Repository Repositioned: a new way of doing things as OER changes

Pat Lockley, No Affiliation
Patrick.lockley@googlemail.com

Abstract
Repositories remain an excellent tool with which to host content. However, as we move away from web 2.0 and into more flexible and distinct topologies for consumption and publishing how do we keep track of where open education resources are published, and if we cannot do this, how can we find them and encourage people to use them? Using a variety of harvesting protocols, we have attempted to address this problem, as well as start to look at how the broadcasting and presentational elements of repositories can be increased.

Keywords
Repositories, web 2.0, downloading, curation, access, translation, network, protocol, RSS, twitter, facebook

Introduction
OER and OCW repositories have traditionally been derivations from institutional repositories whose primary intention is to host published journal articles, examples being dSpace (2014) and eprints (2014). Amongst the members of the OCW Consortium, Content Management Systems (CMS), Plone and EduCommons are other popular alternatives, but remain as such within the field of repositories. Millard (2008) describes Flickr and Youtube as repositories, but suggests that they have interfaces sufficiently different from OER repositories such that OER repositories should borrow from their approaches. As such, the majority of repositories remain depositories, do repositories lag behind the curve of web development?

Even with OER repositories as they are, does their basic functionality technical support indexing accurately by search engines? OCW’s own search tool (OCW, 2014:courses) indexes only 77 of the 239 member organisations (which assumes content is available from each organisation). Do we also assume search engines (Yahoo, Bing, Google) to be the primary driver of traffic to OER, or does social media and the concept of virality now play a part? Given how the research on OER discovery (Dichev, 2012) discovers a rapid window during which OER is searched for – could recommender systems or social media sharing work within this approach? Speed of resource discovery seems paramount, and having access via social media would appear to suit this requirement. McGreal (2010) suggests OER repositories can be catalogued as to whether they are pure OER repositories, or mixed – and again this raises questions over where best to place OER.

If OER is to be used – does placing it solely into a institutional repository make sense. Most OER repositories are based at either a distinct University, or at an organisation such as OER Commons. Some national repositories exist, such as Jorum or Japanese OCW. Many organisations outside of education post their content into other platforms (which may be covered by Millard (2008) but would not be primarily classified as an OER repository, such as YouTube.
or FlickR. Many National libraries deposit to FlickR Commons, but not into associated national or organisational repositories. As such as OER is no longer located in institutional repositories and isn’t a question of OAI-PMH synchronisation or basic indexing, or at least basic indexing of a site the size of FlickR is going to be beyond to the scope of a tool only indexing OER.

Once we accept that we move from repositories with established metadata profiles into the relatively unstructured, unsemantic web, how do we address different metadata profiles and limited metadata quantities. Metadata profiles between different repositories also remains a problem. How can we best allow for differences in profile - say “DC:FORMAT” - and allow for options such as filtering or faceted search? Should we allow for people to submit their own metadata, or allow anyone using the site to contribute using a crowd-sourced approach? How could we allow for this data to represented and what caveats would we need to employ?

Paradata, a concept often tied to the American “Learning Registry” (2014) system, involves plotting metadata regarding how a resource is used. Many sites provide this information in some form, but also it can be extracted from google analytics, LMS / VLEs or other sources. The presence of the data, and how anonymous that data can and should be doesn’t cover how it should he used in search weighting. Is there a case for OER to move towards a more Google like approach.

Around this, various definitions of openness have become expressed via national and not international policies. Varying national or individual definitions of openness manifest themselves in the structure and policies of the repository themselves. For example, the major US repositories are shaped by the copyright defence of “Fair Use”, which is not a concept which can be necessarily applied in other legal systems. There is little research into how different copyright jurisdictions alter and change OER usage. As such in using solely public domain / creative commons licensed content this legal issue is removed.

All of these questions also point towards topological issues around the creation and management of repositories. Repositories are ran by a handful of people, even when analogous technologies such as social bookmarking exist. As such repository functionality is written by the few for the many, and a centralisation of both responsibility and decision making seems counter-intuitive to the Open movement as a whole. Research by Atenas (2014) shows how repositories remain relatively hard to use, and so allowing more – almost WordPress like - flexibility in terms of customising display and usage.

Solvonauts (solvonauts.org) is a new type of repository, both in terms of code and of culture. At present we are very much in development, and we aim to always be in development. Many of the questions posed above are designed to be part of the journey forward, or part of the options which people could configure each repository with. Using an open approach (open source, open data, open education) the repository is built to be reused and repurposed by others, both as a service in itself but also as a site which others could use to build their own versions of it. As such the repository and the underlying code is closer to the spirit of openness.

The architecture of the code is also built to be easily extended and developed. Using a plugin based architecture, the code base is easily extended and fully internationalised. Beyond this it is
being developed to support further the scope for result manipulation and indexing. Scope exists to start cross cataloging content between multiple languages. With approaches like this we hope to allow for increased discovery and usage - again principles and goals which we feel are an explicit part of the OER movement.

We’ve also looked to position OER repositories into the wider web. We can harvest OER from Tumblr, Youtube, Slideshare and using RSS, Atom, OAI-PMH and open data sources. We also provide a series of wider tools for integrating into VLEs, as well as to bring usage data (the aforementioned “paradata”) back to the system. We also have a chrome extension which allows people to check web pages to see which OERs are linked to from that page. Again, this is part of our process in bringing OER infrastructure into the topology of the modern web.


License and Citation
This work is licensed as public domain