Ariel Hammond

Striding towards an institutional repository

Managing staff publications at San Diego Zoo Global

The long, rich, history of staff publications at San Diego Zoo Global (SDZG) is a point of institutional pride. As such, a question like "What has San Diego Zoo Global published on [insert species]?" seems so simple and reasonable, and yet it's proven one of the hardest for us at the SDZG Library to answer.

Managing staff publications, in general, is difficult, and managing them digitally brings new challenges. Thus while the SDZG Library has been attempting to manage staff publications for quite some time, it's become apparent that our previous method is insufficient for our patrons, so we've begun working on a move towards a new direction.

Scientific zoological publications

The institutional knowledge of zoos is both wide-ranging and critical. Tse-Lynn Loh recently assessed the value of scientific zoological publications and found that not only do nonprofit zoos and aquariums publish scientific literature at significantly high rates, these publications are also highly cited by others.[1] Organizations with research-affiliated mission statements had the highest rates of output, as well as those with strong research funding and an established history of publication. This conservation research is a requirement for members of the Association of Zoos and Aquariums, and Loh determined that it contributes to broader ecological knowledge and conservation efforts.[2]

SDZG ranks highly both in publication output and the number of citations.[3] The study by Loh analyzed publications indexed in Web of Science from 1993 to 2013, and under these conditions they found that SDZG had produced 286 publications.[4] This number is far below the total number of publications made by SDZG during the period studied, and by the institution in general. During our 102-year history, we count at least 119 publications from authors whose last name begins with the letter A alone. However, the number cited by Loh still garnered 4,944 citations by others, which equals a rate of 17.29 citations per publication. This high citation ratio demonstrates the value of SDZG publications in the greater knowledge economy, and yet gathering this type of data for all of the publications done by SDZG remains a challenge.

One reason that gathering comprehensive data on all of these publications is challenging is that collection often relies on self-reporting, which is a notoriously unreliable method for populating a staff publications list.[5] The SDZG Library has already encountered this, for staff are understandably more focused on the research itself, than on promoting or cataloging their work.

Ariel Hammond is master of information student at Rutgers University and intern/volunteer at the San Diego Zoo Global Library, email: ariel.hammond@rutgers.edu

Another reason that gathering this information is challenging is that SDZG has a unique institutional structure. Many publications are authored by employees in one specific division, the Institute for Conservation Research (ICR), which maintains a list of publications authored by those currently working there. However, many works are being published by team members in other roles, such as veterinarians, zookeepers, curators, and horticulturalists, which do not fall under the purview of ICR. Thus, the SDZG Library strives to maintain an alphabetical list of all known staff publications, which dates back over 90 years.

Institutional Repository

At the SDZG Library, we believe that it is a curatorial imperative to collect locally produced scholarship and create a single digital container to house this institutional knowledge.[6] For, as Micah Vandegrift argues, these services that support digital scholarship and foster knowledge sharing are not only essential to the community, but indispensable.[7]

One benefit to gathering all of these publications into one place is improved search capabilities, which can match increasingly complex requests. Last year, for example, the library searched for the entire history of SDZGs relations with one specific country. Being able to search in one place for this information, using multiple filters and keywords, would have quickly returned results with multiple layers of depth, and allowed us to focus more time on expertly adjusting the results to the patron's specific needs.

Including digitized archival materials in the IR could also contribute to the layers of knowledge within it, particularly as the SDZG Library has made significant strides in digitization. *The Bulletins of the Zoological Society of San Diego,* for instance, date back to 1924 and are currently available digitally on the Biodiversity Heritage Library website.[8] However, items like board meeting minutes and institutional records contain too much sensitive information to be made public, which is what many digitization grants require. Additionally, some materials are too physically sensitive to be shipped out for digi-

tization. Thus, the SDZG Library has recently built an Archivist's Quill digitization machine. However, none of the currently digitized items are linked to other SDZG publications. Having digital copies of all these materials in an IR can create stronger connections between items, deepen query results, and provide backups in case of emergency.

An IR could also potentially house the data related to our publications, particularly as the need for researchers to have a secure place to house their data is becoming apparent within the organization. However, the sheet size of the datasets at SDZG may prove challenging. One project Wildwatch Kenya has camera trap images, which currently measure at 5.5 terabyte, and will likely measure at 7 terabyte when the data gathering is complete. The size of the entire Ecological Data Initiative, a highly regarded ecology repository, is 8 terabyte. Thus, storage space needs and resources will affect both data selection policy and system design of any future IR.

Indeed, as beneficial as an institutional repository may be, we are not unaware of the challenges associated with creating and maintaining one. Staff publications require managing embargoed publications, which, depending on the software used, may be difficult to locate or receive information about.[9] Author name authority control can also bring challenges, as Lizzy Walker and Michelle Armstrong note in their aptly titled paper, "I cannot tell what the Dickens his name is."[10] Maintaining file formats may also be a challenge, as files will need to be converted to newer formats to maintain accessibility, data needs to be monitored for bit rot, and all changes to data need to be documented to maintain provenance.[11]

Currently

Thus, the SDZG Library is taking measured strides towards repository implementation. First, we are currently uploading each staff publication into the citation manager Zotero, with the intention of transferring them to an IR in the near future.

The philosophy behind the workflow is to batch tasks. Citations are uploaded alpha-

betically by author's last name into a template. They are then separated into sections: those with a unique identifier (DOI or PubMedID), those without, and conference proceedings. Conference pro-

ceedings historically have not had unique identifiers, and as they constitute a sizable amount of scholarship produced by SDZG, separating them out early on streamlines the process.

Citations without an identifier are run through the CrossRef Simple Text Query form, which automatically scans for known identifiers, and these identifiers are added back to the original citation. All items with identifiers are uploaded to Zotero using the Add by Identifier button, which we've affectionately dubbed "The Magic Wand." We check the Zotero citation against its original counterpart for errors, because it is better in the long run to address these errors early on.

Citations without identifiers are checked individually in Google Scholar, as results can be uploaded directly to Zotero from the results page, using its web browser connector. Those not on Google Scholar can occasionally be found through a regular web search, and added directly from the publisher's webpage, yet many must often be entered manually.

This workflow, though slow at times, is already producing results. Searches within Zotero have been swift and fruitful, allowing the library to effectively respond to requests for SDZG-related elephant, rhinoceros, and giant panda publications, which included reports, book chapters, and journal, magazine, and newsletter articles, as well as conference papers and posters. We were able to quickly provide resources on SDZG California condor conservation work, specifically from the 1980s and 1990s, which is a multiple variable search that would have proved extremely difficult before. We were also able to use Zotero to gather citations for open access conservation-focused publications, produced within the last five years, to share with our partners at CITES, the Convention of International Trade in Endangered Species of Wild Fauna and Flora.

Additionally, the SDZG Library has used the information from Zotero to create a physical display of staff publications. The aim of this physical display was to promote institutional research, break down departmental information silos,[12] and encourage self-reporting all at once. Serendipitously, it also allowed one researcher to see one of their works in print for the first time. Thus the benefits of collecting staff publications into one location are already exceeding our expectations, and the ability to export this digital information to an IR via a CSV file, guarantees that the benefits of this work will continue.

In the future

Going forward, the SDZG Library will have much to consider as it selects an IR. Cost and technological resources will factor heavily in the ultimate decision, as will the requirements of each repository, such as accepted metadata schemas and file formats, and the scope of its user base. The SDZG Library is also considering collaborating with other institutions, as well as pursuing grant funding, to ameliorate some of these issues ahead of our planned 2020 implementation.

However, there are issues that affect repositories in general that we can anticipate and try to counter, regardless of the software selected. To encourage content population through self-submission, we believe that the IR partnership model is promising, which is a method that involves working with one branch of the institution at a time.[13] In adopting this model, the library can guarantee that proper procedures are followed while considering the needs of individual departments. To address author name authority control, we can use ORCID, ResearcherID, or Scopus identifiers when available[14] and encourage their use within the institution. When it comes to managing embargoed publications, the SDZG Library can also create alerts for embargoes with information from publication agreements, SHERPA/ROMEO, or the publications themselves.

While there are a variety of materials that may eventually be eligible for inclusion, staff publications are an important place to start, because they are the tip of the iceberg in regard to a repository's value,[15] and housing them can begin to demonstrate this value to the institution as a whole. Ultimately, institutional support in this endeavor is invaluable, because for an IR to contribute to an institution's knowledge economy, it requires strategic investment in information infrastructure and ecology.[16] However, an IR can produce a return on this investment in a number of ways, one of which is to demonstrate the breadth and value of institutional research by analyzing the wealth of data IRs provide, an important feat given that knowledge impact is often hard to quantify.[17]

Stewarding all of this information into a central, searchable, digital container can help support scholarship and knowledge transfer throughout an entire institution, and at the same time enable the library to help pursue larger institutional goals.

In our case, that includes being able to answer those "What has SDZG published on [insert species]?" questions, so that we can do our part to help end extinction.

Notes

1. Tse-Lynn Loh et al., "Quantifying the Contribution of Zoos and Aquariums to Peer-Reviewed Scientific Research," *FACETS* 3, no. 1 (January 1, 2018): 287–299, doi:10.1139/facets-2017-0083.

2. Ibid.

3. Ibid.

4. Ibid.

5. Jiebei Luo, "Using Google Apps to Manage Embargo Records," *C&RL News* 79, no. 3 (March 1, 2018): 137–40, doi:10.5860/crln.79.3.137.

6. Micah Vandegrift, "Designing Digital Scholarship Ecologies," LIS Scholarship Archive (January 2018), doi: 10.31229/osf.io/93zvb.

7. Ibid.

8. https://www.biodiversitylibrary.org/creator/220228#/titles.

9. Lizzy A. Walker and Michelle Armstrong, "'I Cannot Tell What the Dickens His Name Is': Name Disambiguation in Institutional Repositories," *Journal of Librarianship and Scholarly Communication* 2, no. 2 (2014): eP1095, doi:10.7710/2162-3309.1095.

10. Ibid.

11. Wendy C. Robertson and Carol Ann Borchert, "Preserving Content from Your Institutional Repository," *The Serials Librarian* 66, no. 1–4 (May 1, 2014): 278–88, doi:10.1080/0361526X.2014.881209.

12. Dolly Goulart, "Library vs. Library: Recreating and Adding Value," presentation, Special Libraries Association San Diego Chapter Fall Seminar, San Diego, California, September 29, 2017.

13. A. Miller, "A Case Study in Institutional Repository Content Curation," *Digital Library Perspectives,* Bingley 33, no. 1 (2017): 63–76, doi:10.1108/DLP-07-2016-0026.

14. Walker and Armstrong, *Journal of Librarianship and Scholarly Communication.*

15. Heather Joseph, "Next steps for IRs and Open Access," *Making Institutional Repositories Work,* ed. Burton B. Callicott, David Scherer, and Andrew Wesolek.

16. Vandegrift, "Designing Digital Scholarship Ecologies."

17. Miller, "A Case Study in Institutional Repository Content Curation." ꝫ