

Establishing Best Practices in Institutional Repository Organization

Scott D. Bacon, Coordinator of Digital Initiatives
Coastal Carolina University
DCSEUG 2020

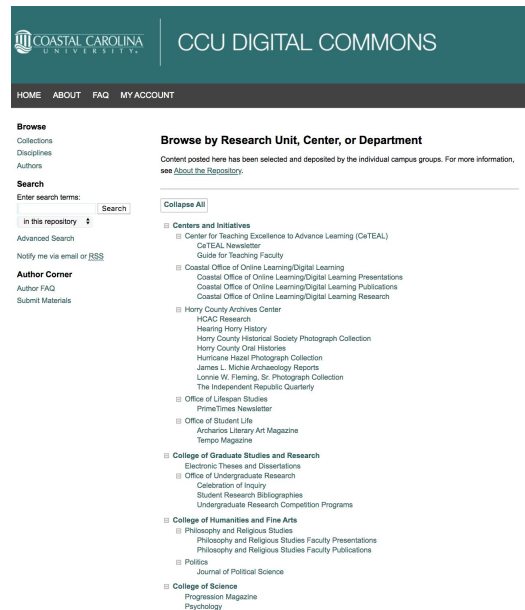
Collections Page Project

Research Questions:

How do Digital Commons repository managers organize their Collections / Communities.html pages?

Do best practices emerge from analysis of multiple Digital Commons sites?

HUGE CAVEAT: I realize that my institution's Collections page needs work, so I get the potential "pot calling the kettle black" reactions. No harm intended, just pointing out suggestions, more like constructive criticism.



I'm Scott Bacon, the Coordinator of Digital Initiatives at Coastal Carolina University, and the coordinator of CCU Digital Commons, CCU's IR.

This past fall CCU Digital Commons had just celebrated its year 1 anniversary, so I wanted to see if there was anything I needed to do maintenance-wise with our Collections Page

Quick overview of the CCU Digital Commons. We went live in August 2018, currently have 4,545 items (as of this week). We have Graduate ETDs and Honors theses, several Digital Collections, 1 conference, 2 journals. Not a lot of faculty works yet.

My Research Question for this project was: How do Digital Commons repository managers organize their Collections Pages? Do best practices emerge from analysis of multiple Digital Commons sites? So I examined how the largest Digital Commons repositories organized their Collections Pages to see if best practices emerged.

Caveat... Our Collections Page is not where it needs to be. I wanted to solicit feedback from other managers to see how I could continue to make our Collections page better. I'm not here to show you a perfect Collections Page, if such a thing even exists.

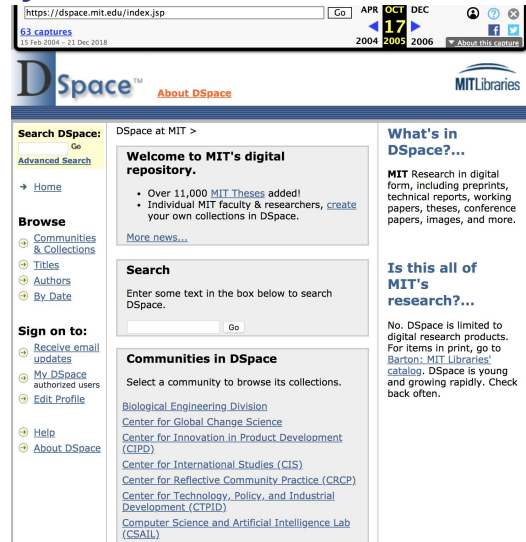
IRs and Communities: History

MIT and HP Labs released DSpace in 2002.

They began using “community” as an organizational tool. (EPrints began in 2000, but seems to use “division” vs. “community”)

Why use community?

- Allows “distributed administration” (SPARC 2002)
- Communities “lend legitimacy” (ibid) to content
- Community-specific workflows



This is not an exhaustive analysis, I didn't want to get bogged down in the early years of repository practices. But it's instructive to see how and why communities became the norm for the division of information on IRs.

So first of all, why use communities to organize?

- Distributed administration: Because “distributed administration” enables differing policies, procedures, roles, responsibilities, metadata schema, etc. expressed among each college/departments/etc. (SPARC 2002, 30). [So it enables discipline-specific control and system functionality as well. Practical.]
- Lends legitimacy: Items that are not peer-reviewed can be uploaded and still be legitimate because they're attached to official centers, institutes, and so on.
- Each community designs a community-specific workflow process that meets its specific needs.

Repository Organization Best Practices

Little to no information on organizing overall repositories, instead: controlled vocabularies and taxonomies for disciplines and genres, subject headings...

Best practices describe objects rather than their communities/containers, so instead:

- Leverage current institutional organizational structures and/or other IR organizational schemes.
- Communities: programs, regional initiatives...
- University Website / University Archives

Common Community Types (Mercer et al. 2007)

- Formal communities
- Informal communities
- Communities of practice

Digital Commons Publication Types

- Series
- ETD
- Journals
- Book Galleries
- Image Galleries
- Event Communities

So now that we have an idea of how communities facilitate the population of IRs, how can we as IR managers organize our communities on the Collections page? What are others doing to effect “good-enough” practices if not “best” practices?

Repository managers seem to leverage organizational schemes seen on existing IRs and tweak them to match their institutions. But communities are based largely on programs offered at the institution and on regional initiatives. So this precludes a one-size-fits-all approach, each institution has its own unique sort of fingerprint. But you also don't have to reinvent the wheel, you can leverage what others have done in the past and then tweak your collections page to suit your own organization. I usually meet or talk with stakeholders on campus to determine the appropriate organization of their materials, but I also:

Studied our university website to help me in initial Collection Page planning. I also spoke with our university archivist, who was helpful in giving me a sense of how univ. Administration organization should be set up for long-term planning.

As far as Common Community Types: I found Mercer's 3 types to be accurate by and large:

- Formal: Associated with academic departments and research units
- Informal: Individuals can contribute without a formalized community structure
- CoP: Interdisciplinary groups that lack a formalized administrative structure

Methodology

Bepress list of all sites, with item counts **n=519**

Removed law, medical and health sciences, professional, special, community colleges, military, government, public library, corporate, professional association, consortia **n=314**

Removed all sites with less than 10,000 items **n=102**

(Data gathered on February 27, 2020)

	A	B
1	IR Name	IR Link
2	University of New Mexico	https://digitalrepository.unm.edu/
3	University of Nebraska - Lincoln	https://digitalcommons.unl.edu/
4	University of Central Florida	https://stars.library.ucf.edu/
5	Kenyon College	https://digital.kenyon.edu/
6	Iowa State University	https://lib.dr.iastate.edu/
7	University of Montana	https://scholarworks.umt.edu/
8	Eastern Illinois University	https://thekeep.eiu.edu/
9	Utah State University	https://digitalcommons.usu.edu/
10	Purdue University	https://docs.lib.purdue.edu/
11	The University of Maine	https://digitalcommons.library.umaine.edu/
12	Georgia Southern University	https://digitalcommons.georgiasouthern.edu/
13	Nova Southeastern University	https://nsuworks.nova.edu/
14	University of Mississippi	https://egrove.olemiss.edu/
15	University of Massachusetts Amherst	https://scholarworks.umass.edu/
16	University of Iowa	https://ir.uiowa.edu/
17	University of the Pacific	https://scholarlycommons.pacific.edu/
18	Cedarville University	https://digitalcommons.cedarville.edu/
19	Missouri University of Science and Techn	https://scholarsmine.mst.edu/
20	Brigham Young University	https://scholarsarchive.byu.edu/
21	University of Pennsylvania	https://repository.upenn.edu/
22	Western Kentucky University	https://digitalcommons.wku.edu/
23	Western Michigan University	https://scholarworks.wmich.edu/
24	University of Tennessee, Knoxville	https://trace.tennessee.edu/
25	University of Kentucky	https://uknowledge.uky.edu/
26	University of Dayton	https://ecommons.udayton.edu/
27	California Polytechnic State University, Se	https://digitalcommons.calpoly.edu/

How did I begin to analyze how other Digital Commons sites organize their Collections Pages?

I contacted bepress, who gave me a list of all the Digital Commons sites in the US, including their item counts. This gave me **519** sites.

I needed to narrow the number down to make this project manageable, so I narrowed the number of institutions by

- Removing the following repository types: Law, Medical and Health Sciences, Professional, Special (Seminaries, etc.), Community Colleges, Military, Govt, Public Library, Corporate, Professional Association (discipline-specific), and Consortia. **n=314**.
- Removed all sites with less than 10,000 records. **FINAL n=102**
- Rationale for cutting off repositories with less than 10,000 items. I predicted that the size of an institution would affect how they organize their repository. The idea is that more items you have, the more likely you've invested time and resources into the IR, meaning increased iteration on the Collections Page, and more detailed institutional organization. Hopefully these larger IRs can serve as better examples of best practices.

Item count data from this study is current as of 2/27/2020.

Top-Level, Sub-Community, Publication

☐ **College of Science**

Progression Magazine

Psychology

☐ **Conferences and Events**

International Gullah Geechee and African Diaspora Conference

☐ **Honors College and Center for Interdisciplinary Studies**

☐ Great Themes: Honors College

Great Themes: Colonialism

Honors 306: Oral History In Action Interviews

Honors Theses

I wanted to quickly make the distinction between top-level communities, sub-communities, and publications as they exist in the Collections Page.

In this example we see College of Science, Conferences and Events, and Honors College as our top-level communities. Sub-communities are and Publications can be contained within top-level communities.

The reason I chose top-level communities is due to time constraints. I couldn't possibly look through every single entry on all of the IRs. Focusing on Top-level entries also kept my focus on how we as IR managers want to showcase the top-level content that stays visible even when levels are collapsed, and to enable a broad to narrow or general to specific browsing strategy to our users.

Community Types

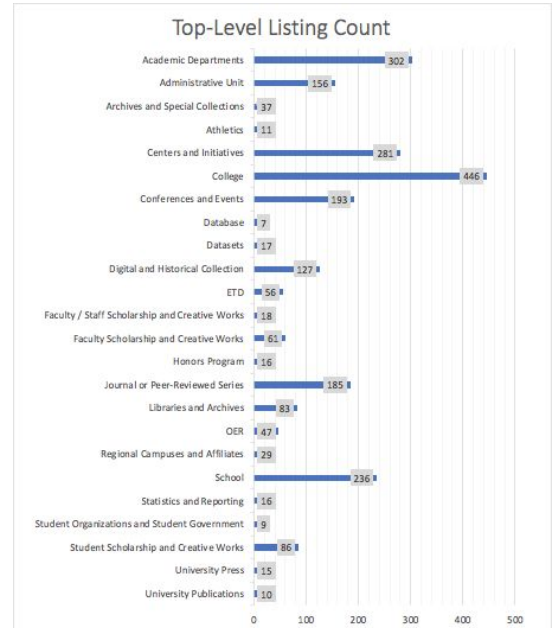
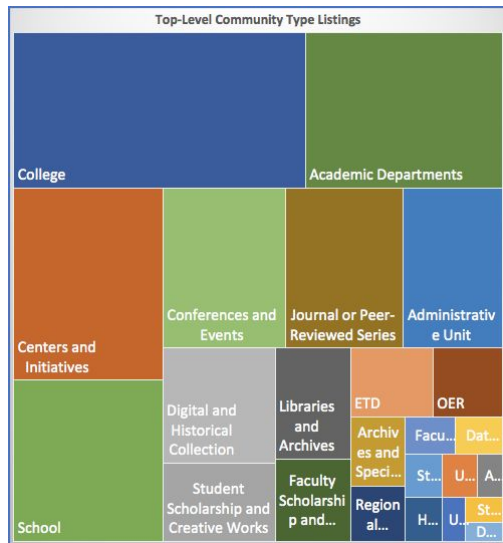
- Academic Departments
- Administrative Unit
- Archives and Special Collections
- Athletics
- Centers and Initiatives
- College
- Conferences and Events
- Database
- Datasets
- Digital and Historical Collection
- ETD
- Faculty and Staff Scholarship and Creative Works
- Faculty Scholarship and Creative Works
- Honors Program
- Journal or Peer-Reviewed Series
- Libraries and Archives
- OER
- Regional Campuses and Affiliates
- School
- Statistics and Reporting
- Student Organizations and Student Government
- Student Scholarship and Creative Works
- University Press
- University Publications

I recorded all of the top-level communities from these 102 IRs, which resulted in 2,468 rows of top-level community entries, and I then coded each top-level community with a Community Type.

There was no overarching taxonomy or controlled vocabulary defining these type labels, it was a totally organic process. I wanted natural practices to emerge, but it was generally easy to categorize most of the entries.

I did not include entries in this list with 5 or fewer instances, as they were not representative of the common best practices. There were also about 15 listings that had to be categorized as “miscellaneous” because they didn’t make sense or were empty or obviously in the middle of testing.

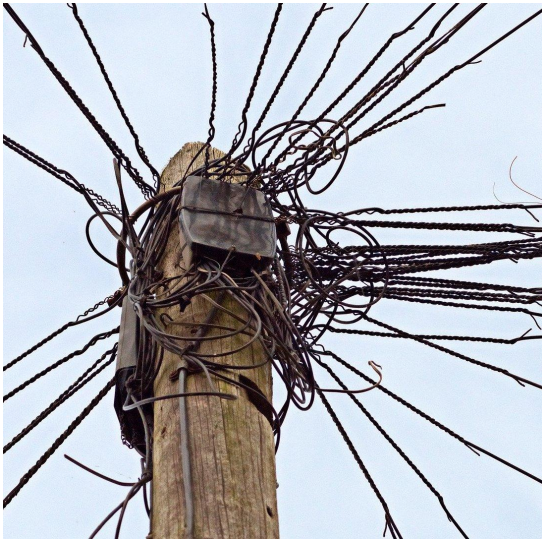
Community Type Findings



I dove into the data on these top-level community types and frankly it did not provide that much useful information in aggregate, but as I was going through all of the Collections Pages of the 102 IRs, I found it very useful to see how others were doing things in common and uncommon ways. That's the most helpful thing I want to share about this project.

The data is available on the conference website to anyone who wants to look through it.

TMI vs. TLI



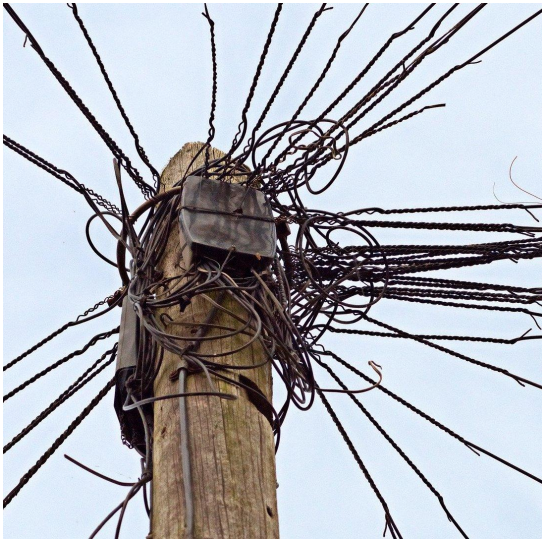
One of the major suggestions that I can make to IR managers trying to spruce up their Collections Pages is TMI vs. TLI: I saw many examples of these, both of which can present challenges to users trying to find resources.

TMI: So much information as to make it difficult for users to find what they need.

TLI: So little information as to make what's there unusable. Image on the right illustrates Man on Horse: This is a common way to describe bad cataloging practices - without context the photo is interesting but ultimately not helpful to potential users if it's described poorly.

Our Objective: Find the sweet spot in between TMI and TLI.

Findings - TMI



Too Much Information

- 25-page Collections page
- 25 digital collections as top-level communities
- Theses hidden in departments
- Not using sub-collections

Google and PDF Cover Pages

TMI Examples:

- One IR had a 25-page Collections page (25 “page down” clicks to get to the end). CTRL+F maybe? Searchers should just use the provided search box, the Collections Page is meant to be for searchers who want to browse.
- One IR had 25 digital collections listed as top-level elements, hard to sort through, esp. since none of them have the word “digital collection” in the title: users unfamiliar with the collections may find browsing hard.
- One IR had Honors Theses listed in 45 top-level departments but didn’t have a collected thesis page, so 9 out of 10 departments only have an Honors Thesis or two in them, which makes for slow browsing.
- One site had 54 total top-level entries, 17 of which are journals, 10 of which are conferences. So half of their entries could be moved to Journal or Conference sub-collections in order to free up eyespace to facilitate browsing of their other types of materials. This is easy to do, just ask your bepress rep!

Google and PDF Cover Pages: While many researchers find our PDFs in Google and never enter our IR, for those who do we need to provide a good search experience.

Findings - TLI

Too Little Information

- Opaque initiatives: Suggest spelling out acronyms and/or providing contextual information in introductory text on department home page.
- University Presentations: Grab-bag with little to no contextual information.
- University Publications: Grab-bag in many cases.



TLI Examples:

- ANEL: You get to the article level and still don't realize what this is, turns out to be a languages department. Adding introductory text on dept home pages is not a Collections page best practice per se but helps to mitigate some of the issues with Collections page constraints.
- University Presentations: One site had a top-level University Presentations community. It must have been a collected page, there was no contextual information to identify conference or event, department or college of the speaker, etc.
- University Publications: Hodge-podge of things like institutional newspapers, magazines, newsletters, Communications & Marketing stuff, journals, yearbooks.

Suggestions for Improvement

Long-term Plan

Revisit periodically, iterate to improve if needed

Label Types (Conferences and Journals)

- Example: TTRA APac. Can't do Travel and Tourism Research Association - Asia-Pacific Chapter Conference, but can do TTRA APac Conference

Departments within Academic Departments
top-level listing (if not under Colleges).

Collected pages and Grouping Tool

Findability vs. Accuracy

- Jargon may be more accurate, but may hurt findability of resources
- A - Z Sorting: Is it more important to have all the colleges together or to call them by their actual names? ("College, Honors" vs. "Honors College", etc.)
- CCU Example: HTC Honors College and Center for Interdisciplinary Studies was changed to Honors College.

Disclaimer: As I mentioned previously, I do indeed know that my own repository's Collections page is woefully inadequate, which is why I wanted to research what others do to either mimic best practices or learn common mistakes to avoid. Here are some suggestions to keep in mind to improve the page:

Plan for the long-term: It's tempting to start filling your repository with items without a long-term plan or strategy. But librarians love to organize, so let's do what we do best. My plan listed all possible communities I could identify at that time, with items in **bold** after they've been added to the IR, and in **red** when i plan to add them as collected pages. I came up with the perfect organization, then will try to accomplish it within the Collections Page constraints.

It's important for IR managers to occasionally revisit this page. We don't want to make it hard to find the great items we spent so much time curating. So revisit this page periodically, maybe annually.

For top-level events put "Conference" or "Symposium" or "Meeting" somewhere in the link text if possible.

Departments: Putting departments as top-level listings can quickly fill up the page and make it hard to navigate. I suggest putting them within a top-level Academic Departments listing if not within a top-level Colleges or Schools listing.

Don't be afraid to use Collected pages and Grouping tool to get things the way you want them.

Findability vs. Accuracy: Jargon may be more accurate, may not facilitate retrieval of items for non-experts.

A-Z suggestions: "College, Honors" or "Honors College"? Is it important to have all the colleges together? Something each manager needs to think about for their specific institution. Something that forces many to choose is the limit of 3-levels of hierarchy on the Collections page.

Example: HTC Honors College and Center for Interdisciplinary Studies - That's a lot to process! You can praise your sponsors all over every other IR page, but the main purpose of the Collections page / breadcrumbs / URL slugs is to aid in navigation.

Suggestions for Improvement (2)

Avoid Repetition. One site had the following as top-level listings

- Faculty Awards and Honors
- Faculty Books
- Faculty Contributions to the Popular Press
- Faculty Dissertations
- Faculty Grants
- Faculty Patents
- Faculty Presentations
- Faculty Publications
- Faculty Research

Mapping out entire organization: Use “Exclude community from communities.html page”

Expand All / Collapse All: Expand All in early years, Collapse All when IR is robust.

Centers and Initiatives

- Centers
- Initiatives
- Institutes
- Labs
- Partnerships
- Interdisciplinary Initiatives

Formal, Informal, Communities of Practice...

Associate with Academic Units when possible

Avoid over-representing: Collected pages are great, allowing you to show things in two or more different areas. But each new line is another piece of info for users to process, so try not to overdo it. Example: Multiple links to the same pages can cause confusion if overdone. I would suggest making a Faculty Scholarship top-level listing and just put these in there.

Mapping out entire organization: I’ve seen IRs do this, where they list all their departments and it looks like they have a ton of items in the IR, but most of the pages are empty. This is frustrating for browsers. So don’t forget to check the “Exclude series from communities.html page” box in the configuration page of each structure if there are no items to show.

Expand All / Collapse All: It makes sense to me to have Expand All as the default when the site is smaller and newer, but as the site grows, I suggest moving to Collapse All to aid in navigation. This is in your overall IR config page.

Centers and Initiatives: This is a good catch-all place for the variety of academic work that goes on at our institutions, and seems to include all 3 of the community types we talked about earlier: Formal, Informal, Communities of Practice. I suggest they be placed within or linked to Academic Units when possible. Sometimes that’s not possible, with things such as interdisciplinary ventures.

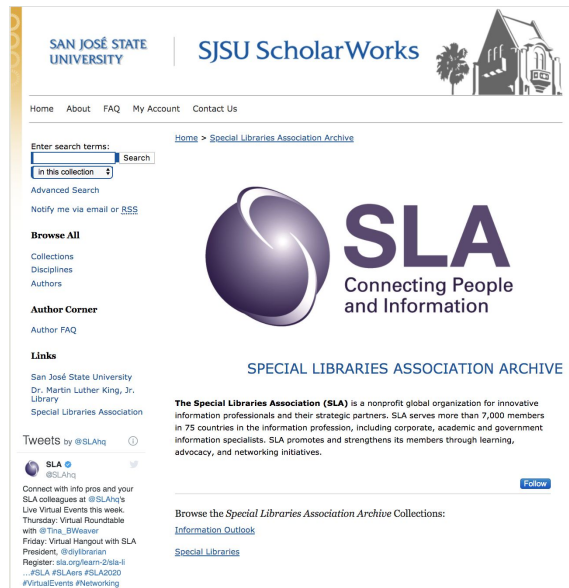
“Miscellaneous”

Special Libraries Association Archive - Contains scans of back issues of two journals.

How would you categorize this for access?

- Digital Collection?
- Journal?
- Affiliate/Partnership?
- Initiative?

SJSU made it a top-level community structure containing two book galleries.



Some materials seem like they could fit in multiple places, so it's difficult to find out where they belong.

“Aboutness” of an item: This seems like a digital collection to me. But it's left open so articles from a future partnership could fit here, as well as current journal in a journal structure.

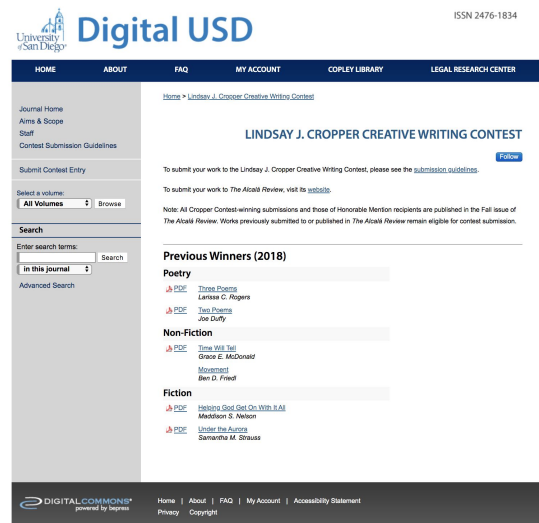
“Miscellaneous” (2)

Lindsay J. Cropper Creative Writing Contest -
Contains materials from previous contest winners, link to submission guidelines.

How would you categorize this?

- Journal?
- Student Scholarship?
- Initiative?

USD made it a top-level series structure.



USD creative writing contest: Is this more a journal, student scholarship, an initiative, or even just a means to submit materials?

Students submit work here, but then publish at a different journal site. Previous winner articles are listed here. We have something similar where we publish winning publications and presentations from our Undergraduate Research Competition in our Bridges journal of student research, but we have Bridges in our IR, so this is a bit different. They have a website for their student journal, which uses Submittable, so I'm not sure why there are two ways to submit, but I like how they feature student work on the IR. But unlike our previous example, it would take some revision to be able to also put an event structure in one all encompassing page, if they start having a Contest Winner Award Event for example.

“Aboutness” of an item: Seems like it might fit in a Student Scholarship community.

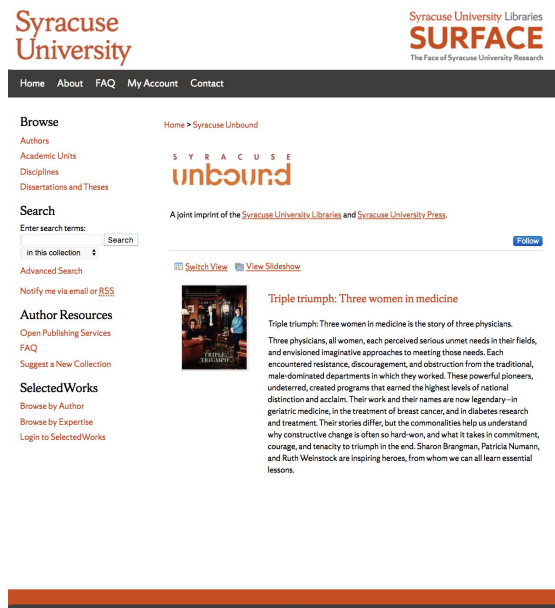
“Miscellaneous” (3)

Syracuse Unbound - Collaboration between the university press and university libraries to share stories of women in medicine.

How would you categorize this for access?

- OER?
- Digital Collection?
- Initiative?

SU made it a top-level book gallery structure.



Syracuse Unbound: Collaboration between University Press and University Libraries to share stories of women in medicine and the barriers they faced before becoming internationally recognized experts. Is this an OER, Faculty scholarship, University Press material, University Publications material, an Initiative? You can see how difficult some of these materials are to categorize.

“Aboutness” of this item: Although this is an initiative to publish an open access book, it seems like a single publication that can later be put under an OER top-level community.

This item is about medicine and doctors so should be easy to put in College of Medicine for example, but for interdisciplinary articles that could be a problem. I’ve seen where managers place the article in the department of the “corresponding author” the one who fills out the submission forms. It’s tricky.

Limitations

Three-level nesting limit

Pages appear once

Hidden collections

Staff / Funding dedicated to the IR

Structure title mirrored on Collections page

Configurable labeling: Can we call something "X" on Collections page and "Y" on the actual collection or sub-collection home page?

Provide multiple means of access (NN/g 2014)

Create your own navigation to your heart's content with university webpages and libguides tables of contents, Collections page may not be the best way to handle robust organization of your repository.

Some of the limitations we deal with in establishing best practices for IR organization are ...

Three-Level Limit: 1) Top-level communities; 2) Sub-communities; 3) Publications. I wonder what would change if we could have, say, 4 levels...? Eprints has 5 or 6 possible levels I think, but it's not ideal style-wise.

Pages can only appear once. We've seen workarounds, with the Collected pages and Grouping tool, but that's still a constraint.

Hidden collections: I was unable due to time constraints to examine all the sub-collections in every heading of every one of the 102 institutions, so there are most likely more things, just hidden in mislabeled headings. Example: One IR's Office of University Advancement listing only contains journals, which is interesting. So again, this project only really focused on Top-Level Communities, and what is useful to know about others' organization of them, and is not definitive but rather working toward a foundation of best practices.

Staff / Funding dedicated to the IR: This varies across institutions. Some limitations seen in this project are no doubt due to a lack of staff time, institutional / faculty buy-in, and many other vagaries of specific institutions.

If you change the structure title it will show exactly that way on the Collections page, so you can't do something like Honors College on the Collections page, then expect

to spell out the entire formal name HTC Honors College and Center for Interdisciplinary Studies on the home page. I wonder if this could ever change in future releases.

I started as a web developer, and so I keep in mind that best practices call for providing multiple means of access, to satisfy the preferences of the variety of search strategies. Some like known item / keyword searching, some want to browse immediately, some are more visual, some like faceted search strategies. Digital Commons provides this with the Collections page as a browse strategy, and also has hierarchical keyword searching, then facets to further narrow results. But you may have to also create LibGuides guides and so on to provide the exact IR browsing experience you want.

Homepage Browse Options

Browse Digital Commons @ ACU [Follow](#)









Types of Works
Explore the types of works available:

Faculty Work [Go](#)

Explore Subject Areas
Browse Faculty and Student Work:

Faculty Publications by College/Departme [Go](#)

PDXScholar makes the research, scholarship, and creative works of Portland State University faculty, staff, and students easily discoverable and available to anyone, anywhere in the world.

 SelectedWorks Faculty Profiles	 Dissertations & Theses
 Schools & Colleges	 PDXOpen Textbooks
 Centers & Institutes	 Conferences
 Special Collections & University Archives	 Journals

BROWSE RESEARCH AND SCHOLARSHIP [Follow](#)



Tackle Box

Each year, students enrolled in EDUC411/412, Literacy in the Content Areas, learn the theories and best practices for empowering all students to read, write, and learn more effectively.

[View More](#)



(ACU, Missouri S&T, PDXScholar, Augustana)

Homepage browse options are becoming more sophisticated too, having more options on the IR puts less pressure on you to design a robust Collections Page.

Drop-down menus

Icon grids

Links with thumbnails

Slideshows

Further Research

Does institution size dictate number of IR items?	Repository manager intentions
How does institution size affect Collections Page organization?	OER participation
Do Collections Page revisions affect access to materials?	No Collections page
Bepress vs. Other IR workflows	Website vs. IR
Digital and Historical Collections types	"Hide from Communities.html page"
	OpenDOAR / ROAR data

The interesting thing about this project is that it led me to ask many questions that would take more intensive research projects to answer. I only needed to spruce up my Collections page, but there are so many interesting things that I now want to find out:

This also could mean, "Does a higher research level result in higher research output in an IR"? This would involve sorting by Carnegie Classification, then determining rubric for "research levels" or something.

Does the size of an institution affect how they organize their IR? Do larger institutions have more Collections page entries, more or less top-level entries? Do R1 institutions organize differently than smaller ones? How many Digital Commons IRs are R1s? etc.

Do Collections Page revisions affect access to materials? So after implementing all of the best practices I learned during this project will it increase the findability of items? This is difficult, you would need to pull numbers from two iterations of the page and compare but not add any items in the interim.

Bepress vs. Other IR workflows: Does a more stringent workflow / policy increase best practices? bepress IRs make you go through an account rep, does this make IR managers hesitant to go against best practices? I predict that if I looked through 100 DSpace sites I'd see more variety of practice, which is not necessarily such a good thing for site users. It would be neat to interview IR account reps and see how internal best practices for Collections Pages have evolved throughout the years.

Digital and Historical Collections: I'm surprised at how many institutions use Digital Commons for cultural heritage collections, given the relative legacy look of the gallery apps like the Slideshow View. These types had a count of 127 in this list of 102, and there are obviously hundreds more "hidden" within subcollection structures. Could this be because these institutions don't have "digital collections" systems, or is there another reason? Our institution preferred at least at this time to put all our collections in the same system vs. having them in Islandora, CONTENTdm or another DAMS.

Repository manager intentions: I'd like to interview repository managers to see what their intentions were for why materials were put in one place or another. It's difficult to find managers' intentions without interviewing them.

OER participation: Only 37 out of the 102 had OER as an entry, which is 1.5% of this list. This seems low, is there a correlation between having an Open Access IR vs. a third-party IR like bepress? Future research between DSpace IR counts of OERs vs. DC counts could be interesting.

No Collections Page: Some institutions disable the page altogether (or at least hide the link to it). Why is this?

Website vs. IR: How closely do entries mirror the organization seen on the institutional website? Or, do web developers organize differently than repository managers, and if so, why?

Hide from Collections Page: What collections/publications are hidden from the page? And why? This would take a lot of detailed research. Would need to start with a data dump of the "Hide from Communities.html page" entries in the config pages, but may surface interesting practice.

OpenDOAR/ROAR data: Further research with non-bepress IR data, such as OpenDOAR repository numbers or ROAR numbers would be interesting to see.

Questions and Comments

sbacon@coastal.edu

References

SPARC and communities: https://sparcopen.org/wp-content/uploads/2016/01/IR_Guide_Checklist_v1_0.pdf

Mercer et al. 2007: www.emerald.com/insight/content/doi/10.1108/10650750710748496/full/pdf

NN/g, Nielsen Norman Group: <https://www.nngroup.com/articles/search-not-enough/>

TMI photo: <https://pixabay.com/users/aitoff-388338/>

Man on Horse photo: <https://www.digitalcommonwealth.org/search/commonwealth:rn301x250>

