

Spotlight on the digital: improving discoverability of digital collections

Paola Marchionni, Head of digital resources for teaching, learning and research, Jisc

LIBER conference 29 Jun- 1 July 2016, Helsinki

The Spotlight on the Digital project

Improving discoverability of digital collections

What is it?

- Partnership project between **Jisc, RLUK, SCONUL** started in 2013, now in phase 2, to address concerns over discoverability of digitised collections

Aims

- **define the discoverability problem** in relation to digitised collections
- **identify practical solutions** to improve their discoverability
- **deliver solutions** to the community

More info at https://digitisation.jiscinvolve.org/wp/spotlight-on-the-digital/

The Spotlight project found that **digitised collections are like a treasure chest at the bottom of the sea**:
it's there but not many people can find it

What Spotlight found

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

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.

What Spotlight found

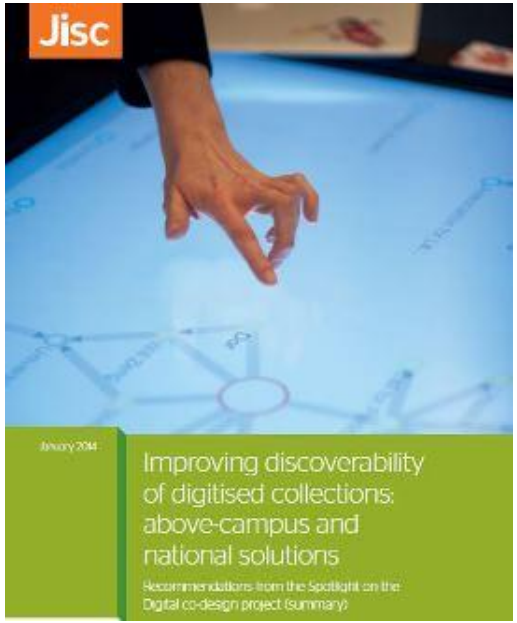
Undiscoverable collections – Some collections become "lost" to the web over time (about 20% of the web assessment sample). Reasons range from poor exposure to search engines to the loss of web access to the content itself to relocation within other collections or aggregation services – which doesn't 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://bit.ly/Spotlight_items

What Spotlight did



Final report phase 1 and other outputs
http://bit.ly/Spotlight_outputs

Phase 2 took forward 3 recommendations:

1 Guidance

2 Training

3 Tool

1 practical online guides

“Make your digital resources easier to discover”

http://bit.ly/Spotlight_guide

Quick guide

Make your digital resources easier to discover

Last updated: 8 September 2015

Universities and other organisations have made considerable investment in creating collections of digitised resources for learning and research. However, these resources are not always as discoverable as they could be by online audiences.

This means that students may miss opportunities to benefit from valuable primary source material. It can also limit the impact of digital collections on education and research, both locally and internationally.

What you can do

As part of our co-design project, also known as spotlight on the digital, we have put together a series of guides to help make your digitised collection more discoverable and easier to use.

9 inter-related guides covering:

- make Google searches work for you
- use social media
- learn to use content aggregators
- make collections available for teaching and learning
- use popular web sites to reach broader audiences
- improve the user experience
- reach academic researchers
- create collection champions
- integrate with your organisation's systems

Make Google searches work for you

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

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

All guides have same structure and some content items are in common.

Contents

Introduction

Structure URLs

Relevancy ranking

Page titles

Meta tags

Sitemaps

Index content

Improve images

Schema.org

Item descriptions

Remove barriers

Heading tags

SEO

Further information

Add structured metadata to your pages using schema.org

Schema.org is an initiative to enable the addition of structured data to web pages in a way that can be extracted and used by search engines. Using more detailed markup for items in digital collections can enable more effective indexing and display of digital items in search engines, and in the future may allow for more powerful search options to be supported.

Examples - The Wellcome Library and Cambridge University Digital Library

- The Wellcome Library **includes schema.org in it's digital object viewer**
- Cambridge University Digital Library **uses schema.org in its item viewer**

Measures for success

More information

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 ad
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 und
traffic to your site as a result of using schema.org:

- This analytics toolkit forms part of the Toolkit for the Impact of Digitised Scholar
- A Google Analytics Health Check from Culture24
- A resource to explain Google Analytics Segments, from Culture24

More information

UNIVERSITY OF CAMBRIDGE Cambridge Digital Library



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

Create resource packs
Common data formats
Use APIs effectively
Public domain metadata
Develop collaborations
Find partners
Further information

In the loop

Linda Newington, University of Southampton

As an example of social media in action, Linda discusses the success of using Twitter to promote the Knitting Reference Library @intheloop:



See more videos at bit.ly/JT_Videos – Academic collaboration in action

2 training offer

- Launching in **Nov 2016**, **2 workshops**
- Based on online guide, pilot with 6 Higher Education institutions and feedback from a free test run

Workshop 1: Using digi collections in learning teaching and research

- Discovery behaviours, promotion channels including social media, tracking use and impact, hands-on session

Workshop 2: Making Google work for your digital collections

- How Google and other search engines work, SEO, schema.org, Google Analytics, hands-on session

Feedback on pilot training

*Good mix of evidence/background, plus **practical approach** and lots of **good research** and **tips** on how to engage with academics etc*

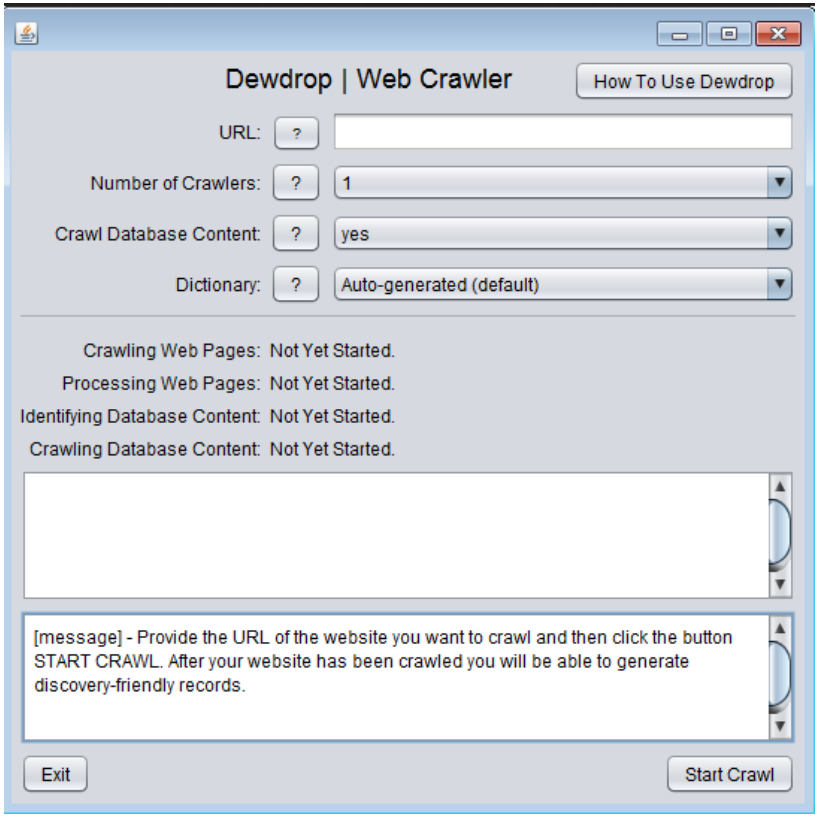
*Content and exercises were **engaging***

*The course as a whole gave me a much broader **understanding of the idea of discoverability***

*Chance to reflect on **best practice** across the sector, meet staff users of comparable collections and particularly **focus in on the use of metrics** to inform progress*

3 tool to support discovery

Dewdrop tool - in development (alpha)



- Uni Sheffield, Historical Research Institute
- Tool to develop **discovery-friendly records** of your collection and make them more easily **accessible to search engines** and third-party content aggregators

- discovery-friendly records act as **signposts**, directing search engines and aggregators to your website
- web crawler with a **text analyser** using among others **Natural Language Processing**
- **suitable for poorly designed websites** that cannot be easily discovered or indexed by search engines
- discovery-friendly records **do not make existing website content invalid.** They are a version of the content that is presented in HTML RDFa, RDF or JSON-LD formats.

Image credits

- Cover slide: Spotlight, Thomas Wood https://www.flickr.com/photos/twoody291/7380317612/ (cropped) Licence https://creativecommons.org/licenses/by/2.0/

- Slide 4: Team by Arnav Sameer from the Noun Project https://thenounproject.com/sameerarnav/collection/social-media-icons/?oq=social%20media%20icon&cidx=0&i=506553

Thank you

Paola Marchionni, Jisc
Head of digital resources for teaching, learning and research
p.marchionni@jisc.ac.uk @paolamarchionni