

Just How Open is Open Access

Exposing Problematic Elements in bepress' Digital Commons Discipline Taxonomy for Institutional Repositories

**by Roger Chabot and Emily Carlisle
University of Western Ontario**

INTRODUCTION

The modern Open Access movement has flourished in part because of the internet as a low-cost and reliable technology for information dissemination as well as through continuing negative pressure from the ongoing serials crisis. Now, open access is built into strategic plans and budgets, encouraged and advocated for by scholarly communications departments, and supported by technical infrastructure, namely library-hosted institutional repositories. All of this is done with at least the following goals in mind:

- 1) To increase the visibility and impact of institutional research
- 2) To make publicly funded research available to communities beyond the Academy
- 3) To advance society through innovation
- 4) And to facilitate equitable access to scholarly materials

Today we will emphasize facilitating equitable access.

By its widely-agreed upon definition, as outlined in the Berlin Declaration, open access has two major provisions:

- 1) to provide unrestricted public access to “a complete version of the work and all supplemental materials,
- 2) and to do so without any meaningful copyright restrictions.

Our focus is on the first provision—the public access element of “open access.” While relaxed copyright and reuse restrictions certainly contribute to freedom and openness in research, the first provision can exist independent of the second, and, is at times treated as such. But are open access materials as accessible as we think? Is research openly accessible simply because it exists without price restrictions? Not exactly. And such was the motivation for our project.

In a 2017 paper exploring open access' two separate but intertwined provisions, Samuel Moore noted: “Forms of OA that prioritise access are often, though not exclusively, associated with repositories.” Moore goes on to state that: “digital technologies [like repositories] enable a more effective way of sharing research such that everyone with a stable internet connection should be able to access it.” In other words, digital

technologies such as repositories are seen as the key to enabling open dissemination of research.

With this in mind, we examined the efficacy with which bepress' Digital Commons repository facilitates public access to research. We recognize that Digital Commons is not the only repository software—with DSpace, EPrints, Fedora, and Islandora also being popular. But the Digital Commons platform is used by 500+ institutions worldwide, with The University of Western Ontario being one of them since 2008. This makes it worthy of study.

Let me preface this talk by saying that our work is not motivated by Elsevier's recent acquisition of bepress, which was met with quite a bit of shock, upset, and commentary from our community last summer. That said, our critical analysis of the platform is timely, given the acquisition and the critical conversations that have followed from it.

Our critical analysis focused on available default metadata fields, namely the bepress discipline list: a three-tiered taxonomy of academic subject areas used to tag content according to subject matter and to enhance discoverability. This is a necessary place to begin an assessment of a repositories' efficacy at facilitating public access to scholarship. After all, as April Hathcock proposed in a 2016 panel on "Open Scholarship for Social Justice," open access to scholarship does not mean findable and discoverable scholarship. She notes:

“... Not all open scholarship is treated equally.... Just because work is open doesn't mean that it will be found, valued, or validated... Open scholarship can and does replicate some of the biases inherent in academia and our society as a whole” (Hathcock, 2016).

We have found these biases in the DC discipline taxonomy, where Anglo-American centrism and imperialist privilege quite literally “other” religious, racial, and ethnic minority groups. Four disciplinary-focused case studies in particular have led us to ask: in our haste to react to rising journal subscription costs and to treat research as a public good, have we failed to properly assess the tools used to forward our public access mission?

AGENDA

The next twenty plus minutes will be dedicated to exploring this question. We begin with a review of the existing literature on open access publishing and the moral nature of classification

This will be followed by our critiques of the bepress digital commons taxonomy where we'll argue that bepress' Digital Commons works against the ideals of open, equitable knowledge sharing. The very ideals it intends to serve. We'll finish with a concluding section where we review the obstacles and solutions to the problem we'll identify.

LITERATURE REVIEW

Two streams of research are important to this line of thought: existing criticism of open access publishing and criticism of library classification schema.

First, commentary on structural and social inequality within open culture isn't new. Criticisms have been made with regard to open culture's perceived ability to catapult international development. In 2000, Ingrid Burkett identified five assumptions that have—and continue to be made—about the impact of information on international development which are noted here:

1. Give the poor a computer and they will move from being information poor to information rich.
2. Information inequality is a North/South issue.
3. Access to more information enriches people's lives.
4. The 'information society' will be more democratic and participatory.
5. Given enough information we can solve all the world's problems. (p. 680)

These assumptions, which have sometimes become embedded in open culture discourse, oversimplify social and economic global inequalities while creating a dichotomy that presupposes superior and inferior ways of knowing. Speaking to this very point, Ulrich Herb wrote in 2010: "The assumption that the mere availability of information would lead to democracy is a myth. It is also a myth to think that...the Internet would lead to a levelling of asymmetric allocations of power."

The literature also suggests that there exists an assumption that access to information (especially the West's information) can solve social inequalities by altering colonialist histories and long-term structural oppression. What this assumption fails to take into account is context—how meaningful the shared information is on a global scale. Instead, it presupposes that Western thought is innately valuable and superior, so as to justify its spread.

Now, on the other side of our issue we have long-standing criticisms of library classification schemes. For library practitioners we have long recognized that "the architecture of classification schemes is simultaneously a moral and informatic" task. The profession's widely-used tools for the organization and access of knowledge, like the Library of Congress Classification, LCSH, and the Dewey Decimal Classification, have long been recognized as problematic in how they refer to (or neglect to refer to) marginalized and vulnerable peoples, or as contributing to the legitimacy of the problematic hegemonic power structures that are a cause of such conceptualizations.

The literature dedicated to library practice is replete with these critiques, which remain an important avenue of investigation today. Perhaps the most notable critique is Sanford Berman's *Prejudices & Antipathies*, in which Berman points out the biases in favour of

white, Western, Judeo-Christian culture in LC Subject Headings. Noting problems with more than 250 terms, Berman offered recommendations for improving them and for generating a more inclusive subject tool. Steven Knowlton's 2005 followup study, notes that while there has been some progress in remedying the issues in LCSH, eighty items from Berman's 1971 list remain unchanged 30 years later.

Other scholars have taken up similar banners, to the present day. Their work provides more pointed analyses concerning representations of particular groups of people in classification schemes, including (but not limited to) women, "East Asian" peoples, transgender peoples, and Indigenous peoples.

These two streams of thought (criticisms of open access and classification schemes) come together in our project. Although intended as a tool to progress open access and to facilitate equitable access to scholarship, the Digital Commons platform comes equipped with an ethnocentric discipline taxonomy with which users are to label and classify works. What does this mean? Works by or about groups that don't conform to the Western norm are rendered less visible on the web, which ultimately works against our community's open ideology.

BEPRESS TAXONOMY

For some context: bepress' three-tiered taxonomy contains academic subject headings used to tag and enhance discoverability of content in Digital Commons repositories. It was initially developed in consultation with existing taxonomies, like the Medical Subject Headings and the Current Index to Legal Periodicals (CILP) among others on the slide.

1. Taxonomy of Research Doctoral Programs from the National Academies
2. Classification of Instructional Programs, 2010 edition, from the National Center for Educational Statistics (NCES)
3. Medical Subject Headings (MeSH) from the National Library of Medicine
4. Law subject headings are informed by the Current Index to Legal Periodicals (CILP) and FindLaw
5. Several Business categories come from Cabell's
6. The University of California's list of departments and programs

The taxonomy is a somewhat living document, with currently more than 1000 unique categories. Suggestions for improving the list are taken on an ongoing basis, though changes can only be made by repository administrators. Criteria for accepting those suggestions, and the frequency with which updates are made to the taxonomy, is not clear.

Bepress touts the discipline field as necessary for search engine optimization, and for good reason. The discipline field is mapped to Dublin Core element "dc: subject." Keywords are also mapped to this element, so both serve a similar function to search engines and other web harvesting tools. Of course, maximum discoverability comes with including keywords and disciplines in an item's metadata. However, the Disciplines

field also ties an item into the Digital Commons Network, which contains the works from all Digital Commons repositories, grouped together by discipline. As a result of this:

- 1) individuals works are displayed in multiple places online, increasing general discoverability, and
- 2) more backlinks to the item are published which results in better search engine optimization.

In short, the Disciplines field is a powerful tool for search engine optimization and discoverability. That's why it's important for the terms that populate this field—which are drawn from headings in the discipline taxonomy—to properly describe a works “aboutness.” They need to do justice to the work's content and should exist in a form that users looking for work on that subject would search. But that can't be done if those terms don't exist in the discipline taxonomy.

Herein lies the problem: There aren't headings in the taxonomy that can appropriately or accurately describe works by or about many non-Western or minority groups. So, it becomes more likely that these works remain in the margins, unable to be easily found on the web because they aren't represented by accurate, appropriate or searchable terms.

CASE STUDIES

Four examples illustrate this problem, although these examples are by no means the only ones that we could have chosen to demonstrate this point. In particular, we'll look at the problems in indigenous topics, religious topics, LGBTQ+ topics, and Occident-focused topics.

Let's start with representations of Indigenous peoples in the repository as identified as being associated with Indigenous works.

CASE STUDY 1: INDIGENOUS TOPICS

Three of those discipline fields were explicitly linked to Indigenous peoples: *Indigenous Studies*, *Indigenous Education*, and *Aboriginal and Indian Law*. *Indigenous Studies* exists as a subdivision of the category *Race, Ethnicity, and Post-Colonial Studies*, such that Indigenous groups are only recognized in the context of their relation to the colonizer. The problems with the language in the topic *Indian Law*, I hope, are evident.

With so few discipline fields explicitly for Indigenous works, efforts to classify works as Indigenous have resulted in many being placed in broad categories that fail to capture the Indigenous elements of the work. Those broad categories include *Canadian History*, *Multicultural Psychology*, *Social and Cultural Psychology*, —or “Other” categories like *Other Anthropology*, *Other Sociology*, and *Other Religion*. Works by and about Indigenous peoples are quite literally “othered”; they are given representation only in headings that suggest that they are distinct and lesser than the norm.

CASE STUDY 2: RELIGIOUS TOPICS

Next, moving on to issues in topics on religious studies. By far, the religious studies subtopic is dominated by Christianity-related categories, reflecting the prominence that Christianity has enjoyed in the societies of both Europe and the Americas. Christianity's prominence here echoes the Dewey Decimal Classification's use of the 200 to 290 classes for Christianity while relegating the rest of the world's traditions to 10 numbers—something that it has long received criticism for.

The taxonomy provides narrower categories for a variety of Christian topics that are not replicated in the categories for other religious traditions. While there exists a subtopic of *Biblical Studies*, there are no parallel categories for studies of holy books such as Talmudic Studies, Quranic Studies, or Upanishadic or Vedic Studies.

Also, prominent Christian sects are given their own unique subcategories in the taxonomy, with the existence of categories for *Catholic Studies* and *Mormon Studies*, and also a general division for *Christian denominations and Sects*. None of the other religious traditions mentioned in the taxonomy are provided with categories for even the most simple divisions of their faiths, such as the division of Buddhism into its Mahayana and Theravada varieties. Many smaller religious traditions are not accounted for at all, like Judaism Studies, which is notably absent as a unique subtopic to parallel Buddhist Studies or Hindu Studies or Islamic Studies (despite being a religious tradition of prominence in the West). The reality is that the great variety that exists within the world's religious traditions are simply not accounted for in the taxonomy and instead are potentially relegated to the category of *Other Religion*, as if to deem them less worthy of study than the others.

Finally, the presence of the subtopic *Missions and World Christianity* among the twenty or so topics within the subtopic of Religious Studies also betrays the Eurocentric, and potentially colonial view of the taxonomy's collective creators, viewing Western forms of Christianity to be “natural” while exoticizing the “other” Christianities in the world.

CASE STUDY 3: LGBTQ+ TOPICS

The bepress taxonomy's treatment of LGBTQ+ people is problematic in two ways.

First, LGBTQ+ peoples are generally invisible outside of the few topics of study which explicitly study LGBTQ+ peoples. To be fair, the taxonomy on the whole fails to consider people themselves as agents in any topic (as the “doers” or “enactors” of education, or medicine, or psychology, for example). This is problematic for all individuals, but vulnerable or minority populations such as LGBTQ+ peoples are especially diminished in the taxonomy.

The broadness in which LGBTQ+ topics find themselves is also problematic. Under the broader topic of *Arts and Humanities: Feminist, Gender, and Sexuality Studies*, LGBTQ+ topics are also combined with feminist concerns. This lengthy subject heading

is perhaps too broad to offer any visibility to the unique concerns and perspectives of queer people. While feminist scholarship is certainly related to LGBTQ+ scholarship, especially in work with an intersectional focus, each subgroup of the popular acronym merits their own section, just as Women's Studies has merited its own as a subtopic within this topic. Doing so would allow for more specificity, with headings that could be used to more accurately capture a work's subject.

Problems are also present in the broader *Social and Behavior Sciences: Communication* topic where LGBTQ+ concerns are present in the subtopic of *Gender, Race, Sexuality, and Ethnicity*. The topic in this case seems to act as a catchall category for all "minority" issues in Communications Studies. However, under Sociology, the subtopics of *Gender and Sexuality* are separate from *Race and Ethnicity* which each get their slightly narrower topics. Again, while minority and vulnerable groups may certainly share similar concerns, they also possess unique concerns which deserves some visibility through the existence of a dedicated category.

CASE STUDY 4: ANGLO-AMERICAN CENTRISM

We'll finally offer examples of the ways in which occident-focused, normalized topics are given privilege of specificity within the taxonomy.

Most glaringly, there's an entire section dedicated to *American Studies*, with further subdivisions for *American Film Studies*, *American Literature*, *American Material Culture*, and *American Popular Culture*. This is space and specificity not afforded to *Basque Studies*, *Korean Studies*, *Japanese Studies*, *Celtic Studies*, *African American Studies*—or any real group of people beyond Americans.

Similar American privilege is shown in the Politics categories. For example, American politics are the only politics recognized in the taxonomy—no other nation is afforded the same space and recognition.

Likewise, in the Law category, there are explicit headings for the First, Second, Fourth, and Fourteenth Amendments to the American Constitution, and another for the Supreme Court of the United States.

As a final example, in the English language and literature category, we see clear evidence of colonialist thought, in that the British Isles and North America are viewed as the only places where English literature is written, despite English being quite a global language. English literature from all other nations is grouped together, as if not worthy of being recognized by name.

DISCUSSION AND CONCLUSION

Now, for some brief discussion and conclusion going forward.

We could say that Western bias in the bepress discipline taxonomy is due to it being created in the United States, and largely used by institutions in the United States. But

that is hardly justified when it is also used by institutions in the United Arab Emirates, Pakistan, Singapore and Hong Kong, among others. Even within North America and Europe, it is naïve to think that research by/and about groups that are minoritized, underrepresented, and even neglected in the taxonomy are not researched and written about.

We could also say that maybe it doesn't matter—that authors can add keywords, should they wish, to describe their work most accurately. Yet, it is the discipline field (not keywords) that yields the more powerful search engine optimization.

Why shouldn't works by or about minoritized groups be given the same discoverability benefits as works by or about topics with a majority focus? As it stands, for searchers, "effective searching for marginalized topics [may] require greater ingenuity and serendipity than searching for mainstream topics." For authors, the onus is on those from marginalized groups or writing about marginalized groups to quite literally write themselves into visibility.

Now, we recognize that Digital Commons is not the only repository software out there. While we weren't able to delve into a detailed analysis of subject or discipline taxonomies used by any and all repositories, a cursory look shows that similar biases exist with platforms beyond Digital Commons. EPrints, for example, is configured by default with a subject hierarchy that is based on the widely-criticized Library of Congress subject tree, while Omeka also provides the option to embed a Library of Congress controlled vocabulary.

These taxonomies and classification systems are supposed to provide a controlled vocabulary that simplifies information organization and retrieval. But as the profession has recognized for decades now, that has come at the expense of minority groups, whose identities are underrepresented and rendered hidden while the classification schemes that replicate hegemonic power structures.

We've failed to be mindful of how the tools that facilitate open access are indeed reproducing systems of marginalization. We've failed to consider how they may, in fact, be failing to facilitate equitable access after all.

So what's the answer?

Perhaps it's abandoning the use of controlled vocabulary within institutional repositories. Sure, it facilitates more effective searching, but if it's contradicting the open access mission, then is it the best option?

Perhaps it's continuing to advocate for equality in representation, so that the bodies in charge of the taxonomies and classification systems have no choice but to make the necessary changes.

Or perhaps, it's exploiting the technology that we have at hand today to its fullest potential, creating our own sets of controlled vocabulary that address the problems we've recognized in existing ones.

These options require an investment of significant time and effort. They may require external consultation, internal and external collaboration, and technical expertise. But they may also improve the visibility of works by/about marginalized groups on the web, so that the tools used to advance open access are not simply spreading what has been deemed "ever-valuable" Western ways of knowing, but actually contributing to true knowledge exchange, and equitable access to all research on a global scale. So on that note, we'll leave you with this thought: When will time, cost, and effort cease to be an excuse for perpetuating archaic, discriminatory systems?