



The challenges of discovery and practical solutions: Spotlight on the Digital project  
Paola Marchionni, Head of digital resources for teaching, learning and research, Jisc  
<http://digitisation.jiscinvolve.org/wp/spotlight-on-the-digital>



The Spotlight project found that the discovery challenge for many digitised collections is a bit like a treasure chest at the bottom of the sea: it's there, overflowing with jewels, but not many people can find it and enjoy the jewels.



The Spotlight team investigated this challenge and identified practical solutions for “surfacing” the treasure chest’s jewels, so that they become more discoverable through the channels most commonly accessed by users .

## » What is it?

- › Partnership project between **Jisc, RLUK, SCONUL** (Jun '13-Jan '14). <http://digitisation.jiscinvolve.org/wp/spotlight-on-the-digital>
- › Part of **Jisc co-design pilot programme** to tackle sector's concerns and develop new products and services  
<http://www.jisc.ac.uk/research/funding>

## » Aim

- › **define the discoverability problem and current status** in relation to digitised collections; **identify practical solutions** to improve their discoverability



- » Project gathered quantitative and qualitative evidence from:
- » **Expert group mtgs** (15 library managers, curators and academics)
- » **Literature review** on online information seeking behaviour  
<http://digitisation.jiscinvolve.org/wp/2013/11/04/top-discovery-channels-in-online-user-behaviour/>
- » **Web-based assessment** (manual and automated tests) of 230 collections digitised between 1998-2003 funded by AHRC, Jisc and the New Opportunities Fund

- » **Technical forum** – 18 technical practitioners and managers
- » **Library focus groups** – involving 24 library and collection managers from HE institutions and National Libraries
- » **Library online survey** – HE libraries with experience of digitisation (31 full institutional responses)
- » **Caveat** – no access to usage figures for digitised collections. Web assessment tests more about testing good practice for web publishing not level of usage or impact of collections as such.



**Global search engines** – search engines (such as Google) represent for the majority the default mechanism for discovering. But surveyed libraries believe key channels are Institutional Repository, Google and the Discovery Layer.



**flickr**

**Popular web-scale channels** – Channels such as Wikipedia and Flickr are regarded as starting points for students and researchers



**Social recommendation** – The impact of recommendation and in particular the roles of experts and peers should not be underestimated; it may become more explicit as online 'social' services achieve critical mass and become more embedded in practice.

**Undiscoverable collections** – Some collections become “lost” to the web over time (53 /230, representing almost 25% of the web assessment sample). Reasons range from poor exposure to search engines to the loss of web access to the content itself to relocation within other collections or aggregation services – which doesn’t necessarily mean that collections don’t exist anymore.



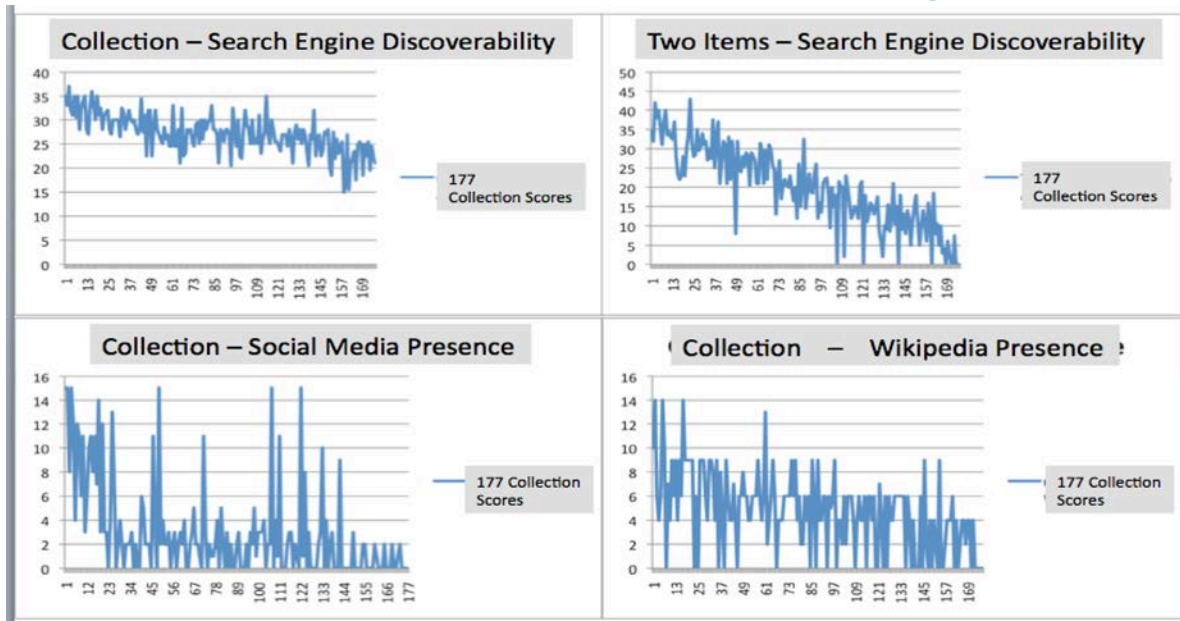
**Error 404**

**Undiscoverable items** – Items, as opposed to collections, are at most danger of being “lost” (only about 50% of items assessed appeared on the first page of Google results using the item name or title).

<http://digitisation.jiscinvolve.org/wp/2013/11/18/discovery-of-digitised-collections-vs-items/>



# Collections and item level performance tests



- Vertical axes indicate highest possible scores, horizontal axes indicate all collections tested (177) From top left clock wise:
- 1) Collections search engine discoverability: generally score well/least badly as illustrated by less steep fall-off across the sample
  - 2) Two items discoverability: difference in scores between best and worst is greater than for collections, fall-off is quite steady
  - 3) Wikipedia: very strong performance of some of the overall best scorers, spikes of those who do/do not work on Wikipedia citations
  - 4) Social media (twitter): strong leading group followed by a sharp fall-off.

*These tests reveal degree of adoption of best practice and where there is scope for improvement – not “absolute” discoverability/use*

In Lorcan Dempsey's words, the discovery challenge that libraries (but also other digital content creators) face is with the shift from **outside-in** to **inside-out** in the provision and access to resources  
<http://orweblog.oclc.org/archives/002206.html>

*"While much of the discovery focus of the library is still on those destination or gateway systems which provide access to its collection, **much of their users' discovery experience is in fact happening elsewhere.** [my bold]*

...

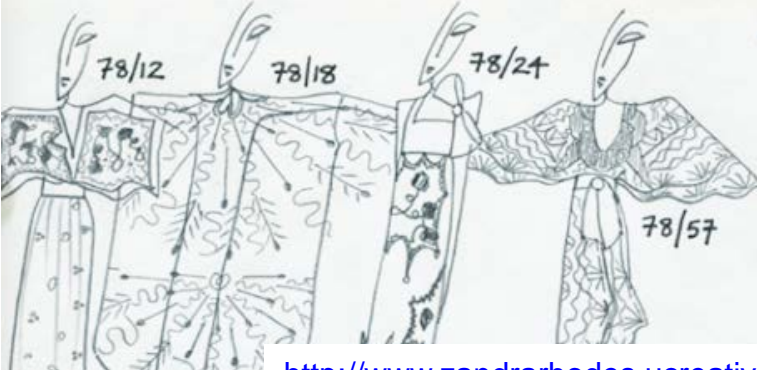
*The challenge is not now only to improve local systems, it is to **make library resources discoverable in other venues and systems, in the places where their users are having their discovery experiences.** [my bold]*

# Some great examples of good practice

## British Library Sounds

<http://sounds.bl.uk>

Listen to a selection from the British Library's extensive collections of unique sound recordings, which come from all over the world and cover the entire range of recorded sound: music, drama and literature, oral history, wildlife and environmental sounds.



<http://www.zandrarhodes.ucreative.ac.uk/>

## The First World War Poetry Digital Archive

<http://www.oucs.ox.ac.uk/ww1lit>



<http://en.wikivet.net/OVAL>

## Cambridge Digital Library

Home Browse My Library About Help

### Newton Papers

“  
Plato is my friend,  
Aristotle is my friend,  
but my greatest friend is  
truth.”

—Sir Isaac Newton  
(MS Add. 3996. 88r)



<http://cdl.lib.cam.ac.uk/collections/newton>

Examples of good practice in discoverability from a variety of institutions, collection types and formats

**Quick guide**

## Make your digital resources easier to discover

7 March 2014

Make it easier for others to find and use your digital collection, enhancing its impact and reach.

Many digitised collections are not as 'discoverable' as they could be through the channels and devices most commonly accessed by users. Clearly this is a missed opportunity for collection owners as it reduces use and the potential impact on research and teaching that a digitised collection can have.

Through a number of clear steps, learn to make your digitised collection more discoverable and easier to use by third parties. These should enhance discoverability within your own institution, and those that will help you extend the reach of your collection internationally.

### Make Google searches work for you

Search engines such as Google, Bing and Yahoo are major discovery tools for all online audiences, including students and researchers. Develop techniques that will ensure your digital collections are easily found through search engines.

### Use social media

Social media allows you to place your digital collections directly in front of targeted, interested audiences. Learn how to develop a social media strategy and select appropriate tools for your campaign.

### Learn to use content aggregators

In academia there are a number of major online services that pull together resources relevant to researchers and students in particular sectors. Examples include Summon, Primo Central, Jorum, Copac, Archive.org and CultureGrid. Learn how these work and how you can use them to promote your collection.

### Make your collection available for learning and teaching

Students of all levels look for recommendations from their teachers to help them discover new resources. Making teachers and lecturers a key path to sharing content from your digital collection. Discover how to help teachers and students find and use it.

### Reach global audiences

Some of the web's most popular sites, such as Wikipedia, YouTube and Flickr, offer an opportunity to reach large audiences with your content. Get to know these and exploit them to reach far larger audiences.

### Improve the user experience

'User experience' refers to the overall experience your users have on your website but is often associated with the ease of use of a web service. Considering how your users access and navigate your collection will help you make it easier for them to find and use.

### Reach academic researchers

Learn about common behaviours such as citation chaining among the research community and discover ways to help researchers find and use your digital collection.

Spotlight produced a guide which identifies practical solutions that can be deployed by managers of digitised/digital collections to improve their discoverability  
<http://www.jisc.ac.uk/guides/make-your-digital-resources-easier-to-discover>

# Make Google searches work for you

Develop techniques that will ensure your digital collections are easily found through Google

## Authors

Ian Chowcat  
Sero Consulting

David Kay  
Sero Consulting

Owen Stephens  
Sero Consulting



[Print this page](#)

## Contents

### Introduction

[Structure URLs](#)

[Relevancy ranking](#)

[Page titles](#)

[Meta tags](#)

[Sitemaps](#)

[Index content](#)

[Improve images](#)

[Schema.org](#)

[Item descriptions](#)

[Remove barriers](#)

[Heading tags](#)

[SEO](#)

## Introduction

Search engines such as Google, Bing and Yahoo are major discovery tools for all online audiences, including students and researchers.

Whilst other specialist search engines exist, Google was repeatedly identified as the main starting point for research by a wide range of users.

This creates a need to improve the way digital collections and their items are found and displayed in search results, a primary mechanism by which the discoverability of resources can be improved for all audiences.



Strategic outcomes



Discovery behaviours

## Too much info?

See our quick guide on [making your digital resources easier to discover](#)



# Add structured metadata to your pages using schema.org

Schema.org is an initiative that allows you to add structured data to web pages in a way that can be extracted and used by search engines. The type of data that can be added is sometimes referred to as 'microdata'. The Schema.org initiative is supported by all of the main search engines, including Google, Bing, Yahoo and Yandex.



## Examples - The Wellcome Library and Cambridge University Digital Library

### Measures for success

The most basic measure you can carry out is to check that schema.org has been added to your website appropriately.

You can carry out spot checks using key search engines to see the impact of adding your items appear.

There are a number of tools that will allow you to effectively use web analytics to understand traffic to your site as a result of using schema.org:

- [This analytics toolkit](#) forms part of the Toolkit for the Impact of Digitised Scholarship
- A [Google Analytics Health Check](#) from Culture24
- A resource to explain [Google Analytics Segments](#), from Culture24

### More information

### More information

#### Skills and knowledge needed

- Web development skills

#### Cost

- £££££ - occasional investment required

#### Resources

- [Schema.org](#)
- [What is microdata?](#) from Yandex
- The [Jisc discovery guidance document Microdata](#)
- [Marking up your site with structured data](#) from Bing
- [Google's advice and FAQs on using schema.org](#)



# Above campus recommendations

- » Spotlight also identified 7 areas for above campus interventions to support and complement institutions' own efforts in improving discoverability:
- » institutional **capacity building**; **foresight**/observatory group; **tools** for managing digital assets lifecycle tasks; content **promotion**; open **licencing** services; **reliable reference service** for automated checking of urls; explore role of **aggregators**

See <http://digitisation.jiscinvolve.org/wp/2014/01/22/discovering-digital-collections-above-campus-and-national-solutions/> for a draft of the above campus recommendations. Final document to follow soon. Phase 2 will tackle these.

# Why is discoverability still a challenge?

- » *“special collections and archives play a key role in differentiating each institution from its peers” (RLUK p15)*
- » *“over 80% [of respondents] agreed that digitized special collections are critical to our current strategic direction” (ARL)*

See: *Survey of Special Collections and Archives in the United Kingdom and Ireland.*

<http://bit.ly/11ljKHa> RLUK-OCLC, 2013

*Appraising our Digital Investment: Sustainability of Digitized Special Collections in ARL*  
[Association of Research Libraries] <http://bit.ly/11ljQOT> Ithaka S+R, 2013

# Why is discoverability still a challenge?

However...

# Why is discoverability still a challenge?

- » Post-digitisation, many projects draw support from host institution typically for storage, preservation, access, but this is **not budgeted or guaranteed** (Jisc/SCA-Ithaka p4)
- » **No much effort** or budget goes **towards enhancement** of resources created (ARL p2)
- » Libraries prioritise **ensuring access to subscribed resources** over maintaining discoverability and access to digitised resources (Spotlight survey)

# Why is discoverability still a challenge?

- » For academics sustaining digitised resources is a **voluntary act**
- » **Libraries can't take responsibility for multitude of digital assets** created through teaching and research activity
- » **Lack of clarity of roles and responsibilities** and availability of human resources are most crucial challenges
- » **Libraries' perception of most important discovery channels** are:  
1. **Inst open access repository**; 2. **discovery layer**; 3. **Google**

# Why is discoverability still a challenge?

Your thoughts...?



# Thank you

Paola Marchionni

[p.mrchionni@jisc.ac.uk](mailto:p.mrchionni@jisc.ac.uk) @paolamarchionni

## Image credits

- » Treasure chest <https://www.flickr.com/photos/63082794@N00/6030879386/lightbox/?q=treasure%20chest%20sea>
- » Social recommendation icon <http://tinyurl.com/q2h2hyr>