City Research Online

City, University of London Institutional Repository

Citation: Stewart, N. (2012). Coming late to the game: how to create a totally integrated (!) repository system. Paper presented at the Open Repositories 2012, 9 - 13 Jul 2012, Edinburgh, UK.

This is the draft version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/1205/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

CITY UNIVERSITY LONDON

Library Services

Coming late to the game: how to create a totally integrated (!) repository system

Neil Stewart, City University London, March 2012

Abstract

This paper recounts the experiences and lessons learned in developing a repository system from scratch, using Eprints and Symplectic Elements, at City University London. The system is called City Research Online (CRO). Contrary to the title, the system created is not "totally integrated" (in fact far from it!). However the author was able to draw upon experiences learned in his previous role as Repository Manager at LSE, as well as those from the broader repository community, to implement automation of processes and systems integration. These lessons were implemented in light of the fact that City was relatively late in setting up its repository, and benefited from lessons learnt by other institutions. The paper details the set-up at City; explains how decisions made on the set-up were informed by past lessons; and examines the level of integration judged by a number of different criteria.

CRO's set-up

The project to implement an institutional repository at City was begun in March 2011. The author began in post as Manager in June 2011. Soon after this, dual implementation of an Eprints open access full text repository and a Symplectic Elements current research information system (CRIS) took place. While this technical implementation was ongoing, various policy decisions were taken regarding the scope of the repository's content, and how the CRIS and the repository would interact with one another. On the basis of the technical implementation and the policy decisions taken, the repository was rolled out to staff in September and October 2011. Development since then has been ongoing and iterative, and has included:

- Establishing a presence on City's main website, as well as on the sites for the dedicated services.
- Using data held in the open access repository to feed various other web services.
- Using the CRIS' API to output bibliographic data, allowing search functionality via City's corporate web presence.
- Developing use of the CRIS to support REF 2014 submissions.
- Developing support for storing and serving City's electronic PhD theses.

Lessons applied & learnt

In implementing the repository system at City, a number of lessons were applied and learnt. Since both sets are now after the fact, it makes sense to consider them together. They were as follows:

- Automating metadata transfer is good: it saves on cataloguing time, and means you can concentrate on doing more interesting things.
- Using hosted services is also (broadly speaking) good, since they tend to be secure and stable, and allow for nimble development.
- Having two systems doing similar things can be difficult:

    - Managing the integration of the two can be a real headache.
    - Managing the tripartite relationship between the two system providers and ourselves is not always easy.
    - Explaining the way in which systems interact to users can be difficult.
- But it also has its advantages:
    - Allows for the automated metadata transfer mentioned above
    - Different systems can be used for different purposes, and there can be a clear differentiation between these systems and purpose(s).
    - As a result of this differentiation, separate development of the two systems can be engaged upon.
    - Though Eprints now has CRIS-like functionality, Symplectic Elements has been designed from the ground up as a CRIS.
- Branding is important. We quickly learnt that our users get confused when we used brand names- using the service's name in a thoroughgoing way is very important. Only refer to software names in the context of technical explanations!

Are we integrated?

The short answer to this question is: somewhat. By the nature of the systems we chose to use, there has been a certain level of "integration", but we're not where we'd like to be. There are three ways in which we can continue to improve our level of integration:

- **"Integration" as a matter of systems development**. We feel we have gone a certain way to achieving this. The real killer app for the service would be a project to create consistent, University-wide staff profiles. We already have experience in outputting publications data via Symplectic's API, and to do so for staff profiles would be relatively easy. Adoption of this project is something that needs to be done at a high level, however.
- **"Integration" as a matter of cultural change**. "Integration" doesn't just mean making your systems talk to one another; it also means getting your systems embedded in the workflow of the rest of the university. The story here is a patchy one- those Schools using CRO for REF 2014 have understandably been keen to engage. Those disciplines with a strong tradition of open access pre-print research distribution have also been receptive. Other parts of the University have been harder to reach.
- **"Integration" as a matter of policy**. Around the time of CRO's launch we lobbied for a mandate to be passed by the University's senior management team; this didn't happen, but it got CRO on the policy agenda of senior managers. We will try again on this- and are conscious that language is important here ("mandate" v. "publications policy").

Conclusion

Implementing the repository late has allowed us to integrate things to a certain extent- but there's always more to do, and integration means more than "making systems talk to one another". Repositories are socio-technical phenomena (pace Les Carr), and while we have done a certain amount of technical integration, the socio- side of things is one we need to continue to work on.