Showcasing Faculty Research with Elements and Tableau

Gabrielle Wiersma and Matt Ramey
University of Colorado Boulder, USA

Abstract

There are various metrics to calculate research productivity, but it can be difficult to collect the necessary data to measure the impact of scholarly works at an institutional level. Information about which books, articles, and other works have been published by faculty is recorded in disparate places such as departmental websites and individual curriculum vitae. Librarians at the University of Colorado Boulder (CU Boulder) have partnered with the Office of Faculty Affairs to systematically collect and analyze faculty publication data. This paper will describe some of the tools that CU Boulder is using, including Symplectic Elements and Tableau, and explain how these software systems can be used to analyze bibliometric data and showcase faculty research.

Introduction

Universities take great pride in the scholarly output of their faculty. Research outputs such as books, articles, presentations, and creative works are evidence of knowledgeable and engaged faculty who are capable of sharing their expertise with the world. Traditionally, research productivity has been measured by counting the number of scholarly outputs produced and how many citations are generated from those works. Research productivity is often used to calculate comparative and competitive rankings for academic programs and institutions. For faculty, scholarly outputs are requisites for achieving positive annual evaluations, tenure and promotion, and awards and recognition. For the university, scholarly outputs are essential for maintaining the overall academic reputation of the institution and recruiting and retaining the highest caliber students, faculty, and staff.

There are various ways to calculate research productivity, but it can be difficult to collect the necessary data to measure the impact of scholarly works at an institutional level. Information about which books, articles, and other works have been published by faculty is recorded in disparate places such as departmental websites and individual curriculum vitae (CVs). Data about how scholarly works are used, such as usage statistics or citations are also disconnected from the works themselves. Furthermore, these data sources may provide evidence of the work of a particular scholar or department, but they do not provide a complete picture of the faculty as a whole.

There are many groups on campus who are interested in collecting and analyzing publication data. Academic departments and administrators need to keep track of scholarly outputs for annual evaluations, tenure review and promotion, grant reporting, program review, and accreditation. In fact, the project at the University of Colorado Boulder started when one of the colleges asked if the libraries could provide a list of recently published books by current faculty, a task which was not easily accomplished.

The libraries have historically purchased books authored by researchers affiliated with the university, but these efforts were limited to purchasing books by known authors or relying on metadata and publication information in the book jacket or online ordering system to identify affiliated authors. Most of the burden of identifying relevant publications fell to subject librarians who were expected to identify and order resources based on their knowledge of the faculty in their assigned departments and by searching for publication information on academic websites or citation databases. Despite spending hours combing through websites and compiling information in spreadsheets or bibliographies, the data was incomplete and ineffectively formatted for reuse.

Understanding institutional research interests is vital to collection development in an academic library. The libraries at the University of Colorado Boulder are interested in scholarly activities to ensure that the collections and services we provide meet the research and teaching needs of our students and faculty. The libraries strive to provide access

to the resources that are produced and used by our faculty and want to share and showcase faculty research outputs with the rest of the world. Knowing that there had to be a more efficient way to collect, analyze, and share publication information with stakeholders, librarians began to contact colleagues on campus to find an alternative solution to manual data collection and static spreadsheets.

Since 2011, librarians at the University of Colorado have partnered with the Office of Faculty Affairs to systematically collect and analyze faculty publication data. "The Office of Faculty Affairs coordinates a variety of activities associated with faculty life and academic programming on the Boulder campus."[1] In addition to creating policies regarding recruitment, hiring, reappointment, tenure, and promotion, Faculty Affairs also maintains the Faculty Information System to track major events related to faculty careers. Faculty Affairs also manages a database of faculty activities related to research, teaching, and service called FRPA port. All tenure and tenure-track faculty on campus are required to submit an online Report of Professional Activities (FRPA) on an annual basis. The FRPA "is designed to serve as an annual inventory of a faculty member's professional activities" and is used by most departments to create reports and documentation for annual merit evaluations.[2] To complete the FRPA, faculty create entries to describe any new scholarly or creative works, teaching, and service activities completed in a calendar year. In addition, most faculty also submit an updated CV to FRPA port, making it a veritable gold mine of faculty publication data.

Access to FRPA port data is typically limited to a few people in each academic department who assist with annual evaluations. However, when librarians approached the office and explained that having access to publication information would help build collections that support faculty research, Faculty Affairs granted a small group of librarians privileged access to the data. The FRPA database was a historic Oracle Database built in the 1990s. Faculty manually entered information about their research, teaching, and service activities into separate text fields with numerical categories. Although the data could be exported by category (e.g., 407—Refereed Journal Articles or Chapters), it was a burden for faculty to enter this information and the free-text forms captured data that was difficult to reuse because the formatting was not standardized. Moreover, the data could not be batch exported from the system. In order to create a list of all publications, one had to open a report for each category and then copy and paste the data into another utility such as Word or Excel. Despite its being a central database for publication information, there was a lot of room for improvement.

When Faculty Affairs decided to upgrade to a new faculty information system, they consulted a group of librarians because they knew that we had been accessing publication data in the FRPA port database. The librarians provided feedback about necessary features for a new system including standardized data entry forms with more defined bibliographic elements and better export functionality. After careful consideration of multiple products, Faculty Affairs decided to pilot Symplectic Elements, a research information management system that many universities are using to collect, manage, and showcase academic research outputs.

Elements was selected because it provides a single site to capture and analyze faculty data and other contributions. Elements was designed to reduce the reporting burden on academics by aggregating publication data from authoritative sources to create structured reusable data for reporting and analysis. Publications can be manually added to Elements through web-based entry forms, which work similarly to the FRPA port system, but are more structured forms to improve data consistency. To initially populate Elements and to fill in retrospectively, Faculty Affairs hired graduate students to manually curate and add records for publications for all tenure track faculty. Going forward, APIs and data feeds will search, find, and ingest metadata for newly published scholarly works. Faculty Affairs has already activated multiple data feeds from major indexes like Web of Science, PubMed, and CrossRef to automatically populate the database. During the implementation phase, librarians suggested additional data sources and helped to configure the Web of Science API to supply citation information. Librarians also shared expertise about citation indexes, classification schema, and search strategies to extract the maximum number of citations from automated data feeds. The Elements system matches publication information to a list of current faculty and pushes citation information into the appropriate faculty member's profile. Faculty interact with the system to claim or reject authorship, link to co-authors, and they can easily export their own publication information in a variety of formats from Elements.

Elements was also selected because of its potential interoperability with other campuswide systems such as faculty profiles in VIVO. Select publication data and files like CVs can be uploaded into Elements and then pushed into public facing VIVO profiles. This would consistently populate faculty profiles and make it easier to keep their profiles up to date. VIVO profiles, which CU Boulder has branded CU Experts, are one example of how research institutions can showcase faculty research and connect researchers from across the institution and around the globe (https://experts.colorado.edu).

As a central data repository, Elements contains a wealth of information about faculty publications that can be used for a variety purposes. The Office of Faculty Affairs primarily collects this data for academic departments, who in turn use this data for annual evaluations. The libraries, however, are interested in publication data to ensure that they are collecting books and other faculty scholarship and to keep abreast of changing research needs and interests. Of particular interest are reports of the top cited publications based on citation counts from Web of Science and a list of journals by frequency (based on the number of articles per journal). Elements can produce a report that lists all of the publications for an individual, department, college, or the entire university.

Elements has built-in data analysis tools including reports, graphs, and statistics for a college, department, or single user. The comparative statistics are particularly useful for deans and departments who can quickly run a list of newly published books, articles, and other scholarly works by their faculty. Moreover, Elements provides structured reusable data that can be exported for a variety of purposes. The libraries can simply use a basic report for the entire organization and export publication information into an Excel spreadsheet or formatted bibliography. Reports in Excel can be sorted, filtered, and graphed. The reports, however, do not answer important questions about whether or not the library provides access to the books, journals, or other publications on the list.

The libraries have started to compare the list of faculty publications with library holdings to verify that books published by faculty are added to the collection and to determine if we are providing adequate access to the journals in which they most frequently publish. Initially, librarians manually searched for book and journal titles in the libraries' catalog and added holdings information to the spreadsheet. This was a very time-intensive process that ultimately produced a spreadsheet that was quickly outdated and designed to suit internal decision making, not to present to external stakeholders.

After connecting holdings, we also wanted to include additional information to evaluate the performance of the materials in our collection such as usage statistics, impact factors, and citation counts. All of this data could be incorporated into a master spreadsheet, but this project would require a significant amount of time. Data from multiple spreadsheets can be combined in Excel, either manually or with the use of formulas, though this is not necessarily the most efficient way to collect and analyze data from multiple sources. There are many types of statistical analysis software that can connect or combine different data sets such as SAS (Statistical Analysis System), Stata, or SPSS (Statistical Package for the Social Science). Multiple spreadsheets or data sets can also be combined in a relational database such as Microsoft Access. Librarians considered all of these options but lacked the technical skills and necessary access licenses to most of these resources. We had been researching different types of data visualization software and decided to try to analyze this data using Tableau.

Collection Analysis using Tableau

Tableau is business intelligence software that many libraries are using to create data visualizations and dashboards. Tableau offers a variety of end user licenses including a freely available version called Tableau Public. Anyone can use the Tableau Public software and create data visualizations by registering for an account and downloading the application (https://public.tableau.com/s/). Tableau Public contains most of the features and functionality of the more robust paid versions except that visualizations cannot be saved locally (they must be saved to the Tableau Public server), and there are certain limits regarding the types of data that can be ingested and the size of the files that can be processed.

Unlike some statistical analysis programs, Tableau is relatively easy to learn and requires little to no technical skills to use. The interface features drag and drop capabilities for manipulating data and users can change the type of visualization with the click of a button. New users can find help on the

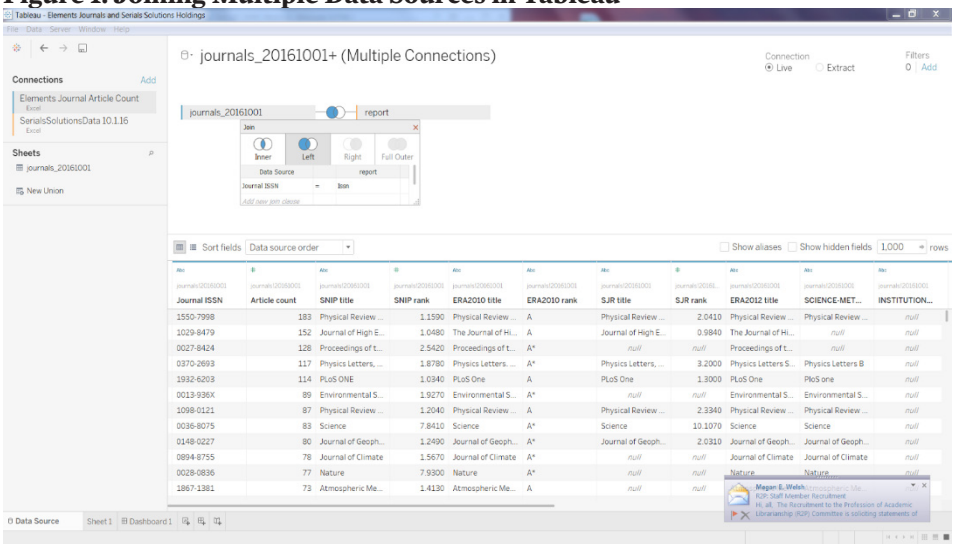
website including video tutorials and forums where questions are discussed and answered by an active user community. There are many other books and guides that describe the features and functionality of Tableau as well as best practices for creating data visualizations.

In addition to creating dynamic data visualizations, Tableau is a useful data analysis tool because it can connect and combine data from many different sources. It can connect to Excel spreadsheets, but it can also connect to Access databases, statistical files (e.g., SAS, SPSS), and many different types of servers (e.g., SQL, Oracle). This is particularly useful for collection analysis because collection data is often available from multiple sources such as an integrated library system, e-resource management system, usage statistics, or an authoritative list. Tableau can combine data from multiple sources as long as there is a unique match point in each of the sources. For example, an International Standard Serial Number (ISSN) is an excellent unique identifier for connecting journal data in multiple sources. It can be easily exported from an ILS and is often included in standard usage reports.

The libraries are using Tableau to analyze the journals in which faculty publish by connecting a variety of data sources and matching them on the ISSN. The Element report identifies which sources have the most journals and highest citation counts and this information can be combined with print holdings data from the ILS and online holdings data from Serials Solutions, usage statistics from COUNTER reports, citations and journal rankings from Web of Science and Scimago. Depending on how Elements is configured, it may already include some of the citation and journal ranking information from Web of Science and Scimago. Other data sources may be connected as needed using ISSN or another unique identifier as a match point.

Setting up data connections in Tableau is similar to creating relationships between different tables in Access or writing a join function in an SQL query (see Figure 1). Tableau can create inner, left, right, and full outer joins to accommodate the availability of data in each table or data source.

Figure 1: Joining Multiple Data Sources in Tableau



After the sources are connected, data fields can be pulled from any of the sheets and analyzed. If the data is not available in one of the sources, then Tableau will display a null value. A basic table combining the journal reports from Elements with holdings data can indicate where the libraries have access or gaps in coverage. Analysis revealed that the libraries provide access to most of the journals in which faculty have published articles including 99% of journals with 10 or more published articles and 100% of the journals with the highest cited articles. These findings demonstrate that the libraries' collections support faculty research and that the libraries are making a substantial amount of scholarly resources available to the university community.

Librarians used Tableau to create different types of graphs to visualize the top 200 journals based on number of articles and citation counts. Bar charts that can be filtered by year are useful for visualizing the journals with the most articles, and a tree map is an interesting way to visualize citation counts by journal or academic department (see Figures 2–3).

Figure 2: Excerpt of the Bar Graph of the Top 200 Journals by Publication Year in a Tableau Story

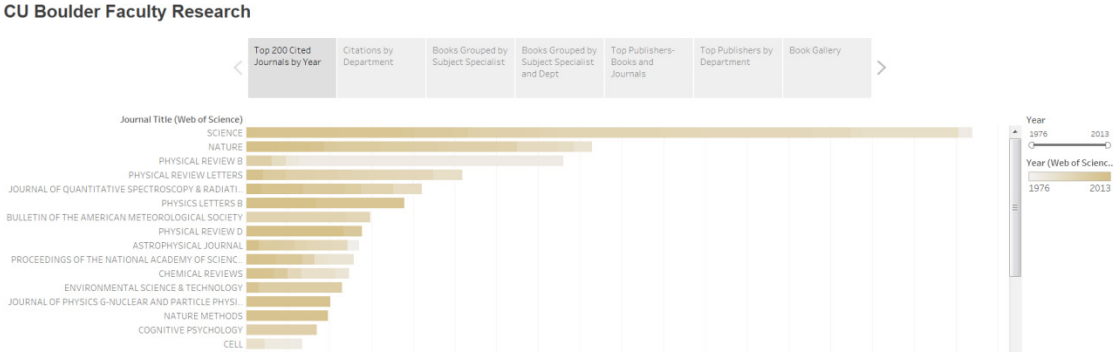
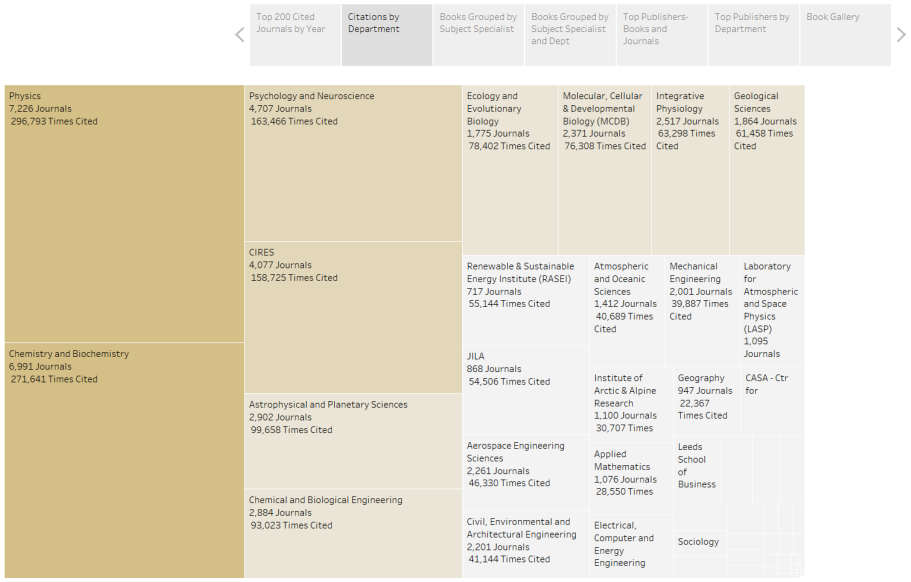


Figure 3: Tree Map of Citation Counts by Academic Department



Visualizations in Tableau are interactive and easy to customize. Visualizing the data makes it easier to see patterns or trends, especially if the data can be charted over time. The visualizations for journals illustrated research productivity for each department and highlighted the most important journals within each field. The libraries are using this information to ensure that we provide access to these resources, but these visualizations could also be published on a publicly available website to showcase faculty research productivity more generally.

Librarians also used Tableau to analyze the list of books from Elements. After combining and comparing the Elements report to ILS data, we found that the libraries have most (75%) of the books published by our faculty. It was still necessary to manually search for the books in the libraries' catalog, but this process was streamlined by inserting a URL for each title that performed a catalog search by title. The book data was added to a dashboard that included analyses by publisher and subject and grouped the data by academic department or subject librarian. These dashboards allow subject librarians to view books by faculty in their assigned

departments and also reported the most important publishers within each discipline. This information can be used to modify the publishers on the libraries' approval plan and will help guide selection of new materials.

Publication data from Elements will also be used for other collection management projects. The books on the list will be used to insert a note in the libraries' catalog record to indicate that a book has a CU Boulder author or editor. This note will display in the public records and will also be keyword searchable so that any user could find a list of books by CU Boulder authors in the libraries' catalog. The libraries may also use these data to create physical and online displays of recent faculty publications. Annotating books by CU authors will also be useful for identifying materials to keep onsite. The libraries typically send books with low circulation to a high density storage facility, but staff can use the notes about CU Boulder authors to selectively keep the books written by faculty and research affiliates onsite.

Connecting the lists of faculty articles to journal holdings and usage statistics could also inform decisions about journal renewals, cancellations, or back file purchases. Filtering the journals by year indicates which journals are most relevant to current scholars and some which may have been important in the past but are no longer applicable. Tableau could also be used to compare and analyze usage for any number of e-journals in the libraries' collection. Usage data could be manipulated to identify high and underperforming journals by title, subject, or publisher.

Next Steps

The authors still have many other types of publications (e.g., conference proceedings, book chapters, and creative works) to review, but these types of materials will require more manual analysis. In addition to scholarship and creative works, Elements also tracks service and teaching activities. These data will not be fully integrated into the system until 2017, but there is a potential for further analysis and an opportunity to learn about the use of scholarly resources in academic courses. The libraries are interested in learning more about which resources are used as textbooks and other assigned readings, and it may be possible to glean some of this information from course descriptions and syllabi if that information is uploaded into Elements.

The authors are also interested in further bibliometric analyses of faculty's scholarly outputs. In addition to analyzing the publications themselves, we are interested in determining whether or not the libraries provided access to the resources that are cited in books and articles. Having structured metadata not only makes analysis of our articles possible, but it could also offer the ability for librarians to do in-depth analyses of cited references to further trace the research lifecycle. For some publications, the cited references may be data mined from the Web of Science API. This would provide further evidence about how well library resources support successful research projects.

The data included in Elements reports are intended for internal use by departments and administrators. As such, many of the initial dashboards were created to inform internal decision making, such as filling in library holdings and identifying research trends. But there have been discussions about creating dashboards that would publicly showcase faculty research in an online gallery with supplementary data. Ideally, book covers or other representative images could be displayed in a digital gallery that could be sorted by department, subject, or year. In addition to Tableau, there may be other tools that might be better suited to showcasing research, such as a Digital Commons Book Gallery or a journal or article gallery in BrowZine.

The libraries also manage an institutional repository (IR) on the Digital Commons platform (CU Scholar). This platform is designed to publish and present various types of scholarship and CU Scholar is "intended to serve as a platform for preserving the research activities of members of the CU-Boulder community, and for promoting that research to the general public."[3] Many other institutions are using Elements to populate an institutional repository by collecting and passing publication data from Elements into the IR. CU Boulder has not implemented a workflow to ingest Elements data into CU Scholar, but this could potentially increase the number of submissions to the IR and would provide additional support for faculty to comply with the university's open access policy.

We are also exploring different ways to keep the data updated. Most of the Tableau dashboards are connected to static Excel spreadsheets and do not reflect changes to the Elements database or other updates. The Elements system is regularly searching,

identifying, and ingesting new publication data from scheduled searches and API calls. In addition, it is updated annually as faculty submit their annual reports of professional activities. It may be possible to connect Tableau to Elements through an API or SQL query, which would provide a real-time data connection and automatically refresh the data.

Lessons Learned

Partnering with the Office of Faculty Affairs was beneficial for many reasons. It allowed the libraries to tap into an existing data source for faculty publications so that the libraries did not have to duplicate efforts to collect this information. They also provided substantial technical and administrative support for the project. Faculty Affairs provided funding, project management, and hired staff with the technical skills to facilitate implementation including programming skills to customize the interface. It is important to note that the mandate for faculty to submit information to Elements comes from the Office of Faculty Affairs with the backing of the provost and executive vice chancellor for Academic Affairs, not the libraries. Faculty must submit a complete FRPA in order to be eligible for merit increases, so they are highly incentivized to submit this data as part of their annual evaluation. The implementation of Elements may not have been as effective if the library alone was asking faculty to voluntarily submit data.

The libraries and Faculty Affairs are using a variety of products to collect, analyze, and showcase research. Each tool has unique strengths and functionality, and it is possible that there is no single solution to accomplish all of these tasks. Further research and discussions need to occur to determine how best to leverage the available tools and creative a cohesive presentation of this information to stakeholders.

Conclusion

Citation data is readily available in many places but libraries need tools to collect and analyze it.

Research information management systems like Elements aggregate citation data from multiple sources and format the data in such a way that it is easier to analyze and export for further analysis. Using Elements to record and report scholarly research activities has improved the user experience for faculty who must enter this data on a regular basis and has streamlined analysis of faculty research outputs. It has also made publication and other data more accessible to individual faculty, departments, and other stakeholders on campus.

Complementary tools like Tableau can support further analysis and present the data to stakeholders in dynamic and informative dashboards. Dashboards can be designed to inform decisions or simply celebrate research accomplishments. Creating these types of custom dashboards to showcase faculty research and measure research productivity is useful for individual faculty, administrators, and librarians alike.

The university libraries have gained a considerable amount of information about faculty research activities due to this access and the improved functionality that the Elements system provides. The libraries now have a reliable listing of faculty publications that is updated on a regular basis. Librarians can log into the Elements system and download publication reports for any department or college as needed. These reports can be combined with other data and customized to present to different audiences using dashboards and stories in Tableau. Interactive Tableau dashboards can illustrate the connections between faculty research and library resources and demonstrate the value of library collections by indicating how well the library supports faculty research.

The libraries are successfully using the Elements data to systematically collect books by CU Boulder authors and subscribe to journals in which our faculty are publishing. As the Elements data continues to expand, the libraries plan to continue to access and analyze publication and other data. This will undoubtedly strengthen and guide collection development as the libraries learn more about faculty research interests and needs.

The collaborative effort between the libraries and Faculty Affairs has made significant progress towards a university-wide shift from citation data collection to research information management across the research lifecycle. Hopefully, streamlining the process for collecting and analyzing faculty research outputs in systems like Elements and Tableau will create more time for us to delve deeper into the bibliometric data and complete more in-depth analyses to determine how these outputs impact research and learning outcomes.

References

1. University of Colorado Boulder, Office of Faculty Affairs, "Overview," 2016, https://facultyaffairs.colorado.edu/welcome/office-of-faculty-affairs-overview.

2. University of Colorado Boulder, Office of Faculty Affairs, "Faculty Report of Professional Activities," 2016, https://facultyaffairs.colorado.edu/faculty/fac-comp-and%20salary/faculty-report-of-professional-activities.

3. University of Colorado Boulder Libraries, "CU Scholar Policies," 2016, http://scholar.colorado.edu/policies.html.