Repositories, content mining and data deposit

Publishing in the Open Access Age,

Society of Biology,

26th October 2012

Yvonne Budden,
E-Repositories Manager, University of Warwick
Chair, UK Council of Research Repositories (UKCoRR)

Talk Summary (1 of 2)

- Yvonne explains the work of UKCoRR – a body that came into JISC funding provided to a number of institutions to get their repositories off the ground. We have a vision of making repository curation a unique discipline within librarianship and other UKCoRR membership bodies (e.g. IT services, research support services etc.). All members must work with repositories, as there are unique challenges for those who are managing and working with repositories. UKCoRR is an entirely independent unfunded body; we don't ask for or receive any subscriptions from our members. We are therefore free to form our own policies and guide our own members, yet this lack of funding does limit our activities. Membership is on an individual, not institutional basis.

- UKCoRR is open to professional and para-professional staff, so we are not only considering the high level guides at management level, but also how to best encourage researchers to use the repository, how can we best gather the data, persuade people that OA is a good idea and worth them taking the time.
- We engage with our membership through our very active mailing list which operates under Chatham House rules, creating a space for discussion and promoting best practice. In most institutions the repository staff is often a single person, often doing the job as part of other work, so the UKCoRR list and membership is the only way for them to get a view of what their colleagues are doing.
- We have a small committee who take a strong view that our remit comes from our members, and does not do any kind of policy setting without first consulting the membership. Some of the big activities involving the membership at the moment are the Finch report, new RCUK policy, WT policy and the growing area of data and research management – which is often assigned to repository staff or contributed to by them.

Talk Summary (2 of 2)

- There are different types of repositories; format specific repositories such as those for Open Educational Resources, subject specific repositories, institutional repositories, where the majority of our members come from, and repositories like PubMed Central which also act as funder's repositories. Repositories tend to operate slightly differently from each other in some way, and all have policies governing the use of papers and metadata that they hold – which is important to note for anyone who is looking to do any text or content mining.

- Each of the repositories are designed to meet the needs of the communities that they serve, and so each face different challenges. For example repositories of creative arts materials, specific subject areas, or those who cover a wide range of different communities under one repository - this requires a clear policy on addressing the needs of specific disciplines.
- Primarily the role of UKCoRR is to provide support for the researchers who use their repository, including OA and copyright policies enabling people to self-archive. We also support researchers in fulfilling funders' requirements and act as a point of expertise for help with interpreting recent OA discussions and reports.
- We spend a lot of time reviewing publishers' policies on self- archiving, to make sure authors are remaining within these terms, while doing what they can to fulfil the needs of the funders. We have found that these policies are difficult to understand if you have no prior knowledge of this type of language. We are trying to encourage publishers to standardise this language, so that it's clear to authors which version of the paper can be archived. There are also lots of different metadata standards, which can cause problems with aggregation of content for text mining or interoperability between services.
- Repositories look to be key to work using data catalogues, and repository staff are investigating publishing data papers and other areas to ensure researchers get credit for making their data available and transparent. This will require using the specialist skills of researchers in different disciplines and the development of training according to the extent to which researchers are familiar with repositories.

UKCoRR : Who are we?

"UKCoRR has a vision of the work of repository management as a professionally recognised and supported role within UK research institutions."

- More than 250 members from 121 institutions
- Independent, unfunded body for repository workers
- Membership is individual not institutional

'Living in interesting times'

- Open to both professional and para-professional staff
- Provides a safe forum for issues and promotes sharing of best practice
- Small committee guided by the membership
  - Finch report and RCUK Policy
  - Research data management
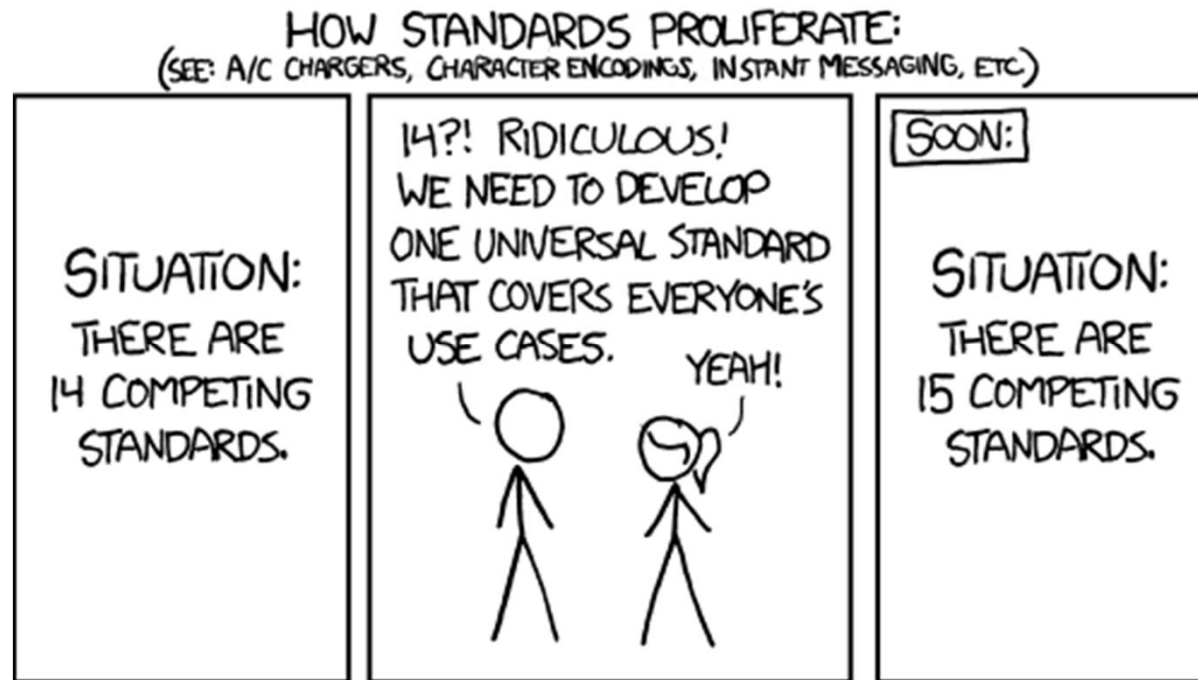
OA and Repositories

- Two main flavours of repository:
  - Institutional, e.g. WRAP, Warwick's repository
  - Subject based, e.g. UK Pub Med Central
- Archive a range of outputs depending on collection policies and institutional need
- All have policies governing reuse both of the papers in the repository and the metadata

Infinite Diversity in Infinite Combinations

UKCORR

- Designed to serve the needs of a community or institution

- Each face unique challenges

- Provide support for researchers:
  - Checking publisher's copyright policies
  - Advising researchers on funder's requirements

- A hub for institutional practice on open access

Speaking the same language?

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.   YEAH!

SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

Randall Munroe, CC-BY-NC (http://xkcd.com/927/)

Data, data, everywhere?

- RCUK drivers for a range of new activities around research data management
- Data Archive vs Data Catalogue
- Place of repositories in this landscape?
- Wide variation in each discipline's level of maturity on this topic
- Further need for specialist skills and training

Contact details

- UKCoRR
  - http://ukcorr.org
  - Blog, details of meetings and committee activity and contact details

- Yvonne Budden/WRAP
  - http://go.warwick.ac.uk/lib-publications
  - y.c.budden@warwick.ac.uk