

2013-05-17

ORCID Task Force Report

Dennis, Scott; Downing, Karen; Hollar, Susan; Kropf, Eryn;
Lalwani, Leena; Mueller, Ellen; Nason-Davis, Jennifer; Nicholls,
Natsuko; Ottaviani, Jim; Rodgers, Emily; Rosenzweig, Merle;
Tschirhart, Lori; Yokota-Carter, Keiko

<https://hdl.handle.net/2027.42/98975>

<http://creativecommons.org/licenses/by/3.0/us/>

ORCID Task Force Report

17 May 2013

ORCID Task Force Members:

Scott Dennis
Karen Downing
Susan Hollar
Eryn Kropf
Leena Lalwani
Ellen Mueller
Jennifer Nason-Davis
Natsuko Nicholls
Jim Ottaviani (Chair)
Emily Puckett-Rogers
Merle Rosenzweig
Lori Tschirhart
Keiko Yokota-Carter

Email: orcid-taskforce@umich.edu

Executive Summary

Open Research Contributor ID ([ORCID](#)) is an open, non-profit, community-based effort to create and maintain a registry of researcher identifiers, and to solve the name ambiguity problems that have hindered the development of interoperable scholarly tools and networks. By creating a registry of unique identifiers for individual researchers, it allows a transparent and reliable method of linking research activities and outputs to these identifiers, thereby enabling not only more accurate attribution in traditional research articles, but also attribution in the new forms of scholarly communication that are arising.

In December, 2012, MLibrary signed a (Basic) Member License Agreement with ORCID which allows us to, among other things, create ORCID records, update existing ORCID records, and use various APIs and the data those APIs can access to perform tasks associated with ORCID iDs.

ORCID is not currently a mature, commercial product. Commercial publisher buy-in is strong — Elsevier Scopus, Nature Publishing Group, and Thomson Reuters are among the launch partners, and Cambridge University Press, Springer, and Wiley are among the sponsors — but right now ORCID is a nascent, open source project.

Regardless, we recommend that U-M, and specifically MLibrary, facilitate and encourage the use of ORCID iDs. Doing so will position us as an early adopter, and we will likely have to commit resources to make things work for us (and by extension, others). The resources we'll need to devote to this are unknown, but early adoption and

active participation in ORCID's development can provide us with opportunities to shape it in ways that will more directly benefit us.

Recommendations

1. The Library should assign ORCID iDs (iDs) to U-M research faculty and staff. We should do this sooner rather than later. *The longer we wait, the greater the chance that we will find ourselves attempting to assign iDs to people who already created one at a previous institution.*
2. The Library should incorporate iDs into Deep Blue for current authors. *We discussed other Library systems as potential venues for iDs, but Deep Blue is the only system where (a) we control the metadata, and (b) records and items are by definition certain to have at least one U-M author. Not all items will have a current U-M author, of course, and at this time we don't recommend assigning iDs to authors who are no longer affiliated with U-M.*
3. The Library should actively partner with the Office of Research and Sponsored Projects (ORSP) and the Medical School Curriculum Vitae system (MCV) to incorporate iDs into their systems. Doing this will speed adoption and use, and prepare the way for using iDs in other systems. *Ideally, we want iDs to be incorporated into and travel along with MCommunity records, so our researchers need only remember their username; this may not be possible in the short term.*
4. U-M should promote iDs as an author disambiguation tool. *The temptation will be to talk about iDs as being useful for creating researcher profiles; iDs' administrative benefits are also clear. However, since there are already commercial tools that do this (and since there have been many efforts at creating research profiles on campus, many of which have met with limited success and adoption rates so far) we should stick with promoting what iDs do best.*
5. U-M should not encourage use of orcid.org's profile feature, at least at this time. *Automatic population of profiles with publications in orcid.org is incomplete, manual population is tedious and slow, and the institutional affiliations module is not in place yet.*

Given the combination of unknown costs and near-term benefits, we should position and frame our participation in ORCID in the same way we positioned initiatives like HathiTrust, Deep Blue, DLXS, and others: as a way to get out ahead of our researchers and meet their future needs. (E.g., we build repositories in advance of funding agencies requiring that such infrastructure exist to support open access mandates.) All of the above implies a final recommendation:

6. The Library should charge an ORCID implementation task force to move forward with the recommendations it decides to commit resources to.

Discussion

The ORCID Task Force was charged with the following:

- Become a team of ORCID experts
Identify and prioritize library systems that are good candidates for ORCID implementation
- Establish workflows and processes that will propagate ORCID
Identify and prioritize administrative systems that are good candidates for ORCID implementation
- Make recommendations for how the Library can engage stakeholders who manage various administrative systems (e.g. Faculty onboarding systems)
- Document and assess the Library's use of ORCID

We have developed some expertise in ORCID's potential and capabilities, and have identified and prioritized systems that are good candidates for ORCID implementation. Our recommendations reflect this first stage in our investigation.

Workflows, stakeholder engagement, documentation, and assessment of our use of ORCID will follow as the Library acts on these initial recommendations.

Note that the discussions we held and the recommendations we make are based both on what ORCID is and does now, and what it's anticipated to do in the future. During the duration of this Task Force, we have already seen rapid change in functionality of the ORCID registry. Currently, as ORCID members we have the ability to

- create ORCID records on behalf of employees or affiliates
- integrate authenticated ORCID IDs into manuscript submission systems
- integrate authenticated ORCID IDs into grant application processes
- ingest ORCID data to maintain internal systems such as institutional repositories
- link ORCID identifiers to other IDs and registry systems
- support the mission of ORCID and maintenance of the Registry

As pointed out in the "Summary and Recommendations", ORCID is currently an open-source project, not a commercial product, so things will change as we and other institutions contribute time and resources to fleshing out ORCID.

1. *The Library should assign ORCID IDs (iDs) to our research faculty and staff. We should do this sooner rather than later.*

Having purchased a membership in ORCID for a term extending through 31 December 2015, we have committed to the project. As an ORCID member, we have the ability to create iDs for U-M researchers proactively, and then promote their use. We could take a more passive route, and simply encourage U-M authors to create their own iDs, but we recommend taking a proactive role here.

The Task Force arrived at the conclusion to proactively create iDs quickly; while we discussed the pros and cons of doing so, that discussion didn't take long, and we were unanimous in recommending it. We then discussed the implications of doing so, and noted that the longer we wait, the greater the chance that we will find ourselves attempting to assign iDs to people who already signed up for an iD.

As of this writing, it is difficult to know precisely how many U-M researchers have iDs, as the "affiliations module" has not yet been activated in the user interface, though it is planned for roll-out this summer. The best number we have is via a query of the ORCID database for registered users whose emails are from @umich.edu: As of 10 May 2013, there are 213 (12 of whom are ORCID Task Force members). This is likely an undercount, as some researchers may have registered using a non-U-M or personal email account. Regardless, adoption rate is currently low.

Note that it is currently possible to populate the affiliation field using the ORCID API, so we can and should create records for our researchers and include information on affiliation.

First step: We recommend beginning to assign iDs this Spring/Summer of 2013.

2. *The Library should incorporate iDs into Deep Blue for current authors.*

We discussed many other Library systems as potential hosts for iDs, including DLPS, HathiTrust, ALEPH, and various local databases (e.g., UM Press, AAEL Comic Collection, etc.). *Appendix 1: Candidate Systems for Implementation for ORCID iDs* provides further details on our evaluation of these systems. We concluded that Deep Blue is the best place to begin. Its flat metadata scheme (Dublin Core) means adding a field (e.g. "identifier.orcid") is simple, and iDs will enhance searching via both internal and external systems.

Doing this will present some challenges, though. Linking iDs to the correct names is time-consuming at best and difficult at worst. Any given author can usually be done quickly by hand, but doing this in bulk and doing it accurately will require a lot of cleanup effort. (Again, by hand.) We have poor authority control in Deep Blue, and many thousands of items deposited by proxy, so we can't use depositor unique name to apply an iD and get good results.

Not all items in Deep Blue will have a current U-M author. At this time we don't recommend assigning iDs to authors who are no longer affiliated with U-M.

DLPS-run systems and ALEPH are also possible venues for applying iDs. Neither is ideal, however. DLPS systems share the authority control problems of Deep Blue, and further, items there do not necessarily represent U-M authors' work in every instance. ALEPH is similarly non-U-M-centric, and while it has authority control, we would need

to commit to local maintenance of authority files to take advantage of iDs. Even if we did so, since VuFind does not support authority functionality, it would be of limited value for Mirlyn users. Since HathiTrust relies on MARC records, the concerns about ALEPH also pertain to it, though this also means of course that solving the ALEPH problems would solve those in HathiTrust.

In the end, we conclude that we can and should first deal with systems where authors and creators supply metadata, Deep Blue being the prime example. Looking toward the future, the newly implemented cataloging rules, RDA (Resource Description and Access), include a section on 'Identifier for the Person' as a core element. So, we do anticipate using iDs in ALEPH and associated systems at some point.

First step: We recommend beginning this work immediately on completion of (1) above.

3. *The Library should actively partner with ORSP, MCV (the Medical School Curriculum Vitae system), and ultimately MCommunity to incorporate iDs into their systems. Doing this will speed adoption and use, and prepare the way for using iDs in other systems.*

We identified a number of lynchpin systems, both internal to U-M and external to it, that, if they implement iDs, will provide the greatest benefit to our researchers, the campus, and to the research community at large.

University of Michigan

- a. ORSP's eResearch Proposal Management (eRPM): Use of iDs in this system, under the Office of the Vice President for Research, would have the greatest effect on campus adoption, and would position us well for any future requirements regarding ORCID use in external grants.
- b. MCV: A natural place to represent and use iDs. Like the eRPM system, use here would have high impact and visibility.
- c. SciVal Experts Governance Committee: Another, and related, natural set of early adopters, and particularly useful in that we can build on the existing groundwork done to create SciVal Experts profiles to populate orcid.org profiles if we so choose.
- d. MCommunity (LDAP): This is a lynchpin system, and if iDs are included here implementation elsewhere becomes more easily automated, if not automatic. Ideally, we want iDs to be incorporated into and travel along with MCommunity records, so our researchers need only remember their unqiuname; this may not be possible in the short term, but is desirable.

- e. Rackham's Dissertation Submission System: Depending on how broadly we assign iDs, Ph.D. candidates may end up with these as part of the Library's efforts outlined in (1) above. If so, adding iDs to Rackham's system would be straightforward and start our doctoral students off on the right foot.
- f. ICPSR: ICPSR has expressed interest in implementing ORCID iDs in their systems, and can provide valuable expertise in working with datasets as we expand the use of ORCID iDs.

Beyond University of Michigan

- a. Library of Congress authority files: Widening the circle, implementing iDs in Library of Congress records would provide tremendous benefits. Inclusion in internationally used authority files for our authors would solve most of the concerns raised in (2) above. We should express our support for using ORCID as a part of such efforts. Note that the timing is good here, given the interest in linked data and associated initiatives around facilitating it. The Library of Congress' Bibframe Initiative has included ORCID in its identifier terminology <<http://bibframe.org/vocab/identifier.html>>.
- b. OCLC: An OCLC Research Task Force, "Registering Researchers in Authority Files" <http://www.oclc.org/research/activities/registering-researchers.html>, is currently working to create a summary of the benefits and trade-offs of approaches to the problem of incomplete authority files. ORCID is one of the approaches the Task Force is evaluating. There is value in starting a conversation with OCLC, to at the very least learn more about its Task Force and if there are other groups within OCLC Research evaluating ORCID and its potential implementation.
- c. Summon: As the underpinning to our ArticlesPlus service, having ProQuest incorporate iDs into Summon via enhancing its records would increase its value.

In any and all of these systems (as well as those managed by e.g. journal publishers), we will want to work on a method to export our iDs into their database records so they can pre-populate their systems with U-M authors' iDs wherever possible. Workflows and APIs will have to be developed (not necessarily by us) to make this happen smoothly.

First step: We recommend that Elaine and Paul (or his successor) reach out to leaders in charge of these systems and begin discussions about benefits and implementation.

4. *U-M should promote iDs as an author disambiguation tool.*

There may be a temptation to talk about iDs as being useful for creating researcher profiles; its administrative benefits are also clear and appealing. In our initial publicity and roll-out we should not emphasize these things, and should probably not mention them at all. Instead, we should focus on the following:

- This is about improving the ability of peers and publishers to associate you with your work. With ORCID iDs, citations will improve and the impact of your publications will become clear.
- iDs are not required by anyone, just as using a consistent form of your name on all your publications isn't required. But publishers and funding agencies, many of whom are already ORCID sponsors and partners, may in fact require it at some point.
- The Library, in cooperation with the Office of Research, the Medical School, and other appropriate campus groups, has taken the step of assigning iDs to authors and researchers in an effort to prepare for any such future requirements, and to begin to improve citations now.

We have provided similar services and done similar promotion before: MLibrary worked with the NIH to aid U-M researchers in complying with the PubMed Central mandate, providing deposit assistance for over three years to ease the transition as our PIs and researchers got used to the federal law. ORCID use is unlikely to become a federal law, but the major academic publishers back this system and will likely provide incentives for using iDs.

However, since there are already commercial tools that do profiling we should stick with what iDs do best. *Appendix 2: Promoting ORCID* provides more details and suggestions on how to begin promoting iDs via an implementation task force charged with a phased rollout. The basic gist is that we recommend creating a variety of promotional and informational materials, piloting them with Library staff, then begin a phased rollout of iDs beginning with targeted groups of early adopters followed by a campus-wide launch.

Again, though the administrative uses — including accreditation, alumni identification, etc. — are compelling we suggest de-emphasizing these aspects when communicating with the campus as a whole, and suggest encouraging schools and departments do the same.

First step: Constitute an ORCID implementation task force, and charge them with creating U-M specific informational and promotional materials.

5. U-M should not encourage use of *orcid.org's* profile feature.

This relates to the previous recommendation, but in addition to de-emphasizing iDs' administrative utility, at this time we suggest consciously turning the spotlight away from *orcid.org's* profiling capabilities. As discussed more fully in *Appendix 3: Testing the ORCID User Interface*, automatic population of profiles is currently incomplete, and

manual population is tedious and slow. The only area where we see immediate value in using orcid.org's tool is to further ensure disambiguation of authors; in other words, we suggest populating these profiles only to the extent that doing so will differentiate between one author and another with a similar or identical name.

If we wish to use or leverage ORCID as a profiling tool, we can consider the revival of the long-dormant CV project (see *Appendix 4: Faculty Publications Database*), or something similar, to get a more complete representation of publications by our authors. This may be especially — or perhaps exclusively — worth consideration for disciplines where SciVal Experts and Pivot are weak now, and are likely to remain weak in the foreseeable future. Per the recommendation above, if we assign an iD and a faculty member goes to orcid.org or Deep Blue, they will typically see and be able to associate only a percentage of their work there with that iD. Assigning iDs gives us an opportunity to make publication lists more complete, and in doing so provide a tangible and immediate benefit and service for those researchers who are enthusiastic about creating a public profile. This would require a commitment of resources that is out of scope for the current Task Force.

First step: Downplay orcid.org and its profiling feature in our promotional efforts (while working behind the scenes to improve it?).

Appendix 1

Candidate Systems for Implementation of ORCID iDs

I. Library systems which are good candidates for incorporating ORCID iDs:

- Deep Blue
- DLPS
- HathiTrust
- ALEPH
- Local databases (UM Press, AAEL Comic Collection, etc.)

Feasibility of adding fields for ORCID storage and use:

- Deep Blue: Easy in principle; DSpace's flat metadata scheme (Dublin Core) means adding a field (e.g. identifier.orcid) is simple, and iDs will enhance searching via both internal and external systems. However, although linking iDs to any given author can be done fairly fast by hand, doing this in bulk will require a lot of cleanup effort. (Again, by hand.) We have poor authority control in Deep Blue, and many thousands of items deposited by proxy, so we can't use depositor username to apply an iD and get good results.
- DLPS: This is feasible, but like Deep Blue, would require significant cleanup. Unlike Deep Blue, many DLPS collections contain work with little or no U-M authorship, so iDs will not end up in every record.
- ALEPH^[1]:
 - Note regarding ALEPH: VuFind is a keyword search system only; it does not support browsing. Users would need to type in the literal string to locate records associated with the string (i.e. a creator's iD). Other universities have played with the "Did you mean" functionality (the spellcheck function) to bypass the lack of authority support, however we would need to include all variants of the known literal string and test the functionality here at MLibrary.
 - Associating an iD with a creator is best done in an authority record to eliminate disambiguation confusion. There are two options for including an iD in an authority record: (1) the MARC field 024 7_ \$\$a[number]\$\$2orcid^[2] or (2) in the \$\$0 of the MARC 100 field. However, since VuFind does not support authority functionality, these references would not point to works with the creators authorized heading, nor are the authority records keyword searchable. Including the iD in authority records would require a fair amount of manipulation:
 - local maintenance of authority records for campus affiliates
 - analyses and recommendation on what should be done where authority records do not exist for campus affiliates

- exploring options for mapping contents of an authority record to a new search mechanism in VuFind (after speaking with J. Rothman, this is probably not feasible)
 - protecting the 024 7 with \$\$2orcid of the \$\$0 from the Backstage overlay process (local Systems work)
 - analysis and recommendation on which unit the work will fall to, whether or not the work will be ongoing or a one-time effort
- Alternatively, we could add the MARC field (024 7_ \$\$a[number]\$\$2orcid) to the bibliographic record. However, providing search ability for the iD would be an involved, and not necessarily consistent, process given that there are often multiple creators responsible for a work. It is not immediately obvious how much effort it would take to manage the addition of an 024 7_ \$\$a[number]\$\$2orcid, as there are multiple issues to address:
 - Mapping any 024 with \$\$2orcid in VuFind to a new search path (much the same way patrons search for ISBN's) and re-indexing ALEPH (not sure how much work this would take)
 - Documentation for catalogers (including guidelines for referencing the list of iDs for campus affiliates, guidelines for disambiguating creators, at what point in the process will the iD be included in the ALEPH record — in other words, where will the work fall to?)
 - Consistency in including the iD in the bib record
 - Protecting all 024 7 with \$\$2orcid from the Backstage overlay process (local Systems work)
 - analysis and recommendation on how many iDs to include in cases where there are multiple creators.^[3]
- There is a chance that we could use the 'changed headings' functionality in ALEPH to trigger the addition of \$\$0 (containing the creator's iD) in authority records to all bibliographic records with the matching heading (whether in the 100, 600, or 700 field). This would eliminate the concern presented by VuFind's lack of authority functionality, since all associated works with a creator's iD would display when a user searched the author field. However, this methodology would need further analysis and recommendations for implementation, at the very least because numeric subfields present in authorized headings do not currently display in VuFind.
- The best case scenario is that Library of Congress adopts the use of iDs (see below) and begins including them in their authority records, which would eliminate the concern above of trying to avoid fields being overlaid (if the master record contains the identifier, there will be no need for manual manipulation, however not every campus affiliate has an authority record). At this time the Library of Congress encourages catalogers to use the 024 7_ in the authority record for inclusion of an iD;

there is no discussion (that we are aware of) to begin including them in master authority records.

- HathiTrust^[4]:
 - HathiTrust sources and uses MARC records from ALEPH for its display, so we'll run into the same issues described above.
 - We'll run into further issues with who and how this would be implemented, now that CDL is managing the records. For those HathiTrust records that are in ALEPH, which display in VuFind, we could update bibliographic records as we did above (adding a MARC tag in the bibliographic record), however these local references won't be ingested into the HathiTrust framework. Further evaluation of whether to include iDs for all known contributors, or Michigan affiliates only, will need to be done.
- Other local databases (UM Press' Title Management Module, AAEL Comics Collection, etc.): Each database/spreadsheet is unique to its local site and will require further analysis, and some may not easily support added fields

Recommendation for inclusion of ORCID:

- Deep Blue: Initiate immediately
- DLPS: Further investigation required to determine costs and benefits
- ALEPH: Local authority records (only if not in danger of loss via overlay); otherwise, bib records
- HathiTrust: Tied in with ALEPH
- Local databases (UM Press' Title Management Module, AAEL Comic Collection, etc.): do not recommend inclusion of ORCID at this time

We may be looking at multiple phases: Phase 1: creator-supplied metadata includes fields for creator or proxy to enter iD. Phase 2: Determine true feasibility/workability for cataloger supplied metadata.

II. University Systems

- ORSP's eResearch Proposal Management (eRPM) (OVPR/ORSP): Recommend implementation; need buy in
- Rackham's Dissertation Submission System: Recommend implementation, need buy in
- MCV (Medical School Curriculum Vitae system): A natural place to represent and use iDs; need buy in.
- MCommunity (LDAP) : Recommend implementation, need buy in; this is a lynchpin system, and if iDs are included here implementation elsewhere becomes more easily automated, if not automatic. Again, buy in is needed at high level

We recommend that Elaine and Paul (or his successor) seek advice on other systems as they make recommendations for implementation in these systems.

III. External Systems

- LC: a lynchpin system; if we can get iDs into internationally used authority files for our authors, that gives us the greatest benefit. We should express our support for using ORCID as a part of such efforts. Note that the timing is good here, given the interest in and associated initiatives around facilitating linked data it. The Library of Congress' Bibframe Initiative has implemented ORCID in its identifier terminology (<http://bibframe.org/vocab/identifier.html>). However, it remains to be seen how the adoption of this initiative will take place, as it is in the very early stages of development.
- OCLC: An OCLC Research Task Force, Registering Researchers in Authority Files (<http://www.oclc.org/research/activities/registering-researchers.html>), is currently working to create a summary of the benefits and trade-offs of emerging approaches to the problem of incomplete national authority files. ORCID is one of these approaches the Task Force is evaluating. There is some chance that ORCID's could be incorporated into OCLC's Linked Data initiative^[4], but we believe the work of their Research Task Force will help inform OCLC's decision to do so. We believe there is value in starting the conversation with OCLC, to at the very least learn more about its Task Force and if there are other groups within OCLC Research evaluating ORCID and its potential implementation. (Note: We found one bibliographic record in OCLC [812311825] with an ORCID [0000-0003-0962-9465] for one of the described item's authors, oddly enough in a MARC 773 0_ \$\$t, with the phrase ORCID: then the number.)
- Summon: As the underpinning to our ArticlesPlus service, having ProQuest incorporate iDs into Summon via enhancing/supplementing OCLC records (?) would enhance its value (and that benefit will not be limited to U-M users). According to our ProQuest support representative, the "Summon team is tracking ORCID closely and plans to use it for author name disambiguation as well as other features in Summon."
- We'd want to export iDs to external systems (e.g. journal publishers) so they can pre-populate their systems with U-M authors' iDs [NIH is beginning to encourage ORCID use, at least passively: <http://nexus.od.nih.gov/all/2013/04/11/taking-on-the-challenge-of-better-biomedical-workforce-data/>]

IV. General Concerns/Issues/Notes

- We can and should deal with systems where authors and creators supply metadata first, Deep Blue being a prime example.
- How can we handle a situation where people who have an ORCID assigned, and then added to one or more of our systems then they opt out/cancel? (And then, possibly, opt back in and generate a new iD? (Ideally, we'd get a feed from

orcid.org to populate and up-date our various databases, but preparing for this case may be difficult.)

- In terms of ALEPH, how do we work effectively with other campuses (particularly Dearborn, with its separate catalog system)?
- The newly implemented cataloging rules, RDA (Resource Description and Access), includes a section on 'Identifier for the Person' (rule 9.18.1), which is a core element. The language of the rule is: *Scope: An identifier for the person is a character string uniquely associated with a person, or with a surrogate for a person (e.g., an authority record). The identifier serves to differentiate that person from other persons. Sources of Information: Take information on identifiers for the person from any source. Recording Identifiers for Persons: Record an identifier for the person. Precede the identifier with the name or an identification of the agency, etc., responsible for assigning the identifier, if readily ascertainable.* The relevant LC-PCC PS for recording this identifier is: *Do not give the identifier alone.*

^[1] For ALEPH/HathiTrust, we're assuming that we'll have a master record of iDs for campus affiliates to look at when creating bibliographic data

^[2] ORCID is an approved standard identifier source code for usage in MARC records, <http://www.loc.gov/standards/sourcelist/standard-identifier.html>

^[3] For ALEPH/HathiTrust, there may be an issue with relation to our committee's agreement on opting in – if there is a master file of iDs for campus affiliates, how often will it be updated, if people opt out, how will the catalogers know?

^[4] The GLIMIR project at OCLC clusters and assigns an identifier to WorldCat records representing the same manifestation. These include parallel records in different languages (e.g., a record with English descriptive notes and subject headings and one for the same book with French equivalents). It also clusters records that probably represent the same manifestation, but which could not be safely merged by OCLC's Duplicate Detection and Resolution (DDR) program for various reasons. *GLIMIR: Manifestation and Content Clustering within WorldCat. Code4lib Journal (ISSN: 1940-5758) Issue 17, 2012-06-01.*

Appendix 2

Promoting ORCID

Preliminaries (first steps):

- *Define stakeholders:* An Implementation Task Force identifies relevant groups/ types (e.g. research administrative staff, Associate Deans for Research, etc.) and develops lists of names to target
- *Develop content:* An Implementation Task Force develops content for use in webpages on the MLibrary website (basic info, compelling benefits and examples, FAQ).
 - Rely on the ORCID-produced FAQ, as well as those produced by other early adopters as models (NYU? Washington? etc.)
 - Come up with a tag line: something that sends a message like “Get credit for your research”
 - Create a succinct list of the benefits of iDs (some directed at different audiences — administrators vs. individual researchers, etc.)
 - orcid.org is not meant to be “comprehensive”: at this stage, ORCID is most useful at making sure we have the right “you.”
 - “It’s a disambiguation engine, you can use it now, and it may be required by funders or publishers in the future. Here’s how to use it.”
 - “Imagine one unique ID that always moves with you that you can use for your research grants.”
 - Especially useful for those who move from one institution to another
 - Develop illustrative examples where ORCID is particularly helpful: common names shared by multiple U-M researchers, unusual names that are easily and regularly misspelled, etc.
 - Create a visual with a short narrative about how ORCID differs from vendor-specific author IDs such as Elsevier’s Scopus ID, Thomson Reuters’ ResearcherID
 - Identify some funding agency uses (or planned/ expected uses) of iDs to use as examples
 - Drawing from content developed for the above, the Implementation TF drafts communication (email and postcard) to be deployed to engage adopters after iDs are assigned
 - Develop a succinct statement of what action adopters should take (e.g., associating a few publications with their ORCID profile is ideal for disambiguation; a more comprehensive list of publications is possible and will likely become easier to realize in the future with more publisher/ indexer adopters)
- *Approach administrative stakeholders:* Elaine and Paul approach ORSP, OVPR, SciVal Experts Governance Group, MCommunity administrators, HR, U-M

Research Administrators Network (RAN), etc. to create partnerships and lay the groundwork for integrating iDs into existing systems.

Phased deployment / outreach:

Phase 1) Pilot with library staff involving:

- Assigning iDs, piloting and soliciting feedback on our informational material, training staff (so they're ready to answer questions when outreach goes further)
- Presenting to MLibrary Public Services Communication Forum
- Encouraging library staff to add their ORCID to their standard email signatures.

E.g.:

Karen E. Downing, Ph.D., M.I.L.S.
Head, Social Sciences,
Education Liaison and Foundations & Grants Librarian
209 Hatcher Graduate Library
University of Michigan
Ann Arbor, MI 48109-1190
kdown@umich.edu
001.734.615.8610
orcid.org/0000-0002-7655-9084

- Discussion with MPublishing staff about incorporating iDs for U-M authors into their publications and processes
- Testing promotional material

Phase 2) Pilot with a select group of early adopters, representing the sciences and engineering, humanities, health science, and social sciences. These units cover a broad range of disciplines and their faculty have many joint appointments, so the knock-on effect and reach will be large:

- SciVal Experts adopters
- International Institute faculty affiliates (at least one of the Centers, perhaps all); this group includes social sciences and humanities and a small number of science faculty. Every faculty member in the Institute belongs to at least one other department, so word-of-mouth will help the ORCID concept spread.
- Faculty from the Center for the Study of Higher and Postsecondary Education (CSHPE) in the School of Education.
- The Department of Ophthalmology has also expressed interest in taking part in a pilot.

It's possible that during this phase profile data from SciVal Experts could be exported into ORCID. This would build on the work already done to create more comprehensive publication lists in SciVal Experts, and provide better results for early adopters of orcid.org's profiling, at least in the sciences.

Phase 3) Deployment with all other campus researchers involving:

- Assigning iDs to all faculty, and perhaps all Ph.D. and masters candidates as well. (Anecdotally, we've had some positive feedback about assigning iDs to students. For example, the Center for Japanese Studies considers this a good way to keep track of alumni, at least those who remain in academia.)
- Promotion
 - Postcard sent to all faculty (as was effective for Deep Blue)
 - Email sent to all campus researchers as identified by Office of the Vice President for Research (OVPR)

Appendix 3

Testing the ORCID user interface

A brief evaluation of the current state of ORCID, from a user's perspective. In general, account creation is straightforward, entering citations manually is inconvenient and riddled with potentials for error, and automatic citation entry and updating is currently of limited use because of incompleteness.

Accounts

- Password re-set works great. Re-set link is sent immediately to email address on account. System did hiccup twice upon first log-in after password re-set.
- Attempts to create a second account were thwarted if the same email address was used. It is easy to create a second account using a different email address. ORCID did not object to identical name and institution being used, though it suggested identification with similar account. Did object to original email address being used as alternate email address however. Looks like the email address is the only element ORCID really checks against.
- When attempting to create a second account with same email address, the following notice is given: "[emailaddress] already exists in our system". Easy to set up second account with different email, even when using same name and copying citations from first ORCID record. As researchers move to new institutions they may set up multiple accounts if unaware of the one we set up.
- X-ing out of the form will lose edits that were not saved.
- When account is deactivated, that same name and email address can never be used again unless you contact ORCID to reactivate it, at support@ORCID.org. Name and email address are kept on file, but hidden. A message is sent to the address on file that an ORCID account will be deactivated. Must click on link. Not known what happens if you do not click on link. We need to test the ORCID help support.
- Once account is deactivated you cannot log in with that email address and password. There is no delay on the account deactivation, it happens immediately. QUESTION: Can U-M reactivate a deleted account on a user's behalf?
- Can register for a sandbox account for experiments that do not affect the database: <http://sandbox-1.orcid.org/register>. Has warning across top that it is a test website. No restrictions on registering the same name, address and password here. Was able to create a third ORCID account in sandbox.

Entering Citations Manually

- Only two fields are necessary to create an entry: title and type (format)
- Publication types will cause much confusion. There is no option for book chapter, for example, only 'chapter anthology' or 'book'. Some formats are combined, e.g. TV Radio. Needs definitions of publication types.

- If an external ID is given, you are required to enter the type, e.g. ISBN. This will cause confusion, as many options will not be recognized by users, e.g. OCLC is written out. User can choose 'other identifier type' however.
- Roles very limited. Does not include co-author, editor, creator, compiler, illustrator, director, etc.
- No formatting options within the entry form. No italics, bold, etc., so there is no way to follow a standard style.
- If an error is made, there is no way to edit the entry. Can only trash it and start again.
- Form allows for infinite errors. Allows wrong page numbers, impossible dates, etc. Can list the same citation twice.
- When changes are saved, form does not acknowledge this in any way. User cannot tell that a citation is saved until they close the Update box and check the list of works.
- Definitions of three publishing options (public, limited, private) are given in linked help box. Not clear how you enable people to view entries with the 'limited' status, or what can be seen with the 'private' status. Status can be changed for individual entries, however.
- Easy to manually add works that are already associated with another duplicate ORCID ID.
- Easy to make up bogus citations.
- Failure to include required fields doesn't generate any response, so users are on their own to figure out why something didn't get saved.

Entering Citations Automatically

Update Works - Search and Add

- This is the default option. ORCID searches CrossRef and researcher can just click on citations to add them.
- There do not appear to be any consequences to adding a citation that is not yours.
- CrossRef searches name keywords. In other words, it offered results from people who simply shared my first name, and not my last name, and vice versa. It didn't pull up any citations that I am associated with in reality.
- Perhaps authors from certain disciplines will have a better time than others finding their works via search.

Import Research Activities

- In this option, Scopus is the only import option for adding citations automatically so far. (When will there be others?) If a match is not found in Scopus, user is left in Scopus with no way to back into ORCID and must X-out. User could easily put in another name, falsely identify oneself and import those works to their ID. Looks like it would be easy to mess with someone else's citations.

Appendix 4

Faculty Publications Database (August, 2009 Proposal)

Summary

A comprehensive database of U-M authored publications would be of considerable general value to the University, serving administrative goals ranging from accreditation to providing broad access to that work via Deep Blue, the Library's institutional repository service.

In 2007, at the request of and with support from the Provost, the University Library undertook an experiment to gather a large number of current faculty curriculum vitae (CVs) from schools and colleges across the University. We received several hundred CVs. Library staff familiar with a broad array of metadata standards analyzed the structure and content of the CVs and estimated the cost of having the data encoded and entered into a database. Specifications were provided to vendors and subsequently refined to create a reliable and cost-effective data entry model.

We received two quotes from these vendors, and their work on the sample data gave us a framework for estimating the cost of the task as a whole. Data entry conformed to specifications and allowed us to load records for the individual publications into a database in a manner that would in turn allow us to search, display, and analyze the information, and to make it publicly accessible if so desired. The quote from SPi, a vendor with a good track record of large-scale keyboarding for the Library, was \$30 per CV, based on the sample provided to the vendor. We are currently exploring other pricing models and options with other vendors as well.

Based on this work, we would need approximately \$140,000 to build the initial database of current faculty publications for the roughly 4,700 scholars here actively engaged in publication. Additional costs, including library staff to manage the process of collecting, organizing and providing materials to the vendor are not included. Nor are the costs of evaluating the vendor's work and putting the content into a local database. We would accommodate these tasks with existing staff.

Scope and cost

As of November 1, 2008, U-M had 2228 staff in the research job class, 2473 tenured faculty, and 687 non-tenured (but tenure-track) faculty, or roughly 4,700 total. [1] To populate a publications database with baseline data from the publication lists in their CVs would cost approximately \$140,000. This is based on a price quote from SPi of \$30/CV. [2]

What is included

The price is for conversion of one complete set of CVs to XML-tagged electronic text per the specification developed by Paul Schaffner and Jim Ottaviani. [3] It would populate a database of U-M publications as up-to-date as the information in those CVs allows. The current specification instructs the data-conversion firm to:

1. **capture basic biographical information** about the faculty member, assuming the CV contains it:
 - § name
 - § UM unickname
 - § departmental affiliation
 - § start and end dates of association with the University
2. **capture basic bibliographical information** about each publication listed:
 - § author(s), in the same order as listed in the CV
 - § title
 - § author and title of any larger unit(s) (book, conference, journal) in which the publication appeared.
 - § authorial role (author, editor, illustrator, etc.)
 - § certain standard numbers (patent nos., ISBN, etc.)
 - § the remainder of the entry, divided between true citation information (volume numbers, dates, page numbers, imprint and edition information), and anything otherwise unaccounted for (to be placed in note fields).
3. **distinguish**
 - § between volume, issue, year, month (etc.) numbers in the citation
 - § between the faculty member and other co-authors.
4. **make explicit or supply** some implicit information (e.g. the author's name, if it is assumed).
5. **classify** publications, so far as feasible, by
 - § format (print, web/electronic, audio, video, etc.)
 - § type (article, talk, chapter, abstract, book, report, book-review, letter, patent, software, etc.)
 - § status (peer-reviewed? invited?)
6. **normalize** the styling of names, the spacing around punctuation, and month- and day-designations to facilitate automated processing.

Though the data as captured may not support sophisticated manipulation of the output, e.g. reliable transformation into multiple bibliographic formats, it should suffice to produce a consistent and attractive display, to allow basic fielded (author/title/date/etc.) searches and sorts, and to serve as a basis for post-processing (e.g. authority control and/or extraction or normalization of author, journal, and publisher names).

The figure quoted above does not include the cost of collecting, organizing, de-duplicating, or (if any are submitted in paper form) scanning those CVs. Nor does it cover the cost of building the database itself, or any interfaces to it, or of maintaining

and updating it. Database construction and interface work would be covered by the Library, while maintenance and updates of the completed baseline database and its contents could be taken care of by departments and/or individual faculty.

Relation to U-M Goals

Accountability and publicity: A complete record of U-M research and scholarly publications provides a quantitative measure of the effectiveness of our faculty and is valuable for both internal assessment and when seeking funding and outside support. Further, by making this information visible (see below re. Deep Blue), consultation and other opportunities for our faculty are more likely.

Faculty CV creation: U-M faculty would be able use this database to maintain an up-to-date listing of their published work, useful for accurately building and maintaining a critical portion of their CV.

NIH compliance: Having a complete record of peer-reviewed publications will allow U-M to assure compliance with the NIH Public Access Policy, which in turn will help assure continued funding.

Deep Blue: With a more complete record of past and present publication activity we can add to U-M's online presence via Deep Blue, the Library's service to provide the broadest possible access to U-M authored scholarship and research. Identifying older work by current faculty is challenging, especially for those who have worked at other institutions prior to joining the University. Further, securing permission to make their work available online can prove difficult and time-consuming. With the faculty publications database in place, identifying work done by U-M authors becomes straightforward. Regarding permissions, we have incorporated a "by request" feature in Deep Blue for cases where the author is allowed to send PDFs of a publication directly to interested readers and colleagues, but is not allowed by contract to make the article freely available via anonymous download. With the request feature, the author/author's designee responds to requests routed through Deep Blue, and can choose to send (or not send) the file. As a result, U-M would have the archival copy that makes the database a complete record of the work done by our scholars and can provide the greater exposure that Deep Blue allows while avoiding the permissions request hurdle that often prevents deposit.

Notes and associated documents

1. Faculty/research staff totals per HR's "Standard Reports" of November 1, 2006 at <http://www.hr.umich.edu/hrris/reports/standard.html>. [via the spreadsheet titled "Faculty Characteristics/Demographics"; including all except 'Lecturers' category]
2. SPi quotation for project available at http://www.umich.edu/~pfs/cvs/SPi_quotation.xls.

3. CV-project data-capture specification and instructions available with examples at http://www.umich.edu/~pfs/cvs/spec_05.html.

4. The NIH Public Access Policy implements Division G, Title II, Section 218 of PL 110-161 (Consolidated Appropriations Act, 2008).