

Citations as Data: Harvesting the Scholarly Record of Your University to Enrich Institutional Knowledge and Support Research.

Leila Belle Sterman and Jason A. Clark

Many research libraries are looking for new ways to demonstrate value for their parent institutions. Metrics, assessment, and promotion of research continue to grow in importance, but they have not always fallen into the scope of services for the research library. Montana State University (MSU) Library recognized a need and interest to quantify the citation record and scholarly output of our university. With this vision in mind, we began positioning citation collection as the data engine that drives scholarly communication, deposits into our IR, and assessment of research activities. We envisioned a project that might: provide transparency around the acts of scholarship at our university; celebrate the research we produce; and build new relationships between our researchers. The result was our MSU Research Citation application (<https://arc.lib.montana.edu/msu-research-citations/>) and our research publication promotion service (www.montana.edu/research/publications/). The application and accompanying services are predicated on the principle that each citation is a discrete data object that can be searched, browsed, exported, and reused. In this formulation, the records of our research publications are the data that can open up possibilities for new library projects and services.

Introduction

At Montana State University (MSU)—as at many institutions—we spend a great deal of time counting research in dollars. What we do not spend much time doing is tracking and celebrating the research itself. We count grants applied for, grants funded, and grants renewed; we track million-dollar grants down to the penny and celebrate their existence. The discussion of grant money, at research institutions, is so pervasive that it is sometimes difficult to see other goals in the value system created around grant funding and tenure.¹ While research output is celebrated in academia, especially in the tenure process, it is often celebrated inasmuch as a publication in a high-impact journal will

**Leila Belle Sterman is Assistant Professor and Scholarly Communication Librarian, and Jason A. Clark is Associate Professor and Head, Library Informatics & Computing, both at Montana State University Library; e-mail: Leila.sterman@montana.edu, jaclark@montana.edu. ©2017 Leila Belle Sterman and Jason A. Clark, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.*

look good on the next grant application. Recently, we found that no one on our campus knew about our research publications in aggregate: the information was piecemeal at best, hidden on department webpages and out of date curriculum vitae (CV). We can get a general sense of our research publications with clues from grant funding and theses and dissertations, but research publications give us a tangible record of the impact on an academic field. We envisioned a project that might: provide transparency around the acts of scholarship at our university; celebrate the research we produce; and build new relationships between our researchers. The result was our MSU Research Citation application (<https://arc.lib.montana.edu/msu-research-citations/>) and our research publication promotion services (www.montana.edu/research/publications/).

Background

Data Gathering

In our first attempts at quantifying scholarship at MSU, we began a partnership between our Library and our Office of Research and Economic Development to collect and disseminate information about research publications from MSU faculty. Initially, the project was focused on the idea of promotion or providing a public story for our researcher's scholarly activities. This served as a useful frame, and we aimed to use this information to highlight faculty research output and gain understanding of the work produced on our campus. As we continued to formulate our idea, we realized there were all kinds of uses for the citation record of our university including: metrics, accreditation numbers, outreach, and facilitating data reuse. Interest from multiple parties continued to grow as these possibilities came into focus.

More specifically, these data about citations would be useful in the library; for example, when our Institutional Repository (IR) was created, the library included a metric in its strategic plan to "Optimize the ScholarWorks Institutional Repository (IR) to hold 20% of scholarly output by 2018." With no specific guide on how to measure what constituted 100 percent, we planned to look to the self-reporting through annual reviews: that process turned out to be delayed at best and hugely incomplete at worst. The university has recently implemented an aggregated Current Research Information System (CRIS) end-of-year review. It seemed that these data could be the answer. This process, however, was very lengthy, based on self-reporting, and incomplete. The volume of information and reporting deadlines delayed yearly data from 2014 until June of 2015 when the library received CRIS data as a large, disorganized spreadsheet. It was unreasonable to work through that spreadsheet quickly given our staffing. These data were disorganized and their quality varied greatly; even in parsed fields we received incomplete citations. Faculty members filling out the CRIS did not use the fields for citation information or used them improperly. Some faculty members did not even contribute their information to the database, although participation in the CRIS was tied to merit raises for the next year. Whole colleges refused to participate, leaving large holes in the dataset. Given the reluctance to participate in university-mandated reporting, it seemed unlikely that other types of voluntary reporting would be successful.

Data Mining

Similarly, we attempted to mine faculty curriculum vitae (CV). We found that citation styles were so diverse that we could not write a script to parse these data that took less time to correct than doing the same task entirely by hand. Some common practices and mistakes slowed this effort down (such as leaving a researcher's own name out of the citation or misplaced punctuation). Additionally, by relying on author-produced reporting, we only had information about forms of scholarship that the authors deemed important on that version of their CV. This method may have understated some forms

of publication such as white papers or overstated items such as short meeting abstracts, which we saw often reported with equal emphasis as research articles.

Planning

To know the scholarly output of the university, we needed a better strategy. The previous attempts to collect these data were inefficient, time-consuming, and incomplete. We needed an aggregation tool that used existing metadata attached to publications to discover, curate, and disseminate this information to our campus. We found no open source tool to collect the citation information from faculty publications (Symplectic Elements does a good job, at a price²), so the Montana State University (MSU) Library built a research citation application to capture citation information from various academic databases using RSS feeds and alerts (<https://arc.lib.montana.edu/msu-research-citations/>). The research citation application is built on the principle of treating each citation as an item that can be searched, browsed, exported, and reused—treating the citation as a discrete data object.

Literature Review

The pressures of the tenure process are often the impetus for academic publication, as much or more than the thirst for new knowledge.³ While knowledge creation is an aspirational goal, budget pressures on the academic system have created immediate demands for many universities. This helped create an accounting system that measures grant dollars monthly, as well as knowledge creation after five or six years through the tenure process. Although “evaluation of research outcomes are increasingly linked to the allocation of research funds,”⁴ at the institutional level, funding is perceived as paramount.

To prevent institutions from counting grant dollars that will lead to no productive research but deepen university pockets,⁵ metrics for counting research output and systematically evaluating that research have increased in recent years, especially in Europe.⁶ In many instances, this task has found a natural home in the library. Research services at many university libraries are centered on helping researchers find, analyze, and use scholarly information. In recent years, this has expanded to include services such as data and software support, data management, and digital preservation of research objects and publications in response to faculty pressure to produce publications and grant applications.⁷ Less frequently, the library is a source of original information about that research and the people who are involved. The possibilities of this meta information are compelling for administrators and research offices. Libraries are working to increase the impact and reach of the publications produced at their institutions; recently, “university libraries, at least in Europe, are also increasingly focused on the development of knowledge and services related to scholarly communication other than simply searching and retrieving scholarly information.”⁸

Metadata about research activities are often compiled in a CRIS or by an Institutional Repository. Researchers themselves are often not interested in this high-level view: Foster and Gibbons observe that “their benefits seem to be very persuasive to institutions, IRs fail to appear compelling and useful to the authors and owners of the content.”⁹ They further note, “The term ‘institutional repository’ implies that the system is designed to support and achieve the needs and goals of the institution, not necessarily those of the individual.”¹⁰ This is largely the justification for libraries performing the bulk of the work for green open access and not relying on self-archiving or self-deposit. It is also a reasonable conclusion for faculty that they should not spend their valuable time working on projects from which they do not see direct or immediate rewards.

To decrease the burden on faculty members and increase both data collection and repository content, multiple projects have automated parts of the process. Projects like IncReASe,¹¹ TARDIS,¹² and DAEDALUS¹³ have all worked to automate the ingest process for repositories, attempting to relieve researchers of a task and increase repository items to a critical mass of items and, thus, importance. SHARE¹⁴ is building a free metadata set that reaches across the research lifecycle to make digital items more easily discoverable and reusable, creating useful and accessible content out of available materials. The library may gain a positive reputation on a campus by becoming a content producer, in addition to a content distributor. McIntyre, Chan, and Gross¹⁵ write about library as publisher and the added value of that work. In their work attempting to understand Institutional Repository success and impediments, Foster and Gibbons¹⁶ studied University of Rochester faculty and found that, regardless of discipline, faculty resented any activity that took time away from research and writing time. Specifically, they resisted duplicated efforts,¹⁷ and, although they used digital tools extensively, they were unconcerned with the tool, just its usefulness. Additionally, faculty do not perceive the benefits of IR deposit or self-reporting, especially for the reasons (preservation, metadata, access, open access/source) that librarians often use to advocate for IR content.¹⁸

While faculty members dislike duplication of effort, services that reduce effort and increase rewards are seen favorably. A researcher's response to appreciation is amplified based on the pressure to succeed, finding "tasks and appreciation related to their publications more rewarding and tasks and appreciation not related to their academic publications less rewarding" compared to colleagues under less pressure.¹⁹ A small task can have a large positive impact on campus culture.

From an institutional perspective, increased knowledge about citations allows for rankings, benchmarking, and metrics to compare universities²⁰ potentially creating competitive goals and fueling increased productivity and efficiency. Increased knowledge of publications also allows institutions to be active participants in the promotion and dissemination of the knowledge created on their campuses.²¹ This promotion could increase impact of research, help attract students and faculty, and improve an institution's reputation.

Libraries are a natural home for this task, as they are usually positioned centrally, unbiased departmentally, and have experience with metadata, citations, and scholarly databases.

Bibliometrics are central to the libraries' role in the current ecosystem of scholarly communication; thus, librarians already have the knowledge and skills to perform these tasks.²² Further, these activities bring attention and prestige to libraries, a factor that some feel libraries have lost as the search and discovery of academic materials increasingly needs no mediation from a trained library and information science professional.²³

Methodology

Collection

This project began with librarians creating e-mail alerts to collect publications with the words "MSU," "Montana," or "Montana State University." We chose databases and journals based on our library subscriptions and open access resources that indexed popular journals. This produced an extensive list of new publications from the past six months. The initial collection was rather time-consuming, as we were sorting through alerts, discerning the quality of the provided metadata, and narrowing the results by appropriate date range. Once we picked the most reliable sources, we set up RSS feeds so the metadata collection was more organized and more easily managed than in e-mail alerts. These RSS feeds are now routed into our application, explained further below.

This project aims to collect metadata about publications based on these criteria: the publication must contain stable, peer-reviewed, scholarly content; be from the current month or the prior six months; and have at least one Montana State University–affiliated author. This means that items such as conference abstracts, book reviews, or letters to the editor are not included. This may vary for different institutions; our goal was to align with the collection scope for our IR. The collection development policy of our IR defines our scope as content that is: scholarly, in a researcher’s field of expertise, an end product (not a work in progress), stable content, deliverable on the web (downloadable from our IR), authored by an MSU affiliate, vetted/refereed, and “a work unto itself” –not short reviews or abstracts.

Database Management

Once we figured out how to gather and sort the metadata effectively, we had to decide how often to publicize the results. From September 2014 to December 2014, we put out weekly e-mails to a small test group of people including the Dean of the Library, the Director of Communications, and the Vice President for Research. This frequency was scaled back to monthly announcements once we promoted this research to the whole community. Weekly e-mails, it seemed, would become annoying to recipients and took a considerable amount of time to produce. We now produce an e-mail that the Vice President of Research, in partnership with the library, distributes to campus monthly. It is a popular and well-read dissemination. Anecdotally, it has encouraged faculty members to congratulate each other and discuss research projects. The monthly output of research from MSU averages about sixty publications. In the early stages, we tried to adhere to the date of first online access or first online publication. Some journals, however, only post the date of physical journal publication, which could be half a year away. This became a problem, as a complex assessment could be necessary to discern when a publication should be promoted. We also found that authors did not care when the “official” date of publication was meant to be—they felt that their papers were published as soon as they were visible online. We modified our monthly inclusion criteria to accept all publication we were notified about in each month. This has worked well.

The library collaborated with the university communications department to produce a template for the monthly e-mails that is attractive and easy to read. The titles in that e-mail link to a communication department–designed webpage that displays the month’s publications and has a searchable database of the collected publications (www.montana.edu/research/publications/). This page also features layperson titles and abstracts for selected publications. These publications are chosen each month based on upcoming limited submission grant applications, university events, or global topical relevance. We highlight any media associated with these publications so that they are eye-catching and easily shared or reused on other university platforms and social media.

Responses to the e-mails fall into two categories: 1) appreciation for the service; and 2) notifications that we have missed a publication. The respondents in the first category enjoy the e-mail based on their new understanding of current research at the university, the pride they feel based on their personal inclusion or their department’s inclusion in the list, or the fact that the university is publicly showing that we value research publications. The second category is not viewed as a negative response. Instead, we encourage faculty members to inform us if we have missed a publication. There are clear indications of how the information is collected and whom to contact should we miss anything. This process has helped us collect publications that we would have missed otherwise. It has also begun to change the institutional culture: some authors now inform the library when they publish a paper without waiting to be missed.

Once we have collected publication metadata, our system easily exports it into a spreadsheet (.csv), text file (.txt), json (.json), or XML (.xml) file by month. This format is useful for compiling the correct information for the next two steps of this process: 1) e-mailing the author(s); and 2) depositing an item into our IR. Before e-mailing authors, we check for copyright information in Sherpa Romeo²⁴ and on journal webpages. With that information, we send an e-mail to the authors of each article asking for the correct copy, as determined in the Romeo service, to post in our repository and congratulating them on their accomplishment. The response rate is just over 30 percent positive responses on the first e-mail. Some authors struggle with the pre- and postprint concepts, or with the thought of posting a less than perfect version of a paper. This learning moment is also a cultural change for many authors and will undoubtedly take time. Additionally, some papers are not able to be posted in the IR. We still send an e-mail congratulating these authors.

Impact of Service

The library was not the only unit on campus interested in increasing our understanding of publication information on campus. During each stage of software development, outlined below, new partners with a desire for these citation data emerged. Our Research Office, the Office of Planning and Analysis, and department heads were all interested in publicizing research output: the citation information we gather will be used to celebrate research, track the outcomes of grants, and integrate with assessment tools. It was clear that these data could add value once collected, but we had to figure out how to make these data harvestable and reusable. We turned toward a structured data and Application Programming Interface (API) solution to help our potential partners.

This service has increased the number of deposits in our IR to just over 30 percent of the university's published research from the past year and broadened the number and disciplines of researchers represented in the IR. Scholarly Communication librarians are often found approaching researchers one by one advocating for open access, collecting CVs, talking to departments at meetings, and imploring researchers to share their three-year-old postprints. This is time-consuming and may create one-time interactions that are difficult to translate into enduring or fruitful relationships. Borrowing the marketing term of a "pain point,"²⁵ we chose instead to contact researchers at a "celebration point." If we can contact, usually by e-mail, an author and ask for the appropriate version of an article at or near the time of publication, we have found we are much more likely to be able to add that item to our IR. The author is more likely to both have a preprint or postprint (if that is the appropriate version to post in the IR) and more likely to be interested in spending the small but burdensome-feeling amount of time to find the postprint and send it to the library. Contributing to the IR seemed to some faculty members to be tedious, extracurricular, time-consuming—in short, it was hard to understand the purpose. Our use of the "celebration point" method has made the process incremental for faculty and simplified the task. This has greatly benefited our repository. And the service has been impactful outside our scholarly communications services. We have also used the citations to provide recruitment leads for our Data Management Librarian, as we can identify research that is likely to have large amounts of data or datasets resulting from the research process. This is a good outreach resource for our Data Management Librarian, since a dataset that is unpublished can be hard to find. By informing our colleague of new publications, she can contact the research team and discuss data management and preservation with them.

It is also worth mentioning that this information populates more than our IR. There is a searchable database run by University Communications that with a few manipulations can also populate college webpages on the university website. This relieves

administrative burden from each of the colleges and ensures that the publications that display on each page are current. Additionally, the database has been queried to facilitate grant applications (for example: *What publications about Yellowstone National Park in any discipline have been produced by MSU in the past 15 years?*), in accreditation processes where colleges were interested in their own publication record and had no other way of easily compiling a current list, and by community members hoping to better understand the research a department is currently producing.

Technical Overview

The app itself is a reworked version of our digital library software.²⁶ The digital library software was built to house and maintain our mostly image- or text-based digital collections. Refining the software to feature citation metadata as the primary digital object required some significant changes to the data model, our data views, and our data management interface. Additional author, affiliation, and database tables were added to provide a more complete citation data model. In our data view work, we built a search-and-browse interface to allow for discovery of the scholarship inventory, as well as a citation item view that displayed a detailed citation record. However, we quickly recognized that the main purpose of the app was administrative and oriented toward metadata management. Our main goals were to ingest data feeds and provide access to these data in various formats with an eye toward the “celebration points” mentioned above; the display aspects were secondary. Underlying the app is an Application Programming Interface (API) that allows us to create these various structured data formats, including: text files, csv files, XML files, and JSON files. A picture of structured data from our API appears below.

It is possible to switch between output formats by supplying a different value to the “format” key in the URL (for instance, “format=xml” would get you an XML ver-

FIGURE 1
Example API Output and URL from MSU Research Citations Application

```

{
  - items: [
    - {
      - item: {
        recordInfo_recordIdentifier: "1304",
        titleInfo_title: "Differential Type I Interferon Signaling Is a Master Regulator of Susceptibility to Postinflue
        abstract: "Bacterial superinfections are a primary cause of death during influenza pandemics and epidemics. Type
        demonstrates that the reduced susceptibility to methicillin-resistant Staphylococcus aureus (MRSA) at day 3 post-
        previously reported was due to interleukin-13 (IL-13)/IFN-γ responses, is also dependent on type I IFN signaling
        for protective IL-13 production. We found, through utilization of blocking antibodies, that reduced susceptibili
        infection was IFN-β dependent, whereas the increased susceptibility at day 7 was IFN-α dependent. IFN-β signaling
        required for MRSA clearance, whereas IFN-α signaling late in infection was not, though it did mediate increased
        time. Type I IFN receptor (IFNAR) signaling in CD11c+ and Ly6G+ cells was required for the observed reduced susce
        infection. Depletion of Ly6G+ cells in mice in which IFNAR signaling was either blocked or deleted indicated that
        for the IFNAR signaling-dependent susceptibility to MRSA superinfection at day 7 post-IAV infection. Thus, durin
        differences in type I IFN signaling increased bactericidal activity of both CD11c+ and Ly6G+ cells at day 3 and
        Ly6G+ cells at day 7. The temporal differential outcomes induced by IFN-β (day 3) and IFN-α (day 7) signaling th
        differential susceptibility to MRSA at 3 and 7 days post-IAV infection.",
        extension: "",
        originInfo_dateIssued: "May 2016",
        citation_url: "https://gateway.ezproxy.knowledge.com/gateway/gateway.cgi?
        QWVersion=26&rcAuth=Alerting&rcApp=Alerting&DestApp=MEDLINE&DestLinkType=FullRecord&KeyUF=MEDLINE%3A27143388",
        relatedItem_identifier: "10.1128/mBio.00506-16",
        relatedItem_identifier_isbn: "",
        relatedItem_identifier_pmcid: "",
        relatedItem_location_holdingSimple_copyInformation_note: "",
        relatedItem_relatedItem_location_holdingSimple_url: "",
        relatedItem_titleInfo_title: "mBio",
        relatedItem_part_detail_volume: "7",
        relatedItem_part_detail_number: "3",
        relatedItem_part_extent_start: "e00506",
        relatedItem_part_extent_end: "e00516",
        relatedItem_part_text: "",
        relatedItem_extensions1: "Chemical & Material Sciences",
        relatedItem_extensions2: "Life Sciences & Earth Sciences",
        relatedItem_extensions3: "",
        relatedItem_accessCondition: "Open Access",
      }
    }
  ]
}

```

(https://arc.lib.montana.edu/msu-research-citations/api.php?v=1&date=2016-06&format=json)

sion of these data). These multiple formats are key to the reuse of our data and are the linchpin that allow our data to travel into other venues such as our monthly e-mail and the HTML pages of the Office of Research and Economic Development.

Our source data-feeds from the various vendors proved to be one of the more interesting challenges of the project. Typically, these were RSS feeds with “msu” and “montana state” pattern matches in the “affiliation” fields of publication metadata. You can see the various source data-feeds in our public listing of the most recent research from MSU in figure 2.

FIGURE 2
Web Page in MSU Research Citations Application Listing Most Recent MSU Research Using Multiple Vendor RSS Feeds

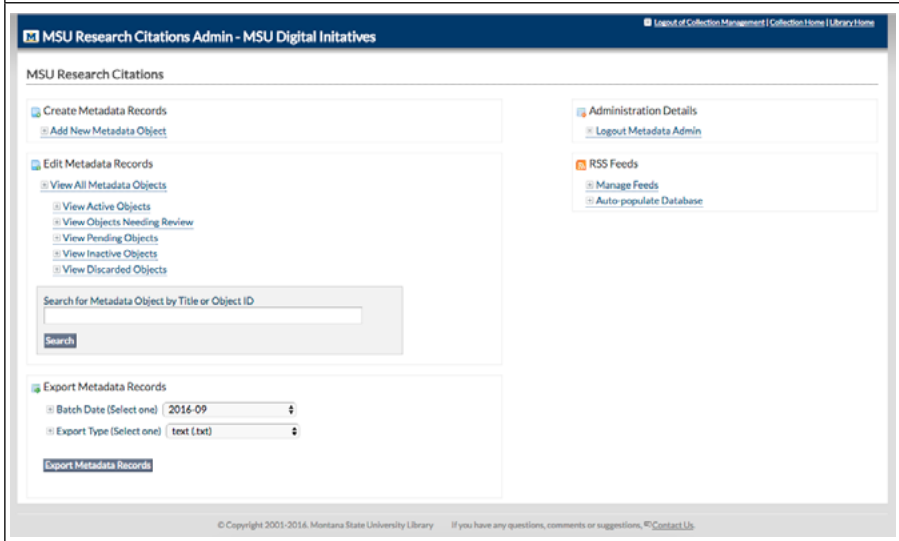
(<https://arc.lib.montana.edu/msu-research-citations/feeds.php>)

The feeds are parsed using a combination of Javascript and PHP scripts and then presented as these lists for a content editor to review. Upon review, each citation is pushed from these lists into a local database where we can store and preserve the metadata as a local record.

These feeds varied in how they applied the RSS standard and sometimes contained duplicate data. We quickly realized that one of the core functions of the app would be normalizing these data, grouping it with date and timestamps, and verifying MSU authorship.²⁷ There are several ways to normalize the RSS feed data. In our initial version of the software, we were pushing the feeds into the Google Feed API, which returns a standard, consistent JSON data format for us to work with as baseline data for the application. Over time, we introduced our own feed parsing script to have a bit more control over how the RSS feeds are normalized and to remove our application dependency on an external service that was discontinued by Google. The review and ingestion interface became a priority, and we brought in one of our lead developers, James Espeland, to work through the design and implementation. The first requirement was an interface that let content editors review, deduplicate, and group citations into batches by date and per a specific output format. Figure 3 is a snapshot of the review interface.

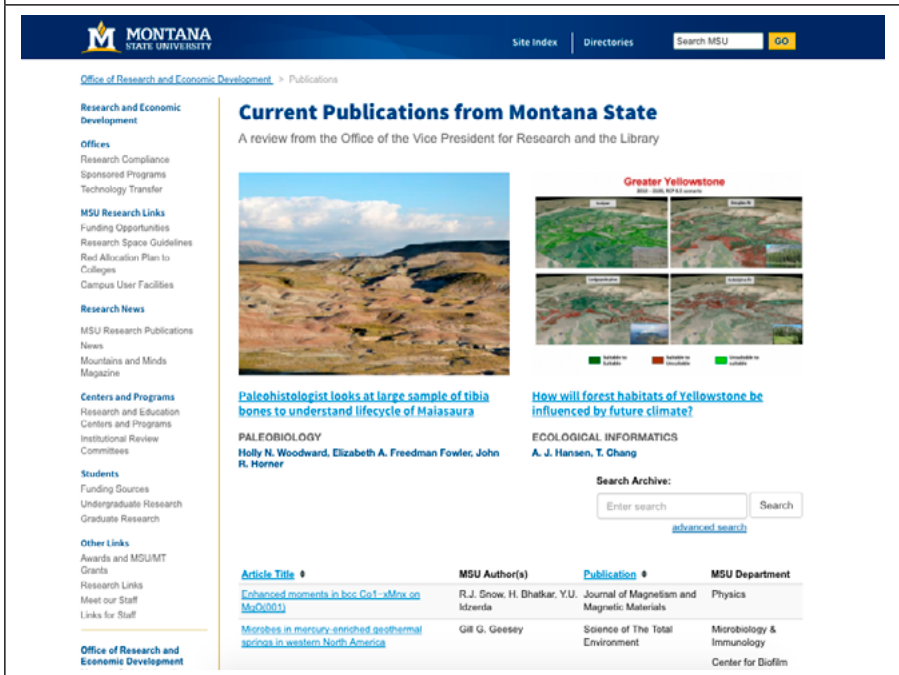
We allowed for human oversight of the deduplicating as a failsafe, but we were interested in lessening some of our review “pain points” as well. We needed a pro-

FIGURE 3
Administrative Review Interface within MSU Research Citations Application



<https://arc.lib.montana.edu/msu-research-citations/manage/>

FIGURE 4
Web Page Showing Library Citation Data (from Our API) Powering Office of Research and Economic Development “Current Publications” Website



<http://www.montana.edu/research/publications/>

cess that provided good data but was not bound to time-consuming human review. To this end, we used a checksum process on the database field for the article title. The characters in the article title are counted and encoded and given a unique SHA-1 “hash fingerprint.” As new items are listed, we can check for this “fingerprint” and automatically flag potential duplicate items and save the time of our content editor. We also prepopulate our data entry forms with institutional information (department, college, research center, and so on) pulled from our university website.

Even with these automated techniques, we recognized the need for manual review activity within the application. Our first step here was to push “skeleton metadata records” — records with basic title, description, author, date, language, and item type fields derived from our feed parsing—into the database and then give the abbreviated records a “pending” status where someone (often a student worker) will add the full metadata and move it to the “review” section. At this stage, there is some manual correction of records: sorting for item type, verification and addition of institutional information (department, college, e-mail address), and a quality control check. Items in “review” get checked by our Scholarly Communications Librarian and added to “active” with a date to align that item with the current month. Once a month is complete, a copy is pulled into .csv files for the repository inclusion process and XML files for transition into the monthly e-mail and public database in the University Communications system.

Discussion

While this is not the first or only attempt to collect and use citation information at the university level, we aim to continue to refine and disseminate our app to simplify the often lengthy, error-prone, or expensive tasks with which scholarly communication teams or research offices are increasingly tasked. Our app brings together multiple information sources in a user- and machine-friendly interface to facilitate the API-driven harvest, ingest, and reuse citation data. We hope that this will continue to increase repository deposit rates from faculty and that we will be able to use resulting citation data to enable new library services that reach broadly across the institution—that citation collection will drive research infrastructure progress and that research products and knowledge generation will be celebrated on university campuses as much as grant activity.

The JISC Publications Router is a promising data feed initiative looking to automate the delivery of research publications info (<http://broker.edina.ac.uk/>). We have considered using the Router as one of our data feeds, but there is a short lag between publication and ingestion into the Router that limits the utility of the Router as a real-time data source for citations. Regardless, this is valuable work from JISC and the University of Edinburgh, and the quality of the feeds is very high (including full citations and consistent metadata). We do anticipate further refinements and integration, with additional data sources that are tailored for “current research information systems” (CRIS). Specifically, we see potential in applying Crossref (www.crossref.org/) data to help us check the veracity of our citation information. Other sources, such as CHORUS (www.chorusaccess.org/), can give us a picture of the scholarly content that has been publicly funded. Finally, additional metadata work to register and integrate our authors with the ORCID ID system (<http://orcid.org/>) will provide a means to disambiguate our authors and provide an identifier that will allow our software to link to external systems. We will take our guidance on other refinements from the data and reporting needs of our primary university stakeholders, the Office of Research and Economic Development and the University Communications Department.

Our project implementation enables the integration of IR activities into broader discussions about research on campus. We can use data that used to be used exclusively by

the IR in new ways that enrich and enable growing aspects of our research enterprise. Our hope is that this is a part of a shift in campus dynamics that creates a culture of interdisciplinary idea sharing and engenders a feeling of community throughout campus.

Conclusion and Next Steps

Taking advantage of this “celebration point” to motivate researchers and make the incremental postprint submission process into a single simple task that is timely and beneficial has greatly benefited our repository. The citation app project has taught us how to harvest and digest data feeds in ways that create value for university partners. Moreover, the library has found a new service in research promotion and advocacy that demonstrates new and emerging roles for research libraries.

These citation data that we have harvested and curated have been useful within multiple projects. This process has allowed less redundancy of gathering publication data. As offices on our campus have become aware of this metadata resource, it has streamlined data collection from scientific institutes on campus when applying to grants and allowed colleges to easily celebrate the publication achievements of their faculty. In fact, our data continue to be reused, and we were recently informed by our MSU Communications department that they are now using the citation data (accessed via our API) to populate department-level pages with publications data through the campus content management system. Even beyond the benefits of data reuse, this project is a way for academics to see the library differently: we want to ensure that the campus is informed about the information and services that the library has currently, not just the books and analog materials they may associate with libraries.

We are positioning citation collection as the data engine that drives scholarly communication, IR deposits, and assessment of research activities. More important, we feel we are in the first stages of this project. These data we have collected have obvious value for research promotion in the library, to the public, and for generating research deposit leads for scholarly communication, but we also are considering further uses. One tangible, workable result we have gained through our citation collection and reuse model is an understanding of what constitutes a good data feed (consistent, structured data) and knowledge of best sources for these data feeds from our assortment of paid and open access databases and journals. We are in the early stages of creating a research network analysis of this collected information as we connect a linked data graph to the citation information. This will enable MSU to investigate the ways that we work together both within the university and beyond our campus. We also now have the opportunity to analyze the text of the metadata we have collected. The abstracts, keywords, and department affiliations we have collected will allow us to get a better understanding of research trends, commonalities, and collaborations here on campus. In many ways, these citation data and our work to codify the citation record of the university is just the beginning. We see new services and partners continuing to appear as the story of our citation data gets told, revised, and analyzed.

Notes

1. Arthur M. Cohen, *The Shaping of American Higher Education: Emergence and Growth of the Contemporary System* (Hoboken, N.J.: John Wiley & Sons, 2007), 1–513.

2. “Elements—Symplectic,” *Elements* (2016), available online at <http://symplectic.co.uk/products/elements/> [accessed 01 September 2016].

3. Pablo de Castro, Kathleen Shearer, and Friedrich Summann, “The Gradual Merging of Repository and CRIS Solutions to Meet Institutional Research Information Management Requirements,” *Procedia Computer Science*, 12th International Conference on Current Research Information Systems, CRIS 2014 Managing data intensive science—The role of Research Information

Systems in realising the digital agenda 33 (2014): 39–46. doi:10.1016/j.procs.2014.06.007.

4. Fredrik Åström and Joacim Hansson, "How Implementation of Bibliometric Practice Affects the Role of Academic Libraries," *Journal of Librarianship and Information Science* 45, no. 4 (Dec. 2013): 316–22, doi:10.1177/0961000612456867.

5. Richard Whitley and Jochen Gläser, "The Changing Governance of the Sciences: The Advent of Research Evaluation Systems," *Sociology of the Sciences Yearbook*, no. 26 (2007): 245–66, doi:10.1007/978-1-4020-6746-4_12.

6. Åström and Hansson, "How Implementation of Bibliometric Practice Affects the Role of Academic Libraries," 316–22.

7. John M. Budd, "Faculty Publications and Citations: A Longitudinal Examination," *College and Research Libraries Unassigned* (anticipated publication Mar. 2017): 1–23.

8. Åström and Hansson, "How Implementation of Bibliometric Practice Affects the Role of Academic Libraries," 316–22.

9. Nancy Fried Foster and Susan Gibbons, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories," *D-Lib Magazine* 11 (2005):1–11, available online at <http://eric.ed.gov/?id=ED490029> [accessed 28 August 2016].

10. *Ibid.*, 5.

11. R.E. Proudfoot, A. Sharma-Oates, M.M. Middleton, and B. Shipman, "JISC Final Report: IncReASe (Increasing Repository Content through Automation and Services)," *Monograph* (May 2009), available online at <http://eprints.whiterose.ac.uk/9160/> [accessed <<need access date>>].

12. P. Simpson, "TARDIS Project Final Report," (2005): 1–13, available online at <https://core.ac.uk/display/32444/tab/similar-list> [accessed 01 September 2016].

13. Morag Greig and William J. Nixon, "DAEDALUS: Delivering the Glasgow ePrints Service," *Ariadne*, no. 45 (2005), available online at www.ariadne.ac.uk/issue45/greig-nixon/ [accessed 01 September 2016].

14. "SHARE" (2016), available online at <https://osf.io/share/> [accessed 01 September 2016].

15. Gordon McIntyre, Janice Chan, and Julia Gross, "Library as Scholarly Publishing Partner: Keys to Success," *Journal of Librarianship and Scholarly Communication* 2, no. 1 (Nov. 2013): eP1091, doi:10.7710/2162-3309.1091.

16. Foster and Gibbons, "Understanding Faculty to Improve Content Recruitment," 1–11.

17. Proudfoot, Sharma-Oates, Middleton, and Shipman, "JISC Final Report."

18. Foster and Gibbons, "Understanding Faculty to Improve Content Recruitment," 1-11.

19. Hendrik P. van Dalen and Kène Henkens, "Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey," *Journal of the American Society for Information Science and Technology* 63, no. 7 (July 2012): 1282–93, <http://onlinelibrary.wiley.com/doi/10.1002/asi.22636/abstract>.

20. Lokman I. Meho and Yvonne Rogers, "Citation Counting, Citation Ranking, and H-Index of Human-Computer Interaction Researchers: A Comparison between Scopus and Web of Science" (2008): 1–35, available online at <http://eprints.rclis.org/11238/> [accessed 01 September 2016].

21. Åström and Hansson, "How Implementation of Bibliometric Practice Affects the Role of Academic Libraries," 316–22.

22. Wayne A. Wiegand, "Tunnel Vision and Blind Spots: What the Past Tells Us about the Present; Reflections on the Twentieth-Century History of American Librarianship," *The Library Quarterly: Information, Community, Policy* 69, no. 1 (1999): 1–32.

23. Jan Brophy and David Bawden, "Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources," *Aslib Proceedings* 57, no. 6 (Dec. 2005): 498–512, doi:10.1108/00012530510634235; Janine Schmidt, "Promoting Library Services in a Google World," *Library Management* 28, no. 6/7 (July 2007): 337–46, doi:10.1108/01435120710774477; Ian Rowlands, David Nicholas, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, Richard Withey, Hamid R. Jamali, Tom Dobrowolski, and Carol Tenopir, "The Google Generation: The Information Behaviour of the Researcher of the Future," *Aslib Proceedings* 60, no. 4 (July 2008): 290–310, doi:10.1108/00012530810887953.

24. Peter Millington, "SHERPA/RoMEO—Publisher Copyright Policies & Self-archiving" (2016), available online at www.sherpa.ac.uk/romeo/index.php [accessed 01 September 2016].

25. Richard L. Oliver, "Measurement and Evaluation of Satisfaction Processes in Retail Settings," *Journal of Retailing* 57, no. 3 (1981): 25–48.

26. The complete data model and the primary files for the MSU Research Citation Application are available on the [msulibrary github](https://github.com/msulibrary) account.

27. We are using the Google Feeds API (<https://developers.google.com/feed/v1/reference>) to standardize and clean up the data in the multiple feeds. The API gives us a standard data format including a title, author, description, publication date, etc. that we can rely on as a primary data source for our application. We are also considering some feed parsing code libraries that will allow us to do the data normalization and clean up locally.