

7-31-2015

Research Data Practices in Veterinary Medicine: A Case Study

Erin E. Kerby

University of Illinois at Urbana-Champaign, ekerb@illinois.edu

Follow this and additional works at: <http://escholarship.umassmed.edu/jeslib>

 Part of the [Library and Information Science Commons](#), and the [Veterinary Medicine Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#).

Recommended Citation

Kerby, Erin E.. 2015. "Research Data Practices in Veterinary Medicine: A Case Study." *Journal of eScience Librarianship* 4(1): e1073. <http://dx.doi.org/10.7191/jeslib.2015.1073>

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.



Full-Length Paper

Research Data Practices in Veterinary Medicine: A Case Study

Erin E. Kerby

University of Illinois at Urbana-Champaign, IL, USA

Abstract

Objective: To determine trends in research data output, reuse, and sharing of the college of veterinary medicine faculty members at a large academic research institution.

Methods: This bibliographic study was conducted by examining original research articles for indication of the types of data produced, as well as evidence that the authors reused data or made provision for sharing their own data. Findings were recorded in the categories of research type, data type, data reuse, data sharing, author collaboration, and grants/funding and were analyzed to determine trends.

Results: A variety of different data types were encountered in this study, even within a single article, resulting primarily from clinical and laboratory animal studies. All of the articles resulted from author collaboration, both within the University of Illinois at Urbana-Champaign, as well as with researchers outside the institution. There was little indication that data was reused, except some instances where the authors acknowledged that data was obtained directly from a colleague. There was even less indication that the research data was shared, either as a supplementary file on the publisher's website or by submission to a repository, except in the case of genetic data.

Conclusions: Veterinary researchers are prolific producers and users of a wide variety of data. Despite the large amount of collaborative research occurring in veterinary medicine, this study provided little evidence that veterinary researchers are reusing or sharing their data, except in an informal manner. Wider adoption of data management plans may serve to improve researchers' data management practices.

Correspondence: Erin E. Kerby: ekerb@illinois.edu

Keywords: veterinary medicine, research data, data practices, data management



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Introduction

Prompted by the rapid growth of eScience and the use of large datasets in research, the U.S. Office of Science and Technology Policy issued a mandate in 2013 designed to expand public access to the results of federally funded research. In practical terms, “results” refers to both published articles and their associated research data. In response to this mandate, several federal funding agencies now require data management plans to be included with some grant applications, and more agencies are following suit. Consequently, research data management has steadily developed into an area of new opportunities and roles for libraries. The research data landscape is complex and constantly evolving, making it a challenge for libraries and librarians to stake out their territory. Nonetheless, examining the details of the landscape with a finer eye, particularly within individual disciplines such as veterinary medicine, can provide valuable information for institutions developing a research data service.

Background

As evidenced in the literature, libraries and librarians have spent considerable time in recent years surveying and analyzing the research data landscape to identify ways in which they can best provide support and services. A small number of academic research institutions have led the way in developing research data services in support of their research community, including the University of Illinois at Urbana-Champaign (U of I). In many cases, the development of a research data landscape within an academic institution means partnering with diverse units on campus, in recognition that data management is acutely complex and that no one unit has the expertise and infrastructure to build a complete service. U of I’s Library eResearch Task Force Final Report states that, “The Library’s public-facing eResearch support services cannot advance substantively unless there is both campus-level technical and personnel infrastructure to support data curation and preservation and domain-specific expertise to support research data management, preservation, and sustained access” (Braxton et al. 2013).

In this case, “domain-specific expertise” refers in part to the subject-specialist librarians, including the author, who act as liaisons to specific departments and units across campus. The expectation is that data service needs will vary, sometimes significantly, depending on the discipline and type of research being done. The librarians can assist in tailoring the service to the need. However, this necessitates that the librarians have a broad understanding of the research being produced in their respective departments or units.

The primary aim of this study is to determine trends in research data output, reuse, and sharing at the U of I College of Veterinary Medicine (CVM). Such a study serves to inform not only the work of an individual librarian — particularly subject specialists — but also of the Library itself as it works toward becoming the public face of the campus Research Data Service. Finally, this study seeks to foster and advance a dialogue on research data management and practices amongst librarians and researchers in veterinary medicine and other related disciplines.

Literature Review

In the veterinary medicine literature, the phrase “data management” first appeared in 1981, referring to the use of an electronic database to store, organize, and recall research data

(Peterson et al. 1981). As this technology advanced, the field of informatics developed in veterinary medicine around the idea of data management, as it did in the broader biomedical sciences and similar disciplines. Informatics in an academic sense is a broad field of study incorporating many aspects of information and knowledge management with technology. Veterinary informatics was first described as “a wide range of efforts to apply advances in computer science, statistics, mathematical modeling, information science, and educational theory to the goal of providing better and more efficient delivery of medical services, whether those be in patient care or other equally important areas of veterinary medicine” (Talbot 1991).

Although specific applications such as practice management and herd-health software and digital-imaging computer systems have received significant attention, veterinary informatics has been slow to gain traction within the discipline of veterinary medicine (Smith-Akin et al. 2007, Smith and Williams 2000). Over the past 20 years, as research in veterinary medicine has grown more collaborative and interdisciplinary, veterinary informatics has struggled to keep pace with this trend (Bellamy 1999, Johnson et al. 2011). This is despite the founding and growth of the One Health Initiative, which according to its website, “is dedicated to improving the lives of all species — human and animal — through the integration of human medicine, veterinary medicine and environmental science” (One Health Initiative 2015). In particular, issues of limited funding and a lack of coordination among stakeholders have hindered progress in veterinary informatics (Santamaria and Zimmerman 2011). Better data management practices on the part of the researchers would help resolve coordination issues, but to date there has been little incentive for them to do so.

The funding agencies’ movement towards requiring data management plans in grant proposals is meant to increase access to research data and thereby help improve transparency and accountability. Most notably in the United States, the National Science Foundation (NSF) and the National Institutes of Health (NIH) now require that grants of \$500,000 or more include a data management plan (DMP). Other federal funding agencies have only recently begun to follow suit.

Although the NIH funds a considerable amount of research in veterinary medicine, the number of awards is relatively low when compared to the overall total funds awarded by the NIH to all researchers — less than 1% in 2011 (National Research Council (U.S.) 2011). Furthermore, it is uncommon for a veterinary research grant to surpass the \$500,000 threshold. For example, in the fiscal years 2013 and 2014, only one U of I CVM faculty member was listed as the principal investigator of a NIH grant totaling more than \$500,000 (National Institutes of Health 2013).

The U.S. Department of Agriculture, another frequent funder of veterinary research, only recently announced a new policy to improve access to the research it funds. Beginning in the next several years, with full implementation and adoption by 2018, a data management plan must be included in all grant proposals (United States Department of Agriculture 2014). It is conceivable that these types of policies from funders will provide the needed incentive for researchers to improve their data management practices, but there are other barriers to consider.

A variety of studies published by librarians in the past several years have detailed the reluctance of many scientists to change their data management practices, for reasons such as

the belief that sharing data is not useful and/or appropriate, concern that their data would be used without proper attribution, and questions of data quality and standardization (Diekema, Wesolek, and Walters 2014, Sayogo and Pardo 2013, Tenopir et al. 2011). Other concerns when considering these new funding agency policies include issues with data storage, sharing, and preservation. While those working in veterinary informatics have spent considerable time and effort developing databases to store and retrieve data, little has been done to address aspects of long-term preservation and data sharing beyond a local context.

Several researchers have established that disciplinary differences in data management practices can have a significant effect on the development of data services (Akers and Doty 2013, Weller and Monroe-Gulick 2014). A few published studies demonstrate how subject-specialist librarians can address data management issues unique to particular disciplines, and bridge gaps to improve data management practices of researchers. Bracke worked with an agricultural researcher to prepare and deposit a data set into an institutional repository (2011). Williams first conducted a bibliographic study to determine trends in data practices and then used that information to identify potential candidates for data services (2012, 2013). Interestingly, Weller and Monroe-Gulick discovered that research methodology can affect data practices as much as a researchers' discipline (2014). This means there might be more commonality in data practices across disciplines than originally was thought.

Methods

Data for this study was collected through a detailed review of articles published by U of I CVM faculty members during the calendar year of 2013. The list of faculty names used was compiled in November 2013 from the college's directory (<http://vetmed.illinois.edu/directory/>), and represented each of the three primary departments within the college: Clinical Medicine, Pathobiology, and Comparative Bioscience. The list included assistant, associate, or full professors, for a total number of 48 faculty members (see Figure 1). For the purposes of this study, clinical and adjunct faculty members were excluded due to the somewhat variable nature of their appointments and publication outputs, as they are primarily focused on

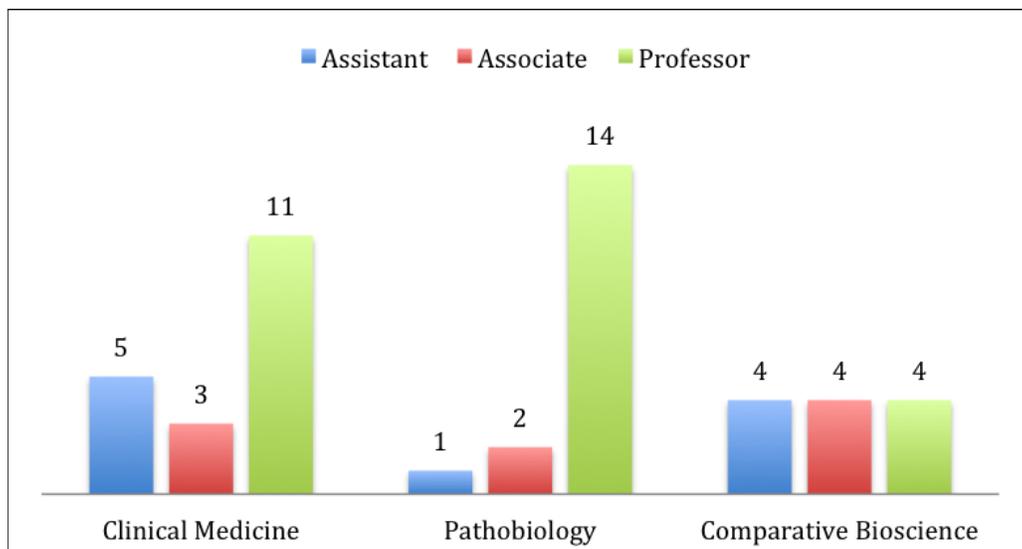


Figure 1: CVM Faculty Department and Rank

teaching. By comparison, tenure-stream faculty members are heavily invested in developing cohesive research agendas with regular and frequent original research publications, particularly assistant professors in the early stages of their career.

Only original research articles or short communications were included in this study, as review and summary articles are poor indicators of research data practices. The author collected the articles by searching CAB Abstracts (via Web of Science) and PubMed, both of which are used frequently to search veterinary literature (Grindlay, Brennan, and Dean 2012). Author last name and affiliation (e.g. Illinois) were used as search terms; in some cases, author first and middle initial were used if further disambiguation was necessary. A total number of 96 articles were retrieved for analysis.

The author read each article in its entirety and recorded findings in the following areas: research type, data type, data reuse, data sharing, author collaboration, and grants/funding. Research types included three broad categories: clinical study, laboratory animal experiment, and laboratory experiment. Clinical study refers to research that directly involved the patient or group of patients. Naturally, in veterinary medicine the patients are the animals. Laboratory animal study and laboratory study refer to research that took place under controlled conditions in a laboratory, either with or without the use of animals. Most articles fell within a single research type, but in some cases, a secondary research type was recorded where the article indicated a mixture.

The types of data recorded included bioassays, clinical records, digital images, digital video, microarrays (DNA), laboratory notebook data, and statistical data. It was not possible to ascertain from the articles the amount of data generated or used in the studies; rather, evidence of a particular data type was recorded as a single occurrence. If the authors of the article indicated that they reused data from an outside source or made provision to share their own data, the mode of reuse or sharing was recorded (i.e. a public data repository). All of the articles were the result of author collaboration, and it was also noted if co-authors were from within the U of I CVM, within U of I but outside of the CVM, or from an external institution. Evidence of one of these types of collaboration was recorded as a single instance.

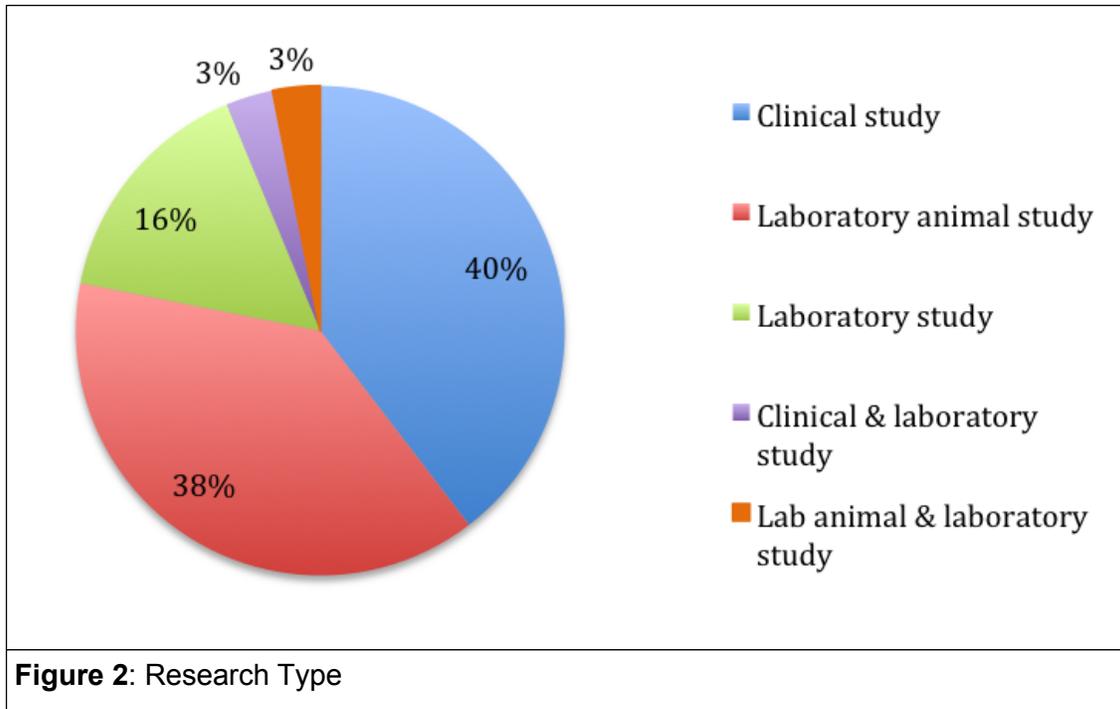
Results

Research type

The articles in this study were the result of three main types of research (see Figure 2). Over three-quarters of the articles were the result of either a clinical study (n=38) or laboratory animal study (n=37). A small amount (n=15) were the result of laboratory studies. A handful of articles very clearly were the product of two types of research, where there were two distinct parts to the study. Several combined a clinical study with laboratory animal research (n=3), and several more combined a laboratory animal study with a separate laboratory study (n=3).

Author collaboration

All of the articles in this study were the result of collaboration between at least one CVM faculty member and another researcher, either within the CVM, within U of I, or with author(s) from an external institution(s) (see Figure 3). Many articles indicated more than one type of author



collaboration, for example, both an internal collaboration and an external collaboration. A slight majority of the instances were amongst the CVM faculty, both within and across the college’s departments. There were a number of external collaborations as well, mostly with researchers from other academic institutions, but also from corporations, the federal government, or a veterinarian in private or group practice.

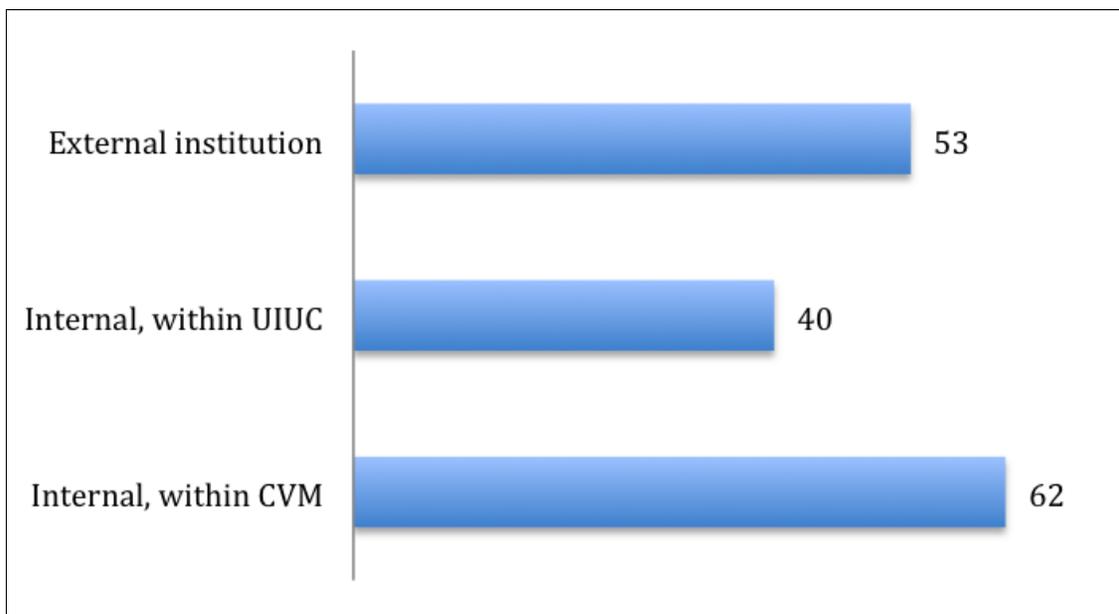


Figure 3: Number of Articles Per Type of Author Collaboration

Data type

Each article indicated that multiple data types were generated and used in the studies (see Figure 4). The most common data type found across all articles was digital imaging (see Figure 5). The types of images encountered included digital photos, CT (computed tomography) scans, MRI (magnetic resonance imaging) scans, ECG (electrocardiogram) images, ultrasound images, radiograph images for the clinical studies, and microscopy images for the laboratory studies. One clinical study indicated the use of digital video. Bioassays were also a commonly found data type and primarily used in laboratory research. This kind of procedure includes various methods for measuring the effects of a substance on a living organism, such as a drug or hormone.

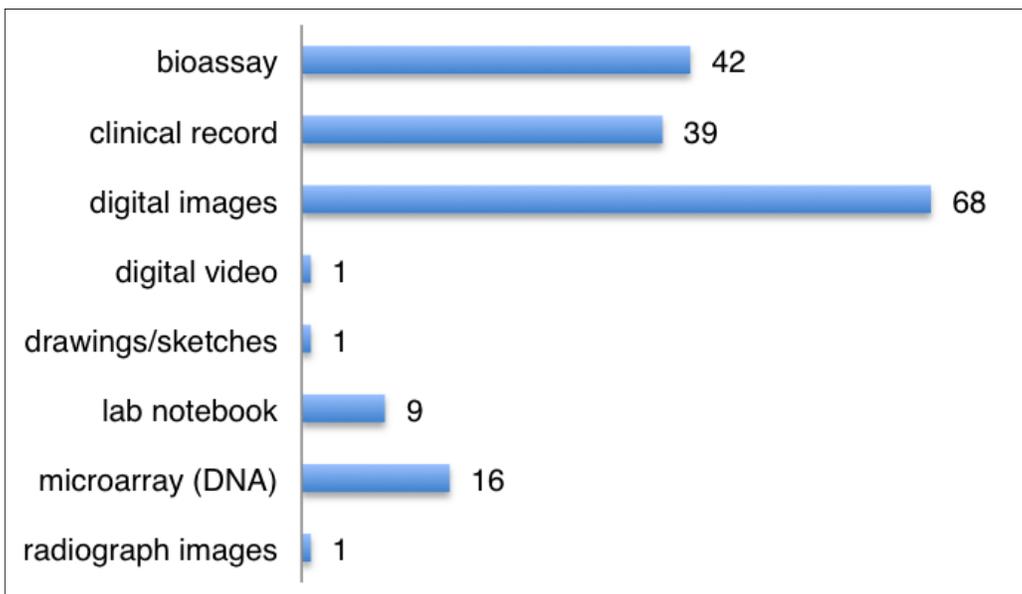


Figure 4: Number of Articles Per Data Type

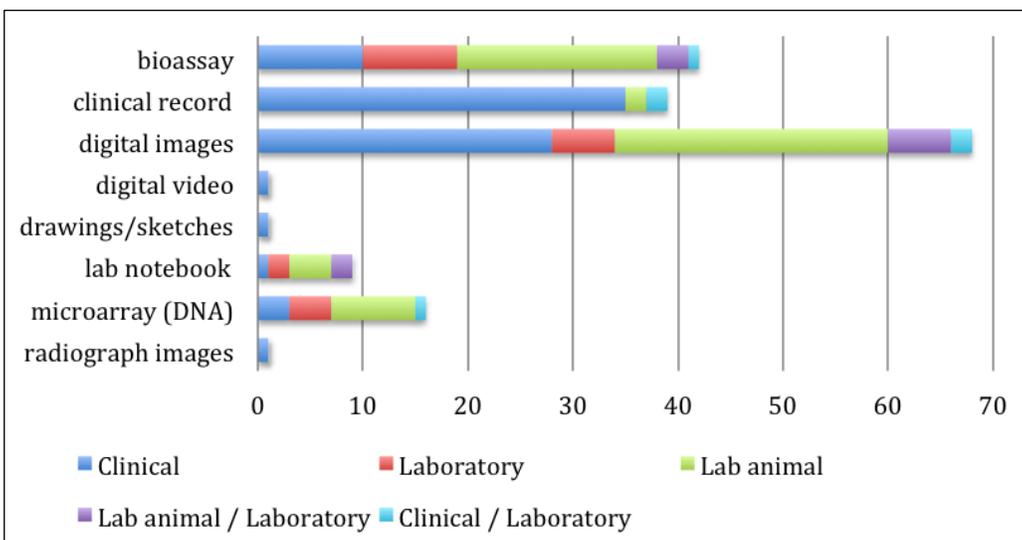


Figure 5: Data Incidence by Research Type

Data reuse

Over two-thirds of the articles gave no indication that data was reused in the study (see Figure 6). However, in nearly 25% of the articles the authors did acknowledge the receipt and use of data from a colleague. Many authors did this more formally in an acknowledgements section at the end of the article, but some included the acknowledgement in the methods section when explaining how data was generated or collected for the study. In some cases, the “data” was a biological substance such as tissue or cell cultures, which obviously need to be handled much differently than a spreadsheet full of numbers or clinical records.

A few articles noted that the authors reused data from a previous study, for example, where the data collected during one study (or even chain of experiments) was analyzed in different ways to produce information for multiple articles. Several articles indicated that genetic sequence data from GenBank was reused in the study. GenBank is an open-access, genetic sequence repository maintained by the National Center for Biotechnology Information (NCBI).

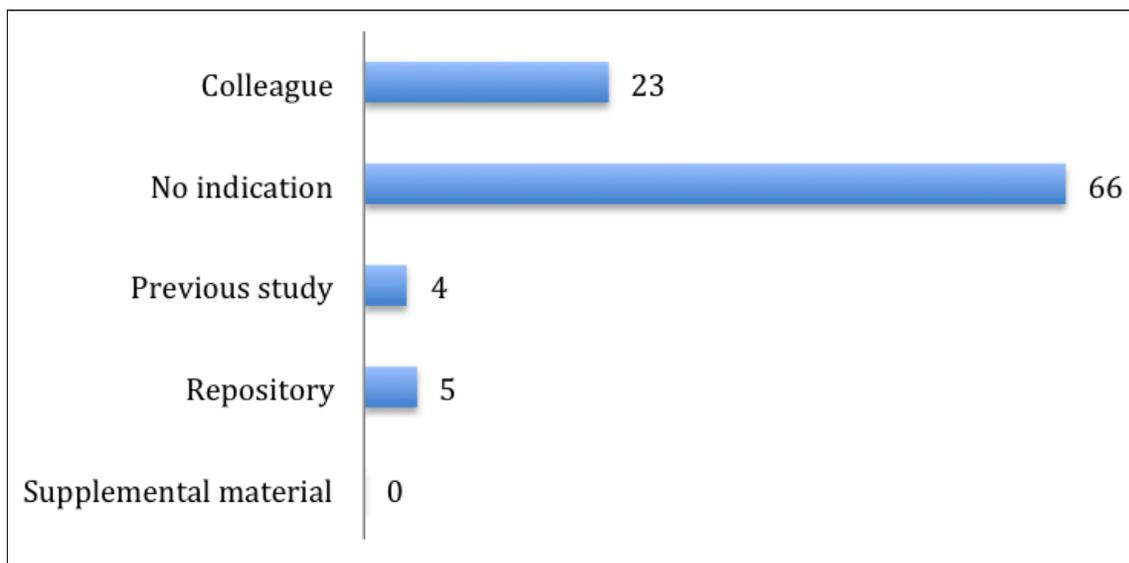
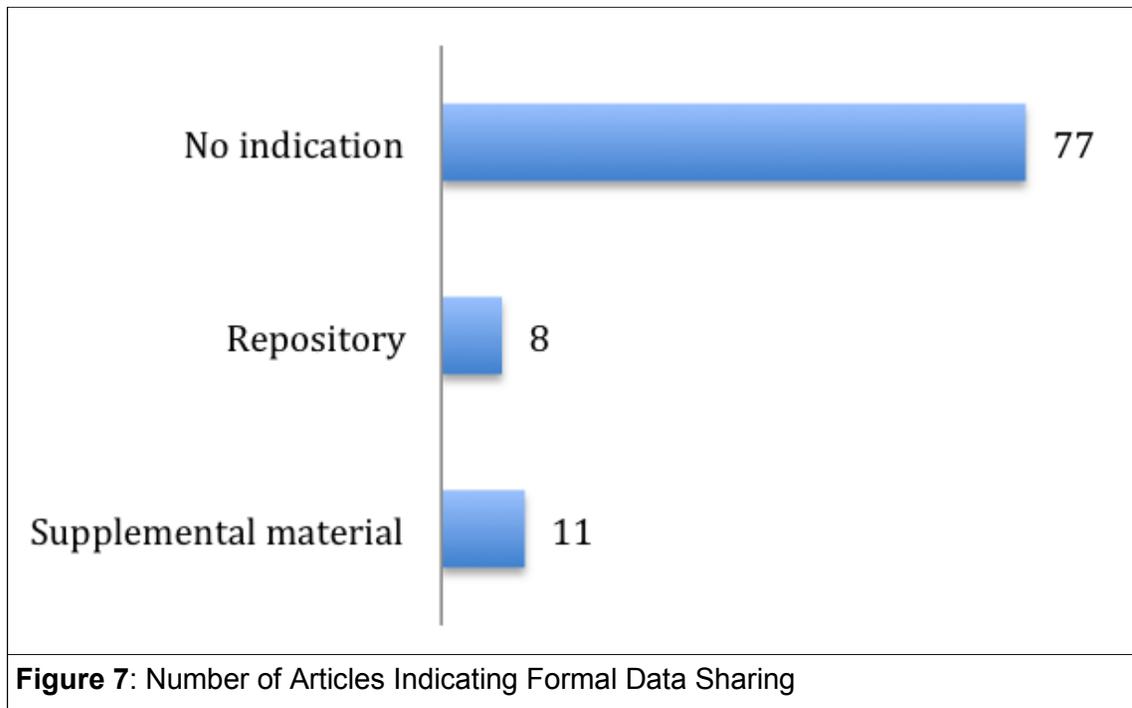


Figure 6: Number of Articles Indicating Data Reuse

Data sharing

Across all articles, there was very little indication that the authors were publically sharing their data (see Figure 7). Only 20% stated that data produced in the study was openly available, either through a public repository or on the publisher’s website as a supplementary file. These were the only two methods of sharing mentioned. The only repository used appeared to be GenBank.



Grants and funding

Of the 96 articles in this study, over two-thirds (n=67) reported that the research had been funded by a government, public, corporate, or private institution in the form of a grant. The types of grants varied from those to support a specific study to those geared towards supporting a specific researcher. Of those 67 articles, 28 indicated there were multiple funding sources, which is important information to consider with regards to data management plans.

Discussion

Veterinary medicine researchers contribute valuable knowledge to the scientific community and society, not just on the health and well being of animals but also humans and the environment. This is done frequently through clinical research involving companion animals, livestock, and wildlife, as well as through laboratory research such as stem-cell experiments or bioassays. The types of research encountered in this study suggest that while clinical studies remain vital to the discipline, laboratory studies are of equal importance and often provide a bridge to collaborate with others in related disciplines. Veterinary data often serves a dual purpose, both as a diagnostic tool in a clinical setting and a discovery tool in a research setting.

This study revealed that veterinary medicine researchers are prolific producers and users of many different types of research data. Such a variety of data makes good data management quite complex and at times cumbersome, not only on an individual level but also at the community level. Software to manage clinical and laboratory data has been in use for several decades, which makes management easier. For example, the U of I CVM uses the Vetstar

clinical management and VADDS laboratory management software, which are produced by Advanced Technology Corporation (<http://www.vetstar.com>). These two pieces of software are designed to integrate into a complete hospital data management system. The software is also customizable with different modules and data points, such as an accounting module to handle billing clients.

A significant amount veterinary data is produced and managed outside of clinical and laboratory software. For example, digital images from a CT scan could potentially be imported into the clinical management software and attached to a particular record. However, not all types of data integrate well with this software. The U of I CVM continues to use paper records and radiograph film, an indication that physical barriers to integrated data management still exist. To further complicate matters, data created and stored in the clinical and laboratory management software often is extracted for use in a research setting, which creates a new set of data management needs.

Information of this nature may help to explain data management practices as they relate to research collaboration. Logically, barriers to managing and sharing data can be overcome more readily within a single institution, yet the development of research data services at many institutions points to a need to improve efforts even further. Nonetheless, the bigger question is how a research data service can contribute to data management and sharing across institutions, especially with so many different types of data being produced. The fact that all of the articles in this study were the result of collaboration suggests that researchers do find ways to manage and share data beyond the boundaries of their own institutions.

Despite the great amount of collaboration taking place, the results of this study provided little evidence of data reuse by the U of I CVM faculty. This does not necessarily mean that data is not being reused; rather, it seems likely that much of the reuse is not being acknowledged in publications. What little data reuse there was in the articles pointed primarily to direct peer-to-peer sharing amongst the authors. This may potentially be the norm for veterinary researchers in general, as a result of the inability to network clinical and laboratory software throughout the discipline. There is, however, one good example of an established computer network for sharing veterinary research data. Banfield Pet Hospital, a corporation with veterinary clinics around the world, uses proprietary software called Petware, a centralized database of medical records (Banfield Pet Hospital 2015). In recent years, veterinary researchers have begun to mine this database and have reused clinical and laboratory data for research.

The results of this study also gave very little indication of formal data sharing via publication or a repository. Several articles stated that genetic sequence data was submitted to GenBank, but no other repository was mentioned. This could possibly point to a lack of awareness on the part of researchers but also to the lack of repositories specific to veterinary data. Only one such repository currently exists for veterinary clinical data, the Veterinary Medical Database (VMDB), which to date contains over seven-million clinical records (Veterinary Medical Database 2015). This may support the theory that simple lack of awareness of data repositories is an issue in veterinary medicine, but a lack of appropriate repositories may also be a barrier.

As was noted earlier, much of the current impetus for improved data management stems from pressure by the funding agencies, although such pressures do not appear to have reached a

tipping point in veterinary medicine. It is likely that including a data management plan in a grant application will be the norm sometime in the near future. Consequently, it is important to begin to pick apart the funding picture in different disciplines such as veterinary medicine. At an institutional level, this information will allow researchers and those supporting them to better prepare for and adapt to developing data policies and mandates. It is conceivable that funding agencies may create data management policies that are at odds with each other, with the researcher stuck in the middle trying to figure out how to reconcile the differences.

Conclusion

This type of case study provides a broad understanding of the types of data being produced in veterinary medicine, as well as clues to how data is being managed and shared. Having this base of knowledge to start from allows subject-specialist librarians to be more targeted in developing data services and to speak more knowledgeably with researchers about their needs. The results of this study revealed that while veterinary researchers have some established methods of managing and sharing their data, there is significant room for improvement. In particular, establishing best practices for data management, improving data storage options, and exploring better options for data sharing should be top priorities.

With the great amount of collaboration occurring in veterinary research, further understanding of these issues can contribute to the development of better data management practices of all stakeholders. There is an opportunity for librarians at this juncture to educate veterinary researchers about good data management practices. Assisting with the writing and review of data management plans and providing expert knowledge about metadata and data storage options are all examples of concrete ways for librarians to get involved. Whether these activities are part of an organized research data service or not, the key is for librarians to learn from researchers what works best for them and to adapt to evolving needs and circumstances.

Disclosure

The author reports no conflict of interest.

References

- Akers, Katherine G., and Jennifer Doty. 2013. "Disciplinary differences in faculty research data management practices and perspectives." *International Journal of Digital Curation* 8:5-26. <http://dx.doi.org/10.2218/ijdc.v8i2.263>
- Banfield Pet Hospital. 2015. "Putting Knowledge Into Practice." Accessed May 16. <http://www.banfield.com/veterinary-professionals/resources/research>
- Bellamy, James. E.C. 1999. "Veterinary informatics — why are we dragging our feet?" *Canadian Veterinary Journal* 40:861-863. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1539868/>
- Bracke, Marianne Stowell. 2011. "Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data." *Journal of Agricultural & Food Information* 12:65-74. <http://dx.doi.org/10.1080/10496505.2011.539158>
- Braxton, Susan, Harriett Green, Caroline Szylowicz, Thomas G. Habing, Karen Hogenboom, Chris Prom, Mary Schlembach, Sarah L. Shreeves, Sarah Williams, and Beth Sandore Namachchivaya. 2013. "eResearch Services at the University Library — Final Report of the Library eResearch Task Force." *The University of Illinois at Urbana-Champaign Library* <http://hdl.handle.net/2142/42561>

- Diekema, Anne. R., Andrew Wesolek, and Cheryl D. Walters. 2014. "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories." *Journal of Academic Librarianship* 40:322-331. <http://dx.doi.org/10.1016/j.acalib.2014.04.010>
- Grindlay, Douglas. J. C., Marnie L. Brennan, and Rachel S. Dean. 2012. "Searching the veterinary literature: a comparison of the coverage of veterinary journals by Nine bibliographic databases." *Journal of Veterinary Medical Education* 39:404-412. <http://dx.doi.org/10.3138/jvme.1111.109R>
- Johnson, Layne. M., Trevor R. Ames, Julie A. Jacko, and Linda A. Watson. 2011. "The informatics imperative in veterinary medicine: collaboration across disciplines." *Journal of Veterinary Medical Education* 38:5-9. <http://dx.doi.org/10.3138/jvme.38.1.5>
- National Institutes of Health. 2013. "NIH Awards by Location and Organization — NIH Research Portfolio Online Reporting Tools (RePORT) ." Last Modified February 28. <http://report.nih.gov/award/index.cfm>
- National Research Council (U.S.). 2011. *Workforce Needs in Veterinary Medicine*. Washington, D.C.: National Academies Press.
- One Health Initiative. 2015. "Mission Statement." Accessed April 21. <http://www.onehealthinitiative.com/mission.php>
- Peterson, Larry L., Daniel L. Ward, David J. Mischelevich, and John A. Shaddock. 1981. "A data management system for veterinary pathology." *Methods of Information in Medicine* 20:1-9. <http://methods.schattauer.de/en/contents/archivepremium/issue/1252/manuscript/14844.html>
- Santamaria, Suzanne L., and Kurt L. Zimmerman. 2011. "Uses of informatics to solve real world problems in veterinary medicine." *Journal of Veterinary Medical Education* 38:103-9. <http://dx.doi.org/10.3138/jvme.38.2.103>
- Sayogo, Djoko Sigit, and Theresa A. Pardo. 2013. "Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data." *Government Information Quarterly* 30:S19-S31. <http://dx.doi.org/10.1016/j.giq.2012.06.011>
- Smith, Ronald D., and Mitsuko Williams. 2000. "Applications of informatics in veterinary medicine." *Bulletin of the Medical Library Association* 88:49-51. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC35197>
- Smith-Akin, Kimberly A., Charles F. Bearden, Stephen T. Pittenger, and Elmer V. Bernstam. 2007. "Toward a veterinary informatics research agenda: an analysis of the PubMed-indexed literature." *International Journal of Medical Informatics* 76:306-312. <http://dx.doi.org/10.1016/j.ijmedinf.2006.02.009>
- Talbot, Richard. B. 1991. "Veterinary medical informatics." *Journal of the American Veterinary Medical Association* 199:52-57.
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* 6:e21101. <http://dx.doi.org/10.1371/journal.pone.0021101>
- United States Department of Agriculture. 2014. "Implementation Plan to Increase Public Access to Results of USDA-funded Scientific Research." Last modified November 7. <http://www.usda.gov/documents/USDA-Public-Access-Implementation-Plan.pdf>
- Veterinary Medical Database. 2015. "About VMDB." Accessed May 16. http://vmdb.org/?page_id=2
- Weller, Travis, and Amalia Monroe-Gulick. 2014. "Understanding methodological and disciplinary differences in the data practices of academic researchers." *Library Hi Tech* 32:467-482. <http://dx.doi.org/10.1108/LHT-02-2014-0021>
- Williams, Sarah C. 2012. "Data Practices in the Crop Sciences: A Review of Selected Faculty Publications." *Journal of Agricultural & Food Information* 13:308-325. <http://dx.doi.org/10.1080/10496505.2012.717846>
- Williams, Sarah C. 2013. "Using a Bibliographic Study to Identify Faculty Candidates for Data Services." *Science & Technology Libraries* 32:202-209. <http://dx.doi.org/10.1080/0194262X.2013.774622>