Joe Cera, Berkeley Law Library

Building wobbly bridges

This project attempts to start addressing the lack of connection between repositories. While proper connections are far down the road, the first step is to be prepared for connections. This session will discuss how Berkeley Law is preparing data through persistent identifiers (author and article) with the hope of reaching out to other institutions to build some wobbly bridges.

How are we preparing for connections?

DOIs assigned to every journal article.

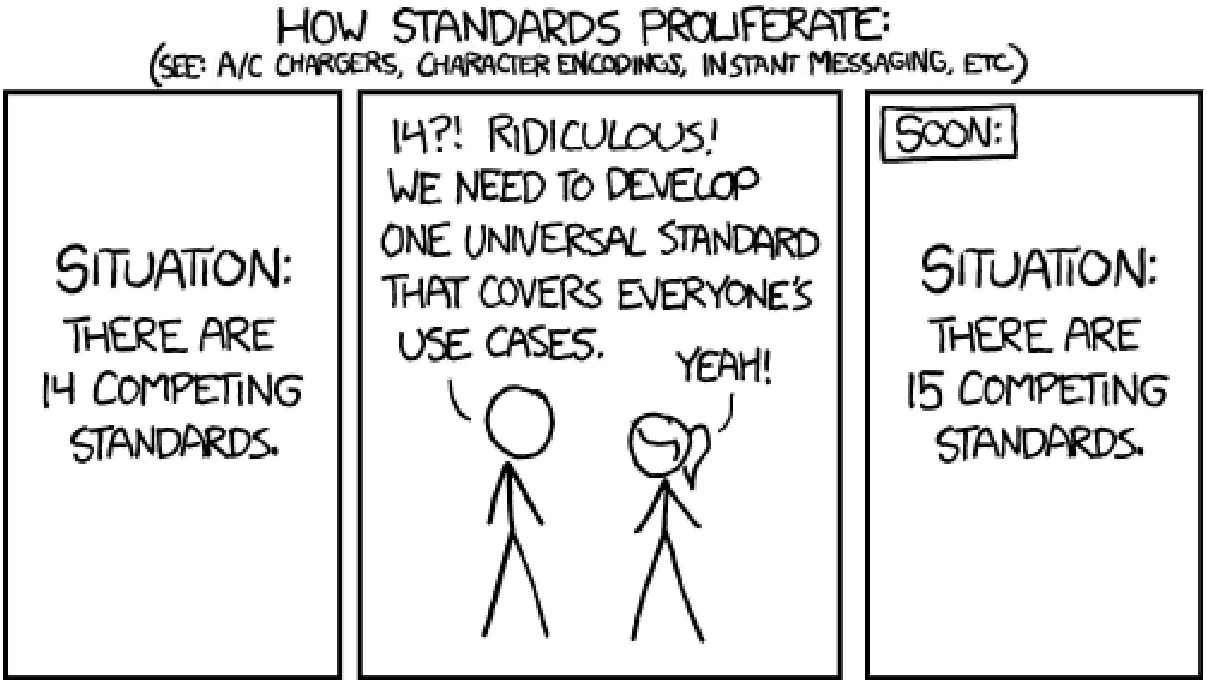
This allows us to have links to articles that are independent of location and platform. We were doing this a couple years before migrating to the Tind platform and it highlighted the importance of having them.

- The DOIs were easy to update and, if institutions and individuals were in the habit of linking using the DOI, our migration would have had no effect on the ability to link to an article.
- We embedded this process in the macro process for each journal so a DOI would normally get added, even if the journal editors weren't sure about the purpose of this.
- This also helps provide access to the article, as published, if there are any changes after a copy was downloaded (and possibly uploaded to another repository).
    - This means that any other repository only needs to include the DOI in the metadata to link back to the published version while still holding one of many full text copies.

As part of the UC system, we have access to the EZID system created by the California Digital Library. This helps us mint new DOIs and we have created a script that only requires a journal editor to follow a link to get a fresh DOI.

The Tind platform also has this capability built in and DOIs can be minted upon record creation but we chose to stick with the workflow that was already working without any problems.

Joe Cera, Berkeley Law Library



https://xkcd.com/927

Identifier(s) for each faculty member.

We chose to use Wikidata as our central identifier and requires us to create or update a Wikidata identifier for all faculty members.

- This was initially done using OpenRefine and then ad hoc for individuals

  *Note: This makes for some potential to create the very errors we were trying to prevent by associating the wrong person with the wrong identifier.

- Added LoC/VIAF/ISNI/ORCID to Wikidata
- Added Wikidata, LoC, and ORCID to IR entries
  - LoC because it was important to cataloging and allowed connection with the ILS (ILS has LoC identifiers only)
  - ORCID because it exists

Why was Wikidata chosen? Because I wanted to play with OpenRefine, it’s open for anyone to edit, LoC grabs data from it for LC name records, and – most importantly – because it could accommodate any other identifier

Does it matter? No, not really. Any identifier would serve all of the same purposes but a public identifier would be best so others can look it up.

Joe Cera, Berkeley Law Library

Wikidata serves as a hook for many other projects:

- We built our own faculty profile pages based on IR API that uses Wikidata ID as the item searched to return all faculty publications in the repository
  - https://www.law.berkeley.edu/library/ir/faculty/
- Annual faculty publications list by Communications is based on a Wikidata ID query of the repository through the API
- Gathering publications for ORCID uploads (Bonus – IR exports as BibTeX)
  - This allows us to pull all the publications by an author, export it, upload to ORCID, and populate the IR with the ORCID identifier
- Quickly added any identifiers for all (ORCID, LoC, etc)
  - If someone decides they really want an identifier included or if we decide that we would like another identifier in the record, it is quick to add it.

Joe Cera, Berkeley Law Library

Planned project:

1. Find some other repository manager who is interested in this or already working on this.
    a) If there are none, start with the other UC law schools.
2. Find all journal publications from that institution that are already in our repository
    a) Create a spreadsheet of this data (citation to items with link if we have it, authors, author identifiers)
3. Ask for the affiliation and identifier to be added for Berkeley Law authors, if it can be accommodated
    a) This would be something like “Article 1 has Berkeley Law Author A already listed. Could you make a note that this person is a Berkeley Law author and add Wikidata identifier Q to the record, if possible?”
4. Ask for the direct link (or even better, a DOI) and any other associated metadata to augment our own records
    a) This way our record will link to the original publication in addition to providing a full text article
    b) Potentially both records will improve and both repositories will have a higher quality of metadata
    c) If there are additional authors, external to both institutions, there is an opportunity to improve additional repositories who likely hold the same file (loop back to step 1)
5. Offer to provide the same services for any items that are published by Berkeley Law journals

This is more or less a trade of data both repositories already have which should make the data better in both repositories. There are definitely more efficient ways to do this but it would take a greater amount of coordination and collaboration (plus time).

Things that I am aware of but am not currently engaging:

1. When new faculty join a new institution, we claim previous works when they were not at our institution. How do we deal with this?
2. When new faculty leave our institution (or join our institution) it would be great to be able to exchange a spreadsheet with the other repository manager to make adding new works easier.