://ludos.leeds.ac.uk/midess/>

³⁸ <http://edina.ac.uk/projects/vsmportal/outputsandactivities.html>

work through the VSM Portal at a faster rate than signature of the licence. Consequently, some of the content currently available through the VSM Portal is unlicensed (although we expect to secure signature of the licence). We have not been concerned by this as the VSM Portal is not currently promoted as a service and thus the risks are low.

As users of a VSM Portal could reasonably be expected to adhere to a single set of terms and conditions for use of the service (as distinct from use of the content accessed through the service), and the terms and conditions of use for the service must mirror those of the licence, it is necessary that all content providers agree to the same licence. Negotiation of licence terms is a trade between costs and benefits. The benefit that a portal could offer to a content provider would be access to a large community of relevant users. As the VSM Portal is at the project stage, without an existing user population and without proof that such a user base will be established, we have limited power in negotiation.

The work involved in negotiating and pursuing signature of the licence is substantial. Whilst the project is restricted to a small number of content providers, this has been possible. For a service, it would be viable only if it were streamlined. The Project team would propose, at service stage, to post the licence and all other required information including the metadata schema and a link to the collections strategy on a 'VSM Content provider's' web page, and invite content providers to contact us. As indicated above, the success of this approach would likely depend on the VSM Portal being regarded either as a conduit through which large numbers of relevant users might access a content-provider's content (and ultimately, the native interface of the content provider), or be brought to the attention of a relevant group of users that they might otherwise find difficulties reaching.

Takedown policies

The project team developed draft policies and procedures for implementing content takedown. Although the licence between the VSM Portal service provider (the University of Edinburgh) and providers of content surfaced through a portal would include a warranty that nothing unlawful was included, provision for removal of metadata records of any content subject to complaint is necessary.

A takedown policy for user-generated content (such as content on the User Forum) is also important – although the issues to be addressed vary from those of the policy relating to metadata records.

User Evaluation

During Phase 2, usability testing was carried out in four institutions, two HEIs (University of Hull and University of Edinburgh) and two FEIs (St Helens College and Swansea College). The results at three of the four sites were largely similar; but users at Hull found the VSM Portal Demonstrator easier to use than those at the other sites. It is not clear why this occurred. There are some differences in the way that the tests were conducted at different sites but nothing that is exclusive to Hull. The fact that subjects at Hull were all sufficiently familiar with their ATHENS passwords that they had no trouble logging in without assistance suggests that they are familiar with online information services. Perhaps subjects at Hull were simply more experienced users and thus were more confident than users at other sites.

The evaluation resulted in a number of suggestions to improve the VSM Portal demonstrator to meet the needs of users in UK H/FE, and guidance on how it should be presented and marketed in an environment where any such service must offer a unique value proposition. The suggestions are documented in Appendix 1, which also includes the project team's initial comments on the desirability and feasibility of the suggested developments.

The aim of the external evaluation was to determine the potential impact on the learning and teaching communities of a VSM Portal and what benefits it could deliver, and its conclusions are reported in the next section.

7. Outcomes

*“... impressed by quality and clarity of metadata”
 “I was able to find some really useful results from my searches”
 “The need for such material is growing”³⁹*

Beneficiaries

The principal beneficiaries of a VSM Portal will be learners, educators and researchers in UK H/FE, and also the providers of content for the portal, who will benefit from exposure of their content to users in the educational community, a potentially new and important audience.

Beneficiaries from the project work will include these groups, and also other teams developing portals or portlets for the JISC IE or other funding bodies. In addition, the outcomes of the project should inform the development of content licensing and IPR models, especially but not exclusively in the area of multimedia materials.

Project achievements

The VSM Demonstrator project met all of its aims. Its highlights include:

- The development of the portal demonstrator itself, developed and tested within the 12-month funded period;
- The development of a portlet, also embedded and tested within the 12 month period;
- A range of collections included in the portal, with good coverage of the different media types;
- Appropriate Web 2.0 technologies (user comments) and content-based image retrieval implemented and tested;
- User requirements for a VSM Portal demonstrator;
- Draft technical and functional specification for a portal service;
- Documentation regarding the deployment of the portlet within a uPortal framework, that can be used to deploy the portlet on any portal framework;
- Policies to support a VSM Portal;
- Recommendations for future development and requirements, including a range of appropriate business and sustainability models (including the level of user support required for full service) to help inform future decisions.

Knowledge and experience gained

During scoping, the project team enhanced their knowledge and understanding of the range of factors that would inform development and testing of an effective VSM Portal demonstrator. During development of the demonstrator, technical standards, metadata, the required and feasible functionality, and the technical implementation issues, were the chief areas in which the team gained experience.

Direct contact with embedding institutions provided valuable information on the end user experience of the portal. The associates also gained knowledge and experience through the usability testing and user evaluation exercise of how valuable their users might find a single national portal platform with a focus on visual and sound materials. The associates also gained technical experience through implementation of the VSM portlet.

The outcomes of the VSM Portal project will be of continuing interest to many communities, both nationally and internationally. The learning of the project team will be disseminated through the project website and other appropriate publication.

³⁹ Comments from HE and FE based interviewees, VSM Portal Demonstrator External Evaluation report, Rightscom Ltd. Feb2008.

The developers on the VSM project gained very useful experience of interoperability issues between portals and multimedia repositories. This differs from the typical bibliographic services environment in a number of ways, as described in the following paragraphs.

Metadata availability

Metadata are generally more freely available from repositories than bibliographic services. This is not for technical reasons, but is essentially a function of the business environment; however it has a considerable impact on the technical solutions that are applicable.

Abstracting and Indexing databases, electronic Tables of Contents and similar services often do not hold the data (*e.g.* full text of journal articles) that the end user wishes eventually to access; rather they act as a mechanism for the end user to find out about content that is available, which the user will often obtain from other sources. The service providers add value to metadata that are already within the public domain, by careful aggregation and indexing of citations in a certain area. This means it is the metadata that are of value, and the end user will usually require a subscription to access the metadata, and portals attempting a search across these services will have to deal with access control.

Multimedia repositories, in contrast, do hold the actual data that the end user wishes to access. Subscription and access control mechanisms, where they are used, are generally applied to the data. The metadata are not the principal items of value, and in fact can be regarded more like a “brochure” that advertises the content available from the services. Consequently the providers of these services are likely to be willing to make their metadata available. Portals searching across these services may be able to avoid involvement with subscriptions and access control. Apart from simplifying the implementation, this means that the developers themselves will have free access to the metadata, which makes it easier to understand the schema, record formats *etc.*

Working with multimedia repositories is considerably more complex than bibliographic services in one respect: it is necessary to deal with issues of the terms and conditions of use of the data.

The experience gained in understanding these issues (though they are not explicitly technical) has been of great benefit to VSM developers in understanding the technology suitable for meeting the needs of all parties involved.

Metadata access methods

There are two main ways in which portals may enable searching of metadata. Firstly, metadata may be searched on the content providers' servers, using a protocol such as Z39.50 or SRW. Secondly, metadata may be harvested using a protocol such as OAI/PMH, and then searched on the portal server. The advantages and disadvantages of each are discussed in Appendix 3. The relative weight of these advantages and disadvantages depends entirely on the context. A distributed search is best suited to large collections, with frequent updates, and in which the metadata are of commercial value; these characteristics are typical of bibliographic collections. For smaller collections, with less frequent updates and freely available metadata, harvesting is likely to be the better option; these are characteristics typical of multimedia repositories.

The VSM project has enabled developers to get to grips with these issues, and develop a good understanding of the circumstances in which various methods should be employed. As well as direct benefits to the VSM team, some of the content providers involved sought our advice regarding the best technical solutions for making their metadata available, and in other cases content providers with existing access methods have been able to improve their implementation by working with us.

Protocols and clients

Implementing client software for the protocols involved has provided very useful experience. Though protocols such as OAI/PMH are relatively straightforward, being based around well understood standards such as HTTP and XML, the VSM project has demonstrated that there is no substitute for “hands on” experience.

The VSM developers now have a much greater understanding of the pitfalls and possible workarounds that can be deployed for a successful implementation.

Metadata conversion and database development

The developers were already highly experienced in handling metadata and creating databases; however, the VSM project has enabled them to extend this experience into new areas. The metadata provided were almost exclusively in XML, and new tools and methods for handling XML were used. The metadata were aggregated and indexed using a new type of Zebra installation, which supports multiple databases within a single installation. The developers also gained more in depth knowledge of Zebra functionality, particularly with features such as relevance ranking. Support for SRU/SRW is now thoroughly understood, from both the client and server perspective.

External evaluation

The external evaluation was intended to determine the potential impact on the learning and teaching communities of a VSM Portal and what benefits it could deliver. The conclusions of the evaluation final report⁴⁰, delivered by Rightscom, were as follows. The project team's comments are given in italics.

1. As we have stated in the introduction to the report, although we understand the motivation for such an early evaluation exercise, we believe that it has probably been premature, in that it is very difficult for potential users to separate their perceptions of a 'discovery' tool such as this portal from that which it is discovering. Adding content, especially where so much of it is highly valuable and expensive to produce, is a time-consuming process of negotiation and integration.

We agree with this conclusion.

2. The presentation of the results under separate collection headings has also been dictated by the collection owners, and this was the aspect of the current implementation most disliked by the majority of users in our evaluation. It was also a negative factor found in the usability testing exercises. In our view, this will have to be changed in any future implementation. Ideally, users would be presented with a single set of results ranked by relevance. At the very least, they would be alerted more strongly by the initial interface to the fact that the content is divided into collections and the collections would be explained and described much more clearly to guide searchers as to which collections were most appropriate to their search.

The issue of categorising results by collection requires further careful consideration. This presentation was not mandated by collection owners, but was decided by the project after guidance from the Steering Group. Presenting results categorised by collection may have no apparent purpose to end users, in which case negative reactions can be expected; these should not be dismissed, but may be outweighed by other factors. See Appendix 1 for this discussion.

3. The confusion by many users as to whether the VSMPD is a search engine or a portal needs to be considered in any redesign. While the clean, uncluttered lines of the initial page and the simplicity of the search were admired, in reality they probably contributed to this confusion. This was also found in the usability testing. It may be that ways can be found to retain this interface, but to make clear to the users that they are not searching the totality of the Internet for particular media types.

We envisage a new design for the portal early in any future development.

4. It is very clear that Google, Google Images, YouTube and Flickr and similar services, for all their limitations in terms of relevance, quality, trust and copyright problems, represent the dominant 'competition' for any discovery resource, especially among students. These tools will only continue to evolve, with Google in particular able to devote massive resources to development.

⁴⁰ <http://www.edina.ac.uk/projects/vsmportal/docs/VSMExternalEvaluation.pdf>

5. The consultations also confirmed that Google conditions people's expectations in a range of ways: what they will see when they carry out a search; the speed with which results are delivered; and a host of other functions such as suggested corrections to misspelled search terms, an advanced search, a search for maps etc. It may be thought that the one way in which a more specialised service would score over Google is in producing a smaller set of more relevant results, but there is not much evidence that this is a major factor affecting user satisfaction. Rightscom's previous research on the use of discovery services, cited above, found that researchers (as distinct from librarians) actively prefer huge sets of results to trawl because they feel they are getting access to everything and because they value serendipity (also see point 8). While it is likely that this applies less to VSM than journal articles, the point remains that researchers and especially students have grown used to putting the balance of their search efforts into sorting wheat from chaff rather than having it done for them.

Participants in the evaluation suggested other possible advantages over Google etc: as well as the trust and copyright advantages, some users found it advantageous to have the content filtered by media type and were impressed by the quality and clarity of the metadata, especially when comparing with metadata offered by the internet services.

6. At present, the most positive response to the VSMPD came from subject librarians in the media, visual arts, design and communications areas, which probably reflects both the current collections accessed by the portal and the demand for images and especially film clips which these users have to try to satisfy.
7. The expressed need for **high quality** and **copyright cleared** images in biosciences and medicine also suggests that the portal should emphasise these aspects of its offer in order to distinguish itself more sharply from the competition.
8. **Relevance** is more problematic as a unique selling point, partly for the reasons outlined in point 5. It also appears from the feedback we have had that there is a demand for subject-based routes into the collections, and this may be needed in order to convince users of their relevance to their particular needs. That would require additional metadata beyond what is supplied by the collection owners, and since they would be unlikely to want to provide this, it would have to come either from professional indexers or through social tagging by the user community.
9. It seems to us unlikely that the portal will meet the needs of FE without considerable changes, both in terms of content and added metadata and navigation. Given that the prospects for revenue from FE would be small, the portal should probably not try to satisfy FE as well as HE institutions, except in niche subject areas which overlap with HE, rather than in the vocational areas.

Yet, the FE evaluators pointed out that there is a particular requirement in FE for practically-orientated video materials and trends in government policy will only increase that demand, because of the skills agenda and the emphasis on vocational elements in courses becoming greater. It may be possible to arrive at a business model that would support FE as well as HE, even if prospects for revenue-raising would be small, but it is agreed that routes into the materials via FE programme areas would be needed (as in point 8).

10. Getting a critical mass of content, possibly in one niche at a time, will be necessary in order to avoid disappointing the initial user groups.
11. Considerable investment will be needed in branding and in marketing the portal; there can be a temptation to see such spending as a waste of money compared with technical development and content acquisition, but that can simply result in a good service passing almost unnoticed by the target user groups.
12. There appears to be no objection from potential users to seeing the portal funded by advertising or sponsorship within reasonable limits; however, its attractiveness to such advertisers or sponsors

clearly depends on its success with users, and it will be in competition with many other online services which hope to sustain themselves through advertising.

It may be possible to obtain interest from sponsors in niche areas to address point 10 as well as point 12.

8. Conclusions

When this project started, we felt it would not be sufficient when judging overall success just to develop a VSM Portal. Potential users would have to be motivated to visit with the expectation that it would be productive and that they would discover useful resources; content providers would have to be persuaded of the benefit of allowing and enabling their resources to be searched and displayed in the portal. To be successful, the demonstrator should offer a working example of what a visual and sound materials portal could deliver and so the demonstrator should:

- Be viewed by potential users as easy to use and to add value over and above a Google search.
- Offer effective engagement with content providers, both JISC-licensed and open access, resulting in sufficient breadth and depth to provide useful results – or at least with the prospect of this beyond the demonstrator.
- Allow embedding in various user environments in order to increase awareness of the VSM Portal.

Having completed the implementation of a VSM Portal Demonstrator the project team has drawn a number of conclusions from:

The technical development process:

- A portal may successfully search all three media types: still images, moving images and sound. Though a degree of effort is required in structuring metadata, designing the search interface, and presenting results, we encountered no problems that could not be overcome.
- Any portal which seeks to provide a search across collections from different providers will face significant challenges, but these will not generally be caused by varying media types. Variations in metadata due to medium are trivial compared to other variations between collections: subject areas; handling of rights issues; methods of providing thumbnail images or “deep” links; and novel interpretations of standards.
- Widespread support for OAI/PMH can safely be interpreted as indicating a willingness to share metadata, but should not be taken as an indication that the process will be easy, or necessarily generate a useful outcome.

The user evaluation work:

- Users found it relevant that a portal should serve both time-based media and images from a single portal platform.
- The user evaluation exercise generated good, specific feedback about the interface and functionality which informed the proposed revised interface and would inform any further development of a service. Users’ perception of the ‘service’ during testing was coloured by the limited content and in some cases the limitations of the service targeted. Furthermore, some of the difficulties encountered during testing resulted from technical problems which have since been addressed.
- In meeting the needs of the UK HE/FE community, a VSM Portal will compete with ‘free’ services such as YouTube, Flickr and Google (YFG). These are easy to access and use. They facilitate discovery, location and accessing of content. A portal that includes subscription content necessarily has an intermediate stage, i.e. it directs users to a native interface which may present an authentication challenge. These native interfaces also vary in their functionality; some are easier and faster to use than others. Users are not tolerant when services fall below the expectations raised by YFG. If a service is not to be compared unfavourably with YouTube and Flickr, users must be made aware of the reasons for differences, e.g. that content is more difficult to access because it is valuable.

- Some collections have, and some lack, persistent URLs. In the least streamlined instances, the user selects a metadata record within the portal, links the native interface where she is required to authenticate and then has to rerun her search in order to find the item of interest (i.e. she must repeat the discovery part of the sequence).
- The user-perceived advantage of a VSM Portal is that it provides access to good quality content that may be readily used for the purposes to which staff and students wish to put it. This is the ‘unique selling point’ and should be clearly communicated to the user both within the service and in all promotion activity. It is based on a collections policy that meets the needs of the community. The scoping study makes clear the difficulty of selecting content that meets user needs within this community. It is essential, also, that the types of use to which content may be put are communicated clearly to users. This project investigated the feasibility of adopting the system designed by Linda Purdy at Sheffield Hallam University which uses simple icons and pop-up boxes to communicate this information to users. It would be further developed for a VSM Portal so that it not only informs users of terms of use but allows them to filter results on that basis. A critical mass of content is required before the types of uses and suitable icons for communicating these could be designed. We see no obstacle to implementing such a system.
- Presentation of results requires revision. The user evaluation suggests that relevance ranking is called for. This is no surprise as it is used by Google and thus is familiar. The utility of relevance ranking in this context for retrieving results that meet user needs requires testing.

Attracting content providers:

- It may be difficult to attract certain groups of content providers: commercial content providers may have little interest in the aims of a VSM Portal; some websites with freely available content that may have already attracted an audience.
- Some publicly-funded content providers may request some assistance with costs to implement a machine-to-machine interface to harvest/query metadata.
- Engaging some content providers may take significant time. Initially this was the case as a working demonstrator was not available until the later months of the project and few other content-providers had signed up to participate in the project.
- Engaging some content providers has been dependent on when they will offer a machine-to-machine interface: their priority is the user interface, and providing a machine interface is often secondary – if on their timetable at all.

Policy development:

- The Code of Conduct and the Takedown Policy have yet to be tested. These are based on policies used for other EDINA services; we are confident that they are well considered and appropriate.
- The draft collections strategy will need to be developed further to outline a ‘minimum’ requirement for inclusion in the VSM Portal. This will ensure that the presentation of collections is useful for the target audience (for example criteria that only collections with thumbnails and persistent URLs will be included).

The external evaluation and sustainability report from Rightscom:

- The demonstrator was of interest to the institutions who evaluated it, more so to HE than FE, although this was likely a result of the limited content available for searching.
- It is too early to assess the sustainability of a future VSM Portal service. A further year at least of experience with users and content-providers would be very worthwhile, as it is clear that there is currently no easy solution to the question of sustainability of this sort of service.
- A critical mass of content, which may be in a focused subject area (Rightscom did not identify a particular area), will be required to sustain any of the business models identified. Some further development will be needed to implement and evaluate a GUI with a "critical mass" of functionality, and to evaluate the potential of the portlet. Taking the portal forward in this way should offer a much more viable market proposition.

- The sustainability/business models suggested by Rightscom broadly agree with those outlined in the JISC sustainability models outlined in the recent ITT Project to support development of service strategies⁴¹.
- Marketing/awareness-raising will be an important and significant aspect for a potential VSM Portal service.

9. Implications

Potential developments

Several areas were identified in the project that could be developed further in future, although many indicated the need for further evaluation as well. These areas are discussed in the following paragraphs.

Refinements of the existing web search interface

Some planned features were not implemented during the project, and some possible enhancements became apparent during the course of the project:

- The initial page of search results, displaying collections with matching records, needs to be reviewed. User testing indicates that end users would prefer simply to see a merged results set, with records ranked by relevance; such a design would however place immense importance on relevance ranking algorithms, the results of which will not necessarily reflect end users' preconceptions of what is relevant. There is a risk of obscuring genuinely useful results from collections that are small. Appendix 1 discusses this issue in more detail.
- Enabling the marking of multiple records in the result set may be useful. This could enable the emailing or saving of records in multiples. This would require the user to complete two steps (mark a set of record; then save/email them) rather than one, so this should not be done without some evidence that end users actually do wish to save or email multiple records. Though this is common practice in bibliographic services, and it is accepted that end users often find several citations of interest during a session, it is not necessarily the case that users of multimedia services often seek multiple items.
- More functions could be added to the "MyVSM" area. These might include multiple levels of folders (subfolders), the ability to save a record into more than one folder, and functions for creating/renaming/deleting folders independent of their contents. More evidence of end users' actual requirements would be useful here. When the functional requirements are more thoroughly understood, an abstract API should also be added.
- Comparison of metadata between records was planned but not implemented. The usefulness of this feature should be re-evaluated in the light of the metadata that is actually available. It would not be very difficult to display metadata for several records for comparison, but significant effort would be needed to design a user interface for selecting records to be compared, and possibly selecting which metadata should be included.
- A feedback function would be a useful feature of a portal demonstrator that is used by significant numbers of end users over an extended period.
- A simple-icon based system should be considered that allows users to filter search results based on the types of uses to which the content may be put.

Advanced features in searching and presentation of results

The version of Zebra being used, and versions released since the beginning of the project, have many advanced features that we only touch upon within the demonstrator.

Relevance ranking has been configured within the database using the simplest approach that provides useful sorting of result sets. However, fine tuning of the algorithms used could be explored and could produce far more useful ordering of result sets for the user. In order to appropriately perform cross-searching of the

⁴¹ http://www.jisc.ac.uk/fundingopportunities/funding_calls/2008/02/sisbusinessplanning.aspx

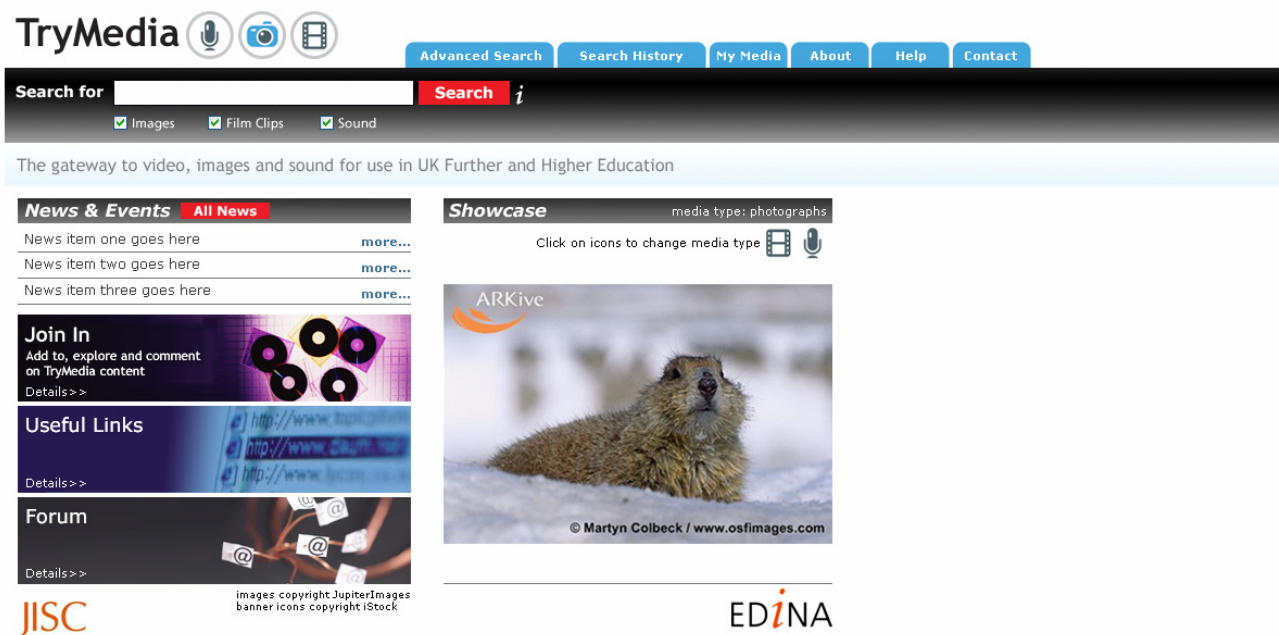
databases, and present a “mixed” result set to the user, excellent ranking algorithms must be achieved. It would also be possible to allow the user to fine-tune certain aspects of the weighting used in the ranking, although this might present more of a challenge to user presentation than technical capability.

Newer versions of Zebra also have a more intrinsic storage of XML data than is currently being used. Getting this functionality to work would improve (the already excellent) performance of Zebra, and provide even greater flexibility in the querying capabilities (again, presenting this increased functionality to the user in a useful way may prove the limiting factor).

Interface design and branding

Designs for a service version of the portal have been produced. These include supporting information and material in addition to the facility to search a number of content sources. The external evaluation identified that branding would be very important for the sustainability of a VSM Portal, while user testing also highlighted that it was necessary to present an interface that was not perceived as another search engine.

The ‘home’ page for a potential service is shown below:



Possible design for new service interface, showcasing media type: photographs

Incorporating metadata from Flickr and YouTube

Since the original ITT for the VSM Portal Demonstrator was issued there has been the emergence and huge take up of a range of services related to visual materials. Flickr and YouTube represent two popular and successful third-party repositories of images and movie-clips, respectively. Whilst these services contain community-contributed objects of a far more general nature than the collections already targeted by VSM, their almost ubiquitous presence on the world wide web and popularity in general would make their absence from an images and movies portal seem anomalous.

Both services provide an API for machine to machine contact. Flickr's API is extensive and available in many different protocols and language bindings. It allows almost the same set of functionality as its web interface for a registered user. Only a small proportion of this might be used by the VSM, in collecting data and their metadata. YouTube's API is marginally less extensive, but should be entirely suitable for usage that the VSM may require.

There are two ways that VSM could go about incorporating metadata from either of these services into the VSM Portal. For all other collections, VSM is currently harvesting the metadata from their sources, essentially a once-in-a-while task, indexing it locally, and using this local database for the user-instigated

searches from the web interface. This allows us much greater control over how the database works, and has been a highly successful approach. However, harvesting the entire contents of Flickr or YouTube is not only impractical, but probably not possible due to a combination of technical and legal limitations. Harvesting might still be an option, though, if only a small subset of the whole collections at Flickr and YouTube were to be used.

Flickr contains some high quality collections from certain users, or corporate users, such as the Library of Congress collections and smaller specialist museums. Some work could go into identifying these quality subsets and through various uses of the API they should be harvestable. Such identification would be an ongoing activity, but the VSM community at large could help to drive this task through use of tagging, or some other mechanism already available in the Flickr interface.

It is less obvious that easily identifiable collections of quality resources reside at YouTube, although there would still be mechanisms within the API to allow a subset of items to be identified for harvesting by the VSM Portal.

Harvesting would of course sit easily within the existing framework of the VSM and most of the work involved in this approach would be in formulating procedures to “tag” or otherwise identify specific items or subsets of items to be harvested by the VSM.

The alternative would be to operate a runtime search, that is, search the Flickr and YouTube databases via their APIs on-the-fly, *i.e.* when the user submits his or her search to the VSM Portal. This does not currently sit directly into the existing framework, but it would not be impossible to adapt the database API to allow for this method of searching. The downside to this would be the extra work in adjusting the current architecture, and an outside reliance on third-parties at the time of searching. The benefits of being able to search the entire corpus of objects at Flickr and YouTube, without the initial and ongoing identification of suitable items to be harvested, would need to be evaluated.

A further use of Flickr is demonstrated by the National Library of Australia in its Picture Australia⁴² web-site which incorporates images deposited in specified Flickr groups within its search results.

Portlet development or embeddable HTML

Any further development of the portlet should include adding the two stage approach to searching by having asynchronous multiple searches of the collection level databases, and the collection level choice of viewing as given to the user of the main VSM web portal interface. This would require extensive development of the portlet code, and significant design work to fit the web portal functionality and presentation into a portlet presentational framework.

An alternative to the portlet would be embeddable HTML that would provide, on an external web page, a simple search box that would then perform a deep search of the VSM Portal and present results to the user using the existing VSM web portal. This is akin to the approach used by Google to provide small Google search buttons from within other web pages (often used to search either just the third part site or the whole web). The effort to develop such HTML would be minimal compared to developing the portlet, and would return a solution that would be as functional as any alternative, require no (extra) maintenance and would not be dependent on an institution using a portal framework.

Content Based Image Retrieval (CBIR)

Some direct evaluation of this feature is necessary. We suggested one potentially useful CBIR-based feature (“find more images like this one”), and added this to demonstrate the use of CBIR within a portal. This has novelty value but it is not clear if it is genuinely useful in retrieving records that end users wish to find. Anecdotally we know that at least some of the positive remarks about the CBIR feature are because end users actually welcome the element of apparent randomness of the results set. If serendipitous discovery of interesting items is what users actually want, this could be implemented with considerably less time and expense (in development and ongoing server load) than CBIR.

⁴² <http://www.pictureaustralia.org/>

If further development of CBIR is required, it may focus on these areas:

- *Allowing users to upload a “seed” image.* CBIR searches are based on colour, shape and texture parameters; the easiest way for an end user to set these parameters is by selecting a seed image, from which these parameters may be measured. Our current portal implementation only allows images within harvested collections to be used as seed images; an interesting alternative would be to allow the user to supply the seed image, thus (potentially) allowing him/her to find images similar to one he/she has already obtained from some other source.
- *Re-examination of the software implementation.* Scaling and performance issues mean that the current implementation is not of service quality. It may be that performance can be improved with some minor development work, or it may be that major work and/or alternative software is required.
- *Allowing users to manipulate the colour, shape and texture parameters used in the CBIR search.* For example, the end user could be allowed to specify whether they are searching for images that match primarily the colour, the shape or the texture of the seed image; this might be done with a simple switch to include/exclude each of the three parameters, or could possibly allow the user to assign varying priorities to each. For example, a user with a picture of a London bus might exclude colour from the search parameters, if he/she is seeking pictures of buses that need not be red.

Metadata harvesting procedures

Metadata harvesting updates should be automated. All the necessary code has been written and structured in such a way that this may be done, and only a little extra development would be involved. On the current platform this would require configuration so that existing scripts are executed on an appropriate schedule, probably using the unix *cron* daemon.

It would also be useful to identify features of collections that are common and may be usefully dealt with by configuration rather than code. For example there is duplication of Perl code in some cases where similar XML formats occur in different collections, in particular XPATH declarations for elements that are directly incorporated into VSM metadata elements; this would more elegantly be implemented with shared code and configuration files.

Support for OAI/PMH is adequate but was forced to include provision for workarounds where OAI servers provided new and exciting variations on the PMH protocol. The implementation does not permit some features, in particular harvesting by specifying a date range; this is of no practical consequence at present, but it may trigger the need for additional development in the future. The most desirable way forward would be to work with content providers to ensure all the OAI servers correctly support the protocol, so that a standard PMH client (there are several open source options) may be used for the VSM Portal.

A lower priority but worthwhile development would be fixing the XML DOM support in the Zebra database installation.

VSM metadata scheme enhancement

There may be scope for enhancing the VSM metadata scheme. In the current implementation, the metadata are restricted to a relatively small number of fields that are used consistently throughout all, or most, of the collections. We were unable to include potentially useful metadata, such as date or coverage.

It would be possible to include more fields in the VSM metadata scheme if a greater degree of inconsistency between collections could be tolerated, but this would introduce problems that would have to be addressed.

One of the major problems lies in the design of the user interface. If, for example, a certain type of date information is available in only a half of all collections, then how should the user be informed of the effect of adding a date parameter to a search? If collections that do not support a date search are excluded, this falsely implies that no relevant records are to be found in those collections; but if they are included, which can be done only by ignoring the date parameter when searching those collections, irrelevant records will be included in the result set. If the user is not to be misled, the user interface must make the effects of using certain indexes clear to the user; this may be difficult to reconcile with the desire for a simple and intuitive design.

In the longer term, it would be advisable to include this area in negotiations with content providers. It may be that metadata are genuinely not available, but we have anecdotal evidence that useful fields have been excluded because it was too difficult to reconcile the “native” metadata scheme with Dublin Core⁴³; it is certainly the case that DC metadata routinely exclude useful information. There is little practical value in DC, and it may be to the mutual benefit (and relief) of both parties to abandon it, or at least to offer a version supplemented by non-standard elements.

Other additional developments

The user evaluation work suggests that consideration should be given to an editorial board that includes the input of teaching practitioners to further develop the collections policy and to select collections that fulfil that policy.

With the widespread use of social book marking, the options for enabling use of services such as del.icio.us⁴⁴ etc within the VSM Portal Demonstrator should be considered.

Possible roles for a future service

The project team suggests that there are a number of roles for a future VSM Portal service:

- A VSM Portal could offer a much needed user interface to an “image reservoir network”, as proposed by the JISC Images Working Group.
- A VSM Portal could offer a single focus and stepping-off point for the outputs from the JISC digitisation initiative⁴⁵.
- A VSM Portal could offer a platform for content providers of smaller, specialist collections relevant to UK HE/FE to reach a wider audience.
- A VSM Portal could offer a window onto images and time-based media held in repositories in institutions within UK HE/FE. It may be that harvesting of metadata held in any institutional repositories is done by a group such as Intute⁴⁶, or it could be that the VSM Portal Demonstrator harvests this data, and besides making use of it for the portal, also offers a m2m interface for other services to use.

⁴³ The frequency with which Dublin Core (DC) is used is hard to explain, given its unpopularity with developers. It could be that because it is mandated by OAI/PMH, a DC implementation is unavoidable, and developers are unwilling to maintain an additional metadata scheme. Alternatively it is possible that the decision was made without an awareness of the practical implications.

⁴⁴ <http://del.icio.us/>

⁴⁵ http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation.aspx: 22 projects, at least 10 of which are primarily related to images, film or sound.

⁴⁶ <http://irs.ukoln.ac.uk/> - Intute Repository Search, currently searches across academic eprints repositories.