

Catalogue and Index

Editorial

June 2012, Issue 167

Those of us who don't work in academic institutions might find repositories a bit of an unknown entity. In this issue we hope to solve your dilemma – we give you everything you always wanted to know about repositories but were afraid to ask. We start with articles by Gareth Johnson and Jackie Wickham that provide some background information, and follow up with more personal articles from a variety of practitioners that give a flavour of what it's like managing a repository, dealing with classification problems, and moving from cataloguing into repository management. We hope these will answer most of your questions. Finally, we include a review of a seminar and a book review.

Cathy Broad

(library@ethicalsoc.org.uk)

Heather Jardine

(heatherjardine402@hotmail.com)



Contents

2-4 UKCoRR – At the Heart of the UK Open Access Repository Landscape *by* Gareth J Johnson

5-8 The Repositories Support Project *by* Jackie Wickham

9-15 Managing a Creative Arts Institutional Repository *by* Anne Spalding

16-19 Repository metadata for diverse collections *by* Alex Clarke and Anna Lawson

20-23 Cataloguing, classification and the University of Northampton Institutional Repository *by* Janet Milner

24-27 Implementing a Current Research Information System (CRIS) with an existing Institutional Repository (IR) *by* Dominic Tate

28-32 Why not send a cataloguer? *by* Janet Aucock

33-34 Reviews

Repository Metadata for Diverse Collections

Alex Clarke and Anna Lawson, Library Services, University of the West of England

The UWE Research Repository was launched in March 2010 as a way to consolidate the University of the West of England's research outputs and, where possible, to make them publicly available. There are two members of staff who work on the repository, a Repository Manager, who is responsible for development of the repository, strategic planning and the training of users, and an Administrator who checks the records entered by users and maintains the data in the repository. Neither of us is an experienced cataloguer. However, a significant part of our job is cataloguing repository entries so we do have a basic awareness of metadata schemas.

As with a number of UK universities, the repository uses the EPrints software produced by the University of Southampton. UWE allows researchers to submit their own metadata records, with attached versions of the published output where possible, for inclusion in the repository. The metadata is then checked by the Administrator to ensure that it is accurate before being made live. UWE researchers work in a wide variety of disciplines and research output is in a range of formats, and this has made it particularly difficult to create a consistent style of cataloguing. This in turn can make it difficult for users searching the repository to find relevant items.

About seventy percent of traffic to the repository comes via Google. When Google searches the repository it looks at both the metadata record and any full text item attached to it. This means that repository items appearing in search results will usually be the result of terms found in the full text item, rather than in the metadata record. As such, it would appear at first glance that formal cataloguing of a repository entry isn't a particularly important part of this process. However, there are many items in the repository where the full text item doesn't lend itself to this kind of searching.

UWE's Faculty of Arts, Creative Industries and Education produces many items that defy full text searching because they are image-based, rather than text-based. *Hedgewitch* (Banks, 2008) (<http://eprints.uwe.ac.uk/15950/>), a short film, is one example, as Google would be unable to search the video file attached to the record. This could be remedied, at least in part, by providing a copy of the script as supplemental material. However, other video items in the repository defy this approach: for example, '*Computation by competing patterns in cellular automata, not-majority binary adder implementation*' (Martinez, 2010). The full text of this item is a series of tracks that are simultaneously shaded (Figure 1. <http://eprints.uwe.ac.uk/8364/>). Clearly, providing a transcript for an item such as this would be impossible.

Some items are even more difficult to classify. "idonthaveyourmarbles" is an e-bay account that sells a variety of eccentric items (Figure 2). It is officially listed in the repository as a Show/Exhibition. However, while the physical exhibition connected with this project had a specific date attached to it, the project continues and is

constantly updated with new items.

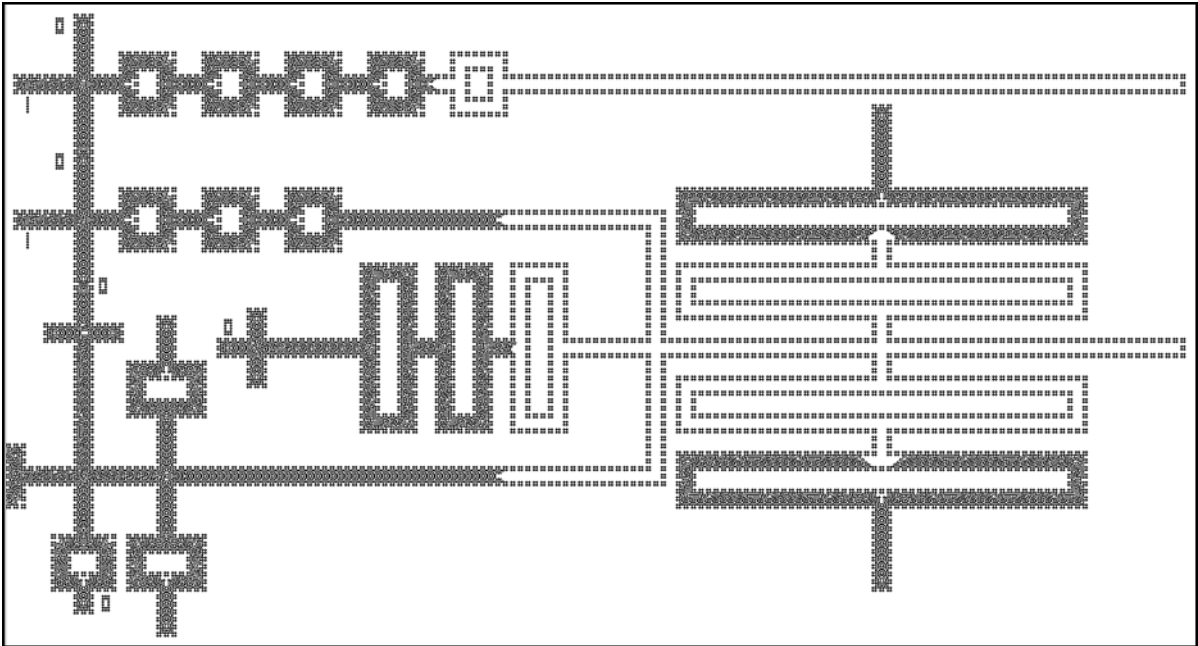


Figure 1

Creating metadata for such diverse material within a single environment presents a significant challenge. If we want our users to be able to search within the entirety of our repository's data and locate the full range of our resources from a single search, there needs to be at least some level of consistency between the records.

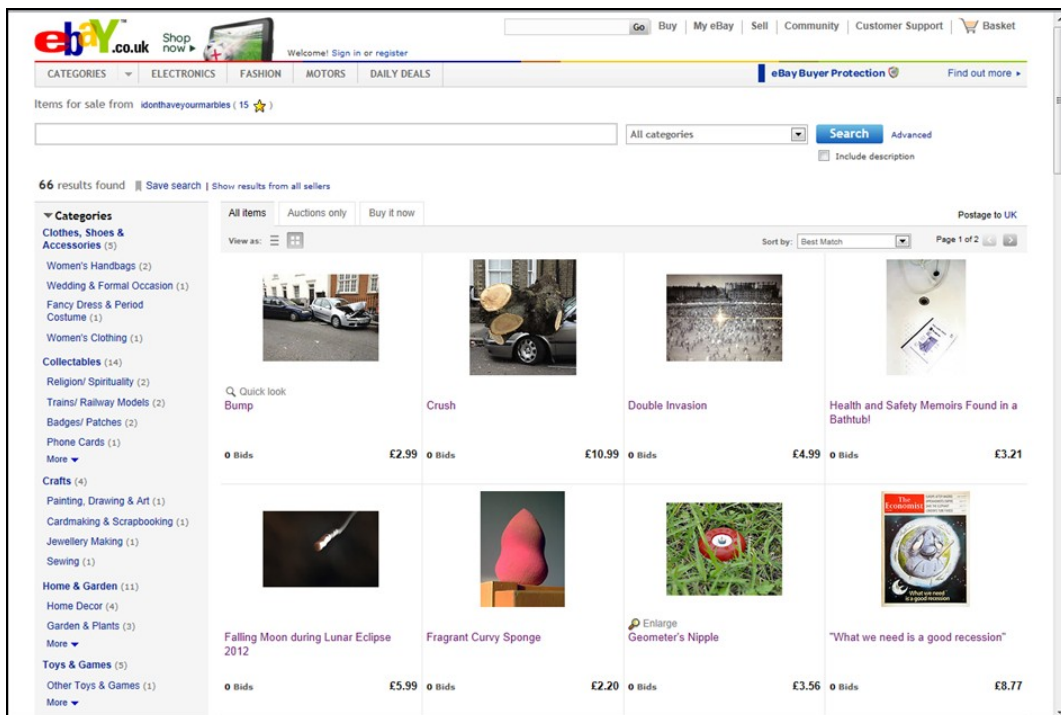


Figure 2

To achieve this consistency, the UWE Research Repository uses the Dublin Core (**Dubin Core Metadata Initiative, 2010a**) metadata schema. Dublin Core has the advantage of being flexible while providing a minimum metadata standard for all records within the repository, ensuring a basic level of consistency between records. Most EPrints based repositories use Dublin Core as this is the schema that the software uses out of the box. As such, these repositories effectively operate at Level 1 of the Dublin Core interoperability Levels; Shared Term Definitions. This is useful as it allows for sharing of data between these repositories. So if an academic moves from one institution to another then their data can be moved with them.

Of course, not all repository staff across institutions use Dublin Core in the same way. Some may, for example, record papers published in conference proceedings in a similar way to books, while others might record them alongside the details of the conference itself. On top of this, many repositories add or amend fields to record additional data. This is because the Dublin Core metadata schema currently in use in many repositories does not, on its own, provide enough fields to cater for the wide variety of resources they hold. This has resulted in most repository administrators creating additional fields and terms to support their entries. Universities have different needs to one another and some will need additional data whilst others will not. This can result in data being difficult to import if the receiving repository has no appropriate field to store it in.

There are a number of potential solutions to these problems. For art and design materials, there are a range of metadata schemas that may be appropriate. For example, the Categories for the Description of Works of Art (CDWA) (J.Paul Getty Trust 2010) defines categories for use in describing and accessing information about art, architecture or cultural objects, along with any related images (Grant & TASI, 2006).

Alternatively, a more appropriate metadata schema to use may be the one developed by the Kultur project (<http://kultur.eprints.org/>). The Kultur project was a JISC-funded project that ran from 2007 to 2009. It was set up to create a model institutional repository for use in the creative arts, and aimed to showcase outputs such as films, photography and paintings. In order to do this, the project established a metadata schema that was appropriate for arts outputs. Repositories that hold arts materials are generally aware of the Kultur project, and if they were all to adopt this metadata schema there would likely be more consistent metadata, and therefore greater potential for interoperability, across these repositories.

However, this only solves the problem for art and design metadata. For interdisciplinary repositories such as the UWE Research Repository, there needs to be a metadata schema that can cope with the wide range of materials the repository holds. It might therefore be worth considering a different approach. Adopting Level 2 of Dublin Core Interoperability; Formal Semantic Interoperability, (Dublin Core Metadata Initiative, 2010b) and utilising the more recent DCMI (ibid) metadata terms within repositories, rather than the original Dublin Core schema, would mean that every record would have a core set of metadata terms. This would ensure that

the range of terms that could be utilised, both for describing items and effective searching, would increase. These terms would need to be entered in a precise manner to describe the item. This would provide a basic standard between records across repositories and allow records to be imported and exported more smoothly. However, this approach may result in too restrictive a set of metadata terms, and once again result in repository administrators feeling the need to create additional fields and terms to support their entries.

One current development within the repository community is to use an application profile. Application profiles allow particular user communities or organisations to define a more specific metadata schema that is fit for a certain purpose. A repository community can define a metadata schema by selecting metadata terms from one or more existing metadata standards. Application profiles can have new elements added to them, but may not add terms to an existing standard. A number of application profiles have been developed by building on the Dublin Core standard, but they are all intended to support the description of a specific type of work. For example, there is an e-thesis application profile and an images application profile (JISC 2012).

It seems that when it comes to metadata schemas for repositories, there is no obvious one-size-fits-all approach. Dublin Core was originally adopted by repositories for its simplicity and flexibility, whilst still providing a minimum standard. The updated DCMI metadata terms may improve the way in which repository items can be catalogued, as they provide a richer use of metadata. However, there is no guarantee that even this schema will accommodate the range of entries in a truly interdisciplinary institutional repository.

References

- Banks, L (2008) *Hedgewitch* (2008) [video]. Directed by Liz Banks. <http://eprints.uwe.ac.uk/15950/> Accessed 11 June 2012
- Dublin Core Metadata Initiative (2010a) *Dublin Core Metadata Element Set, Version 1.1* [online] <http://dublincore.org/documents/dces/> Accessed 11 June 2012
- Dublin Core Metadata Initiative (2010b) *DCMI Metadata Terms* [online] <http://dublincore.org/documents/dcmi-terms/> Accessed 11 June 2012
- Grant, G & TASI (2006) *CDWA* [online] <http://standards-catalogue.ukoln.ac.uk/index/CDWA>. Accessed 29 May 2012
- JISC (2012). *JISC infoNet Application Profiles* [online] <http://www.jiscinfonet.ac.uk/infokits/repositories/technical-framework/application-profiles>. Accessed 29 May 2012
- J.Paul Getty Trust (2010) *Categories for the Description of Works of Art* [online] http://www.getty.edu/research/publications/electronic_publications/cdwa/categories.html Accessed 11 June 2012
- Martinez, G (2010) *Computation by competing patterns in cellular automata, not-majority binary adder implementation* (2010) [video]. Directed by Genaro J. Martinez <http://eprints.uwe.ac.uk/8364/> Accessed 11 June 2012.