

New Approaches to Describing and Discovering Open Educational Resources

Martin Hawksey, Phil Barker and Lorna M Campbell, Cetis

martin.hawksey@strath.ac.uk, phil.barker@hw.ac.uk and lmc@strath.ac.uk

Abstract

This paper will report and reflect on the innovative technical approaches adopted by UKOER projects to resource description, search engine optimisation and resource discovery. The HEFCE UKOER programmes ran for three years from 2009 to 2012 and funded a large number and variety of projects focused on releasing open educational resources (OERs) and embedding open practice. The Cetis Innovation Support Centre was tasked by JISC with providing strategic advice, technical support and direction throughout the programme. One constant across the diverse UKOER projects was their desire to ensure the resources they released could be discovered by people who might benefit from them; if no one can find an OER no one will use it. This paper will focus on three specific approaches with potential to achieve this aim: search engine optimisation, embedding metadata in the form of schema.org microdata, and sharing “paradata” information about how resources are used.

Search Engine Optimisation (SEO) is the process of improving the visibility of resources in search engine results in order to make the resource more discoverable. Discoverability also relates to the ability to find resources in appropriate places, for example, in curated collections, institutional repositories and through web services. In terms of open educational resources, SEO interventions can be made at the level of the individual OER, e.g. as described by projects such as OpenSpines and SCOOTER; or at the collection management level, e.g. Triton’s use of WordPress to optimise SEO.

While SEO focuses on human readable, textual descriptions of resources, presented in a structured or semi-structured format; an alternative approach to resource description is structured, machine-readable metadata. The two can be combined in approaches such as microformats, RDFa, and microdata which bridge the gap between human-oriented resource description and machine readable metadata. This paper will report on activities undertaken throughout the UKOER programmes to identify what metadata is really required for OERs, challenges in formalising metadata to describe educational characteristics of OERs, and efforts to address some of these issues through the Learning Resource Metadata Initiative (LRMI).

It has long been acknowledged that publisher-created resource descriptions and formal metadata records are not the only useful sources of information about learning resources, particularly OERs. Often more useful, contextually sensitive and extensive information can be created by users, both incidentally as they interact with resources, and through the conscious actions of reviewing, tagging, discussing and recommending OERs. “Paradata” offers a new approach to gathering, surfacing and sharing this information, which may offer potential solutions to some of the more intractable problems around describing the educational characteristics of resources. We will report briefly on the activities of the Learning Registry and other projects that are exploring the use of “paradata”.

We hope this paper will highlight the importance of effective resource description to the discoverability of OERs, explore innovative approaches to old problems and provide pointers to where future efforts might be directed to maximise the benefits of open educational resources.

Keywords

Resource description, resource discovery, search engine optimisation, SEO, metadata, LRMI, schema.org, microdata, paradata, Learning Registry, UKOER, Cetis

Introduction

The HEFCE UKOER programmes, which ran for three years from 2009 to 2012, funded a large number of projects focused on releasing open educational resources (OERs) and embedding open practice (<http://jisc.ac.uk/oer>). Over the course of the programme conservative estimates indicate that over 10,000 individual OERs were released and a significant number of institutions have embedded sustainable open practices as a result. The programmes were managed by the JISC and the HE Academy, and the Cetis Innovation Support Centre was tasked with providing strategic technical support and direction at both programme and project level. A recurring theme that emerged from the programmes, which is relevant to the release of OERs more generally, is the challenge of making resources easily discoverable. This paper focuses on three innovative approaches to resource description and discovery that address this challenge: search engine optimisation, embedding metadata in the form of schema.org microdata, and sharing “paradata” about how resources are used.

UKOER Approaches to Resource Description

At the outset of the UKOER programme, Cetis was commissioned to provide technical advice and guidance on approaches to resource description. Reflecting the innovative nature of the programme, Cetis recommended a new approach to metadata and resource description:

“Rather than mandating a formal application profile based on a single open standard we are instead identifying the type of information that projects must record for the resources they create without mandating how this should be done. Hopefully this will give projects considerably greater flexibility as to how they describe their resources and ultimately we hope that this will result in richer descriptions that are of value to end users” (Campbell, 2009)

The expectation was that projects would identify their own resource description requirements and those of their stakeholders and think through the issues they would need to address. It was hoped that by encouraging this methodology, native practices and authentic approaches to open educational resource description and discovery would be surfaced.

Although Cetis did not mandate the use of specific metadata schema and controlled vocabularies, a minimum set of information was mandated that projects should record to enable users to find their resources:

- The title of the resource.
- The author, owner or contributor of the resource.
- An indication of the date that the resource was created or published.
- The URL at which the resource could be found.
- Technical information such as format, file size, etc.
- The programme tag, ukoer.

Other information such as description, subject classification, keywords, tags, comments and the language of the resource were recommended as desirable. The Pilot Programme guidelines were modified only slightly for subsequent phases of the programme; licensing information was added to the mandatory list and technical information demoted to recommended.

Search Engine Optimisation

Search Engine Optimisation (SEO) is the process of improving the visibility of resources in search engine results in order to make them more visible and discoverable. Discoverability also relates to the ability to find resources in appropriate places other than search engines, e.g. curated collections, institutional repositories and through web services. In common with other types of web resources, most people will use a search engine to find OERs, therefore it is important to ensure they feature prominently in search engine results (OCLC, 2011).

A number of UKOER projects focused on SEO techniques to increase the discoverability of OERs; Multimedia Training Videos explored the potential of purchasing Google AdWords to boost search engine ranking; SCOOTER, and its follow-up HALS OER, investigated various aspects of OER discoverability including the use of social networks and SEO techniques; and Triton and Great Writers Inspire examined the use of WordPress as an easy way to optimise SEO.

Collection level SEO

In addition to more traditional learning resource repositories, such as ePrints, DSpace and EdShare, UKOER projects also explored the use of third party services and applications such as YouTube, Slideshare and WordPress. The Triton project specifically used WordPress to optimise SEO and generate and promote links to OER learning pathways in the domains of politics and international relations. The project highlighted the benefits of the platform, and particularly the utility of WordPress plugins and themes, to optimise the discoverability and search engine ranking of collections of OERs (Mansell, Lockley, & Robinson, 2011). One of the plug-ins used by Triton automatically generates encoded sitemaps that direct search engines to new and existing content so it can be indexed and ranked. While many repository platforms have the ability to generate sitemaps it is unclear how many institutional implementations routinely have this feature enabled, or how effective they are in directing search engine traffic to resources and resource collections.

Resource level SEO

The Triton project also produced guidance that highlighted the importance of considering the search terms people use and making sure these words appear in the title and body of individual resources (Robinson, 2011). This point was also covered by the SCOOTER project's "Guide to Search Engine Optimisation", (Rolfe and Griffin, 2011), which suggested:

"website description and website keywords need to reflect the relevant words".

Whilst placing search keywords in page headers was common practice at the time, subsequent changes to the way search engines interpret this data now means that this practice could potentially have negative effects on search engine ranking and optimisation (Sullivan, 2011). Although ranking algorithms are rarely transparent, it appears that under some circumstances some search engines, e.g. Bing, may now interpret header keywords as an indicator of spam and negatively rank pages accordingly. The HALS OER project that followed on from SCOOTER also acknowledged that:

"Another notable difference to our previous project SCOOTER was the increased rankings of news, social networking and personal profiles." (Rolfe, Fowler, & Williams, 2012)

Embedding metadata in the form of schema.org microdata

Recently there has been a general trend in SEO away from the use of metadata and keywords in hidden document headers and towards the use structured data markup in the body of resources. This is partly a response to changes in search engine ranking algorithms, which have attempted to prevent the manipulation of ranking results by the unscrupulous use of hidden metadata. The solution has been to move towards combining human-oriented resource descriptions with machine-readable metadata using approaches such as microformats, RDFa, and microdata.

A widely used example of RDFa can be found in the Creative Commons license chooser, a tool that generates HTML licence code that can be embedded in resources prior to distribution (<http://creativecommons.org/choose/>). In addition to human readable text and icons it also generates machine-readable markup that includes the `rel="license"` attribute shown in Figure 1.

Figure 1 Example of RDFa markup used in Creative Commons license

```
<a rel="license" href="http://creativecommons.org/licenses/by-sa/3.0/us/">  
Creative Commons Attribution-Share Alike 3.0 United States License</a>.
```

The inclusion of `rel="license"` allows search engines to identify that a resource might be released under a specific license, allowing this information to be used as a means to facet search results. One significant development that has adopted this approach is schema.org, a project involving Google, Yahoo, Yandex and MS Bing that aims to:

"... improve the web by creating a structured data markup schema supported by major search engines. On-page markup helps search engines understand the information on web pages and provide richer search results." (schema.org, 2013)

There are two aspects to schema.org; a syntax for encoding the markup; and a shared schema of item types and their properties. The Learning Resource Metadata Initiative (LRMI) is also working to extend the schema ontology so that selected educationally significant characteristics may be marked up. This does however raise the issues of which educationally significant properties of a resource should be described. As noted in the Learning Material Application Profile Scoping Study (Barker, 2008):

"metadata for education was one of the domains where the issues were least well articulated and where solutions were least well developed."

In other words, while there is a common assumption that it is useful to describing features such as the "educational resource type" or the "educational level" of a resource, there remains a gap between this desire and defining exactly what these terms mean and how these characteristics should be described.

As the schema.org and LRMI specifications were still under development during the UKOER programmes, few projects had the opportunity to implement them. However Great Writers Inspire did discover that they could not include schema.org in their resource description as a result of the way in which their resource management system, Drupal, operates (Mansell, Robinson, and Berglund Prytz, 2012). Throughout the programmes Cetis tracked the development of schema.org, contributed to the LRMI Technical Working Group and disseminated relevant developments to UKOER technical developers.

Paradata

The term “paradata” was first used by the NSDL in early 2010 to describe data about user interactions with learning. The term was proposed to:

“distinguish between traditional, relatively static metadata that describes a digital learning object and the dynamic information about digital learning objects that is generated as they are used, reused, adapted, contextualized, favorited, tweeted, retweeted, shared, and all the other social media style ways in which educational users interact with resources” (NSDL, 2012)

The concept of paradata was subsequently adopted by the Learning Registry, a US initiative initially funded by the Departments of Education and Defense, that aimed to create a technical infrastructure for “capturing, sharing, and analyzing learning resource data”. The Learning Registry project has created a network architecture composed of multiple nodes that can receive, share and distribute both metadata and paradata, regardless of schema.

On the basis of Cetis’ recommendations and engagement with the Learning Registry project, JISC commissioned Mimas to build a Learning Registry test node as part of the third phase of the UKOER programmes. The JLeRN Experiment successfully built a Learning Registry test node that was capable of ingesting data from a range of sources, including the Jorum national repository (see Campbell *et al*, 2013).

A number of UKOER projects explored using the JLeRN test node including ENGrich, visual media search facility for engineering education, SPAWS, sharing paradata across widget stores and RIDLR, which tested the release of contextually rich paradata to the JLeRN node and harvested back paradata. The SPAWS use case neatly illustrates the potential of the Learning Registry. The project connected several educational web app stores together and used the Learning Registry to syndicate information to all the stores on the network whenever a user downloaded, embedded or commented on a particular widget or app in any one of the stores (Wilson, 2012). While this information could have been shared directly between the online stores, using the Learning Registry provided a shared interface that also had the potential to enable the wider community to reuse the data.

Although paradata, and the technical approaches for sharing paradata, developed by the Learning Registry have aroused considerable interest in the UK F/HE community, these are still relatively experimental and immature technologies and it is debatable how much impact they will have in the immediate future. Whilst the Learning Registry offers a solution for sharing paradata, issues relating to the collection and analysis of this data are yet to be fully addressed.

Conclusion

This paper illustrates only some of the innovative approaches to resource description and discovery explored by the UKOER Programmes. A common theme connecting these approaches is the influence that search engines, like Google, have in making open educational resources discoverable. This places the onus on the individual and the institution to ensure the OERs they create are appropriately described and that metadata is optimised for search indexation. Whilst the impact of projects like the Learning Registry is unclear, the concept of using paradata to aid resource discovery is already being explored by Google and other search services and thus may have a considerable impact on how users discover open educational resources in the future.

References

- Barker, P. (2008). *Learning Material Application Profile Scoping Study – final report*. Edinburgh: Heriot-Watt University. Retrieved on 20th February 2013, from <http://www.icbl.hw.ac.uk/lmap/lmapscopingreport.pdf>
- Campbell, L. M. (2009, March 30). Metadata Guidelines for the OER Programme. Message posted to: <http://blogs.cetis.ac.uk/lmc/>
- Campbell, L.M., Barker, P., Currier, S., & Syrotiuk, N., (2013). *The Learning Registry: social networking for open educational resources?* In *Proceedings of OER13: Creating a Virtuous Circle*. Nottingham, England.
- Mansell, L., Lockley, P., & Robinson, P. (2011). *TRITON Final Project Report*. Oxford: University of Oxford. Retrieved on 20th February 2013, from http://openspires.oucs.ox.ac.uk/triton/resources/triton_final.pdf
- Mansell, L., Robinson, P., & Berglund Prytz, Y. (2012). *Great Writers Inspire Final Project Report*. Oxford: University of Oxford.
- NSDL Network. (2012). *Paradata | A place to build and strengthen your connections with the NSDL Community Network* Retrieved on 20th February 2013, from <http://nsdlnetwork.org/stemexchange/paradata>
- OCLC. (2011). *Perceptions of libraries, 2010 : context and community : a report to the OCLC membership*. Dublin, Ohio USA: OCLC.
- Robinson, P. (2011, February 4). OER Discoverability – Top Tips for Search Engine Optimisation (SEO). Message posted to OpenSpires: <http://blogs.oucs.ox.ac.uk/openspires/>
- Rolfe, V., & Griffin, S. (2011). *SCOOTER Project A guide to Search Engine Optimisation*. Leicester: De Montfort University. Retrieved on 20th February 2013, from http://www.sicklecellanaemia.org/OER/Resources/SCOOTER80b_SEO_Guidelines.doc
- Rolfe, V., Fowler, M., & Williams, J. (2012). *HALS OER Final Project Report*. Leicester: De Montfort University.
- schema.org. (2013). *Frequently asked questions*. Retrieved on 20th February 2013 from <http://schema.org/docs/faq.html#0>
- Sullivan, D. (2011, October 14). The Meta Keywords Tag Lives At Bing & Why Only Spammers Should Use It. Message posted to <http://searchengineland.com/>
- Wilson, S. (2012, October 19). Sharing usage data about web apps between widget stores. Message posted <http://scottbw.wordpress.com//>

License and Citation

This work is licensed under the Creative Commons Attribution Licence <http://creativecommons.org/licenses/by/3.0/>. Please cite this work as: Hawksey, M., Barker, P., and Campbell, L.M., (2013). *New Approaches to Describing and Discovering Open Educational Resources*. In *Proceedings of OER13: Creating a Virtuous Circle*. Nottingham, England.