Improving discoverability of digitised collections: above-campus and national solutions

Recommendations from the Spotlight on the Digital co-design project (summary)

David Kay and
Owen Stephens
Contents

Foreword 2.

Introduction and Findings 4.
  » Findings
  » Recommendations
  » Solutions

Solutions 6.
  » 1 - Institutional capacity building
  » 2 - Role of aggregations
  » 3 - Technology foresight group
  » 4 - Reliable reference service
  » 5 - Reusable software tools
  » 6 - Strategic content promotion
  » 7 - Open content licensing

Concluding observations 14.

Acknowledgments 16.
Foreword

As one of the very first projects within Jisc’s pilot co-design programme, Spotlight on the Digital has been a beacon project, and its progress has been watched with special interest.

Working together, Jisc, Research Libraries UK (RLUK) and the Society of College, National and University Libraries (SCONUL) have sought to understand in detail how researchers, teachers and students hunt for digitised resources online and what hampers their efforts. We also wanted to think in detail about how content needs to be curated if it is to become easier to find and then remain discoverable for the long term. It’s a timely piece of work as ever greater emphasis is placed on integrating digitised and born-digital research, primary sources and data, and also one in which we, as co-design partners, have particular interest and expertise. As the voices of the academic library community RLUK and SCONUL are fully committed to making sure that quality resources remain readily available and reusable. We hope that our involvement has speeded up the project, brought wider insight to the information gathering process, and contributed to achieving really robust conclusions.

Over an eight-month period the project team gathered evidence in a variety of ways, by consulting with more than 40 institutions that are actively involved in digitisation, engaging in focus groups with information professionals, academics, web developers and commercial research organisations, surveying SCONUL and RLUK members, and sampling over 200 digital collections to see how easy
(or difficult) it is for people to find them. One stand-out message is that discoverability problems typically occur at item level. While whole collections may be relatively easy to find, single items present a much bigger problem, and a staggeringly high proportion of these may not come up in a standard Google (or similar) search. As custodians of digital content we need to think more from the researchers’ point of view and less like curators: people come looking for resources in ways we often haven’t even imagined.

This report identifies seven key areas where work is needed - including the need to build institutional capacity and the need to exploit aggregations - and it proposes some practical solutions that we believe can best be implemented at an above-campus level.

In addition, creators of content and managers of collections will need to take their own steps to support and sustain discoverability. The Spotlight project explored these too and Jisc has produced a new quick guide, “Make your digital resources easier to discover”.

Digitisation of content is just the start of an ongoing process if we don’t want valuable resources to slide into oblivion. They must be looked after carefully, and we must all be ready to respond to fresh challenges to discoverability in the future. The interventions have to keep on coming.

With that thought in mind, funding is now in place for the next phase of the Spotlight project. Work begins soon to prioritise the recommendations, start on some quick wins and develop and test new tools and services.

Kitty Inglis

David Prosser

Make your digital resources easier to discover
jisc.ac.uk/guides/make-your-digital-resources-easier-to-discover
Introduction

Jisc, Research Libraries UK (RLUK) and the Society of College, National and University Libraries (SCONUL) undertook Spotlight on the Digital as a co-design partnership project between June 2013 and January 2014.

The project emerged out of the concern that digitised collections are not as easy for users to find as they could be, which inhibits their potential for use by researchers, teachers and learners and limits the impact that collections can have on education and research.

Spotlight sought both to assess challenges to the discoverability of digitised collections and to identify practical solutions that will improve discoverability. It delivered:

» recommendations for above-campus and national solutions, which are outlined in this report and are aimed primarily at Jisc and the project’s co-designers

» a web resource with practical advice and guidance intended for creators, curators and managers of digital collections within institutions

  jisc.ac.uk/guides/make-your-digital-resources-easier-to-discover

The project ranged widely, working with managers, library practitioners, academics and technical experts to understand the challenges faced by institutions in ensuring the sustainability of digital collections for teaching, learning and research and the practical responses - both strategic and tactical - to these challenges.

In an intensive period of research, consultation and synthesis, the project benefited from the input of around 50 institutions (including three national libraries) and 60 individual experts. This involved four meetings of the Spotlight Expert Group, three specialised focus groups involving managers, curators and technologists, a survey of academic libraries, systematic assessment of 217 collections digitised with public funds and a literature review.

Findings

The key themes arising from the published research around online information-seeking behaviour underpin the project’s recommendations. These themes are:

» global search engines represent the main mechanism for discovering and locating content

» web-scale channels such as Wikipedia and Flickr are increasingly the starting points for discovery

» the impact of recommendation and in particular the roles of experts and peers should not be underestimated, especially as online "social" services become more embedded in practice

Taking account of these underlying behaviours and expectations, Spotlight undertook a systematic web assessment of 217 digitised collections to assess their level of discoverability.
Improving discoverability of digitised collections: above-campus and national solutions

Introduction

This highlighted that:

» some collections become lost to the web over time for a variety of reasons, while users and their citations require content to be persistently available, with appropriate redirection where applicable (a total of 40 web sites, amounting to almost 20% of the web assessment sample)

» while content is typically digitised under the umbrella of a collection, such groupings are not directly relevant to user behaviours, where users require discrete items to be discoverable (almost 50% of items were not discoverable on Google search results by item name or title). This finding was, however, counterbalanced by notable examples of best practice.

Recommendations

The above-campus and national solutions presented below focus on ensuring that digitised resources are directly and persistently discoverable on the web, can be amplified through recommendation, and are easy to find for the widest range of learners, teachers and researchers.

These solutions are based on powerful levers emphasised by managers, practitioners and technologists as well on the evidence gathered throughout the project. They can be grouped under three over arching themes:

» Build institutional capacity - Using national support to develop local institutional and practitioner capacity and capability (see #1 below), while recognising the value-added and fail-back roles of aggregation (#2)

» Exercise collective watchfulness - Engaging experts to ensure the sector is not caught out by technology changes (#3) and implementing automated checks to identify assets that are becoming undiscoverable (#4)

» Leverage expertise at scale - Identifying demanding lifecycle activity that can be optimally delivered and repeated at scale and therefore above-campus, such as software tool development (#5), high profile promotion (#6) and licensing advice (#7)

Solutions

The seven above-campus and national solutions are (in no particular order):

1. Institutional capacity building - Projects delivering explicit capacity development outcomes for the project partners and resulting in open best practice documentation

2. Role of aggregations - Engaging aggregators in content rehabilitation, transfer and mirroring projects to ensure that key digitised content is sustainably discoverable

3. Foresight group - Expert group meetings, assessing socio-technical trends and posting open advice

4. Reliable reference service - Development and operation of an automated online link checking and notification service and supporting website

5. Software Tools - Small open source software projects to develop tools, building on existing code

6. Content promotion - Strategically targeted press and social media activity, working in tandem with institutions, seeking national recognition and community amplification

7. Content licensing - Legally informed work to establish open licensing with rights clearance for both previously and newly digitised content where there is an appropriate level of interest

These recommendations combine to form a cohesive package. However, it will be helpful to recognise that core investments in capacity building (#1), exploiting aggregations (#2), foresight (#3) and oversight (#4) can create an environment within which the additional recommendations can be introduced as time and funding allow.

A web resource with practical advice and guidance

jisc.ac.uk/guides/make-your-digital-resources-easier-to-discover
Solutions

1. Institutional capacity building
2. Role of aggregations
3. Technology foresight group
4. Reliable reference service
5. Reusable software tools
6. Strategic content promotion
7. Open content licensing
Improving discoverability of digitised collections: above-campus and national solutions

Solution 1 - Institutional capacity building

Problem

Building local capacity offers a route to sustainability, but few institutions have the expertise and capacity in-house to develop the skills and processes required in the digital asset lifecycle.

Nonetheless, most prefer developing their own capacity to relying on external resources and being poorly equipped to support researchers involved in ongoing digitisation.

Use case

As a library director/stakeholder with a commitment to digitised collections, I want to:

» develop internal capacity to ensure easy long-term access to assets
» be able to advise and support researchers involved in digitisation
» underpin practice with clear policies and responsibilities
» endorse this commitment with evidence of use and impact
» ensure that future generations of curators are appropriately skilled and motivated to manage digital content

Solution

We need to assist institutions to develop and embed skills across their teams so they can meet the requirements of both human and programmatic discoverability and access. This can be enabled through facilitated ‘joint action’ projects, which will leverage digital library models and digital humanities practices. In these projects, an institution (or a group of institutions) will work in partnership with expert ‘ambassadors’ to undertake valued digitisation tasks and associated technical developments (such as implementing APIs) as a springboard for developing/embedding local skills and processes. The scheme could involve:

» developing ambassadors - typically experts assigned to this scheme under a partial secondment arrangement
» approving projects that create opportunities for ambassadors/technical experts to work with creators and curators in beneficiary institutions to enhance services and to develop and embed skills
» making a more conventional training offer, fuelled by these exemplar projects, available to all institutions
» linking the ambassadors and the training content to curricula for digital information professionals
Improving discoverability of digitised collections: above-campus and national solutions

Solution 2 - Role of aggregations

Problem
Some creators and curators don’t possess the right level of skills, or lack the mechanisms and infrastructure to safeguard content and to make it discoverable.

Others are looking for supplementary channels to promote their locally curated content more widely. Aggregators may be in a position to help with relevant collections but may not be making opportunities for hosting or mirroring clear to those responsible for digitised content.

Use case
As a creator/curator, I want to:

> maximise the reach of my resources whilst minimising the effort needed to do so, balancing efficiency and effectiveness
> use external services to overcome my lack of web presence
> find a good home for those assets that we don’t have the resources to curate, and protect them against technical or organisational change
> ensure at the same time that the user experiences this content through an appropriate interface

Solution
Our solution is to mobilise trusted aggregators to play a wider and more proactive role, especially adopting and mirroring digitised collections. We will need to:

> First, review the roles that might be played by current aggregations
  a. National - e.g. the Archives Hub, Copac, Culture Grid, Jisc Historic Books, Jisc MediaHub, VADS
  b. International - e.g. Europeana, Hathi Trust, Internet Archive
> Support pilots that establish feasibility and sustainable operational practice
> Consequently, assess any requirement for a new aggregation service to cover collection types that have no natural aggregator interest
Solution 3 - Technology foresight group

Problem
The discovery environment is controlled not by universities, nor by the education or GLAM (Galleries, Libraries, Archives, Museums) sectors more broadly; change is driven by global and vendor developments.

Furthermore there is a tendency in the face of change to focus on technical responses rather than to consider the socio-technical ecosystem.

Use case
As a creator/curator, I need to:

» be aware of changes in web and vendor discovery environments and to be advised of key risks and remedial options to keep my content discoverable
» understand technical change in the context of the user experience

Solution
We propose to establish a "foresight" (or "observatory") group that might collaborate virtually twice annually in order to:

» track key technical and market trends in discovery and related areas
» embrace the range of scenarios and use cases prevalent in the sector
» make recommendations to stakeholders and the community
» pre-emptively guide development of related practice, notably influencing other activity proposed here, such as training and tools

Partners might include the British Library Lab, the Digital Curation Centre, the New Media Consortium and Wikimedia Commons alongside leading institutions and services, though it is important that we don’t lose focus by attempting to draw in all stakeholders.
Problem
Websites can cease to be discoverable by humans and/or by machines for a variety of reasons, many of which can be remedied given a timely response, but this may not be forthcoming from creators or curators unless they are alerted.

The headline reasons for failure identified from the 40 cases in the web assessment sample (almost 20% of 217 tested) were as follows:

- 26 cases where the Site no longer exists or is marked as closed without redirection;
- 8 cases where the assets have been incorporated elsewhere or the named collection cannot be identified within the site;
- 6 cases where the site still says it is beta / work in progress or site functionality is broken so the assets are not accessible.

Use case
As an interested party (researcher, creator, curator, institution, investor), I want:

» citations of assets to be reliable over time
» mechanisms in place to check automatically on the health of digital assets, to ensure they do not disappear unnoticed and are consistently visible to web scale discovery mechanisms over the long term
» the lead curator to be alerted of any detrimental changes

Solution
Our proposed solution is to provide a discoverability oversight mechanism. This will use software that can - automatically and continuously - validate that digitised content at registered URLs is discoverable by machines and therefore also by humans. This will be directed through a “collections” reference list, which might loosely be called a “directory” or “registry” but without the implied technical characteristics. The objective is to provide automated health checking rather than necessarily resolving identifiers to resources, thereby providing registered users with notifications. This type of targeted and fundamentally straightforward service is highly feasible in the current web environment. It will:

» ensure resources can be found via search engines, citation and social media
» generate confidence through regularised automated checking of URLs and other key features of registered content
» alert lead parties of issues
» build reputation as a place to discover well-curated, web-aware collections
» contribute to community awareness of good practice regarding search engine discoverability

Initially this can be built from the Spotlight Web Assessment, which covered 217+ collections, though there is interest indicated more broadly.
Solution 5 - Reusable software tools

**Problem**
Technically overseeing and managing the digital asset lifecycle can involve frequent and repetitive actions over large numbers of files, web pages and metadata records; there are similar technical challenges in assessing performance and impact.

**Use case**
As a creator/curator of large and diverse resources, I need to:

» maintain web pages, files and metadata as efficiently as possible, to keep all my content in a highly discoverable position

» automate, where possible, review of the web performance and impact of my content

**Solution**
We propose to identify and/or develop open source tools and services for local and above-campus use, working internationally where possible. Specific requirements will be identified though other activities recommended here, for example:

» rehabilitating assets that are “orphaned” (i.e. no longer being looked after/no longer discoverable)

» describing and surfacing item level content on the web

» inserting appropriate page descriptions

» linking to authoritative identifiers (e.g. places, names, subject headings)

» measuring impact, through usage data and building on existing resources - eg the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR)

This may involve working with specific systems vendors (e.g. Adlib, Calm, OCLC ContentDM), with Open Source products such as ePrints and Fedora, with sector partners, and commissioning expert developers (e.g. British Library Labs, Cottage Labs, Knowledge Integration).
Solution 6 - Strategic content promotion

**Problem**
The efforts of creators and curators, and the focus of institutional PR teams, are often restricted to the development and launch of digitised collections. It is therefore difficult to sustain ongoing effort through activities such as press releases and social media.

There are also challenges in working effectively with Wikipedia’s editorial expectations and adopting optimal strategies with other web scale amplifiers.

**Use case**
As an interested party (researcher, creator, curator, institution, investor) I want:

- assets to be promoted over the long term in order to stand out from the ordinary and to build their scholarly reach, research and teaching impact and reputational value
- broader amplification and contextualisation to supplement my local communications focus
- to use high impact web services like Wikipedia to create powerful links for audiences potentially interested in my content

**Solution**
We propose sustained advocacy and promotion of collections, their distinctive parts and their thematic connections through a range of techniques and channels:

- press releases at key moments (e.g. launch, nth page impression, etc)
- providing broader contextualisation and thematic bundling (e.g. WW1, Commonwealth Games)
- social media campaigns and persistent exposure
- positioning through events (e.g. Jisc Digital Festival)
- visibility at popular web destinations (e.g. Flickr, Wikipedia)
- potentially selective dissemination “push” services
Solution 7 - Open content licensing

Problem
Licensing of digitised content, rights clearance and associated stakeholder relationship management can demand experience and effort that are unavailable from the creators or curators. Orphaned collections, or the ones that become ‘lost’ to the open web, also present difficulties, perhaps more problematic than when they were originally digitised.

This is because licensing terms and conditions might not have been a consideration at the time and negotiations with rights holders may have been poorly recorded.

Use case
As a creator/curator, I want to:

» be sure that my digitised content is licensed as openly and permissively as possible to maximise freedom for reuse and to enable the unexpected
» ensure rights are appropriately cleared to maintain positive relationships with associated rights owners and benefactors
» support academics in becoming aware of, and adopting, essential practice as they undertake further digitisation

Solution
We propose to facilitate centrally the open licensing of collections, especially those that are currently unlicensed, orphaned or unwanted. This may be undertaken proactively (e.g. for key collections) or by institutional application. The service would include:

» Creative Commons licensing for content and metadata
» rights clearance where required, including consideration of any copyright disclaimer that might have been published with the original digitised items
» ongoing review of fit with prevalent licensing models and emerging interpretations (e.g. CC advice on high resolution images)
The Spotlight study was undertaken to a very tight timescale and there remain several areas that would benefit from further research to maximise the opportunities and to develop better understanding of the challenges highlighted here and in the advice for institutions.

Potential roles of aggregations
As highlighted in these national recommendations, feasibility needs to be understood with reference to specific collection types and aggregations (not exclusively in the UK)

Maximising assets in the UK cultural sectors
While this project was focused especially on the perspective of the academic library function, it is essential to recognise the range of curatorial roles at play - archives, galleries and museums within HE and across the wider public realm, and also research groups and individual scholars. Several of the recommendations here could be enhanced through application to the broader UK cultural sectors, working with galleries, libraries, archives and museums. Public libraries represent a particular case as curators of digitised content on account of their ongoing restructuring

Exploiting the web as a community
Given the wider community of contributors and curators, dangers could potentially arise from competing in the style of the commercial entities for primacy on the web if massively distributed content for any given area of interest is to be readily discoverable. This should be a consideration of ongoing work, taking account of initiatives such as the Digital Public Space
Policy development

The majority of institutions highlighted a lack of clear policies and designated responsibilities relating to digitised content (and also for born digital assets). Work with RLUK and SCONUL to develop and disseminate policy and role templates would be of clear practical value.

Digital library models and digital humanities practices

The recommendations relating to capacity development (#1), foresight (#3) and oversight (#4) need to be closely linked with initiatives to define effective digital library models and to develop best practices in digital humanities. While the UK’s HE sector aspires to be a world leader, this should take full account of international developments.

Links with OERs and other digital learning assets

It was noted during the consultation that other types of digital content suffer from discoverability problems very similar to those identified with digitised assets. Consideration of this wider field may uncover extended service opportunities.

Discovery behaviours of lecturers on further and higher education

Spotlight recognised that this group is not covered in the literature and therefore undertook some high level conversations. Lecturers play a crucial role in recommendation and therefore it will be important to develop a better understanding of their interaction with peers and students.

Further consultation with academics would also add value to the targeting and impact of the proposed solutions and to the Spotlight institutional advice.

Other behaviours

For reasons of timing and resources, Spotlight limited its focus to analysing the information-seeking behaviour of researchers, students and lecturers. There are other important groups that would merit consideration in the light of the Phase 1 findings and recommendations, such as professional researchers, and teachers and students in further education and schools.

Understanding workflows

The expert and technical groups both highlighted the importance of workflows beyond the user’s discovery of initial touch points. A better understanding is required of how users navigate the range of tools and systems available in institutions (e.g. Virtual Learning Environment, reading lists system, discovery layer, digital library) leading to enhanced discovery and eventually to access.

Identifying impact measures

No survey respondent reported “considerable experience” in measuring and analysing use or impact of digitised content (with a similar position likely applying to born digital assets). Early work towards identification and development of factors that can be measured in practice would attract widespread engagement.
Sero Consulting, with the collaboration of Ian Chowcat, and Owen Stephens Consultancy undertook all the research, consultation and evidence gathering activities for Spotlight and authored this report.

Acknowledgments

Co-design partners and project team

» RLUK - David Prosser (RLUK executive director)
» SCONUL - Kitty Inglis (university librarian, University of Sussex, and chair of the SCONUL academic content and communications strategy group)
» Jisc - Catherine Grout, Paola Marchionni, Peter Findlay, Liam Earney, Andy McGregor

The project wishes to acknowledge the extensive support, time and imagination applied to the project by contributors within and beyond the higher education sector.

The Spotlight expert group

» Chris Batt - PhD researcher, University College London
» Julie Carroll-Davis - vice president Global Content Alliances, ProQuest
» Pat Christie - director of libraries and academic support services, University of the Arts
» Paul Ell - director, Centre for Data Digitisation and Analysis, Queen’s University, Belfast
» Ed Fay - digital library manager, LSE
» Tim Hitchcock - professor of 18th century history, University of Hertfordshire
» Kitty Inglis - university librarian, University of Sussex, and chair of the SCONUL academic content and communications strategy group
» Mike Mertens - deputy executive director, RLUK
» Catherine Moriarty - curatorial director, University of Brighton Design Archives
» Andrew Prescott - head of Department of Digital Humanities, King’s College London
» Angharad Roberts - British Library/University of Sheffield
» John Stack - head of digital transformations, Tate
» Grant Young - head of content programme, digital services, Cambridge University Library
» David Zeitlyn - professor of social anthropology, University of Oxford
HE institutions and services taking part in consultations (43)

» AIM25 (Kings College London)
» Archives Hub (Mimas)
» Birkbeck College, University of London
» Brunel University
» Cardiff Metropolitan University
» Cetis
» City University London
» De Montfort University
» Durham University
» EDINA
» King’s College London
» London School of Economics
» Mimas
» National University of Ireland, Galway
» Queen Mary University of London
» Queen’s University Belfast
» Royal Holloway, University of London
» Southampton Solent University
» University College London
» University of Birmingham
» University of Bristol
» University of Cardiff
» University of East London
» University of Hertfordshire
» University of Huddersfield
» University of Hull
» University of Leeds
» University of Leicester
» University of Manchester
» University of Newcastle
» University of Northampton
» University of Nottingham
» University of Reading
» University of Salford
» University of Sheffield
» University of Southampton

» University of Surrey
» University of Sussex
» University of the Arts
» University for the Creative Arts
» University of York
» York St John University

Other contributing organisations (8)

» British Library
» Curtis + Cartwright Ltd
» Hedtek Ltd
» Knowledge Integration Ltd
» National Library of Scotland
» National Library of Wales
» Wellcome Trust
» Wikimedia UK

Further information and resources
digitisation.jiscinvolve.org/wp/category/spotlight/

For any enquiries please contact
Paola Marchionni
p.marchionni@jisc.ac.uk
Share our vision to make the UK the most digitally advanced education and research nation in the world

jisc.ac.uk