Clever Crosswalking - what do you take from one system to another?

Yanan Zhao[1], Kim Shepherd[1], Leonie Hayes[1] & Andrea Schweer[2]
[1]The University of Auckland, New Zealand, y.zhao@auckland.ac.nz
[2]Library Consortium of New Zealand (LCoNZ), schweer@waikato.ac.nz

Our story started here …

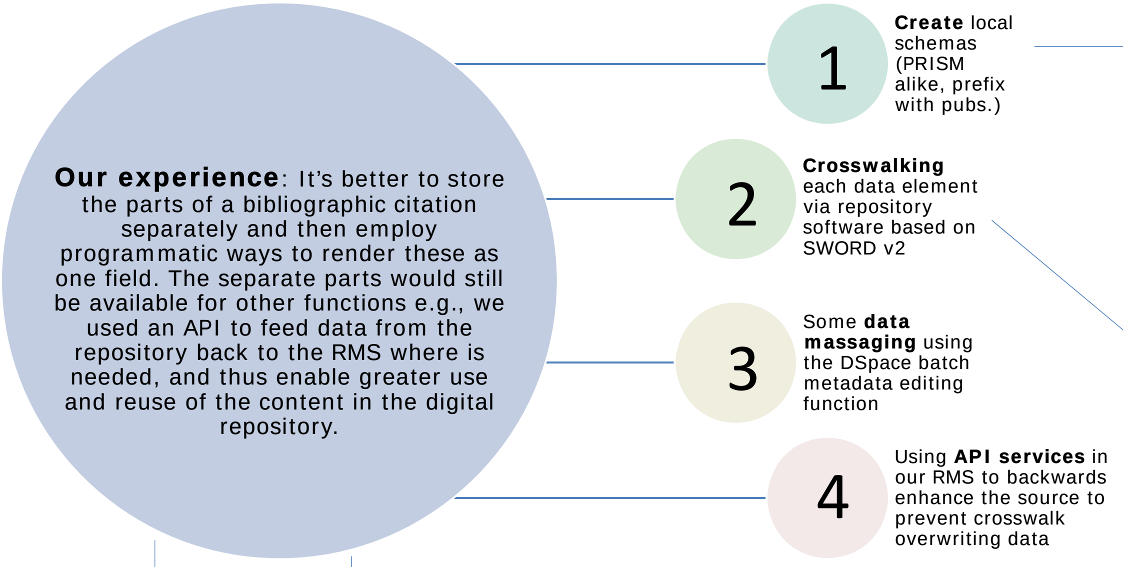
We needed to provide a solution for a complex project that involved melding the existing repository content which contains mainly theses content to becoming the institutional collection for research outputs.

- We had a schema that was based on Dublin Core which works well for simple and generic resource description; but it isn't sufficient for describing publications
- We had well defined workflows in the repository, but we needed to make the Research Management System (RMS) the primary place for data input and then push content from the RMS to our repository

This presented a number of questions:

- How could we take a very granular level of metadata from our RMS and expose it in our repository which uses less granular level of metadata schema 1) without losing anything and 2) with the least amount of work, highest amount of accuracy and facilitate the maximum reuse possible?
- Do we put the publication citation data in one single field in the repository? Or do we store the parts of a bibliographic citation separately, and how?

Our solutions



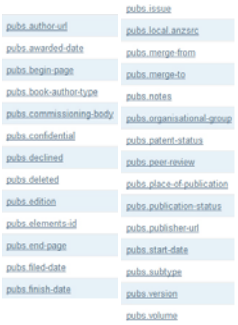
Our experience: It's better to store the parts of a bibliographic citation separately and then employ programmatic ways to render these as one field. The separate parts would still be available for other functions e.g., we used an API to feed data from the repository back to the RMS where is needed, and thus enable greater use and reuse of the content in the digital repository.

1 **Create** local schemas (PRISM alike, prefix with pubs.)

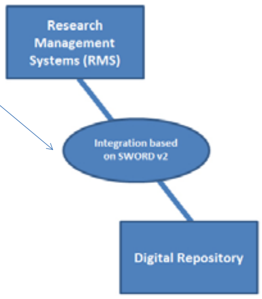
2 **Crosswalking** each data element via repository software based on SWORD v2

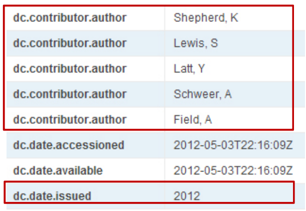
3 Some **data massaging** using the DSpace batch metadata editing function

4 Using **API services** in our RMS to backwards enhance the source to prevent crosswalk overwriting data

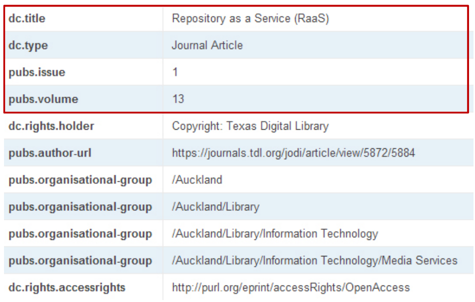


	pubs.issue
pubs.author-url	pubs.local.anzsrc
pubs.awarded-date	pubs.merge-from
pubs.begin-page	pubs.merge-to
pubs.book-author-type	pubs.notes
pubs.commissioning-body	pubs.organisational-group
pubs.confidential	pubs.patent-status
pubs.declined	pubs.peer-review
pubs.deleted	pubs.place-of-publication
pubs.edition	pubs.publication-status
pubs.elements-id	pubs.publisher-url
pubs.end-page	pubs.start-date
pubs.filed-date	pubs.subtype
pubs.finish-date	pubs.version
	pubs.volume

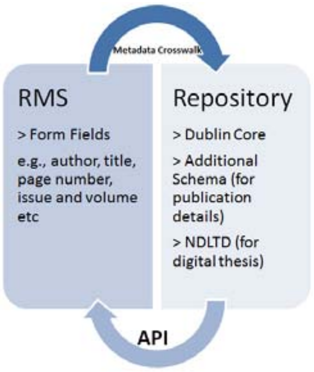

Research Management Systems (RMS)
Integration based on SWORD v2
Digital Repository



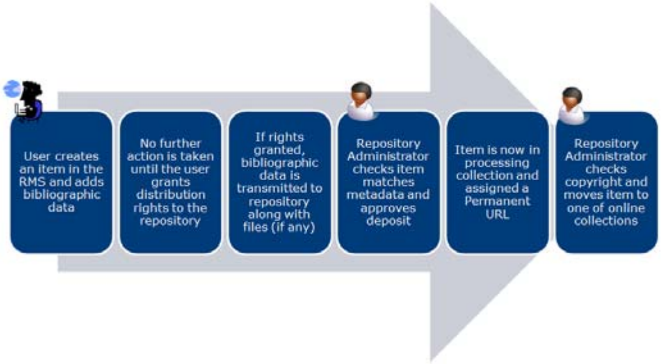
dc.contributor.author	Shepherd, K
dc.contributor.author	Lewis, S
dc.contributor.author	Latt, Y
dc.contributor.author	Schweer, A
dc.contributor.author	Field, A
dc.date.accessioned	2012-05-03T22:16:09Z
dc.date.available	2012-05-03T22:16:09Z
dc.date.issued	2012



dc.title	Repository as a Service (RaaS)
dc.type	Journal Article
pubs.issue	1
pubs.volume	13
dc.rights.holder	Copyright: Texas Digital Library
pubs.author-url	https://journals.tdl.org/jodi/article/view/5872/5884
pubs.organisational-group	/Auckland
pubs.organisational-group	/Auckland/Library
pubs.organisational-group	/Auckland/Library/Information Technology
pubs.organisational-group	/Auckland/Library/Information Technology/Media Services
dc.rights.accessrights	http://purl.org/eprint/accessRights/OpenAccess


Metadata Crosswalk
RMS
> Form Fields
e.g., author, title, page number, issue and volume etc
Repository
> Dublin Core
> Additional Schema (for publication details)
> NDLTD (for digital thesis)
API

Workflow: A simple deposit process @ UoA



1. User creates an item in the RMS and adds bibliographic data
2. No further action is taken until the user grants distribution rights to the repository
3. If rights granted, bibliographic data is transmitted to repository along with files (if any)
4. Repository Administrator checks item matches metadata and approves deposit
5. Item is now in processing collection and assigned a Permanent URL
6. Repository Administrator checks copyright and moves item to one of online collections

Our story continues …

- Why do we need citation information in the repository items' metadata? Is it about locating the item/article? OR is it about knowing how to cite it?
- Flat metadata schemas – is it the easiest way to crosswalk?
- Complex relational models are in most cases too challenging to implement for re-use, and how do you crosswalk non metadata elements?

We would love to hear what you think of these questions; contact us at y.zhao@auckland.ac.nz or schweer@waikato.ac.nz