discovery

A metadata ecology for UK education & research
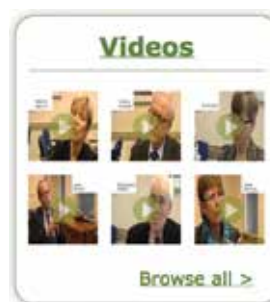
open data open doors

Issue 09 - October 2012

_Mimas   JISC   RLUK

From ideas and experiments to embedding and services

Videos

Browse all >

Over the past two years the adoption of the Discovery principles has noticeably shifted gear from ideas and experiments to an emphasis on embedding in everyday institutional and above-campus services. This has, for example, been apparent in the work of the Phase 2 JISC-funded service projects, reported in this issue.

In order to inform and support these developments, we have recently published a suite of online materials that emphasise the business case for Discovery and uncovers accessible approaches to adopting and embedding the open metadata and technical principles across a wide range of organisations and services.

» The Business of Discovery films – reflections of UK academic library directors
» Case studies – focusing on practice and outcomes in diverse organisations across the spectrum of museums, libraries, archives and aggregators
» Guidance materials – clarification and exploration of key topics associated with Resource Discovery

Discovery Phase 2 Projects: Enabling services

JISC has funded a second round of projects to establish feasible service models based on the Discovery open metadata and technical principles. In the last issue we looked at projects releasing open metadata. In this issue we track the work of projects that are building services on top of open metadata to address researcher and learner needs.

All 17 open metadata and service projects met in Birmingham in September to share updates and ideas as they conclude their work. It was a very stimulating meeting, not least because the shared 'Discovery dialogue' seems to have developed significantly since 2012.

The Phase 1 projects undertook some very useful experiments, but the Phase 2 projects have taken things up a notch. Here are a dozen recurrent themes from the meeting that were highlighted in the Discovery blog:

Theme 1: People
» Community is a vital driver – open communities maintain momentum; specialist enthusiasms and ways of working provide strong use cases
» For embedding new metadata practice, start where the workers are – as illustrated by add-ins to widely used systems such as Calm and MODS
» More IT experience and skills are required on the ground
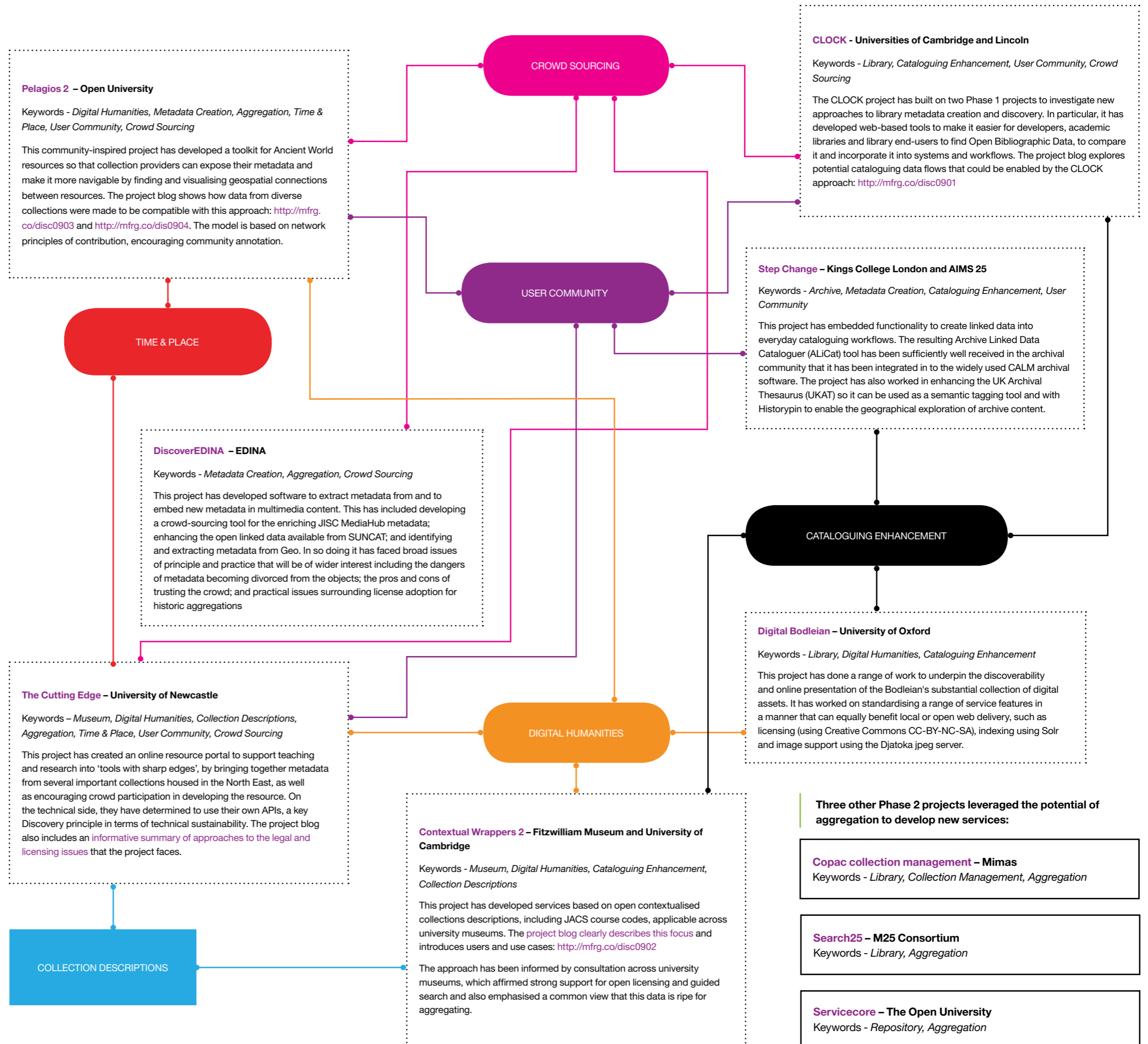
Theme 2: Data and access points
» Time and Place are priority access points
» Consistent and persistent URIs offer an effective base level linking strategy
» Collection level descriptions have potential as finding aids across domains

Theme 3: The way the web works
» Google's strength insists 'Do it my way' – and we should take heed (but we do need a both/and strategy)
» Aggregators need to know what is where to build or add value so an open metadata registry could be a valuable service
» No man is an island – It's a collaborative world with requirements to interact with complementary services such as DBpedia, Europeana, Google search, Historypin, Pleiades, UKAT and VIAF

Theme 4: Tools and technology
» There is opportunity / obligation to leverage authoritative data and vocabularies – examples as above and more, such as Victoria County History
» APIs are essential – but little use without documentation and example code
» OSS tools have been built by several projects - but how do we leverage them?

CROWD SOURCING

Pelagios 2 – Open University

Keywords - *Digital Humanities, Metadata Creation, Aggregation, Time & Place, User Community, Crowd Sourcing*

This community-inspired project has developed a toolkit for Ancient World resources so that collection providers can expose their metadata and make it more navigable by finding and visualising geospatial connections between resources. The project blog shows how data from diverse collections were made to be compatible with this approach: http://mfrg.co/disc0903 and http://mfrg.co/dis0904. The model is based on network principles of contribution, encouraging community annotation.

CLOCK - Universities of Cambridge and Lincoln

Keywords - *Library, Cataloguing Enhancement, User Community, Crowd Sourcing*

The CLOCK project has built on two Phase 1 projects to investigate new approaches to library metadata creation and discovery. In particular, it has developed web-based tools to make it easier for developers, academic libraries and library end-users to find Open Bibliographic Data, to compare it and incorporate it into systems and workflows. The project blog explores potential cataloguing data flows that could be enabled by the CLOCK approach: http://mfrg.co/disc0901

USER COMMUNITY

TIME & PLACE

Step Change – Kings College London and AIMS 25

Keywords - *Archive, Metadata Creation, Cataloguing Enhancement, User Community*

This project has embedded functionality to create linked data into everyday cataloguing workflows. The resulting Archive Linked Data Cataloguer (ALiCat) tool has been sufficiently well received in the archival community that it has been integrated in to the widely used CALM archival software. The project has also worked in enhancing the UK Archival Thesaurus (UKAT) so it can be used as a semantic tagging tool and with Historypin to enable the geographical exploration of archive content.

DiscoverEDINA – EDINA

Keywords - *Metadata Creation, Aggregation, Crowd Sourcing*

This project has developed software to extract metadata from and to embed new metadata in multimedia content. This has included developing a crowd-sourcing tool for the enriching JISC MediaHub metadata; enhancing the open linked data available from SUNCAT; and identifying and extracting metadata from Geo. In so doing it has faced broad issues of principle and practice that will be of wider interest including the dangers of metadata becoming divorced from the objects; the pros and cons of trusting the crowd; and practical issues surrounding license adoption for historic aggregations

CATALOGUING ENHANCEMENT

Digital Bodleian – University of Oxford

Keywords - *Library, Digital Humanities, Cataloguing Enhancement*

This project has done a range of work to underpin the discoverability and online presentation of the Bodleian's substantial collection of digital assets. It has worked on standardising a range of service features in a manner that can equally benefit local or open web delivery, such as licensing (using Creative Commons CC-BY-NC-SA), indexing using Solr and image support using the Djatoka jpeg server.

The Cutting Edge – University of Newcastle

Keywords – *Museum, Digital Humanities, Collection Descriptions, Aggregation, Time & Place, User Community, Crowd Sourcing*

This project has created an online resource portal to support teaching and research into 'tools with sharp edges', by bringing together metadata from several important collections housed in the North East, as well as encouraging crowd participation in developing the resource. On the technical side, they have determined to use their own APIs, a key Discovery principle in terms of technical sustainability. The project blog also includes an informative summary of approaches to the legal and licensing issues that the project faces.

DIGITAL HUMANITIES

COLLECTION DESCRIPTIONS

Contextual Wrappers 2 – Fitzwilliam Museum and University of Cambridge

Keywords - *Museum, Digital Humanities, Cataloguing Enhancement, Collection Descriptions*

This project has developed services based on open contextualised collections descriptions, including JACS course codes, applicable across university museums. The project blog clearly describes this focus and introduces users and use cases: http://mfrg.co/disc0902

The approach has been informed by consultation across university museums, which affirmed strong support for open licensing and guided search and also emphasised a common view that this data is ripe for aggregating.

Three other Phase 2 projects leveraged the potential of aggregation to develop new services:

Copac collection management – Mimas
Keywords - *Library, Collection Management, Aggregation*

Search25 – M25 Consortium
Keywords - *Library, Aggregation*

Servicecore – The Open University
Keywords - *Repository, Aggregation*

Mimas is Re-Discovering World War One

© IWM (Q 20028)

2014 marks the centenary of the outbreak of the First World War. The 'war to end all wars' to this day retains a large place in the education syllabus at all levels, and is a focus for research across disciplines.

JISC has previously funded a team from King's College London to identify and synthesise WW1 content and collections in the context of teaching and research priorities. Their recently published report, 'Digital Content for World War One', provides an opportunity to reappraise received notions of the experience and legacy of the conflict across disciplines. For example, previously under explored aspects of the war such as the global nature of the war, medical and nursing history, and the study of wider economic and social issues are identified as potential topics for further research.

Following this Phase 1 work Mimas is building an aggregation API and discovery layer for WW1-related content to support research and teaching. The WW1 Discovery project aims to make resources about the First World War more discoverable, and find new and innovative ways to present this content for the benefit of education and research. The team are working closely with the King's College team who have identified suitable First World War data sources and are now developing an online resource 'UK World War One Collections'.

It's an important initiative for Discovery. As Joy Palmer from Mimas puts it: "Ultimately what we want is to prove the concept that through adhering to the Discovery principles, good things will happen for teaching and research. We really want to be able to showcase the possibilities of pulling together content from different sources. This represents a new, lighter approach to resource management, with less emphasis on curation and more on enabling people to find exciting connections for themselves."

Joy is under no illusions as to the magnitude of the task ahead: "One of our biggest challenges is managing expectations. Users are not interested in the technical issues, they want a seamless, coherent and meaningful search experience across various LAM datasets. But we need to overcome these technical and licensing issues in order to demonstrate benefit. We also need to be able to present a robust business case for senior stakeholders. And all this before the end of the year!"

How will success be measured? "We just need to be able to access a small number of high-value datasets to demonstrate what's possible, and make this whole crucial concept quite compelling. But I hope we can go further than that, and demonstrate rich and usable connections between different datasets, and show that this approach can help us to aggregate different content types in meaningful ways."

Developments will be tested and evaluated by subject experts who will be asked to use the API to explore ways into, across and around the various resources. "We will be asking for honest reactions", says Joy. "This is not just about proving a technological concept. WW1 means a great deal to many people across all walks of civic and academic life. There's a wealth of knowledge still to be discovered. This is where we hope to be by October 2012, but it's not the end of the story."

Work is progressing well and Mimas have identified existing APIs, particularly those using Solr, for accessing relevant WW1 content

such as the National Maritime Museum, Culture Grid, Europeana and perhaps most importantly the Imperial War Museum. Other relevant resources include the People's Collection, Serving Soldier and Welsh Voices of the Great War, and The Oxford Great War Archive.

A prototype API based on the Solr technology, is undergoing testing at Mimas and should be ready for public consumption before long. In the last few weeks more data sources have been added from the list provided by King's College, and the Mimas team have just commissioned two companies to work with the API to develop the discovery layers that end users will experience. These are We Are What We Do, the company behind HistoryPin, and also Mickey and Mallory. The team are expecting to learn a great deal from their perspective about the realities of working with the API, with each company pursuing very different ideas about the user interface.

There's not much time left for the project to bring in more data, but they're doing what they can to help make the most of this important commemoration. If they succeed it will add an essential element to the Discovery vision that will allow us to interpret events from 100 years ago from the perspective of a modern audience that has few links to the conflict itself, but which has been profoundly affected by it politically, economically and socially. The legacy of the war that shaped subsequent generations is at risk of slipping from public memory, making it essential that today's educators and heritage custodians are able to reinterpret, re-engage and re-create a new social memory 'lest we forget' the experience and lessons of the war to end all wars.

Book your place on a Discovery Licensing Clinic

M	T	W	T	F	S	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

Friday 30th November, London

An opportunity for Managers and decision makers in libraries, archives and museums to consider the practical implications and operational options for open licensing, to address concerns and to develop an institutional action plan in response to the Discovery principles.
Register here.

Stay in touch

To endorse the Discovery Open Metadata Principles visit:
http://discovery.ac.uk/businesscase/principles
To make contact with members of the project team:
http://discovery.ac.uk/contact
To sign up to this newsletter simply send an email to:
rdtf-discovery@sero.co.uk
Discovery website: http://discovery.ac.uk
Discovery blog: http://blog.discovery.ac.uk