Exploring the Relationship Between Health Sciences Libraries and the Research Lifecycle

---

A Capstone

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

By

Bart Ragon

B.S., Lander University

M.L.I.S., University of South Carolina

May 2018

ABSTRACT

Health sciences libraries have traditional been the gatekeepers of medical knowledge. However, exogenous shocks in information technologies are challenging the role of libraries in supporting biomedical research. To maintain both legitimacy and relevance within the academic health sciences centers, libraries need to adapt quickly to the shifting landscape of academic biomedical research. Research lifecycle models have emerged in an attempt to define the research workflow from project inception to completion. Many libraries are evolving their services to better support biomedical research; yet, questions remain about what roles are appropriate for libraries in supporting biomedical research.

This study explored the changing demands on health sciences libraries created through evolutions in biomedical research workflows and investigated avenues for libraries to expand their role in the research lifecycle. The study was guided by the following research questions:

- What are the key activities in the research lifecycle for biomedical researchers?;
- What aspects of the research lifecycle are evolving, if any, due to emerging practices in biomedical research?;
- In what ways do health sciences libraries support the research lifecycle and emerging practices of biomedical research?
- What skills and practices might health sciences libraries cultivate in prompting new roles in the research lifecycle?

This research strengthens what is known about the research lifecycle and the support provided by libraries. Combined with awareness of organizational needs, results can provide a useful framework for health sciences library leaders to guide their organizations in discovering new roles for libraries in addressing the emerging needs of biomedical research.

Bart Ragon, MLIS

Education Leadership, Foundations and Policy department
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE CAPSTONE

This capstone, Exploring the Relationship Between Health Sciences Libraries and the Research Lifecycle, has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Education.

Brian Pusser, Ph.D., Chair

Catherine P. Bradshaw, Ph.D., Committee Member

Aaron F. Pannone, Ph.D., Committee Member

Justin B. Thompson, Ph.D., Committee Member

_______________________Date

DEDICATION

Dedicated to my wife, who stands with me even as I pursue Gatsby's green light. Dana, unlike Gatsby, we will run faster, stretch our arms farther, and not be eluded by our orgastic future. And so we beat on, boats with the current, borne forward, ceaselessly into the great unknown.

ACKNOWLEDGEMENTS

It would be ludicrous to start my acknowledgements anywhere but with my parents. Primarily I blame them for this degree. As a child, their relentless desire for me to read, matriculate, and evolve really took its toll, forcing upon me an aspiration to never stop learning. Thanks mom and dad for all that you have ever done. I love you both!

To my committee, I cannot thank you enough. First and foremost, Brian Pusser, who is a very, very patient man. Thank you for your mentoring, dedication, and humanity. On numerous occasions I didn't make this easy on either of us, but you got me through it, nudging me along with your positive demeanor and patiently waiting for me to understand the path forward. I am truly grateful to all my committee members. Catherine Bradshaw, who is far too impressive to have time for my committee, but found the time anyway. Justin Thompson, who I have come to appreciate for his kindhearted individual strength and thoughtful professionalism. Aaron Pannone, who I also have the pleasure of working with in my job, and who put up with my numerous methods and analytics questions. Along with my impressive committee, it is important to recognize the contribution the Curry School of Education has made to my development. All of the faculty who taught my courses were fantastic and challenged me to see issues in higher education from different perspectives. Curry also helped support my research with funding through the Curry Innovative, Developmental, Exploratory Awards (IDEA) program. The IDEA award allowed me to gather more data and strengthened my overall project.

Over the last six years many of my colleagues have also helped me along the way by listening to me and even challenging my assumptions. I couldn't have done this without their support. My director, Gretchen Arnold, has been extremely supportive throughout my career and has helped me grow into the professional I am today. I can't imagine a more supportive boss. Jeff Fox and Arlyn Burgess my friends, colleagues, and comrades here at UVA. Jeff for being the nicest person you could ever meet, and Arlyn for always arguing the opposite of whatever I just said, and I mean always. You both helped me sharpen my understanding of what I was trying to say. Andrea Denton for being the best librarian I know and for her willingness to travel down the rabbit hole with me. Stephen Turner for helping me understand a new path for libraries and the value of data science. Ron Hutchins for always challenging assumptions and never giving up. Michele Claibourn for being awesome and lighting the way through her leadership. Andrew Sallans for being the genius before the rest of us were ready. Pete Alonzi for his quantum way of thinking. Troy Buer, who at the beginning of all of this, encouraged me to start the program. Phil Bourne for helping me validate my selection criteria and for his passion for open data. Sandra Burks and Karen Johnston for chasing excellence. David Martin and Marieke Jones for embodying the example of a new kind of information scientist (p.s. thanks David for the script). Gabe Rios, my friend, mentor, and Jedi Master, thank you for helping me understand the type of librarian I wanted to become and for teaching me the ways of the Force. Thanks also to Wallace McLendon for his futuristic vision of libraries and for the wisdom that scotch made one thirsty and could be quenched by beer. Lastly, my gratitude to Kristi Holmes for her shy awesomeness and for being the instrument of change for the rest of us to follow.

Along with my fantastic colleagues, there have been others who have supported me in the wider aspects of life or by just being there when I needed a release. Thank you Jaron Bushman for weeknight beers, they help me stay sane. You heard about my project from start to finish. Russ Manley for his friendship, willingness to ask creative questions, and love of hiking. Roger that Roger for helping me build the keezer and for being such a good friend. Finally, thank you to the three souls who have helped me stay balanced and compassionate. The impact of the love from a good dawg cannot be overstated. Annie Baby Puppy Dawg, who demanded only what she wanted. Bogie, who is and will always be my girl. Weezy, rambunctious, crazy, and sweet lovin' Weezy who lays next to my heart. Oh how we love you, now go get me a toy!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The creation of new knowledge has long been associated with innovation and social well-being (David & Foray, 2002). Libraries have supported access to knowledge since the classical cultures of Babylon, Assyria, Greece and Egypt to the establishment of the first academic health sciences libraries at the Faculté de la Médecine of Paris in 1395 (Jayne, 1916).[1] Libraries grew as medical education transitioned from medical philosophy to medical science during the second half of the 19th century. The first library in the United States devoted entirely to medical knowledge was established in 1762 at Pennsylvania Hospital in Philadelphia (Jayne, 1916). In the United States, an estimated 15,000 young and ambitious American physicians traveled to Germany between 1870-1914 to learn new techniques that focused on the experimental method and laboratory science, accelerated this transformation by bringing back with them university-based medical training with science at the core (Bonner, 1963, p. 23; Kaufman, 1976). In 1910, the Flexner Report reviewed all medical schools in North America and ushered in a new wave of reform after it described medical education in the United States as lacking adequate foundations in science (Flexner, 1910). The transformation of medicine during the twentieth century made access to libraries and laboratory research essential to medical education (Ludmerer, 1999, p. xxiii).

---

[1] Health sciences library is a general term used throughout this paper in lieu of other terms such as medical library or biomedical library. To some readers the term medical library may imply a more clinical functions, where biomedical library may imply scientific functions. Health sciences library is a broader term that encompasses all aspects of library support in academic health sciences centers.

In the United States, the National Library of Medicine (NLM) helped pave the way for health sciences libraries supporting access to knowledge. From 1887 – 1962 the NLM expanded access to knowledge through national loaning services for library materials and by enabling the adoption of emerging technologies, such as microfilming and photocopying ("A Brief History of NLM," n.d.). During this time, the number of health sciences libraries at academic institutions increased significantly (Birchette, 1973). Throughout the second half of the twentieth century, libraries leveraged the technology of the day to enhance access to knowledge. According to Wayne A. Wiegand (1999), as early as the 1950s libraries utilized computers to provide document reference. In 1971 a significant advancement occurred when *Index Medicus*, a paper-based information retrieval system, was transformed into the computerized bibliographic system the MEDical Literature Analysis and Retrieval System (MEDLARS) (Wiegand, 1999). MEDLARS was a precursor to web-based systems that would emerge in the 1990s. During the 1990s, publication formats transitioned from print to electronic, altering the way information was accessed and retrieved (Odlyzko, 1999). Card catalogues, print indices, and print material transitioned to online books, journals, and bibliographic databases. In the early part of the new millennium, libraries evaluated how users sought to access knowledge and transformed their collections, services, and space to accommodate internet-based paradigms (Brown, 1999; Haines, Light, O'Malley, & Delwiche, 2010).

In the new millennium, information technologies inspired new concepts like the knowledge-economy and knowledge society as society attempted to put into context the role of digital information as sources of knowledge (David & Foray, 2002). New technologies allowing information to be “born digital” began affecting the collection of biomedical knowledge, later expanding to include the scientific workflows in medical research. As the ability to generate and analyze data has expanded, biomedical research has become increasingly computational. Thus, leaving a skills gap for medical experienced researchers not trained in new technologies and techniques. Additionally, external forces such as federal funding policies, open science initiatives, and data science techniques are altering the dissemination of knowledge.

Research lifecycle models have emerged in an attempt to define the research workflow from project inception to completion, addressing each step that occurs during the research process (“Research Lifecycle | e-Science Portal for Librarians,” n.d.). While there is no standard definition of the steps in the research lifecycle and terms used to describe individual activities vary, most models document the major components to include planning/proposal, research and analysis, and dissemination. Additionally, most research lifecycle models include information on the data lifecycle. Data lifecycle models address the stages of research data from its collection during the project phase in the research lifecycle to the management and curation of data after project completion, including possible reuse of the data by other researchers (“Data Lifecycle | e-Science Portal for Librarians,” n.d.). Organizations such as the Association of Research Libraries (ARL), DataONE, and National Network of Libraries of Medicine New England Region (NN/LM NER), have created educational programs and information resources to enlist librarians as support partners for researchers in the management of their research data (“Data Access, Management & Sharing | Association of Research Libraries,” n.d., “Data Lifecycle | e-Science Portal for Librarians,” n.d., “Librarian

Outreach Kit | DataONE," n.d.)(Association of Research Libraries, n.d.). To a lesser degree, the research lifecycle has been used as a framework for mapping the scientific workflow to library support services. Within the research lifecycle, some health sciences libraries have sought to more effectively support knowledge creation through new services that include data management planning, funding discovery, federal public access compliance, and metadata support.

Research support provided by contemporary health sciences libraries can be characterized as traditional, transitional, and evolutionary. Traditional services are those services in which health sciences librarians have historically had a role, such as literature searching. Transitional services apply the existing skills of librarians through different media. For example, many libraries have transformed skills in cataloging books or journals into metadata services describing digital objects. Evolutionary services are services that require and provide new skills and training, such as bioinformatics support (Geer, 2006). Many health sciences librarians lack the skills to fully address the emerging information needs of many biomedical researchers, despite advances in transitional and evolutionary services for biomedical researchers, Yarfitz and Ketchell (2000) found that less than 20% of librarians, who did not possess a strong background in molecular biology or specialized training, could answer bioinformatics consultation questions, potentially illustrating a support gap for researchers adjusting to changes in scientific workflows. Some libraries have replaced vacant reference librarian positions with Ph.D. life scientist to address gaps in information needs related to bioinformatics (Li, Chen, & Clintworth, 2013).

In the profession, questions remain about what roles are appropriate for libraries supporting biomedical research. Researchers may have a traditional view and see libraries only as the repositories of knowledge. Further, librarians accustomed to roles as information gatekeepers may lack training for the development of evolutionary services, or may not believe support or bioinformatics is an appropriate direction for the profession. This study investigated the changing demands placed upon health sciences libraries created through evolutions in research workflows. It addresses ways in which health sciences libraries can expand their role in supporting knowledge creation within the research lifecycle. Findings from the study aim to strengthen what is known about the types of support needed by biomedical researchers and the efficacy of support services provided by health sciences libraries. The primary research questions in this study are:

1. What are the key activities in the research lifecycle for biomedical researchers?
2. What aspects of the research lifecycle are evolving, if any, due to emerging practices in biomedical research?
3. In what ways do health sciences libraries support the research lifecycle and emerging practices of biomedical research?
4. What skills and practices might health sciences libraries cultivate in prompting new roles in the research lifecycle?

Exploration of library support for the emerging needs of the research lifecycle help to define the changing relationship between researchers and health sciences libraries and reveal potential new roles for health sciences librarians. Data collected from the experiences of biomedical researchers and health sciences library leaders provide a direct lens for understanding the phenomenon of the research lifecycle and provide insight into how health sciences libraries support knowledge creation within this paradigm. The research questions help to explore the potential gap between the needs of researchers in the research lifecycle and the current or planned support for the research lifecycle by health sciences libraries.

CHAPTER 2

LITERATURE REVIEW

Origins of the Health Sciences Library

Health sciences library history is closely linked to the accessibility of medical knowledge. The earliest forms of medical writing date to 2000 B.C. (Birchette, 1973). When German medical science began utilizing the scientific method and university laboratories during the 19th century, access to medical knowledge became a more significant aspect of supporting scientific research. The impact of laboratory medicine led to increased demand to capture knowledge and make it accessible. William Welch noted from his visit to German medical universities and laboratories the importance of a working library with important books and periodicals in the field (Bonner, 1963, p. 109). From 1898 to 1958 health sciences libraries experienced unprecedented growth in the United States, expanding from 120 to 506 libraries with the average collection size increasing 22 percent (Birchette, 1973). The role of the library increased as physicians moved away from personal collections to organizational collections. Between the period of 1862 – 1866 the National Library of Medicine (NLM) began sharing medical knowledge by loaning materials and providing support for research questions ("A Brief History of NLM," n.d.). In 1898 the Association of Medical Librarians was created as the first professional organization supporting access to medical literature (Birchette, 1973). Prior to World War II, the Association of Medical Librarians, later renamed the Medical Library Association (MLA), primarily supported the establishment of specialized medical libraries and the exchange of medical literature between these libraries (Birchette, 1973; Connor, 2011).

Early research support. Literature on the health sciences librarianship as a profession during the first half of the 20th century highlights the need for a defined role for libraries in supporting access to knowledge. Within that literature, support for research is a recurring theme. In 1925, James Ballard suggested a curriculum for health sciences librarians that included "methods of research" as a core component to the training of librarians (J. Ballard, 1925). Again in 1943, Jennie Greenbaum analyzed the functions and duties of health sciences library staff, stating that knowledge of current research projects and trends was a primary skill needed by librarians so that they could identify overlooked publications to researchers (Greenbaum, 1943).

Professional librarianship. Post World War II, two factors began to influence change within health sciences libraries. First, MLA redefined its mission from that of an information exchange run by physician leaders, to an organization led by and supporting librarians (Connor, 2011). Although MLA was founded in 1898, it did not include a professional membership category for library workers until 1929 (J. F. Ballard, 1998). A more significant change occurred in 1946 when the MLA constitution was amended, placing physician leadership in a subordinate role to working librarians (Connor, 2011). Second, massive increases in federal spending supporting scientific research amplified the need to make knowledge accessible. Increases in federal funding fueled the notion of an "endless frontier", one that was partially designed to increase standards of living by utilizing research universities as the intellectual centers for biomedical science (Guston & Keniston, 1994, p. 1). From 1969 to 1979 the number of specialty health sciences libraries, such as medical society and industry libraries, decreased overall during this period, academic health sciences libraries increased by 25% (Crawford, 1983). Growth of academic health sciences libraries was also fueled by the federal government through the Medical Library Assistance Act of 1965, which directed the NLM to create a system of regional libraries supporting research, scientific publication, and training of the next generation of librarians (Cummings & Corning, 1971). Between 1964 and 1980 federal support for NLM increased by more than a factor of ten, growing from $4,055,000 to $46,350,000 (Crawford, 1983).

Qualitative crises surrounding health sciences library identity. With increases in federal spending, the proliferation of health sciences libraries, and the emergence of the health sciences librarian as a profession, it might be expected that this time was the golden era of health sciences libraries. However, preoccupation with recruitment and librarian skills demonstrates the challenge for the profession (Anderson, 1989). Another consistent theme documented in the library literature of the time centered on the changing qualifications needed for professionals. Libraries questioned how closely their qualifications should have a foundation in medicine or science. Kronick Rees, and Rothenberg (1972) state that there was not a quantitative problem in terms of manpower, but a qualitative crisis, noting that 75% of librarians had little or no training in sciences related to health. During the 1980’s, concerns over the adaptability of the professional and education persisted, heightening awareness about the professions ability to adjust to advances in information technology and computing. Robert M. Braude (1997) used an evolutionary metaphor to describe how 500 years of support for biomedical knowledge was being disrupted by rapid changes in information technologies. Health sciences libraries feeling the impact of technology realized that a professional transformation was underway. Erika Love (1987) lists systematic continuing education and "participation in research on the fundamentals of the information process" as the means for political survival. Love links science to information technology while urging the profession to rethink its priorities, adding, "unless we accelerate our own approach to knowledge and information, our profession is in danger of becoming extinct, or at best, a servant of other disciplines. Strengthening our research component today is a matter of political survival."

Janet Doe lectures. Alison Bunting's (1998) analysis of the Janet Doe Lectures from 1967 – 1997 provides an insight to how the profession considered questions related to changing roles, educational needs, and credentialing. Janet Doe lectures are delivered annually at the MLA's annual meeting and are considered to be a landmark event in a health sciences librarians' professional career. The focus of the lecture is on the history and philosophy of health sciences librarianship, encouraging the lecturer to consider where the profession has been and where it is headed (Bunting, 1998). As such, the Doe lectures are an indicator of what luminaries in the field believe is important to the profession. Bunting's analysis illustrates how professional development of health sciences librarians was a major theme, including changing roles of librarianship and library education as significant sub-categories (Bunting, 1998). For example, in his 1996 Doe lecture, Robert M. Braude sees education as "our adaptive strategy, our process of selective differentiation if you will, whereby we changed our beak to meet the changing conditions of our territory" (Braude, 1997).

Consumer-Based Knowledge Systems

On June 26, 1997, Senator Tom Harkin announced that free web-based access was available through the NLM medical bibliographic databases, PubMed and Internet Grateful Med (“NLM Technical Bulletin. May-Jun 1997,” n.d.). Prior to this time, librarians acted as gatekeepers to the digital form of this knowledge through the MEDLARS (Nicoll, 1992). For the first time, direct consumer access to digital medical knowledge was provided to the library user without the need of librarian assistance. Opening up access to medical knowledge was the first step in “consumerizing” medical research to the masses. The consumerization of information technology is characterized as the impact of consumer-oriented technology on an enterprise organization (“Consumerization - Gartner IT Glossary,” n.d.). However, health sciences libraries continued to see themselves as gatekeepers of knowledge. Robert M. Braude (1997) states “The basic role of the health sciences librarian has not significantly changed throughout history. It has been-and remains-to collect information and organize it for effective use. What has changed is the environment in which this role is carried out and the tools used to accomplish the tasks”. What Braude could not predict in 1997, was how exponential growth of the internet would alter the expectation of access to information by library users. In 1995, 14% of Americans used the internet, as compared to 50% by the year 2000 and 72% in 2005 (“Internet Use Over Time | Pew Research Center,” n.d.). Consumer-based access to medical knowledge created questions about the skills needed to support access to knowledge. Librarians sought to transition their ability to located scholarly print material to internet sources. Exponential growth of the internet during the early 1990’s fostered the concept of “digital libraries” as the library of the future. For health sciences libraries, digital libraries were seen as a way to aid physicians in rapidly locating answers that supported clinical care (D’Alessandro, Galvin, D’Alessandro, Erkonen, & Choi, 1999). Digital libraries were perceived as innovative ways to deliver new services as libraries transitioned from being “information place” to creating “information space” (Lucier, 1995).

The Informationist Model

Despite the increasing growth of the internet, time constraints and the complexity of efficiently locating information presented challenges to clinicians who struggled to integrate evidence-based information in clinical care. Health sciences librarians were stunned in 2000, when the Association of American Medical Colleges (AAMC) proposed the need for a new health profession, the informationist (Davidoff & Florance, 2000). Davidoff and Florance (2000) acknowledged the disappointing reality that physicians did not consult medical literature in the course of clinical care. They proposed a national program designed to train a new kind of information specialist, one who would be included as part of the medical team. Reaction from Kronenfeld (2000), was that Davidoff and Florance were proposing a profession currently served by clinical librarians. Kronenfeld went on to add that what was needed was greater advocacy for the role of the librarian in supporting knowledge-based information. Regardless of backlash by health sciences librarians, the article inspired a new dialogue about the role of the profession in supporting access to knowledge. In 2002, the National Library of Medicine’s Lister Hill Center hosted a conference that sparked a national discussion promoting the development of librarians as informationists (Shipman, Cunningham, Holst, & Watson, 2002). At the conference, the informationist concept was expanded to include skills beyond clinical support, inspiring a new role for librarians, the research informationist (Shipman et al., 2002).

Growth of data and informatics. Hersh (2002) suggested medical informatics training as an alternate path for informationist. Medical informatics involves acquiring, storing, and using information specifically through the application of information technology (Hersh, 2002). While this definition appears to be very close to how health sciences libraries function, it is broader in the sense that it can include types of information not typically supported by libraries. At the same time, academic medical centers began collecting and storing more digital information, such as electronic medical records, in the form of data. Projects such as the Human Genome inspired thousands of research projects by using standard data models and making data freely available on the internet (Kohane, 2000). In biomedical research the application of information technologies and machine learning techniques were seen as the means for extracting knowledge from clinical databases (Kohane, 2000). As more information was collected and became available, the information support needs increased, generating the potential for new types of information stewards.

NLM administrative supplement awards. In 2005, the National Library of Medicine offered a prediction for the future of health sciences libraries (Lindberg & Humphreys, 2005). Lindberg and Humphreys imagined librarians working outside the walls of the library as contributing members of health care teams, helping to efficiently extract information from the wealth of information created by the post-Google information world. To incentivize collaboration between librarian informationist and the National Institute of Health (NIH) grant awardees, the NLM began sponsoring administrative supplement awards in 2012 that seek to embed librarians within teams of biomedical research teams. Health sciences libraries who received the awards explored new roles including creating a data dictionary, analyzing data collection workflows, and data management planning (Henderson, 2014, p. 409).

A common theme identified by informationists who received the NLM supplement program awards was the value of librarians embedded within research teams. Lisa Federer (2013) explained how attendance of the research team weekly meetings provided context for an informationist who lacked formal training in the subject matter. Hasman, Berryman, and McIntosh (2013) described how their experience led to a better understanding of the grant submission process and the research team's culture. Sally Gore (2013) noted that becoming integrated with a research team allows the library to provide greater value than traditional library support roles. Hansn, Bakker, Svirsky, Neuman, and Rambo, (2013) saw their informationist role as case study for understanding the data management needs of researchers and as a potential funding model for future grant proposal.

Goode and Anton (2013) noted that in addition to technical complexities, their work as informationists required a systematic review and other forms of literature searching. According to Hasman et al., (2013) they were surprised to discover that traditional librarian roles were needed throughout the project. Two of the supplements awards had traditional library services, such as systematic reviews and the management of bibliographic information, as specific aims written within the grant proposal (Gore, 2013; Surkis et al., 2013). Almost all of the projects described in the literature reveal libraries defaulting to traditional services of literature searching and systematic reviews.

It is unclear if traditional services were the true need of the projects or a reflection of the skills brought to the research team by librarians. Many of the projects included providing transitional services by creating, aggregating, or managing data (Federer, 2013; Goode & Anton, 2013; Gore, 2013; Hansn et al., 2013; Hasman et al., 2013; Surkis et al., 2013; Whipple, Odell, Ralston, & Liu, 2013). Federer added that as the data sharing requirements that accompany federal funding become more stringent, libraries can become essential members of the research team, but that they should consider how to prepare staff to provide specialized services for researchers (Federer, 2013).

External Forces

Demographics of the workforce. Libraries are challenged in transitioning their workforce to support data science workflows by a workforce that was trained in a pre-internet era. Many mid to late stage career librarians pursued library degrees in the print era and are consequently inadequately trained to support next generation research tools in an era of collaborative networked science (McGowan, 2012). A study by Cataldo, Tennant, Sherwill-Navarro, and Jesano (2006), suggests that health sciences library patrons value librarians who have subject area expertise. The need for greater subject knowledge has led some libraries to recruit non-librarian's skillsets. For some librarians this was viewed as an affront to the profession, but one survey found that 40% of academic library directors who hired non-MLS applicants did so as a response to a need to expand the nature of the library work and enhance the applicant pool to include additional skillsets (Simpson, 2013). If health sciences libraries are to succeed in defining new roles supporting biomedical research, they must combine their established services supporting knowledge with skills that entice researchers to seek out their services.

Background and curriculum. Health sciences librarianship has been slow to respond in providing professional development for librarians without a science background (Petrinic & Urquhart, 2007). A lack of formal training in biomedical sciences can lead to a lack of confidence, both by the librarians and the patrons they support. This presents a challenge for health sciences librarians with liberal arts backgrounds (J. Lyon, 2003). William Hersh (2002), suggested a medical informatics curriculum without the need for a MLS that included computer programming and quantitative methods to aid in supporting research and analyzing data. Hersh acknowledged the "fuzzy boundaries" between the training needs for librarians and medical informatics, leaving questions on how to train librarians for new roles requiring informatics expertise. Petrinic and Urquhart (2007), examined whether librarians with generalist backgrounds could transition into roles that demanded expert knowledge in the health sector. They discovered that continuing professional education was needed to meet rapidly changing needs, because library school curriculum was inadequate in preparing librarians to support scientific research.

Open access to scholarly content. As the curators of academic knowledge, traditional library research support has revolved around information discovery, collection development, and information management (Corrall, Kennan, & Afzal, 2013). As health sciences libraries investigated meeting new demands to support bioinformatics, advances in computing and networked technologies also created new opportunities for libraries. The widespread growth of the internet during the 1990s had a dramatic impact on modes for publishing and accessing scholarly articles (Wren, 2005). According to Laakso, Welling, Bukvova, Nyman, Björk, and Hedlund (2011), the internet enabled low-cost distribution of digital content. Academics dissatisfied with existing publishing models were among the first to take advantage of low-cost internet publishing in an effort to make knowledge more freely available through an open access model of distribution (Björk, Shen, & Laakso, 2016). The rising cost of journal subscriptions created incentives for institutions and researchers to consider peer-reviewed open access publishing models. According to the Association of Research Libraries (ARL), between 1986 and 2001 journal expenditures for its member libraries increased 210 percent (Liesegang, Schachat, & Albert, 2005). As a result of the sharp increase in journal subscriptions, by 2005 it appeared research libraries would no longer be able to afford to serve as the repositories of scientific knowledge (Liesegang et al., 2005). Björk et al. (2016) define open access in the realm of scholarly publishing as the unrestricted access to scholarly journals online. The growth of the open access movement began in the late 1990s (Corrado, 2005; Liesegang et al., 2005). Notable open access initiatives in biomedical publishing include PubMed Central, Public Library of Science, and BioMed Central (Liesegang et al., 2005). Efforts like the Public Library of Science (PLoS) have sought to challenge traditional publishing models through the creation of its own journals, which utilize an open access model (Brower, 2001).

Open access and federally sponsored research.[2] The federal government is the largest investor in academic biomedical research with the National Institute of Health

---

[2] Open access and federally sponsored research section adapted from researcher's prior work (Ragon, 2013). Adaptation is allowed under Creative Commons Attribution—Non-Commercial—Share Alike License ("Journal of eScience Librarianship: Final Manuscript Preparation Guidelines," n.d.).

contributing 57% of support for university-based research (Calhoun, 2006). Since NIH's investments are made on behalf of taxpayers, the relationship between federally funded research and scholarly publications is complex. In a simplified model, the federal government sponsors research for the benefit of citizens by providing grants to investigators. Findings from research are generally disseminated though publication in scholarly peer-reviewed journals. Since most publishers of scholarly journals charge for access to published content, taxpayers do not have access to the findings for research they helped fund. In 1999, then director of NIH Harold Varmus, proposed the creation of an electronic archive for biomedical research data – later to become PubMed Central (Bloom, 1999). In 2005, NIH created a policy requesting investigators deposit an electronic version of publications to PubMed Central within 6-12 months of publication ("NOT-OD-05-022: Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research," n.d.). In 2008, President George W. Bush signed the Consolidated Appropriations Act which contained a provision requiring the mandatory deposit of peer reviewed articles for NIH funded projects ("NOT-OD-08-033: Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research," n.d.). In 2013, key issues emerged reigniting debate over open access. The White House Office of Science and Technology Policy (OSTP) released a policy requiring federal agencies to ensure publications are freely available to the public (Stebbins, n.d.). Further, language in NSF Biosketch access requirements was changed from "Publications" to "Products" ("NSF Grant Proposal Guide," n.d.). NSF clarified that products of research might include, but was not limited to publications, data sets, software, patents, and copyrights.

Role of Librarians in Scholarly Communication

As the debate about open access continued, libraries began to expand their roles in supporting access to scholarly publications by assisting researchers in understanding and participating in open access publishing. The American Library Association, Association of College & Research Libraries, and the Association of Research Libraries worked to educate librarians on open access issues, offered suggestions on how they could participate in advocacy (American Library Association, n.d.; Association of Research Libraries, n.d.). In a systematic review of new roles for health sciences librarians, Cooper and Crum (Cooper & Crum, 2013) described a new role of Scholarly Communications Librarian as one that encourages participation in open access and develops open institutional repositories that disseminates the results of research. Cooper and Crum noted the departure from a traditional view of scholarly communication focusing on the dissemination of scholarly work, to one that includes the creation of new knowledge. In their words, “it encompasses the entire process by which faculty, researchers, and other scholars share and publish their findings within and beyond the academic community.” Changes in biomedical research combined with open access initiatives became even more complex as technology altered what was possible in the creation of new scientific knowledge, primarily new forms of data.

Scholarly literature and data. Today’s information technology allows for the capture of scientific processes as they are created during the scientific workflow. Hey, Tansley, and Tolle (2009) refer to evolutions in science as the Fourth Paradigm, where scholarly literature and its underpinning data are unified online. Access to scholarly literature and data is becoming increasingly important to the advancement and reproducibility of science. According to Corrall, Kennan, and Afzal (2013), national and international policy developments have “created opportunities for libraries to create value by extending their stewardship and service activities to the management and sharing of research datasets as an increasingly vital dimension of the global research knowledge base”. Since data can be collected, described, and linked, health sciences libraries have an opportunity to support access to these objects as sources of information. Efforts by health sciences libraries to expand research support have increasingly altered how libraries supported knowledge.

Support of the Research Lifecycle

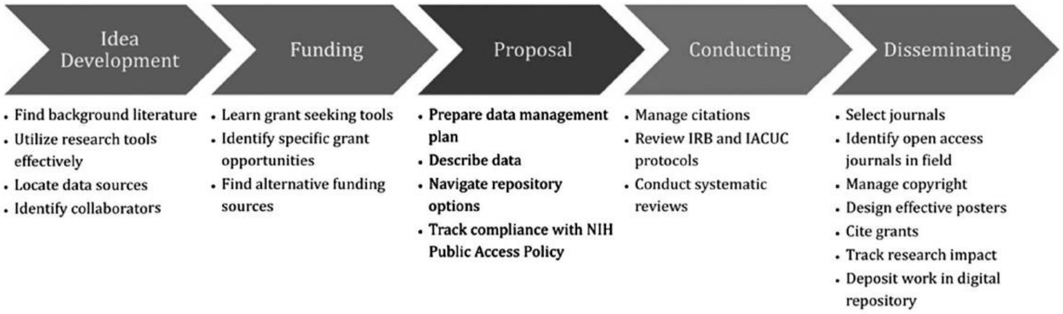
Bioinformatics support. In the early- to mid-2000s, new models for health sciences libraries to support research began to emerge. One model included an expanded role in supporting bioinformatics through librarians holding advanced scientific degrees. Lyon, Tennant, Messner, & Osterbur (2006) describe how their backgrounds in basic science allowed them to create niche bioinformatics support models at four universities. Activities in these models enabled them to explore new methods for supporting research. New areas included the development of workshops and seminars related to bioinformatics tools, course-integrated instruction, as well as reference and consultation (Lyon et al., 2006; Robison, 2008) Lyon (2003) suggests that health sciences librarians do not have the luxury of avoiding bioinformatics if libraries wish to maintain their role as information provider and contributing partner.

Research workflows. As health sciences libraries continued to explore new roles supporting research, language used in the health sciences library literature describing new activities evolved. Informationist, bioinformationist, and bioinformatics librarian, have all been used to describe new and emerging activities (Rankin, Grefsheim, & Canto, 2008; Robison, 2008; Tennant, 2005). Anna Gold (2007), states that the key to librarians having a more significant involvement is being positioned as support partners throughout the research process. Gold's concept is based on what D. Scott Brandt (Eden, 2012, p. 104) refers to as libraries "working upstream" in the research process. Brandt's reasoning is that if research workflows are imagined as a river, then libraries have typically worked downstream collecting scholarly materials as they are published. Brandt suggests that libraries can expand their role by engaging in research with researchers, the opposite end of the scholarly communication spectrum.

Opportunities in the research lifecycle. Library services supporting the research lifecycle have typically aligned with activities at the beginning and end of the lifecycle. At the beginning of the research lifecycle is the discovery phase, where researchers seek out the foundational literature that their work will build upon. At the end of the research lifecycle librarians have collected, described, and made available the published scholarly work. In the 1990's, as publication formats transitioned from print to electronic, library information systems evolved from print indices and bound journals to electronic bibliographic databases and online journals and books. In the new millennium, librarians began increasing their research support role by creating new services along the continuum of the research lifecycle. New activities for libraries include data management planning, funding discovery, public access policy compliance for federal funding, scholarly communications, bibliometric, and metadata support.

Numerous research lifecycle models are documented on academic library websites describing to researchers how library services support their work. In biomedical peer-reviewed literature, only one model was discovered connecting library support services to the research lifecycle. Vaughan et al. (2013) addressed how niche services by individual librarians could be developed into a standard service model supporting the entire research lifecycle. The authors used concept mapping and an interactive poster session with the research community to discover which services were acknowledged as the most valuable and to identify potential new roles. The resulting model from the study outlines 5 major areas of support; Idea Development, Funding, Proposal, Conducting, and Disseminating. Support roles for libraries are listed under each major area (see Figure 1).

Figure 1. Library services across the research lifecycle



Reprinted with permission. Vaughan, K. T. L., Hayes, B. E., Lerner, R. C., McElfresh, K. R., Pavlech, L., Romito, D., … Morris, E. N. (2013). Development of the research lifecycle model for library services. *Journal of the Medical Library Association : JMLA*, *101*(4), 310–4.

Within the model, several services are identified including those in which health sciences librarians have historically had a role, such as background literature discovery, systematic reviews, or helping others in understanding copyright. Other services are highlighted that utilize a librarian's existing skills in a different medium to support a research process. For example, discovery and retrieval of datasets illustrates a new service role that leverages the existing skills of librarians in retrieving journal literature. Other services may require new skills and training. Support for research networking systems, seeking grant funding, NIH compliance support, IRB participation, and bibliometrics are examples that may require libraries cultivate new skills. The results of the study by Vaughan et al. are limited in that they did not outline specific new skills needed for the profession and did not consider how research workflows themselves are evolving.

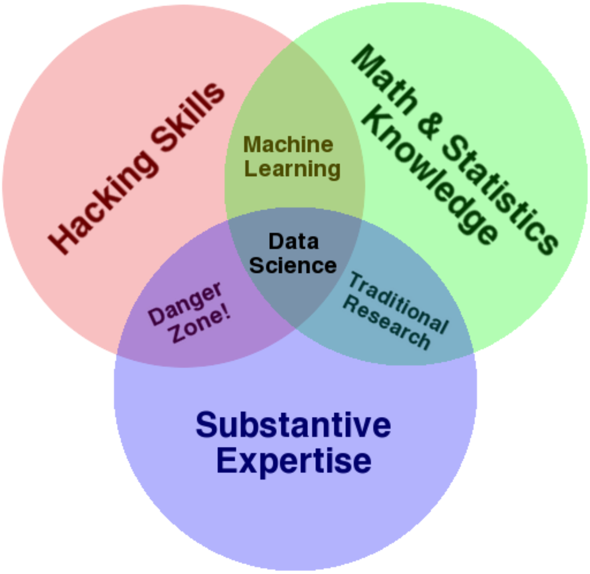
One library's approach to defining the research lifecycle was to document activities in the lifecycle and the departments who supported those activities. The University of Central Florida (UCF) developed a map of the Research Lifecycle based on the OpenWetWare model, a group of researchers interested in increasing the amount of organization, dissemination, and communication in biological research." ("OpenWetWare:FAQ - OpenWetWare," n.d.) OpenWetWare developed a lifecycle model that consists of four subcycles. The subcycles are Research Planning, Experimental, and Publishing ("OpenWetWare:Headquarters/Research Pathway - OpenWetWare," n.d.). UCF modified the OpenWetWare Model to include four subcycles: Planning, Project, Publication, and 21st Century Scholarship ("Overview: Research Lifecycle - UCF Libraries," n.d.). The UCF model was developed by the library to document the steps of the research lifecycle and identify the services available to researchers. The model also identified several gaps in support including data curation, data sharing, data visualization, analysis support, and long-term preservation.

Data Science

Data science is an emerging field having a dramatic impact on many domains, including biomedical science. The concepts of bioinformatics and computational biology are closely linked to data science. The NIH defines bioinformatics as, “Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data” (Huerta, Downing, Haseltine, Seto, & Liu, 2000). NIH defines computational biology as, “The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.” Thus, bioinformatics describes the application of data science techniques for statistical analysis, where computational biology describes the application of data science techniques for scientific modeling. The terms are so closely linked that NIH’s National Institute of Dental and Craniofacial Research administers’ request for application announcements are under the program “Bioinformatics, Computational Biology and Data Science” (“National Institute of Dental and Craniofacial Research: Requests for Applications,” n.d.).

While there is no single definition of data science, Drew Conway created a Venn diagram of data science that has been commonly used to represent the components of data science (see Figure 2). The diagram contains three major components; Hacking Skills, Math and Statistical Knowledge, and Substantive Expertise (“The Data Science Venn Diagram — Drew Conway,” n.d.).

Figure 2. The data science Venn diagram



The Data Science Venn Diagram available at http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagramunder under Creative Commons licensed as Attribution-NonCommercial.

Hacking skills, as connected to data processing, uses programming techniques to prepare and analyze data. In biomedical research, open source programming languages such as R and Python, have been embraced by the biomedical research community who actively writes and shares code to address domain challenges. Mathematical and statistical knowledge is associated with skills to test and validate research results. Substantive expertise is specialized knowledge and training in the scientific domain. In Conway's model, knowledge in any one or two of the components represents an incomplete or even dangerous application of data science. The potential gap between hacking skills and statistical knowledge is an opportunity for health sciences libraries to create new roles within research lifecycle by integrating their support during the phase where researchers are collecting data and analyzing results.

Like a book or journal, research data is a source of information. Researchers use and generate datasets, but receive little or no training when it comes to organizing, managing, describing, and archiving their data. Health sciences libraries have traditionally specialized in organizing, managing, describing, and archiving information resources. Significant challenges exist in preparing data prior to analysis, even though much of the emphasis is placed on insights offered by the data science techniques themselves. According to the *New York Times*, 50% to 80% of time is spent on what data scientist call “data wrangling,” “data munging” and “data janitor work” (Lohr, 2014). As service entities with a long-standing commitment to supporting the biomedical research community, health sciences libraries could position their services to consult with researchers on how to clean and analyze data used in scientific research.

Research Software Skills

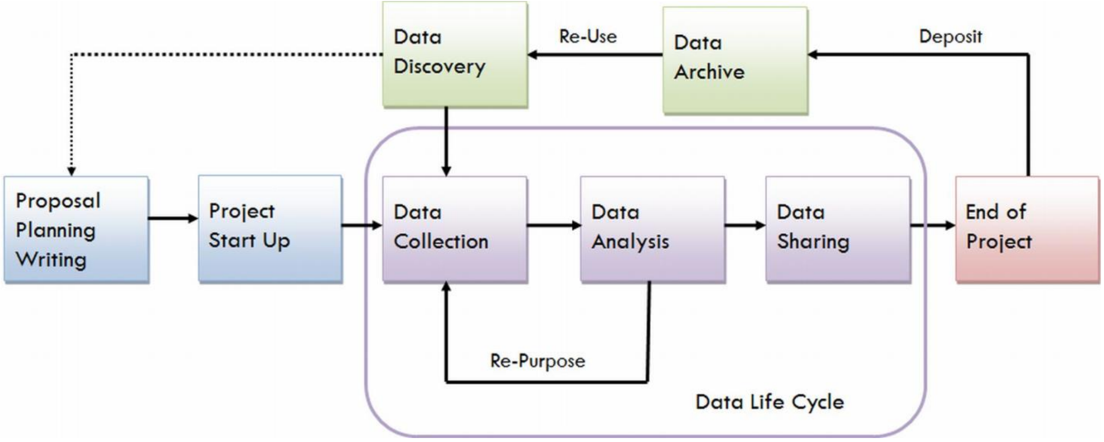
Software tools have increasingly joined the microscope and other equipment as essential instruments for biomedical scientists. However, scientist are often required to write software themselves to accomplish their goals. Hannay, Langtangen, Singer, et al. (2009) found that scientists spend approximately 30% of their time writing software; yet, 90% or more are self-taught (Wilson et al., 2014). There are a growing number of workshops and short courses intended to address these needs directly with biomedical researchers. The Software Carpentry Foundation (SWC) is a non-profit organization whose goal is to make the work of scientist more productive and reliable through software skills training (“Software Carpentry: Frequently Asked Questions,” n.d.). Software Carpentry uses volunteer scientists to teach other scientists how to use programming in analyzing and visualizing their data. Crouch, Hong, and Hettrick (2013) found the Software Carpentry model to be effective in providing scientists basic skills in computation and enabling scientists to adopt techniques that are considered standard in the software industry. The limitation of the SWC model is the reliance on the altruistic nature of scientists to educate their peers. Often SWC instructors need to travel from other institutions where there is no local expertise, taking precious time away from their own research.

Software Carpentry skills have the added benefit of increasing the likelihood that science is reproducible and replicable. For over ten years this group has run in-person “boot camps” as well as hosting online material and videos teaching scientists about the UNIX command line, programming with R or Python, version control, and more. R and Python are open source programming languages that have been adopted by the biomedical research community. Scientists write code packages specific to their domain needs and then make them available to fellow scientists, incentivizing the use and further development of these techniques by their peers. Other courses focus on teaching biologists the skills necessary for data analysis in specific domains. Health sciences libraries are accustomed to teaching a variety of educational workshops and have access to classroom and consultation spaces. Libraries might benefit from providing data science training by positioning themselves to demonstrate value to researchers who need assistance in preparing, organizing, and analyzing data. Support of data science workflows would require that they develop services beyond the services described by Vaughan et al. (2013). Some academic and health sciences libraries have begun to provide consultations and support to the development of hacking skills and statistical knowledge.

Data Science Support by Libraries

University of Virginia, University Libraries. The University of Virginia (UVA) University Libraries, Research Data Services + Sciences department offers workshops, short courses, and consultations in R, Python, and statistical analysis. The University Libraries - Research Data Services + Sciences offers a different conceptual model for the research lifecycle. The framework describes the research data life cycle as a subset of the research lifecycle and contains similar themes expressed by Vaughan et al. (see Figure 3).

Figure 3. The research lifecycle



Reprinted with permission. The research lifecycle is available at http://data.library.virginia.edu/data-management/lifecycle/.

The model presented by the UVA University Libraries does not attempt to outline the library's support role at each stage of the research lifecycle. However, the description of library services mirrors what was observed by Vaughan et al. Similarly, services include data management services for locating, collecting, documenting, organizing, storing, and preserving data.   One key difference between the two models is the connection between library services and data workflows as part of the research lifecycle. The model provided by Vaughan et al. describes new roles for the library supporting data through locating data sources, preparation of data management plans, and describing data. The model presented by UVA intentionally denotes the data lifecycle as its own set of processes. Their model breaks down the use of data in research into two major areas. The first area called the data life cycle has three processes; Data Collection, Data Analysis, and Data Sharing. The second area occurs after the end of the project and entails Data Discovery and Data Archive. Investigation of research workflows coupled with defining the future needs of biomedical research will assist health sciences libraries in discovering the connection between traditional support roles and the development of new roles within the research lifecycle.

Conceptual Framework

Open systems. Open systems theory provides a useful model for examining how health sciences libraries function as an organizational unit within the ecosystem of academic health sciences centers. According to Lawrence and Lorsch (1967), when applying an open systems view organizations are not isolated, but related to their specific environments. Salancik and Pfeffer (1974) state that organizations operate within social systems where they compete for vertical or horizontal power with other subunits. In a vertical power structure, one unit directly manages subunits. Health sciences libraries may operate in a vertically aligned power structure through reporting relationships to health systems, schools of medicine, or university libraries, but do not often oversee non-library organizations. According to Salancik and Pfeffer, horizontal power is the “use of influence among coacting peers to obtain benefits for themselves" and is important for the allocation of resources. Academic health sciences centers are complex organizations where revenue is generated by clinical services, tuition, and research grants. Health sciences libraries, as a service unit, generate no meaningful revenue and act as cost center within the organizational ecosystem. As such, it is important for libraries to develop horizontal power by adapting to contextual demands from the stakeholder subunits in order to ensure institutional legitimacy.

As previously noted, health sciences libraries are subject to internal and external forces that challenge their traditional role as the gatekeepers of knowledge. Hickson, Hinings, Lee, Schneck, and Pennings (1971) propose that the success, or power, of an organization is connected to its ability to cope with uncertainty. Libraries are a common good to the organization and derive power from the ability to align within the organizational workflows, especially in times of great change. Power for a library is not its ability to influence or direct another organizational subunit, but its ability to provision and secure resources valuable to the organization. Salancik and Pfeffer (1974), argue that a persistent problem for universities, and by extension academic health sciences centers, is the uncertainty for the provision of resources required to sustain operations. They contend that organizations, as open systems, depend on continuous resource acquisition for their survival. Libraries as a subunit can garner more power to the extent they provide resources for the organization that are critical, important, and valued.

Natural selection and resource dependency. Aldrich and Pfeffer (1976) examine the relationship between organizations and their environments. They propose a natural selection model that asserts that environmental factors select the organizational characteristics that that best fit the environment.   This view aligns with resource dependency models that maintains that organizations seek to strategically adapt to their environments. Aldrich and Pfeffer assert that organizations are not able to generate all of the resources needed to maintain themselves. Subunits within the organization are dependent on resources and survival is dependent on a subunits ability to fit within the context of organizational needs. As stated previously, libraries as a nonrevenue generating organization cannot compete with revenue generating units, which could jeopardize their legitimacy within the organization. In contrast, the advantage for libraries is that they are not dependent on revenue streams that can alter by external market forces.

Evolution and adaptation. Aldrich and Pfeffer explain that it is important to modify natural selection models when applied to organizational analysis. They maintain that unlike ecological evolution, where a species may be eliminated through selection, within an organization it may be found that most survive after making significant transformations. The responsibilities of transformation of health sciences libraries resides with library directors and other leaders. Aldrich and Pfeffer explain that structural contingency theory relies on, “managerial adaptation to find the fit, rather than on change being accomplished through differential survival.” They go on to add that the deliberate introduction of variation can alter customary modes of behavior for innovative organizations. This view suggests that adaptation provides a better model for understanding libraries and organizational change.   According to Hannan and Freeman (1977) using an adaptation perspective, “usually managers or dominant coalitions, scan the relevant environment for opportunities and threats, formulate strategic responses, and adjust organizational structure appropriately.” Understanding the emerging environment for academic health centers and developing iterative strategies to meet these needs can position libraries to demonstrate value to the organization and ensure their legitimacy within their organizations.

CHAPTER 3

METHODOLOGY

This study used a mixed methods approach, a methodology involving the collection of qualitative and quantitative data and integrating both forms of data (Creswell, 2014, p.4). According to Creswell, the core assumption for this method of inquiry is that the combination of qualitative and quantitative approaches provides a more comprehensive understanding of the research problem. The study utilized a convergent parallel design (see Figure 4), which collects qualitative and quantitative data independently and then compare or relate the results (Creswell & Plano Clark, 2011, p. 69). In a convergent parallel design model, data is prioritized equally and the results are combined in the overall interpretation (Creswell & Plano Clark, 2011, p. 70). Greene, Kreider, and Mayer (2005, 276) refer to this as a component design, where, “data retain their original form and character throughout, and conclusions and inferences seek harmony and connection rather than full blending of integration.

Qualitative data were collected from one population, biomedical researchers, in tandem with quantitative data collected from a second population, health sciences libraries leaders. Triangulating data assisted in identifying any connections between health sciences library services and the needs of the research lifecycle. According to Marshall and Rossman (1989, p.252-253), this method of analysis can enhance a studies generalizability by using multiple data sources to address the research questions. Triangulation can be used to corroborate, elaborate, or illuminate information about the research questions (Marshall & Rossman, 1989, p.252-253). This study used an elaboration design, intended to, “illuminate different facets of the phenomenon of interest” (Rossman & Wilson, 1994) and helps to extend our understanding of the information needs of the research lifecycle and the connection to library support services (Rossman & Wilson, 1994). Utilizing this multiphase process allowed for a more comprehensive understanding of the research questions, including the ability to compare differences in perception between the two populations (Creswell & Plano Clark, 2011, p. 73).

Figure 4. Convergent parallel design

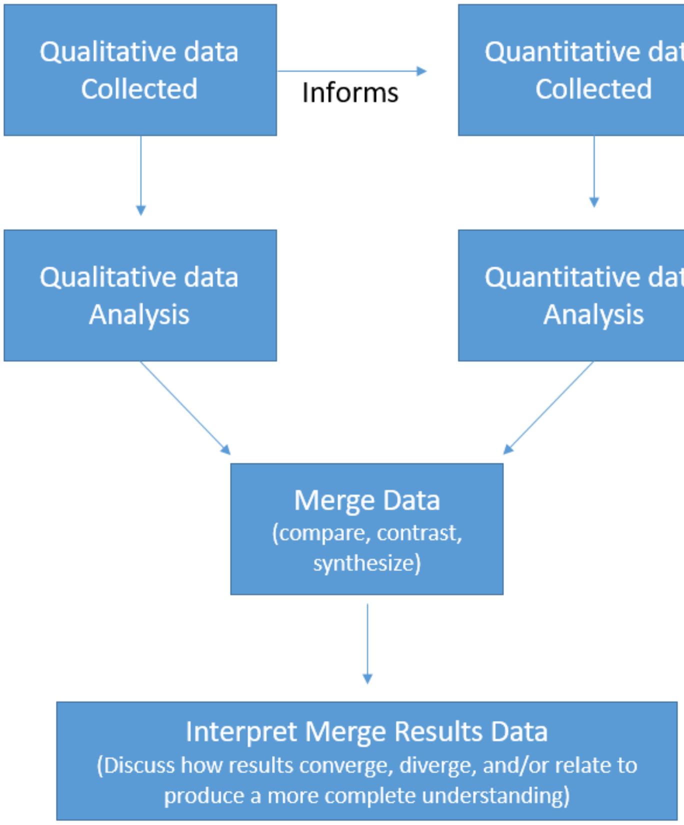


Figure 4. Convergent Parallel Design. Adapted from *Designing and conducting Mixed Methods Research*, by J. W. Creswell and V.L. Clark 2011, p. 79. Reprinted with permission.

Study Design Overview

Document analysis. A document analysis of research lifecycle models discovered in scholarly literature or used in institutional research support websites was conducted in the first phase of the study. Analysis of the various models provided the context for creating a consolidated list of relevant themes and activities within the research lifecycle and later explored in the qualitative and quantitative data collection phases. Documents analyzed included research lifecycle models collected from academic library websites and scholarly literature created by universities in the United States. Exclusion criteria included research lifecycle models created outside the United States, data lifecycle models, and sponsored program lifecycle models since the scope and culture of these models may focus on elements outside the scope of this study. A list of themes was created from the activities listed in the research lifecycle models reviewed. Themes within the research lifecycle were identified and used as a basis for initial codes for the qualitative interviews in phase two. Codes derived from the document analysis were triangulated with data from phases two and three in order to identify alignment and gaps in library service models along the research lifecycle.

Qualitative interviews. In the second phase of the study, qualitative data were collected from in-person interviews designed to capture individual perceptions of the research lifecycle and emerging practices in biomedical research. The goal of the qualitative phase was to collect data about scientific workflows from problem identification to dissemination of knowledge. Data collected by the biomedical researcher interview protocol was informed by the document analysis of research lifecycle models. Themes extracted from the document analysis focused on the individual researcher's workflow, emerging scientific practices, library support services, gaps in current library support models, and potential library support that might demonstrate value to the researcher. A dialogue between the researcher and the participants was constructed to increase responses that are consistent with the research questions. Initial codes were first developed based on themes discovered through the phase one document analysis and expected themes derived from the research questions. The interviews were then coded using the initial codes and used open coding techniques to ensure that emergent themes are allowed to surface. Participants were allowed to deviate from the semistructured questions, so that unanticipated themes could be explored. After each interview the data were analyzed to capture emergent themes for use in subsequent interviews. Data were collected until saturation of themes from study participants was achieved. Qualitative data provided insight into how current library models of the research lifecycle align with research workflows of biomedical researchers. This includes how external forces or barriers to conducting science might translate to potential roles for health sciences libraries. The qualitative data were examined for themes and informed the design of the survey instrument that was be administered in phase three.

Quantitative survey. In the third phase, a survey was administered to the leaders of academic health sciences libraries. The Association of Academic Health Sciences Libraries (AAHSL), an organization that supports academic health sciences libraries in advancing the missions of their academic health sciences centers, maintains database of member libraries that was used as part of the selection criteria. Library leaders were asked to complete a survey on existing research support services, plans for expansion of current research support, and the what new skills should health sciences libraries cultivate in order to more fully support biomedical research. The health sciences library leadership survey protocol was initially informed by the document analysis of research lifecycle models and later by emergent themes identified during the biomedical researcher interviews to assist in the development of an effective survey tool. Data were collected independently and combined during the analysis phase for interpretation.

Sites for Data Collection

A purposive approach was used to identify biomedical researchers and health sciences library leaders to ensure a diverse pool of data. The Blue Ridge Institute for Medical Research (BRIMR), a non-profit organization with its own research mission, produces an annual ranking of NIH funded medical schools in the United States ("Blue Ridge Institute for Medical Research," n.d.).   Information in the BRIMR rankings is obtained from the NIH Research Portfolio Online Reporting Tools (RePORT) and modified to account for discrepancies in the NIH raw data. Use of the BRIMR rankings for this study assisted in categorizing potential site locations by using NIH funding as a metric to gauge research activity level. The BRIMR data were cross-referenced with AAHSL member libraries to create a unified list of potential site participants for phases two and three of the study. The potential sites were divided into five equal groups, or quintiles, so that a diverse sample of participants could be assembled based on level of NIH funding. This this study applied was approved as IRB-SBS Protocol Number 2017-0156.

Participants

Biomedical researchers. Awardees of NIH Research Grants (R series) awards, Career Development Awards (K series), and Research Training and Fellowships (T and F series) awards from the unified list of potential site participants were used as criteria for participation in the study. Other similar awards, such as U series or BD2K awards, from NIH were also considered as long as they were significantly funded and met the general criteria of principal investigator, career development awardee, and research training awardee. NIH RePORTER was mined for potential participants and email listservs researchers were likely used were used to solicit participation. Personal and professional contacts were also used to identify potential participants and solicit participation from biomedical researchers.

Participant selection provided a contrast of experience within the research lifecycle and ensure a broad range of data from trainee to established researcher. However, it is worth noting that the selection criteria limited participation from researchers from institutions with schools of medicine and that tend to be prominent, prestigious, and well-resourced and institutions. This study did not collect data from other types of institutions that conduct biomedical research or from institutions outside of the United States. It should be noted that biomedical research occurs at institutions who do not have schools of medicine or have high level of funding from sources other than NIH. This also does not account for biomedical research that occurs in private industries, such as pharmacological companies. Therefore, it is unknown to what extent the data collected about the research lifecycle is applicable to these other types of institutions.

Health sciences library leadership. In phase three, participation from health sciences library leaders was solicited from AAHSL members identified on the unified list of potential site participants. Membership of the AAHSL is comprised of 155 health sciences libraries in the United States and Canada affiliated with accredited medical schools belonging to the Association of American Medical Colleges ("About AAHSL," n.d.). According to the AAHSL website, the organization was founded in 1977 and supports its members through programming and services for, "benchmarking; advocacy; partnerships with like-minded organizations; and the development of a community of colleagues". Library directors who met the selection criteria were emailed directly to solicit their participation. Participation in the study allowed for directors, deputy directors, associate directors, or other leaders likely to be involved with current or planned services supporting biomedical research to complete the survey. Potential participants who were solicited and did not initially complete the survey received two reminders asking for their participation.

Research Paradigm

The research paradigm chosen for this study was pragmatism. Pragmatism provides an umbrella paradigm to the research study and allows for the use of diverse approaches that valuing objective and subjective knowledge (Creswell & Plano Clark, 2011, p. 43-44, p. 78). This approach is useful when merging multiple perspectives to gain a greater understanding of the research question. The nature of the research itself was interpretive and an attempt to understand how the future needs of biomedical research may impact the role of health sciences libraries is supporting the research lifecycle. This study helped to address the lack of prior research investigating the role of the library in supporting the research lifecycle and attempt to fill the void by exploring the vision that guides academic medical centers as they strategically realign their research computing infrastructure to accommodate data science workflows. This study examined how health sciences libraries might transform their services to support the research lifecycle through the perspectives of the participants in the study.

Transferability

Given the diversity and background of participants, unexpected themes were likely to emerge and led to additional insight (DiCicco-Bloom & Crabtree, 2006). Data from the survey were analyzed to determine the state of the profession and provide recommendations for the purpose of strategic planning. Broader implications were evaluated, providing insight to health sciences libraries for the design and delivery of services. It is hoped that insights gained from the study would provide evidence for health sciences libraries seeking direction for organizational and professional evolution. Narrowly focused, findings could have implications for the education needs of librarians who support biomedical research communities.

CHAPTER 4

FINDINGS

This study explored the association between the biomedical research lifecycle and health sciences library support. Many research lifecycle models are posted on websites created by libraries and other academic departments. Among the models analyzed, no standard model was uncoverd to draw from as a framework for this study. This study sought to understand the biomedical researcher workflow and the perspective of health sciences library leaders on how they support or planned to support biomedical researchers. Biomedical researchers were invited to participate in a qualitative interview designed to explore the major components of conducting biomedical research and explored themes that emerged from the process. The study conducted a document analysis of research lifecycle models created by libraries and other academic sites to aggregate a list of activities associated with the research lifecycle. Activities within the research lifecycle were consolidated into a unified list to be used as probes during the qualitative interviews of biomedical researchers. The survey instrument administered to health sciences library leaders was developed to include the research lifecycle activities and was modified to include themes that emerged during the analysis of the qualitative interviews. The four research questions that guided this study are:

1. What are the key activities in the research lifecycle for biomedical researchers?
2. What aspects of the research lifecycle are evolving, if any, due to emerging practices in biomedical research?

3. In what ways do health sciences libraries support the research lifecycle and emerging practices of biomedical research?
4. What skills and practices might health sciences libraries cultivate in prompting new roles in the research lifecycle?

Description of Respondents

Biomedical researchers. This study solicited participation of biomedical researchers working at institutions of higher education. Selection criteria included biomedical researchers working for institutions ranked on the Blue Ridge Institute for Medical Research's list of medical schools in the United States receiving NIH funding, and whose libraries are members of the Association of Academic Health Sciences Libraries. The BRIMR complies a list of 139 NIH funded United States medical schools. AAHSL contains 142 member libraries and is affiliated with the American Association of Academic Medicine. The combined list yielded a total of 112 institutions as potential participants. Among the 112 institutions that met the selection criteria, 52 individual researchers received personal emails to participate in the study. Researchers were asked to participate in a study of library support of biomedical research workflows that was investigating how health sciences libraries can expand their role in supporting knowledge creation within the research lifecycle. Contact information of the researchers was assembled by searching the funding database NIH RePORTER, through professional contacts from peer health sciences libraries, professional contacts at the University of Virginia, and the Software Carpentry email distribution list. Eighteen biomedical researchers agreed to participate in the one-hour interview. One interview did not occur due to scheduling difficulties, leaving 17 total interviews completed. To gather diverse perspectives, this study solicited researchers based on their stage of career, type of biomedical science, and institutional level of NIH funding.

The inclusion of researchers who conducted wet lab and dry lab research produced some notable differences in the research processes, but did not create discrepancies in how they navigated the research lifecycle[3]. For example, wet lab researchers whose process included the collection of biological samples from patients had a clearly defined protocol that they followed that considered not only patient privacy, but informed consent. These processes were defined as part of the IRB application and often administered by research trainees. The added complexity of working directly with patients required that the researchers follow a more prescriptive process. However, outside of the inclusion of informed consent, navigation of the research lifecycle by biomedical researchers was generally uniform. Most of the researchers interviewed for this study discussed some aspect of privacy when it came to working with their data and when appropriate included processes that protected the privacy of patient data.

This study sought to achieve parity among the categories of researchers selected for the study; however, low participation rates within the categories prevented an even distribution of participants. After the 17th interview, data saturation was determined and no further interviews were administered. Table 1 summarizes the distribution of biomedical researchers interviewed in the study.

---

[3] Wet labs contain the "appropriate plumbing, ventilation, and equipment to allow for hands-on scientific research and experimentation." Dry labs make use of "computer simulations or for data analysis." ("Dry Lab Medical Definition | Merriam-Webster Medical Dictionary," n.d., "Wet Lab Medical Definition | Merriam-Webster Medical Dictionary," n.d.)

Table 1[4]

Distribution of biomedical researchers interviewed

Participants

		N=17	%
Gender			
	Female	7	41
	Male	10	59
NIH funding by quintile			
	1st	4	24
	2nd	10	59
	3rd	1	6
	4th	0	0
	5th	2	12
Career Stage			
	Trainee	7	41
	Early-career	2	12
	Established	8	47
Science			
	Wet lab	5	29
	Dry lab	12	71

Due to the small sample size of participant data statistical comparisons among types of researchers is not feasible. As a result, biomedical responses are presented in descriptive terms. The presentation of the findings are made with a description within the context of research questions 1 and 2.

Health sciences library leaders. This study solicited leadership at health sciences libraries to participate in a survey designed to investigate support provided by the libraries in the research lifecycle. Selection criteria of institutions was the same as for biomedical researchers including libraries supporting institutions on the BRIMR NIH Funding US Medical Schools ranking list and who are members of the Association of Academic Health Sciences Libraries. Library leaders were asked through email to

---

[4] Throughout this paper percentages are rounded to the nearest integer.

participate in a study of library support of biomedical research workflows that was investigating how health sciences libraries can increase their role supporting the creation of knowledge within the research lifecycle. The online survey requested by email one response per health sciences library from a professional at the Director, Deputy Director, Associate Director, or comparable leadership level. Of the 112 libraries solicited, 51 completed the study. Table 2 summarizes the distribution of health sciences library leaders who completed the survey instrument.

Table 2
Distribution of health sciences library leadership participants

		N=51
NIH funding by 5th quintile		
	1st	12
	2nd	12
	3rd	9
	4th	9
	5th	9
Position		
	Executive Director, Director, or similar level of responsibility	44
	Deputy Director, Associate Director	3
	Assistant Director, Department Head, or similar level of responsibility	4

Analysis of Results

Dedoose, a web-based qualitative and mixed methods tool for analyzing data assisted in the analysis of transcribed audio interview data collected from biomedical researchers. Dedoose allowed the researcher to code the raw data and to expose patterns uncovered in aggregate data. Due to time constraints or the nature of the dialog with the interview participants, it was not possible to explore all probes in every interview with biomedical researchers. Interview probes helped to explore aspects of the research lifecycle, but biomedical researchers were encouraged to describe their workflow from their perspective. Frequency data reported in the tables below is not intended to be a representation of all aspects of the research lifecycle, rather a description of key activities from the perspectives of the study participants. Code application and code co-occurrence data helped to identify meaningful patterns that exist in the research lifecycle among the researchers. Code application represents the frequency of code use and code co-occurrence represents when a two or more codes were applied to the same excerpt.

Due to the lack of consensus for a standard definition of the research lifecycle, the first research question asked biomedical researchers to define the major activities of their research process. Participants were encouraged to describe the major activities involved in each subcycle; planning for research, conducting experiments, disseminating knowledge, and assessing impact. Researchers had latitude to describe their workflow from their perspective and the protocol probes assisted in guiding the researchers to fully define their key activities. In most cases, the activities described matched the probes within each subcycle, but it was also apparent that independent research activities can and do occur throughout the entire research lifecycle. As a result, this study did not find that it could develop a unified research lifecycle model that accurately captured the workflow for all of the researchers interviewed. However, key activities did emerge from the data and are identified in the analysis. A description of the cycles and code definitions is provided in Appendix A. Deductive techniques were used to first read and then code the data for themes matching the initial research lifecycle probes. Subsequent reading utilized inductive techniques by combining codes for redundant themes and establishing new codes that emerged during the analysis.

Qualtrics, a web-based survey tool, administered the instrument to health sciences library leaders. Because of cultural differences between terminology used by biomedical researchers, the design of the survey instrument did not allow for the quantitative data collected from library leaders to be mapped to the research lifecycle probes used in the qualitative interviews. For example, the concept of data management carries different meaning for researchers than it does for libraries. Three open-ended questions on the survey were coded using inductive methods to identify themes. Data from the qualitative interviews and the survey were analyzed using SPSS version 24.0, Dedoose, and Microsoft Excel 2016.

Research question 1: What are the key activities in the research lifecycle for biomedical researchers? Analysis of the interview data yielded 1196 coded themes. Multiple codes were applied to excerpts when individual themes were described in connection to other research activities. Codes were organized into five subcycles: general, research planning, experimental, dissemination, and impact. Table 3 illustrates the prevalence of codes from researchers within each subcycle.

Table 3, (N=17)
Distribution of coded themes that occurred

Subcycle	Cycle Activity	Total Codes	# of References	# of Participants Referencing
General Themes	Mentor/Mentee	280	73	17
	Non-library support		50	16
	Library support		39	17
	Reproducibility and replicability		14	10
	Competition		8	6
	Licensing and Venture		5	5
Research Planning Subcycle				

	Grant funding	262	66	16
	Literature searching		40	17
	Methodology		35	15
	Identify collaborators		30	16
	Grey literature		17	9
	Citation management		17	14
	Data literacy		14	12
	IRB/IACUC		13	9
	Biosketch		12	10
	Systematic Review		10	9
	Bioinformatics tools		8	7
Experimental Subcycle				
	Data analysis	388	59	17
	Collaborating		57	16
	Data collection		47	16
	Data management		41	15
	Open source software		35	14
	Organizing and storing information		34	14
	Proprietary software		26	14
	Data privacy and security		18	8
	Project management		18	10
	Lab notebook		17	12
	Statistical Methods		17	12
	Pilot experiment		13	11
	Metadata		5	3
	Ethics		1	1
Dissemination Subcycle				
	Open access	231	53	16
	Writing		33	13
	Social media		32	15
	Journal selection		29	15
	Attending conferences		21	13
	Data preservation		17	12
	Presentation		15	10
	Compliance		12	6
	Conference selection		11	11
	Author rights and copyright		4	4
	Preprint		3	2

Impact Subcycle	Citation styles	58	1	1
	Citation metrics		30	16
	Altmetrics		5	4

The frequency of code occurrence by subcycle helps to illustrate the adoption of behaviors within each subcycle, but does not indicate why one researcher's activities differs from another. The frequency of code application by activity is helpful in understanding the prevalence of that activity among the researchers interviewed. The number of individual researchers referencing an activity helps to illustrate to what extent an activity was referenced by the study sample. In some cases, it was found that an activity could have a high frequency of references, but that the frequency of references occurred from a low number of participants. Indicating that key activities in the research can vary among biomedical researchers. Deeper analysis of the data within each subcycle is explored in the sections below.

General themes. Table 4 summarizes the frequency of activities referenced by biomedical researches for general theme. General theme codes were used for activities that did not fall within the normal context identified by the document analysis. The code *mentor/mentee* emerged as a significant concept within the research lifecycle during the data analysis process and was not among the original predefined probes. Similarly, *non-library support* was not identified as a predefined code, but quickly emerged as an important component within the research workflow. *Library support* was included as a predefined code to help explore the relationship between health sciences libraries and the research lifecycle. *Reproducibility and replicability* was coded as part of general themes, because it is a collection of other activities that occurs at different stages in the research lifecycle. *Competition* emerged conceptually as part of the biomedical research culture, but not as a specific activity. *Licensing and venture* was not heavily referenced in the document analysis and had no natural fit within the subcycles defined by through the document analysis. *Emerging practice* and *influencing science* were coded as part of the general themes, and are analyzed during research question 2.

Table 4, (N=17)
Frequency of coded themes that occurred within general themes

Cycle Activity	Total Codes	# of References	# of Participants Referencing
Mentor/Mentee	280	73	17
Non-library support		50	16
Library support		39	17
Reproducibility and replicability		14	10
Competition		8	6
Licensing and Venture		5	5

Mentor/mentee. All 17 participants identified the relationship between the mentor and mentee as an important concept in the research lifecycle. The co-occurrence of the *mentor/mentee* with other activities helps define the importance of the mentor and mentee relationship in connection with other activities. In most cases, the mentoring as an activity was discussed within the context of other research lifecycle activities. The most frequent activities mentioned in relation to *mentor/mentee* were *collaborating, data analysis, grant funding, literature searching,* and *non-library support*. Even though the *mentor/mentee* emerged as an important concept in the research lifecycle, only one established researcher spoke of the process of mentoring itself and the value to his lab. He shared his perspective on the importance of training students to become good scientists:

> So I really feel like I'm training them to become independent scientists. I think some labs manage their graduate students to be more like really, really good technicians. It's a different philosophy about how you would manage the group.

When asked about the value student involvement brought to his lab he stated:

> They'll remember details better than I will, they'll have a much better sense of exactly what we've done or can do and the practicality of things. They're just smarter than I am. And so I bring them along to keep me in check, to be honest.

For this researcher including his trainees within the process of decision making not only made them better researchers but strengthened his research.

In other parts of the interview, this researcher described how he used the relationship between the mentor and mentee to create not only the next generation of scientist, but also for transitioning his mentees into mentors themselves. His process was to mentor his students at different stages of their development. According to him, "I will often start with a student who is already working in a related area so I know that they're gonna have some familiarity with some of the problems and concepts around this new idea. In my lab I will often have a student, a more junior student, working on a project that a more senior student's a part of." Later he sends an email to the students and then they, "get together and kind of talk about well, what has been done in this area, what are some questions that we've talked about that are maybe kind of connected. And then I think the next step is probably trying to distill down to what is that simplest key experiment or analysis that would get at the heart of that question that we can pilot something with." In his case, he is using mentorship to not only train his students, but to as a way to helped to sustain the viability of his lab's research.

Early-career researchers described the process of mentoring in terms that help them meet their career objectives. The career path for biomedical researchers is reinforced by funding mechanisms at NIH, which incentivize mentoring through career development grants (K awards). An early-career researcher expressed the importance of selecting the right mentor, since it impacted her ability to establish her professional reputation:

> … you have to tell the story about what you wanna become an expert in and what this mentored grant is gonna do to prepare you to become the expert. And you need to convince NIH that not only do you personally have the capacity to do it, but a larger part of the story is what training you need and how are your mentors. Are they well-funded? Do they actually have the time? So, picking somebody with four R01s might not be the best mentor. If it's somebody with two that had more time – finding mentors who have a K24, which is specifically a mentoring mechanism. And so, that's what I've done now, I've kinda put together a mentoring team, and I'm still looking for that critical piece of who's gonna actually pay for the science that I wanna do.

For early career researchers, mentor selection is an important aspect of their professional development. Mentoring is established early in the development of biomedical researchers and multiple mentors may be involved. One biomedical research trainee referred to her whole dissertation committee as her mentors. Another trainee spoke of the research team she worked with as mentors. From the interview data, it was clear that the mentoring was a critical part of training, but it was also a critical component in how the labs sustained their operations.

Library support and non-library support. The concept of support is prevalent throughout the interview data. Aside from support created from the relationship between mentors and mentees, researchers receive support from a variety of places. Administrative support staff were identified as a crucial component of lab operations who helped PI's and their trainees navigate the process of writing and administering grants. Administrative support staff were frequently described as a lab staff member or as reporting to the lab's department or school. Support was critical to the biomedical researcher's ability to navigate the research lifecycle. Evidence of *library support* for biomedical research was found throughout the research lifecycle. The most frequent co-occurrence of *library* and *non-library support* included *literature searching*, *systematic reviews*, *biosketch creation*, *citation metrics*, *data analysis*, and *collaboration*. *Library* and *non-library support* are described in more detail during the analysis of individual components of the research lifecycle.

Reproducibility and replicability, competition, and licensing and venture. Code application for *reproducibility and replicability*, *competition*, and *licensing and venture* occurred at low frequencies among the study's participants. The concept of *reproducibility and replicability* was most frequently connected with *open access*, *open source software*, and *data analysis*. Among the researchers who discussed *reproducibility and replicability*, most believed that these practices were becoming an important part of biomedical research and one established researcher described the need for "reproducible research workflows". In some cases, teaching reproducibility was integrated with the mentoring process. One established researcher directed his students to find a scientific paper and a model to see if they can reproduce the same figure and data. Most of the participants identified *reproducibility and replicability* as an emerging trend in biomedical research, which will be discussed in more detail in the finding for research question 2.

When the concept of *reproducibility and replicability* was mentioned they were often connected to the concepts of *open access* and *open source software*. Researchers described sharing not only data, but also other aspects of the workflow including *methodology* and software code. An early-career researcher said, "You can't cook the data quite as easily if everybody has access to it, and you're gonna have to be very clear in your methods of what you actually did." However, some researchers felt that open data created a risk. At least one research trainee expressed concern over being too open with data, noting privacy concerns when using protected data.

> The people calling for open data are the ones – tend to be the ones who benefit the most from it so, the ones who have heavy infrastructures in place, so they can quickly analyze it and publish on it, so I think that's one thing, but I think the other thing is that there's a lot of risks, especially with patient level data, there's a lot of risks making it open and I don't – if I wanted to say "Okay, I have this – my electronic health based cohort, which is all I [*sic*] already approved and everything but it's a 15,000-person cohort, patient level data. So, making it open that only puts – it only puts me at risk, and I don't know who and how – who will be able to appropriately deidentify it, how they'd be able to appropriately deidentify it and the process of deidentifying I could probably lose a lot of valuable patient-level information that's important to understand in disease genotypes for them. So, - it's a challenge, I think it's important but I think there's no absent [*sic*] universal resources and with patient data it's really hard.

Thus, concepts of *reproducibility and replicability* and openness were generally described as positive trends in biomedical research, but not without caution. Researchers were aware that making their data open could jeopardize the privacy they were required to protect.

Competition emerged as a significant concept, but was not included as a probe during the interviews. As a result, frequency data on competition is likely lower than concepts specifically asked about during the interview. Despite low frequency data, *competition* was a recurring theme discussed by researchers. Concerns included data theft, other 'scooping' their science, and competitiveness for grant funding. *Licensing and venture* was not a strong concept that resonated with the researchers interviewed. It is possible that the sample population of researchers were unlikely to produce research products viable for commercial development or that the researchers themselves were not interested in commercializing their research. Two researchers mentioned commercial aspects of their science. In one case, the researcher had previously started two companies, while the other managed a unit that produced potential commercial products not connected to his specific research.

Summary. In general themes, concepts connected to support were established as the most prominent themes. Support for researchers come from labs, administrative personnel, and the library. The relationship between mentors and their trainees was identified as a critical component necessary to sustain operations of research labs. Both mentors and mentees relied on this relationship to accomplish their individual goals. Administrative personnel helped researchers sustain lab operations and *library support* most frequently supported aspects of the research lifecycle including *literature searching*, *systematic reviews*, *biosketch creation*, *citation metrics*, *data analysis*, and *collaboration*. *Reproducibility and replicability*, *competition*, and *licensing and venture* did not emerge as dominant themes, although r*eproducibility and replicability* was connected to emerging practices in research that may have a deeper impact on researchers in the future.

Research planning subcycle. Table 5 summarizes the frequency of activities referenced by biomedical researchers in the research planning cycle. *Grant funding*, *literature searching*, *methodology*, and *identify collaborators* were the most frequently applied codes.

Table 5, (N=17)

Frequency of coded themes that occurred within research planning subcycle

Cycle Activity	Total Codes	# of References	# of Participants Referencing
Grant funding	262	66	16

Literature searching	40	17
Methodology	35	15
Identify collaborators	30	16
Grey literature	17	9
Citation management	17	14
Data literacy	14	12
IRB/IACUC	13	9
Biosketch	12	10
Systematic Review	10	9
Bioinformatics tools	8	7

Grant funding was the second most frequently coded category within the entire research lifecycle. Established researchers expressed an overall ease with finding funding for their science, as well as familiarity with current funding trends. They acknowledge receiving funding opportunities though listservs, departmental emails, and search alerts, but most stated that confidence in locating opportunities because of their years of experience, professional networks, and domain knowledge. On the importance of networking and domain knowledge, one researcher emphasized:

> When scientists hear the word networking sometimes they feel a little icky, “Ah, I gotta go talk to somebody.” And many scientists are introverts. Some of them are extroverts, but the reality is if you don’t know what’s going on in your field and you don’t read the literature or you don’t get updates on a fairly frequent basis, it’s hard to stay abreast of what the funding is, right?

Researchers may have entered the field because of their passion for research, but as entrepreneurs are also required to build professional networks that allow them to establish successful collaborations. Several established researchers mentioned information overload as a distraction when it came to funding opportunities. An established researcher insisted that it is, “hard to describe how much we get inundated with grant information”, and “I think, almost a better question than how do I find grants, is how do I filter? To figure out which ones work for me. And that's a question that applies across the research lifecycle."

Publishing was viewed by some established researchers as a way to stay competitive for funding opportunities. When one established researcher was asked how he kept aware of funding trends he responded, "I do think when you're regularly publishing that you're staying cutting edge and so then the funding kind of goes along with that. It's not been something I've consciously been deliberate about."

Early-career and research trainees often expressed frustration with discovering and applying for grants. Several appeared to be frustrated with funder websites, the numerous amount of opportunities available, and narrowing down opportunities appropriate for their research. As one research trainee phrased it:

> I think the issue is there's too much information and there's not enough information that's focused on me, focused on things that I'm eligible for or interested in. And it's hard to sift through the general information to find funding opportunities that are specific for me.

Another trainee found the NIH website difficult to navigate and relied on networking as the means for identifying appropriate opportunities. This was echoed by an early-career researcher who felt that the best method for discovering opportunities occurred at conferences where, "…if you go to a conference and then you hear people talk, 'I applied for this grant' and then, you ask them, 'What's that?' Part of the frustration in obtaining funding was tied to early-career researchers and trainees needing to establish their professional reputation.

Building a professional reputation was paramount to early-career researchers:

> … something like 30 percent of all the funds go to the top 5 percent of their researchers – don't quote me on those numbers, I think they're off. But I think the idea is that, the very top researchers are getting the lion share of the funds that are

> available, and there's just a small proportion of those people. So, I'll be willing to state that that small fraction that's getting all the money, are not newly graduated PhD students trying to make anything for themselves, so, I would say that the experience of being able to frame your research and submit it to get funded is maybe more difficult in this environment, because you have lots of older researchers than you, a lot more younger researchers going out as well, and still the money, it's still directed to that couple of researchers, so it's really difficult to get younger, unknown names the money.

Early career researchers and trainees were not only competing with other early career researchers and trainees, but with established researchers for the available funding opportunities. Training grants to help early-career researchers build their reputation were described as only partially helpful for career development. An early-career researcher described career development awards as a stepladder that is helping her, but only part of the way. In her words:

> It's because the research budget on a K01 is so small, you need to find a sponsor. You need to find a senior researcher who's actually willing to pay for your research because you're only bringing your salary – about $25,000.00 – to the table. So, this has been about a year-and-a-half-long process for me. As soon as I got my K12, I started working on what my external pay would be. And figuring out what I wanted to – because the biggest part of a K01 is telling a story about what you wanna become. But you have to tell the story about what you wanna become an expert in and what this mentored grant is gonna do to prepare you to become the expert. So, it's actually a lot of pressure, there's a lot of pressure.

The career awards only paid for part of her funding, including salary, so she needs to connect her research to established researchers with funding while she builds her professional reputation. Her success depended on locating an established researcher willing to fund her and mentor her to the next level.

<u>Literature searching and systematic reviews.</u> Knowledge of the literature was noted as a key function within the research lifecycle by researchers at all stages of their careers. Most of the participants demonstrated a heavy reliance on PubMed, though many referred to Google Scholar as their second most preferred database for discovering

literature. During the interviews, each participant was asked to describe their comfort level in searching for and retrieving literature. All researchers expressed a high degree of comfort conducting searches themselves. Many expressed the importance of library collections and ability to access remotely. Several researchers, while first expressing comfort in conducting their own searches, also spoke of the value of working with librarians on more complex search needs. One early-career researcher stated:

> I wasn't able to create the query myself, so I did consult the librarians on that, and it was infinitely helpful trying to speak it out, how to create the query, just even to be held as it was, it was really helpful. So, you, I think can come to the librarian in that sort of situation. Otherwise, the kind of less formal searching, I felt pretty comfortable with using the filters and to search for things like that.

Working with a librarian on complex searches was valued by researchers, even if they felt comfortable with their own search skills. Another described what she learned from working with a librarian, "I do all the searching myself. I've used the library once for my research, and I didn't know we could do that … they (the library) were great. I learned that I should probably refine the scope of what I'm looking for a little bit more".

Researchers indicated that they used library literature support services for three primary reasons: expertise, grant application support, and training. One researcher stated, "it was easier to go to the library and ask them to find stuff than for me to go in that rabbit hole." The process of searching the literature as an activity supporting grant applications emerged as an important practice to the success of biomedical researchers. Researchers described the activity as doing their "homework" and more specifically, "…you do need to sound like you know what you're talking about. You don't want to sound ignorant."

Two established researchers described using library systematic review services to support a grant application. The value to the researcher was that it helped them to identify them how in a grant proposal how their research added to the body of knowledge and supported their argument to the funder on why their proposal was worthy of consideration. The rationale expressed by one established researcher illustrates the point:

> So we usually try to double dip actually so we'll often design a systematic review that will be on its own, a standalone paper that kind of defines the field that we're interested in at the time that we're writing the grant about. Try to get that out in a little bit of a hurry so that it kind of sets the stage for our reviewers. But then also incorporate that directly into the grant because as a reviewer, I have found I'm much more impressed if someone can show me a formal systematic review than if they tell me they think they know what to worlds literature is on this, just kind of casually. So I think for a variety of reasons I've come to really appreciate the power of a systematic review.

On the importance of library support in this process, another researcher noted:

> So, we do use that library service for systematic reviews, and I've done quite a few…I can't believe I've done my first one without the same kind of structured support because not only do they (the library) help with the search; it's all the logistics, the software, the infrastructure for screening the articles. It's getting the full text. I mean, I can't even think about doing a systematic review nowadays without that kind of help.

The value of library assistance with systematic reviews not only helped researchers create a search strategy, but helped them efficiently collect, review, retrieve, and access knowledge. Despite the two examples listed above, most researchers did not state that they conducted systematic reviews. One researcher was even unclear on the difference between a systematic review and a deep literature search stating, "We call them just the review papers. It can – You reveal what happens before your report and then – like you read such-and-such and they'll tell people what's going to happen."

Methodology. *Methodology* was often expressed by biomedical researchers as a significant component of their workflow. The application of *methodology* code was closely associated with *data collection* and *data analysis*. Methodological techniques described by wet and dry lab researchers varied widely. A common theme expressed by researchers for wet and dry lab researchers was the soundness of power in statistical tests. An established researcher referred to methodology as, “the way you make sausage” and that the methods and experiments he used help him know what kind of “flavor” he would get. Not setting up the experiment properly in advance could have catastrophic consequences. An established researcher stressed, “Have you heard the quote from Fischer, I think, the famous statistician that if someone comes to him with the results of an experiment in which they didn’t have enough sample size, they can do an autopsy and say what the experiment died of. You can’t resuscitate it.”

Accessibility of large data sets and the ability to process them was having an impact on the methods developed by scientist. One established researcher, whose research focused on methodological development, referred to methods themselves as a science. He summarizes the context as:

> I really don’t feel like there was awareness in the field of biology of the problems that were going to be caused by lots of data and very little software to analyze it. I think people said once we get the sequences, we’ll know the answers, and they didn’t realize that there's a lengthy data analysis phase. The more your data grows, the more complex your data grows, the lengthier your data analysis phase becomes. So, I would say that I had to argue quite a bit in a social context within the universities that what I was doing was… what I was thinking about, what I was focused on was valuable and interesting. Now, I think everybody sees the problem, and they need less convincing that the general area is valuable, right? There's a lot more hiring of people with my background and skill sets now than there was when I started.

His approach was to start big using “heavy weight methods”, later going back and trying different methods to see if he could replicate the answer he already knew were right. In contrast to this approach, a wet lab trainee stated that she started with a small and general idea and then moved to more specific questions. These two approaches illustrate the importance of methodology to different domains of biomedical research while highlighting that researcher’s needs often vary.

Identify collaborators**.** Experienced, early-career, and research trainees all expressed the importance of conferences for establishing collaborations. The process for starting or discussing potential collaborations often starts from conference presentations. An established researcher remarked that, “I’ve gained collaborators just from presenting at conferences and having conversations”. Attending professional conferences provides two major functions for the researcher. By attending presentations of other researchers they identify expertise of other researchers as potential collaborators or by hearing ideas that that could strengthen their own research. A research trainee found that, “…attending conferences is probably the other big resource (for finding collaborators). Not only for networking with people who are presenting there, but also from hearing the ideas and the studies that are being presented.” Second, by presenting their research findings other researchers have the opportunity to view their work, and may reach out for a potential collaboration.

Early-career and research trainees also expressed the importance of conferences for identifying collaborators, although their focused was in building their professional network. In one trainee’s view:

> …if I know someone who’s remotely interested, I’ll talk to them and then – and honestly a lot of it was in person, I was – I realized I was very interested in this

> area and there was not a lot of people doing work in it so, starting about four years ago at the beginning of the fellowship I was pretty aggressive about contacting people, whether is through cold emails or through friend of friend or going up to them at meetings

Another trainee’s statement illustrates her awareness that she needs to establish a network to support collaboration, but is not fully aware on how to do so. She states, “I just got back from two conferences where I presented a couple of posters at each one and I met a lot of people through that, and I don't know if that is actually how collaboration in the real world works. No, I honestly don't know.”

Although many institutions maintain research networking platforms containing information about faculty expertise, only a few study participants mentioned having used these systems. One established researcher did note that use of such a system led to successful collaboration with another researcher, but he also stated that more effective technique was to discover collaboration through “word of mouth”. Another established researcher suggested that relationship building the most effective way to build collaborations.

Mentoring was connected to the process of identifying collaborators and was bidirectional in nature. According to an established researcher, “I always bring a student with me to those conversations. So go into that conversation kind of sharing that idea, and then trying to kind of pitch that experiment. Yeah, and then looking at what kind of resources or techniques do we have expertise in or we’d need expertise for to develop.” In this case, having the trainees who run the day-to-day experiments and techniques was helpful to progressing the conversation and solidifying the collaboration. For trainees the mentor-mentee relationship helped to lay the foundation for identifying collaborators in the future. Mentors help trainees to establish partnerships, but they also realize that this is a process that they will need to learn to do for themselves. One trainee states:

> I just graduated from my PhD a year and a half ago. So, up until this point, a lot of the collaborators have been through my mentor; she was a quite well-established researcher in the field, and was quite great about introducing me to people that might be good collaborators. But going forward, that's a good question…and you know, have been trying to reach out to different departments, and divisions, and get a sense of who is doing what, but to be honest, it's difficult to know all the work that people are doing on campus.

An early-career researcher's approach to solving this problem was to locate and establish a mentor to assist her to establish her career, but she recognizes that there is also a benefit to the mentor. She states, "So, yeah, I think senior mentors do look for you to really contribute to their research as well as help set you up but it's not all altruistic."

<u>Grey literature, citation management, IRB/IACUC, bioinformatics tools, and biosketch creation.</u> Additional activities in the research planning cycle include *grey literature*, *citation management*, *IRB/IACUC*, *bioinformatics tools*, and *biosketch creation*. No significant information was uncovered in discussions from NCBI bioinformatics tools, other than some researchers used them as part of their preliminary investigation.   For most researchers, creating a biosketch was described as a routine activity, though a few linked them to *citation metrics* and *altmetrics* discussed further during research questions 2.

Of the 17 participants, nine of the participants noted the role of *grey literature* using Google, YouTube, and Wikipedia. Most of the use cases described involved searching for information that could not be found in the peer-reviewed literature. Early-career researchers and trainees had a higher frequency of referring to these as resources. This is likely because they are more directly involved in using the equipment and software needed to conduct experiments. Specific examples included looking up technical information provided by companies who produce the equipment used, working papers from Medicaid, and specifications sheets on chemicals. An early-career researcher also expressed that Googling was important aspect of her continuing education. Established researchers also saw the value of grey literature as part of the learning process, but connected it with the need for authoritative evidence. One stated, "To make sure it's good, but those are, I think, often great starting points to then go into the literature and see well, what's really been validated." Another described similar value of grey literature for the development of trainees:

> Well, I think that there are several approaches to learning: One, is to read books. One, is to read articles. One, is to use online tutorials for particular preliminary language, for example, or for particular software packages. One, is to use forums, and there are online communities that have millions of people such as Stack Overflow that are available for people who are learning. So, another is to study documentation for a particular software package. So, when you add all these sources together, there are many resources available for people to train.

The internet has provided various ways for researchers to locate information online in support of their research. Social media, forums, video, and manufacture's websites were used as resources supporting the research lifecycle.

All interview subjects expressed comfort with obtaining and managing literature citations. Most stated using Endnote as their primary software used to manage this information, although Zotero, Mendeley, RefWorks, and Papers were used by some researchers. In almost all cases, researchers were satisfied with how they managed their citations. An exception was an established researcher who used the library:

> I usually in the early days of having a new onboarding postdoc or other trainee, I always send them to the library for at least one of the EndNote courses and often both so they get a good deep sense of how to use it, and don't make their early

> mistakes that many of us made in our career when we didn't think those were useful things.

A few researchers stated that they did not use any citation management software, preferring to use PubMed (My NCBI), Google Drive, or their inbox. Two researchers stated that they used a shared library to manage and share citations with their colleagues. Two researchers used Rich Site Summary (RSS) feeds designed to push and store relevant content through a web browser or mobile app. One was an established researcher who used the online service Feedly and a trainee used Read, a service specifically designed for medical and scientific literature.

When asked about navigating IRB and IACUC protocols, researchers spoke of the need for data privacy and security. Most researchers either expressed familiarity with IRB/IACUC protocols or stated that they received non-library support from their department.

***Summary*.** For all participants, locating and obtaining grant funding was a key activity of the research lifecycle. Established researchers stated that they were comfortable with this process. Early-career researchers and trainees were less confident. *Attending conferences* was noted as an activity that helped researchers identify collaborators. Establishing a professional reputation and publication of research findings were mentioned as activities that helped assist researchers establish successful collaborations. As a group, researchers felt comfortable in searching the literature, but some relied on librarians when it came to complex searches and systematic reviews. Participants felt comfortable searching the grey literature and managing citations themselves. *Methodology* was closely associated with *data collection* and *data analysis*.

IRB and IACUC protocols along with *biosketch creation* were viewed as routine activities often required as part of the research lifecycle.

Experimental subcycle. Table 6 summarizes the frequency of activities referenced by biomedical researchers in the experimental cycle. *Data analysis, collaborating, data collection, data management, open source software, proprietary software,* and *organizing and storing information* were the most highly coded activities.

Table 6, (N=17)

Frequency of coded themes that occurred within the experimental subcycle

Cycle Activity	Total Codes	# of References	# of Participants Referencing
Data analysis	388	59	17
Collaborating		57	16
Data collection		47	16
Data management		41	15
Open source software		35	14
Organizing and storing information		34	14
Proprietary software		26	14
Data privacy and security		18	8
Project management		18	10
Lab notebook		17	12
Statistical Methods		17	12
Pilot experiment		13	11
Metadata		5	3
Ethics		1	1

Collaborating. Grant funding is a foundational element to the research process that incentivizes collaboration and ensures that researchers can accomplish the milestones prescribed in their grant applications. Most researchers acknowledged that recent emphasis from federal funders has incentivized external collaboration through their funding practices. According to an established researcher, “I think we’ve seen for several years now that collaboration is what’s getting the funding, I think sort of especially interdisciplinary collaborations. It seems the bigger funding agencies certainly want to give their money to as many people as possible. So I think that's changed a bit, a shift from sort of the single PI research project." At the same time, internal collaboration with biostatisticians and other forms of support, link grant funding closely to the researchers who rely on these services. Although grant funding appeared to be a major driver in incentivizing collaboration, at least one researcher described the process of establishing collaboration as hurried. According to her, "the collaboration actually work is to work out some of the details of data management and sharing ahead of time. That doesn't always happen. Often it's just a quick rush of oh, you can do this one thing that I need for my grant; hurry up and send me a letter and we'll push this through."

Collaboration was commonly linked to other elements of the research lifecycle. These included *data analysis*, *statistical methods*, *data management*, and others. Internal collaboration was often expressed through the mentor-mentee relationship. Established researchers needed the trainees to navigate the research workflow, while trainees needed internal collaborators in the lab to guide them through the process. This symbiotic relationship helped keep the project on track and to coordinate disparate activities that might be occurring from different lab members. Routine lab meetings were most often used to assist in the management of this process. An established researcher described his process as:

> In my lab we have kind of three main research thrusts and so everyone in the lab is a part of one of those three. And each of those three we have our own little mini group meetings every couple of weeks or so where it's really on the ground like hey, here's some data, I don't know what to do about this, or here's another experiment I'm thinking about, or I keep having this error show up. And then we have weekly lab meetings where the whole lab gets together. So I'd say every student has their own independent project and then some of those projects are, I think, stitched together as part of a larger vision. I really feel like – and students in my lab, I'm training them to be – and this is a philosophical difference between labs. So I really feel like I'm training them to become independent scientists. I

> think some labs manage their graduate students to be more like really, really good technicians. It's a different philosophy about how you would manage the group.

Another established researcher described the process in a way that leveraged the individual strengths of his trainees. He used a metaphor that they were all looking at different parts of the "elephant". In his research:

> …there are people who are very good functional genomics people who are looking at epigenetics and DNA methylation and how that relates to trajectories of cardiovascular health across the life course. There are other people who are actually creating and describing those trajectories from big data where we followed people for long periods of time and they're actually knitting together different cohorts to create one synthetic life experience. There are other parts to the team who are drilling down on the preclinical need before there's actually a heart attack or a stroke. And so we're trying to knit together multiple different looks at the same problem using different tools. And so we have people whose hammer is omics, and we have people whose hammer is data science, and we have people whose hammer is sort of clinical imaging.

From small labs to large labs, research was frequently described as multiple moving parts being conducted by different researchers. Part of the management of the research process was ensuring that research activities were coordinated and later combined when findings were disseminated.

Trainees often referred to collaboration as the ways to refine their research ideas, enhance their professional reputation, and leverage services they need to conduct their science. One trainee stated that when he sought out collaborators, "it's really just figuring out who does what I'm looking for best and going to them and figuring out." He needed to collaborate with an institutional center to plan a clinical trial and he stated that he did not know, "how I would have planned it that as well without them." Other trainees expressed the value of collaborating with biostatisticians and research cores. The value of collaborating with others allowed the trainees to accomplish tasks that were beyond their current skillset or expertise. As one trainee stated when working with a flow cytometry group, “I don’t know what to do because I know that there’s bleed over. And I took the flow course before, but they actually sat down with me and they were like okay, well, for this one, like which one do you care about the most – like we’ll eliminate the bleed over.” Finally, leveraging each other’s expertise and collaborating with other researchers helped to build their professional reputation. As one trainee stated, “like when we help those people with their experiments, they put my name on their paper, so then it’s beneficial for me because I get my name on a publication which publications are your currency, more or less.”

In contrast to the evidence about the value of internal lab collaborations, some research trainees felt disconnected from others in their lab. One trainee felt that her lab was, “not super connected. It’s kind of like we have our own little bubbles of projects.” Another trainee did not see working with others as collaborations, but as transactions. He stated that, “…a lot of people they view it as a transaction and it’s what can you provide me with, what can you help me with and for the mentors who view it that way, which is a lot of them, you need to come in with something to value.”

Trainees and early-career researchers expressed needs to collaborate with internal support like biostatisticians and administrative staff. A trainee stated that, “So we have a statistician that works with us on every project. And so the way that it works is I'm the student and so I'm supposed to figure out what the statistical method is and then I go to my statistician and I say, "Is this correct?" And he'll tweak it a little bit. I guess that they would be considered part of our lab. They're usually on (funded by) the grants that all these projects are funded through.” An early-career trainee described how her funding worked within her lab as, “It’s kind of like I contribute a little bit of my money to this shared thing and then, they contribute big money and then we’ll take money from there and then do research – whoever needs that. It’s kind of like a shared pool or something.”

Data collection, data management, and data analysis. Participants’ description of *data collection*, *data management,* and *data analysis* were interconnected and not described as distinct activities by researchers. *Data analysis* was the activity most frequently referenced by researchers in the experimental subcycle, but it was also closely associated with the collection and management of data. Research trainees exhibited a closer connection to the tasks of collecting, managing, and analyzing data than did established researchers. While the techniques differed, this was true for wet and dry lab science. A research trainee described the process as, “I design it, I implement it and then I collect the data, whatever it may be, whatever the redo and then I analyze it and try to figure out what it means.” Early-career researchers also appeared to be closely tied to collecting, managing, and analyzing data with one key difference, whereas the research conducted by trainees was connected to their mentor’s lab, early-career researchers explored their own research agenda. Activities described by trainees and early-career researchers included creating data dictionaries, retrospective data collection, and collecting samples from patients.

Established researchers described how evolutions in data analysis were affecting biomedical science. One researcher stated:

> I think the impact of next generation sequencing has been pervasive, and will probably accelerate. And so, I think that the growth curve always seems rapid…So, I think there are a number of measures of this as we look at the number of basis of DNA in Genbank, or in NCBI that are housed, and we can see the exponential rise. So, I think that there – this is impact in the kind of science that is done.

Another established researcher described three scenarios for how data collection and analysis was affecting his science. The first was repeatability, a term he preferred over reproducibility or replicability. In this case, the question he stated that his science tried to answer was whether he could take another researcher's data and use a different method produce the same result. The second scenario he described was applying new sequencing techniques in an effort to create a benchmark for future research. He described the process as:

> So, there's a new type of sequencing that been coming out a lot and we're starting to try and identify datasets that would be good that new datasets to adopt for benchmarking purposes. And there, we sort of flowed on top of the literature reading a few things as they come out before really deciding that we're going to tackle a particular dataset. And there, that's driven by reading the literature. I read a paper that says, Oh, we developed this new dataset, and I go, Oh, that'll be perfect for me to swipe and look at.

In the third scenario he described aggregating data from multiple sources so that he could ask his research questions across multiple datasets.

As researchers spoke of managing their data they addressed concepts of storage, privacy, security, and cleaning. No researcher referred to the need for data management plans, such as the type required by some funders. Instead, the challenges researchers face tended to be logistical in nature. One established researcher spoke of using local servers, large institutional storage systems, and the need for physical and cyber security. His research involves clinical and basic science so one issue he described was the need to move the data so that he could have relevant conversations with biostatisticians. An early-career researcher described a different challenge, working with antiquated equipment that utilized proprietary software. Her solution was to design a workflow that allowed her to manage the data so that it could be analyzed. She states, "You can move that to another computer because the computer has the drive reader for that and then you take the data from the ultrasound machine with this drive and then move that to the computer with the card reader. And then you can analyze the data over that computer.” The development of scientific workflows helped researchers manage their data while creating efficiencies. An established researcher connected the development of the data management workflow to the process of creating reproducible research stating:

> We put a lot of effort into developing our workflows and our pipelines for each paper so that they're executable. So that there's no manual steps involved or a minimal number of steps, and everything is scripted or automated. And this naturally lends itself to pushing the data out in a form that can be… You know, we prepare our… We write methods that take raw data and then prepare it the way that our pipelines need it to be prepared and we script all of that. I don’t know if that makes sense. So, then it's all there and then we can push either the raw form or the processed form to public locations because we have both and we can tell people exactly what was done to get from point A to point B.

By considering the reproducibility in the creation of his scientific workflow, this researcher is providing transparency and enabling others to more quickly build on his work.

Not all researchers had established processes for how the lab managed and documenting their data. A research trainee described his method of documenting data as writing a sentence or two so that it would be easy to find by him though a key word search. He went on to add that everyone involved in his lab had his or her own individual method. When asked about how often he uploaded data and documentation on lab servers he said, “I’m actually bad about putting stuff on there, for a long time.”

Organizing and storing information and lab notebooks. The concept of *data management* naturally led to discussion of how researchers *organized and stored information*. Researchers’ management of data was entwined with how they documented and organized that information. There was no standard process among researchers in their approach. In most labs, established researchers left it up to the individual trainees to determine how they wanted to document their data analysis, although many spoke of the challenges of managing and accessing that information when needed. Some research trainees used paper, while some used software such as Microsoft OneNote. Several of the trainees mentioned difficulty with keeping up with their documentation, one stating, “I should be writing it down in a lab notebook. I try to work my lab notebook. Honestly, I get behind on that. I eventually have days where like I’ll catch up and put everything in.” One trainee did not trust storing her information electronically and thought that it was safer to keep her notebook in a locked drawer. Only one established researcher stated using electronic lab notebook software and even he expressed some difficulty in coordinating the system with his lab team. He stated that he was slowly “pushing” everyone to put everything into the system. The advantage to using an electronic lab notebook was not having to track people down when he wanted to look at their data.

When asked about how they organized and stored information, researchers interpreted this question to apply to many types of activities within the research lifecycle. One major activity identified was the organization of journal literature. As previously discussed, many researchers used tools like EndNote or Zotero. One established researcher paid for extra storage so that her lab would have a central location for articles. An interesting point uncovered in their descriptions of the use of these tools is that even though these tools are designed to assist in the organization of citation information, it was clear that most researchers were still struggling to manage information overload, especially when using these tools in a team atmosphere. An established researcher acknowledged that they used EndNote to keep their citation data organized, but also stated that he emails articles directly if he wants to highlight something in the article to his lab.

The other form of organizing and storing information presented by researchers was where to store collected data. In some cases, this information was placed on shared storage servers, while others used cloud-based resources. Use of cloud-based storage tools, like Dropbox, Box, and Google Drive, were identified as a resource when connected to institutional subscriptions. Another established researcher associated storage with cloud services with data analysis, noting that he used Google Genomics and Amazon Web services to process and store data. Several researchers did not trust cloud-based resources, believing that storing their data on the cloud would increase the likelihood that their research could be stolen.

The most common way that researchers described organizing their information was by date. An early-career researcher noted that version was less important than denoting what was new data. In her view:

> The version is not very important for us because we basically just keep adding new stuff. It’s like this version of data is for this animal; next day, new animal. So, it’s kind of more adding up things. And the version becomes important when I write a paper. And then – because I use Word to write paper; the data won’t work for that. So, I have to say: Version 1, Version 2, Version 3.

Another early-career researcher stated that she always keeps a version of her original data untouched and annotates her software code to indicate where and how she made changes to the data. Researchers at all career stages expressed a high degree of autonomy when it came to organizing this information within the resources they used. Wet lab researchers were more likely to describe a formalized process for structuring their collected samples. One established dry lab researcher described the method for his lab as, “There’s latitude, and we try to define best practices. So, there are resources that describe, for example, how the structure directories within a Linux setting. And we don’t have a single requirement for how information is stored and disseminated.” For a wet lab that is collecting and storing numerous samples, the process can be more complex.

> And so then we also have barcodes that are made up, so it's attached to a study ID. When I go over to the hospital and I have various patient swabs or helped her look at swabs, they're all barcoded. So, after I get my samples, I put the barcodes on and then I take them over to the lab. And then the lab will scan them in, and they'll run their tests, and then they'll enter those results into a shared database that we have, which is just an access database… So everything is on paper form that gets transferred to the access database. And so I keep those paper forms so that if something doesn't make sense down the road, I can go back to the paper form.

Collecting HIPAA protected data required extra care to protect the privacy of patients. She noted that organizing and storing her data required her to audit the process.

The latitude afforded to individuals within the lab let them develop their own process that support the general needs of the research. For the trainees this was an important part of the mentoring process. However, the lack of formalized structure created some information management issues recognized by the established researchers. One established researcher described the problem that led her to create more structure:

> I had one particular PhD student who was absolutely wonderful at generating tons of data but after he left, we don’t know what tubes are what. So it’s made me a lot more cautious but given me a lot more oversight – or trying to have more oversight in terms of how they’re doing things. So we’ve gotten a little better at that. And basically having them have master lists of what samples they have and what they are, and corresponding back to their lab notebooks. That’s our biggest hurdle, really, is forcing everyone to take the time to do that and keep track of everything.

Another established researcher noted how the library assisted in the creation of a process that assisted in tracking the organized and stored data. She offered:

> They (the library) also helped a lot of us with data management so they helped us create some just standard Google forms every time we did certain experiments so the students could just sort of fill out the data and have it stored in a database the same way every time. It sounds simple but it was a big help.

Day to day activities of collecting and processing data creates challenges for researchers in managing their data. Well defined processes and procedures assist researchers in managing their data over time and need not be complicated to demonstrate value.

Project management. Obtaining grant funding required management of the project to meet budgetary and reporting requirements from institutions and funders. Grant administrators provided crucial support in the management of research during the experimental cycle. Some researchers had access to departmental staff who helped them prepare and manage their budgets, while others received help from their departments or other central services. One established researcher, also holding a senior positon supporting research at his institution, was building a research support program designed to help researchers manage administrative work associated with grants. He stated that:

> …our institution is relatively young…so it's not been around very much, and early on, that kinda worked because that meant that they had a chance to focus on areas they thought were important. But now as research enterprises got bigger, that doesn't work because you're missing opportunities. And so the goal now is to broaden this out. Used to be I was an independent operator. I was essentially like a small business owner. I had to get my grant money, employ my people, create the product, launch it, market it, get people to digest it and then tell me whether it sucked or not. You know, I mean it was kind of brutal. And it still is in ways. I don't think it's gotten that much easier, but that's what I'm trying to do in this position is to really facilitate that, get a lot of those barriers down where our faculty can actually be the scientists that they need to be.

He went on to add that graduate school does not teach students how to manage a research program and that what he was trying to do with a postdoc program was to build that skillset. An established researcher agreed with the view of utilizing departmental and institutional staff to assist in the management of grants, stating that he took a lean approach and he kept his staff small, tapping into the institutional resources that were available to him.

Data privacy and security. *Data privacy and security* was closely associated with IRB or with lab processes that helped researchers manage and protect their data. Researchers stored their data either on institutional resources, departmental or lab servers, or locally on their hard drive. Storage methods for wet lab and dry lab science varied widely due to the nature of the data. Open access to the data was a concept many researchers were aware of, but many struggled to find the balance between trends in openness and compliance with data privacy requirements. When referencing IRB protocols, researchers expressed that obtaining IRB and compliance with IRB standards for data protection were the primary drivers for ensuring privacy. As long as researchers adhered to their IRB, then most researchers expressed comfort with how the data were being protected. In some cases, researchers provided unclear information on if their data were de-identified. One researcher described her data as “pretty much de-identified”.

Types of data had an important impact on how it was handled and stored. At one institution an early-career researcher and a research trainee, both wet lab researchers from different labs, expressed divergent points of view of where their data could be stored. The research trainee, whose research involved human subjects, stated that, “we have to be very careful, obviously, with how our research is presented, coding everything, where things can be saved to – like we can’t have any patient information on our computer. It has to be on a like health center drive.” Another researcher referred to a manual paper-based system for handling confidential information:

> … he seals it up in an envelope. I know it sounds like some game show but literally, that’s exactly what he does. And he puts in our division secretary, in a safe that she’s got. And we just roll. And everybody just runs the experiments and we keep on track and we know what the timeline is. And then we get to the end and we let and all the data goes to him, and then he unblinds it, and he assigns a code to the data. And then we send it off to a person that does stats to analyze it individually. So that way we really do get an idea of what the hell’s going on.

In contrast, the early-career researcher whose data were animal samples maintained a backup of her data on her local computer. Wet lab researchers had a need to collect or obtain specimens to do their research. For that reason, they had more complicated collection and storage of wet lab data were closely linked to project management procedures. One wet lab researcher who worked with human samples described her process for managing her data collection.

> So we have data collection forms that I've made. We have a protocol. We work with a lab. So I guess the process is I make these data collection form packets. Every day I get an email from – it uses Epic, which is the hospital's database. And it tells me every day who is positive for VRE in one of the ICU's. I have a unique identifier for that person that is like a study ID. It is tied to their medical record number, but you would have to go through many layers to figure out who that person is. And so then we also have barcodes that are made up, so it's attached to a study ID. When I go over to the hospital and I have various patient swabs or helped her look at swabs, they're all barcoded. So, after I get my samples, I put the barcodes on and then I take them over to the lab. And then the lab will scan them in, and they'll run their tests, and then they'll enter those results into a shared database that we have, which is just an access database.

Another research trainee described a similarly complicated process for protecting her research data because it involved multiple sites. She too had developed processes and technologies that helped her manage the day-to-day operations of her research.

Statistical methods. The application of statistical methods was a key component of the experimental cycle. Although all researchers spoke of a reliance on reliable techniques, their definition of what was reliable varied greatly. Most researchers spoke of a reliance on biostatisticians to ensure that their research findings were accurate. In most cases, this support was provided by institutional centers, statistics departments, or through public health science departments. The role of biostatisticians was generally described as a for-fee service, where funding for the use of their expertise was built into the grant application, or where the biostatisticians was an investigator on the grant. One established researcher mentioned the library as a potential source for statistical support and described the value of this approach in connecting to other needs within the research lifecycle. According to him, “I’ve seen some institutions that kinda couple the library and biostatistics together. Which is a great idea I think. And then also in the context of actually preparation [*sic*] of manuscripts or preparation of documents and getting them better integrated into a form that would allow the information to be disseminated.”

In general, established researchers described their own understanding of statistics as adequate or more than adequate, but also described the value of biostatisticians for establishing and understanding the result of complex designs. Several established researchers expressed the importance of utilizing biostatisticians during the planning phase of the research lifecycle while the idea was still being formulated. As such, *statistical methods* were a critical component of the experimental design. One researcher believed that the more complex the study, the more time should be invested early into the study design.

Early-career researchers and trainees described using statistical support more frequently. This may be in part because these researchers are more likely to be conducting the experiments and analysis themselves. The use of a biostatistician was often described as a benefit of working in their lab, a form of collaboration, or an aspect of their training. According to one trainee:

> Yeah. I think the one thing that I did during my F32 was finish a masters in clinical investigation to be able to get a basic handle on whether it's Stata or SPSS. Doing pretty basic analysis. And then the goal will be to be able to do the preliminary basic work myself. And then based on that, transition to using or working – collaborating with a bio statistician for most of the analysis.

In this scenario, she described the importance of her own knowledge of statistics so that when she worked with the biostatistician she knew what questions to ask and how to ask them. Another trainee stated, “I'm the student and so I'm supposed to figure out what the statistical method is, and then I go to my statistician and I say, ‘Is this correct?’ And he'll tweak it a little bit.”

In contrast, some trainees or early-career researchers did not express a need for the help of a biostatistician for their research. For these research trainees, use of tools like Excel or GraphPad Prism allowed them to run the statistical test they felt they needed to use. According to one researcher:

> I took two semesters of statistics when I was in grad school and I think I have two semesters of statistics when I was in college too. So, I think my statistics is pretty strong. And then, if I do have questions, I can Google, right? And then, if you don’t know how to do that, you can YouTube.

Another stated that she could run the test herself using software, but that if she did not have access to software she would not be able to do the test herself. An established researcher provided a drastically different viewpoint on this subject. He stated that at the beginning of the project researchers needed to know that the experiment would have appropriate levels of statistical power, to not introduce batch effects within the design of the study.

Open source software and proprietary software. Researchers use a variety of open source and proprietary tools to analyze their data. Tool selection was influenced by the researchers lab, which may have relied on a particular software. In some cases tools were prescribed by the lead researcher, in other cases trainees were able to select the tool themselves. As one established researcher stated, "everybody in the lab is an expert in one of three, so Python, R, or MATLAB… So they're all experts in at least one of those, but then they end up developing expertise in multiples."

R and Python were the open source tools most used by researchers, although Perl and Java were also mentioned. Use of open source software for analysis was closely connected to sharing code online with other researchers. One advantage of working with an open source programming language was the community of knowledge that helps to support the use of the tools. Many of these communities are comprised of other researchers, where users can learn from and help each other accomplish their research. A trainee, who was originally trained in proprietary software, explained the value of communities supporting open source software:

> I think R is better partly because it's free; it's open source. There's actually tons of great online support; every time I have used it for – like, there's certain things that it just does better, and so I do use it every once in a while. I don't see a downside to it aside from that I think in academia, people are sort of set in their ways a little bit. And the sort of old school statisticians all use SAS. There's sort of this connotation that SAS is like what the real statisticians use even though – you know, for most things, R is just as good. I think SAS is a little better at managing very large data sets. But I'm not even sure how true that is anymore.

Websites like Stack Overflow were mentioned as key resources to gain access to these online communities. GitHub was most frequently mentioned as a place where researchers shared their code, although one researcher maintained a place on Bitbucket where his entire lab for public and private repositories of code. One established researcher, a systems biologist, stated that in his field there was a culture of publishing a link to the GitHub site with the journal article.

> So we have a model of metabolism in a hepatocyte for rat and for human and that model has a standardized – like scientific community format. That model we host on a GitHub site so anyone that's read the paper can pull that from that – that's the same version that we use. And in there, there are conventions for how you name genes and metabolites and everything that we use. And then mapping that to some publicly accessible data, their gene IDs and other kinds of formats that are kind of community accepted standards.

He went on to add that in a role as reviewer for a journal, he expects to see that researchers have published their code and noted the research community in his domain hold researchers accountable for sharing their code and that some journals now require it.

Some labs exhibited a culture of openness that influenced how they did their work. An established researcher stated that almost everything he did was released open sourced. His lab took the extra step of creating virtual machines that allowed others to not only to download the code, but also the entire computing environment in a single package, including the operating system, software stack, and data. Another early-career researcher described trying to use open source software to create a pipeline for reproducible research. He described the problematic example of working with a cohort study where he was making choices in his data manipulation and analysis. To help circumvent this he made thorough notes in his code and saved iterations of his files so that he could retrace decisions made in his research. In contrast, an established researcher expressed discomfort with her trainees using open source software for analysis. She stated, “I have a student who’s really good at R but I’m not that good at it, so I don't like him to use it that much because I don’t know what he’s doing.” Her preference was for her research team to use the proprietary software that she was trained in, Excel and GraphPad Prism.

Although several researchers either used or recognized the need for open source software tools, many still preferred to use proprietary tools for analysis. The major tools referenced were SAS, Stata, Excel, and GraphPad Prism. Reasons for using these tools included ease of use, comfort using the software, and use of the tool by their labs and colleagues. For some trainees, use of proprietary software was connected to their level of comfort with statistical analysis. One trainee explained why she used Prism, “If I didn’t have software that did it for me, I would not be able to do it. I think – I don’t have the background in statistics that would be necessary for that.” Another trainee used Excel because he only needed to produce a P value in his research, while a third referred to the use of the software as “dummy-proof for doing statistics.” The point and click nature of proprietary tools was a motivating factor for some trainees in allowing them to quickly produce results. As one trainee states, “You can have this huge data set and not know what’s publishable in it or not, and then you can just take ten minutes in Stata and all of the sudden you have real science, and real publishable results.” In one case, a researcher received data formatted for SAS, so it was easier for her to use that software for her analysis. Other factors important for researchers in selecting proprietary software was the ability of these tools to easily produce attractive graphics for publication.

In most cases researchers or their labs used a combination of open source and proprietary software for data analysis. The degree to which each type of tool was used varied based on the utility of the tool and culture of the lab. An established researcher whose lab had a strong degree of openness stated that they used open source software for 98% of their work. Other researchers used a combination of open source and proprietary tools to leverage the individual strengths of each tool. For example, proprietary tools like SAS and Stata may have a greater ability to process large data sets, while packages like R allowed for more finesse when producing graphics. Still, others sought to balance a strong connection to the proprietary tools they were familiar with and the need to use tools that were emerging within their domain. At least in one case a researcher saw a disadvantage with using proprietary software because of shifts within her scientific domain:

> SAS has gotten better but a lot of people still run that data in R. More and more of epigenetics, microbial research, that's all genetic epidemiology – that's all run in R, for the most part, anyway. So, I feel like I'm at a disadvantage only using SAS.

Other researchers acknowledged that the shift in biomedical research to open source tools was driving the need to use and learn both types of tools. One SAS user reflected, "So, I guess the reason why I use SAS is because that's what I was taught in school, so that's what I'm most comfortable with. And also, everyone in the division here also uses SAS, so we share a lot of SAS code. That being said, if I were to start again and choose…I think R is better."

The current shift in science to use open source tools was often noted by the trainees, who realized that expertise in using these software environments for their analysis is helpful in advancing their career. An established researcher observed that his lab mostly used R, Python, or MATLAB to conduct analysis. To him, any of these tools can be used to conduct analysis and he expressed that he did not care which tool his trainees used. He described one of his trainees' interest in learning open source software:

> I had this one student who has a gene expression data that she's working with and she's got a lot more MATLAB background, but she just really wanted to develop a more R – expertise with R, so she took (a library) workshop…and then she just forced herself, like I'm gonna do this in R and I'm gonna teach myself to do it.

The library was also mentioned by another established researcher as a place to learn software skills. He stated that his Ph.D. students and postdocs often take course either in their Master's program or at the library.

Pilot experiments. *Pilot experiments* were not identified as a component of the research lifecycle through the document analysis, but emerged as an important theme in the interviews. Researchers described pilot experiments as a way of discovering the next iteration of their research and as the means of remaining competitive for future funding opportunities. Researchers embedded pilot experiments within their current experiments. A wet lab research trainee described the process as, "investigating gene X but you see gene Y change at the same time. So, it's like let's go look into gene Y a little bit". In this case, exploration of gene Y provided insight for potential directions of the research. An established researcher stated that he did not wait on funding to start preliminary research because when he did secure funding it gave him a head start. The constant exploration provided a direction of the current research to build upon, and preliminary data that could be used in a future grant application. Another established researcher stated more plainly that the grant application process, "almost always requires some sort of preliminary data and some sort of a justification for the risk, and the outlay, and those types of things." A third established researcher described a process where he used pilot experiments to establish a body of work, but this was not the only way he used pilot experiments. He stated:

> So sometimes from the genesis of the idea, of those initial pilot experiments, we'll keep following it through because it'll be related enough to existing funding that we'll just kind of keep going along. And then when it feels like there's a body of work that provides a really good, solid foundation of preliminary data that we then really wanna expand in this direction, that's when I would then go to a grant. And in that case it's 'here's the idea, let me go see if I can find what funding is gonna be right to support it.' But I think the other thing – the other kind of inverse often happens which is I'll just see RFAs come out and I'll say you know, what we're doing fits there. And so it's kind of the reverse.

In this case, the pilot experiment helps the researchers prepare in advance and remain competitive for future funding opportunities and to target specific requests for applications by funders.

<u>Ethics and metadata</u>. The concepts of *metadata* and *ethics* were rarely mentioned of by researchers. As previously discussed, researchers either used some organizational structure to help them organize and store information or it was left up to the individual trainees to develop their own system. When asked about metadata and her research, an established researcher described the file structure and readme files she used with her lab for her data.

The most interesting information provided on *metadata* was from a researcher who was investigating indexing and search methodologies. His research made him acutely aware of issues regarding poor use of metadata in research. He stated, "The metadata on these things sucks because the standard suck and people aren't motivated to fill them in properly anyway. Plus, people are often mistaken about what's actually in their datasets." In describing the focus of his research around one data set, he stressed:

> We needed to update the dataset that those authors had communicated to us, the metadata in some sense. And so, for this new paper, what we have now is sort of,

> I would argue, a pretty beautifully broken down set of the metadata that’s been updated from public data bases that’s now relevant to this, the raw data, and it's all in a form that is well documented and sitting in a nice place.

In his view, the assignment and use of good metadata practices would not improve while it relied on human intervention. His interest was in building systems that look inside of the dataset and automate the process while conforming to a standard.

Ethics was not directly mentioned by researchers as part of the research lifecycle. When prompted, researchers referred to ensuring that they complied with IRB or IACUC protocols. The concept of *ethics* was generally expressed through activities that ensured privacy, rather than ethical issues connected to biomedical research. When asked about the role of ethics in her research, a researcher trainee stated, “I mean I work with the patient samples. So, we have to be really careful in terms of how the information is distributed, how things are shared.” Even though *ethics* is clearly an important consideration in biomedical research, specific questions about ethics did not yield substantive responses from the researchers. For example, one early career researchers response to the questions was “Like you’re in Buddhism and then you have to kill rats? Is that it?”

Summary. Data collected on researchers indicated that funding trends from federal funders is incentivizing external collaborations. *Methodology* was connected to the process of *data analysis*. In the experimental subcycle, the collection, management, and analysis of data were a continuous process and not mutually exclusive. Researchers used both proprietary and open-source software tools to analyze their data. Choice of software was influenced by lab culture. *Data privacy and security* was frequently described as complying with IRB and IACUC protocols. Pilot experiments that occurred during the research process were noted as a driver for preparing and competing for future funding opportunities. Data provided by researchers did not reveal a standard method for how they organized and stored information. Some researchers saw project management for their research as analogous to operating a business. Grant administrators were identified as the most utilized resource when it came to *project management*.

Dissemination subcycle. Table 7 summarizes the frequency of activities referenced by researchers pertaining to the dissemination cycle. *Open access, writing, social media*, *journal selection*, *attending conferences*, *data preservation*, and *conference selection* were among the most frequently referenced activities. The high frequency of references to *open access* is likely due to the interviewer's interest in understanding the connection of open access practices as an emergent theme in biomedical research. Researchers generally mentioned that *presentation*, *compliance*, *author rights and copyright, preprint*, and *citation styles* as part of the research lifecycle, but did not indicate that they were critical parts of this process.

Table 7, (N=17)
Frequency of coded themes that occurred within the dissemination subcycle

Cycle Activity	Total Codes	# of References	# of Participants Referencing
Open access	231	53	16
Writing		33	13
Social media		32	15
Journal selection		29	15
Attending conferences		21	13
Data preservation		17	12
Presentation		15	10
Compliance		12	6
Conference selection		11	11
Author rights and copyright		4	4
Preprint		3	2

Citation styles	1	1

Open access, compliance, and data preservation. Biomedical researchers were aware of open access compliance policies associated with NIH funding. Researchers supported by NIH funding are required to deposit published work into PubMed Central. Most references by biomedical researchers to this policy were connected to what they were required to do by NIH, rather than a general sense of making their published work available to the masses. One trainee stated that she deposited her work because she was "required". She went on to quote the policy, but had difficulty recalling the specific requirement. She stated, "Yeah, within like 60 days or six months or something – something with six – yeah. So, I know they have to do that, but I really haven't thought too much of it." An established researcher stated that when it came to compliance with NIH policy she, "just let the publisher do that."

Despite these statements, many researchers were excited about the possibilities open access created. This excitement was expressed more frequently around not only open access to publications, but also open access to data. One trainee stated:

> In principle, I think it's fantastic. I'm all about transparency, and sharing of data when possible. I just think that for me, I think it's something you need to consider early on in your process. It should maybe even be built into the grant application. I think it's something you need to plan for at all stages, and get all of those relevant – you know, your IRB or whoever the funder involved early on to say, 'Here's what we can do'.

It is possible that the low frequency of references to open access publishing was connected to the nature of working with sensitive data. More than one researcher discussed the need to balance the need to protect sensitive data with open access concepts.

Data is an integral component to the concept of open access. The collection, management, protection, analysis, and preservation of data are connected to the concept of *data literacy*. In total, codes related to data were used 196 times, accounting for 18% of the codes used within the research lifecycle. *Open access* had 274 codes that co-occurred with other themes and concepts related to data.

To the researchers, the definition of "open" data was not entirely connected to publically available datasets online. They also spoke of open data they could obtain without paying a fee. This included publically open data sets, but also data that was made available to them after negotiation and signing of a data use agreement. An established researcher believed that data literacy was an important part of his role as mentor. When using other researchers' data he included his trainees in the signing of data use agreements, because they needed to understand the requirements attached to the use of data. Researchers' views on data use agreements varied greatly. An established researcher stated, "I operate under the auspice of buyer beware." Still another established researcher choose to avoid data using agreements entirely, even if the data were in the public domain. He stated:

> …licenses for data use and reuse are… it's a mess of different beliefs that haven't really been tested in a legal system and that nobody is quite sure…So, I just decided to ignore the whole thing. If it shows up in the sequence read archive, it's fair game for downloading because usually it's been published somewhere and at least the rules… So far with non-human genomic data, with non-privacy protected genomic data, is if it's been published, it's reusable.

He went on to add that if he had to click through a license, then he did not use the data because there was plenty of other freely available data. In contrast, another researcher stated that he signs no documents without the review of administrators at his institution, citing, "I'm not qualified to know what is appropriate to sign."

Despite caution for using open data, most researchers saw the value in using open data and making their own data available. One established researcher described his view of data literacy and open data:

> I do worry about data literacy and also rigor of citizen scientists. We (his lab) spend an awful lot of time thinking about what the limitations of our data are, and we don't just kind of go in there wander around and look for positive findings. We actually still believe in hypothesis-driven research. And I worry about – there's a huge role by the way for unbiased, untargeted discovery in data science. I completely believe that. But it's fraught with peril too in false discovery and type 1 errors. Sorry. So I'm cautious, perhaps cautiously optimistic about what open data and open science means.

He added that his lab attempted to provide good examples for how to use the data they made open. To help others understand how to use his data, he spends time documenting his data through white papers and primers. Although there were differing views on open data and data use agreements, it was clear that access to data was influencing scientific practice. An established researcher stated:

> I don't know what that's gonna look like in 15 years. I think it's really cool. I think I'm probably a little further along in my career, unfortunately, to maybe take advantage of that. I keep trying to tell my students and my son, "You should make hay with this because there's so much you can do." So I don't know. We'll see what happens.

This researchers view illustrates the excitement around access to open data and is an example of its effect on the emerging practices of biomedical research.

Collecting data from publically available sources can assist researchers in advancing their research. They can generate ideas for their own research by utilizing open data and reading the scientific literature. One established researcher described how he used a publically available toxicogenomics database to test his lab's research models, run experiments, and validate predictions from their own data. Another established researcher stated that in the sequencing domain there are only three or four centralized databases that operate as an archive for his type of science, but that the genesis of his research idea came from a paper that he read. Other researchers can use open data to ask different questions of a sample. A research trainee stated, "People are taking tumors from people and we look at what the different gene expression changes are in that tumor".

When it comes to biomedical research, one limiting factor is the need to ensure proper data protection and privacy. An established researcher described the importance of open data to reproducible research, but noted that, "it's not happening if the consent form isn't set up properly." He went on to add that, "I think that's an education issue in which the library system can have a crucial role in teaching people such as clinicians or researchers, what their responsibilities are in order to do research in the genomic era." While many researchers spoke of using open data and concepts like reproducibility, there was not a lot of evidence when it came to the management of their own data. It was more likely that they published their methods, or software code that used an already existing open dataset, or that they did not have the right to publish an open dataset that they used. An established researcher stated:

> We have talked about housing it there (institutional repository) or in the Cloud to facilitate collaborations with investigators at other institutions. The main limitation in our case is data use agreements. Because almost all of these in this example are government funded studies, and while we could easily put out via identified sets, we don't own them. We just use them. So data use agreements do not currently allow us to just park it out there where everybody can see it.

Another researcher did state interest in the Center for Open Science's Open Science Framework, but she was concerned about competition from other researchers. She stated:

> I definitely am noticing the push for opening data sets and more open science I think from the early stages instead of just when everything's done. We've dabbled in looking at and using open science framework a little bit… I think it could be a

> really powerful tool because you can read in from any of the storage – cloud storage and organize it in there. But I think everyone's a little hesitant to jump onboard, especially to make things open.

The push for openness and reproducibility were often described as at odds with each other within the research lifecycle. Several researchers spoke of the need to consider openness in the planning of the research project, yet one established researcher added that privacy and an understanding of data usage rights also needed to be considered. None of the researchers mentioned specific open access repositories where they deposit their data. Some believed that the published literature was their data set, while others spoke of internal or cloud-based servers where they stored their data. As a result, data preservation was a poorly defined concept.

An established researcher acknowledged the need to comply with NIH guidelines, but stored his data on institutional servers. He stated, “We need to know how we’re going to make that available. And that’s almost a computer science problem.” A trainee simply stated that researchers should contact him about the data connected to his publication and then he would give it to them. Additionally, several researchers were aware of institutional data repositories, but none stated that they used them.

Despite the lack of examples of researchers preserving their data in open access repositories, many did use open access data. An established researcher described access to data as “an embarrassment of riches” and that there was so much data available today that data owners are begging researchers to analyze it. Researchers were aware of how the access to data was affecting biomedical research. Many saw the early-career researches and trainees as the ones who would leverage access to bulk datasets for the purposes of research. Researchers acknowledged that there were many issues with open data, such as poor descriptions on how the data were collected or organized. Proper training and data literacy were described as ways to prevent misuse of open data.

Writing. For researchers, the process of writing is a critical component of both getting a grant and disseminating the findings through publication. All of the researchers interviewed seemed aware of the importance of writing for publication, but did not describe writing for publication as an area of difficulty. Instead, they described *writing* as the natural output of the other components in the research lifecycle. One research trainee stated, “I think if you want to be an established researcher, you have to worry about two things – not worry, like, think about two things: First is your publication; second is your funding.” An established researcher described the importance of continued publication as his mean for staying on the “cutting edge” and that it was a deliberate part of his process.

Journal selection. When it came to selecting a journal in which to publish research findings, most researchers defined the audience as the most influential factor, but also acknowledged journal metrics like impact factor. An established researcher stated:

> Where you choose to publish it is important because that’s the audience of who’s going to digest the material and likely to cite it later. And so if you publish it in a paper where – or excuse me, a journal where nobody gets it and nobody understands it, of course, they’re not going to cite the paper, right? It’s not going to matter. But if you publish it in a journal where the whole field understands it and reads it and they go ‘Oh, that’s really important.” Even if it’s not a hugely impactful IF (impact factor) journal, they’re gonna cite the hell out of it.”

This sentiment was expressed by researchers repetitively throughout the interview process. One established researcher stated:

> We didn’t do that because it was a higher impact factor. We did it because we wanted the work to be viewed by a larger audience, and so, if we’ve got

> something that we want a larger audience to view, we’ll push that, but if we’ve got something we want our more focused audience to view, then, we’ll push it there.

Many researchers described the selection of the journal to submit their publications to as a function of the need to tell their story, but most had difficulty in ignoring impact factor entirely. Trainees and early-career researchers shared the sentiment that publishing in journals that had the best audience for their work was important, but seemed also to consider the best journal to be the one with the highest impact. An early-career researcher stated that, “first you find a relevant journal and then, the impact factor, the higher is better. That’s pretty much that.”

For many of the trainees and early-career researchers who were trying to establish their own reputation, just getting their work published appeared to be the priority. Because they did not have the professional reputation of established researchers, many had developed their own method for selecting the journals to which they would submit their work. To find their target audience, many scanned the contents of journals and tried to find ones whose scope was appropriate. One trainee described the process as, “Oh, they wrote about this and this in this journal. Ours are similar to that. Maybe we should think about going to that journal.”

For others, the time it took to publication was a critical factor. A trainee stated, “At some point, it feels like it's less important what the impact is, but more important how quickly it could get published.” Another trainee described it a trade-off, because publishing in a high impact journal might take two years and she did not have that much time while establishing her career.

A few trainees and early-career researchers tried to strike a balance between the need to publish quickly to publish in high impact journals, using the Journal/Author Name Estimator (JANE). Journals discovered in JANE allowed early-career researches and trainees to produce a list of journals to which they might submit their work. During one video interview an early-career researchers turned the camera around to the list of potential journals she kept on her wall stating, “this is the list of infectious disease journals in rank order” that she used for selecting her journals.

Author rights and copyright, preprint, and citation styles. *Author rights and copyright* were not presented as key aspects of the research lifecycle. Most researchers did not understand the question or were unconcerned about the concepts. Many researchers felt that the university (through the library) purchased the copyright or owned the article, which is not the case. One trainee described her experiences signing publishing agreements as “pretty standard” and that she did not have any issues or concerns with the process. Only one established researcher stated that his lab used the library for assistance with publishing agreements, stating that the library had been very helpful in that regard. Likewise, researchers in the interviews did not heavily reference *preprints* and *citation styles*. Researchers, when asked about preprints, described them as an emerging trend or with trepidation. One established researchers stated that, “I’m hesitant to get onboard with that but yeah, I just don’t know if it’s a good idea or not.” For *citation styles,* researchers relied on software like EndNote or Mendeley to produce citation formats like APA, MLA, or Chicago and expressed a high level of comfort in using the software.

Attending conferences and conference selection. *Attending conferences* and *conference selection* combined were mentioned frequently by researchers, but as previously described, conferences were most often described in the context of professional networking. Many researchers stated that they knew which conferences were most appropriate to their field. As one early-career researchers stated, “It does matter, because I only go to, American Diabetic conference, because my research is only about diabetes, right?” This seemed true for another trainee who had worked in a few different domains. She maintained:

> …when I was doing TB research, there was one main TB conference. It was an international event; it was quite large. So, I went to that one. And then, the group that I work with here is very involved into national professional societies like, the Society for Hospital Epidemiology of America. And then, the Infectious Disease Society, as well, so everyone tends to go to those two meetings.

Researchers were confident in their ability to identifying conferences to present their work. In most cases, researchers noted that the domain of their research was often supported by a society or organization that sponsored the major conference in that field.

When there were options for selecting conferences, the two primary factors influencing conference selection were size of the conference and timelines for presentation submissions. Some early-career researchers and trainees stated that they selected smaller conferences because the intimate setting gave them a greater chance to network. Others spoke of selecting a conference because their submission deadlines aligned with when they had results to share from their research. However, for a few researchers the selection of a conference was tied to the results of their research. An established researcher stated that, “We don’t have one rule, we’re really – we’re not following one meeting, or one particular way of doing this. So, this is more haphazard.”

Another experience researcher stated that he selects conferences based on the degree to which it can help his project move forward, even if it is not a conference he routinely attends.

Presentation. When it came to creating the graphics needed to present their research, most researchers used software like R, GraphPad Prism, and Excel. None of the researchers expressed a need for assistance in creating a poster or using PowerPoint. One earlier career researcher did state that if she needed help she could easily find it on YouTube. Another early-career researcher described his process of using D3 software to create infographics. Mostly, researchers' concerns over presentation were related to translating their research properly to the public, rather than in conferences. A research trainee described the process of communicating science without using jargon as "extremely difficult". Another trainee stated, "ABC, NBC pick up these stories from, like, research papers, but they get a lot of it wrong. But it's shared with the public. And it's like eating yogurt will prevent depression – I mean (my institution) found was that probiotics have an influence on mental state. However, eating yogurt is not gonna cure depression."

Social media. Researchers described use of social media as a communication tool to promote their research online. Specific social media tools used included ResearchGate, LinkedIn, Twitter, and Instagram. ResearchGate and LinkedIn were also referenced as ways to manage their professional identity. Despite this professed need to manage their online profiles, almost none of the researchers described using ResearchGate or LinkedIn to discover, or connect with, other scientists. Instead, it appeared to be a tool to control and curate the information about them on the web. As one researcher stated it, “I recently set up a LinkedIn page, just because I wanted to make sure I was controlling what was popping up”. The one exception was an established researcher who used ResearchGate to help some researchers who had contacted him using the site. Not all researchers felt that maintaining their information online was important, as one early-career stated that she started her profile on ResearchGate and “gave up in the middle”, citing dissatisfaction with the process.

The uses of Twitter included general use, as well as monitoring Twitter hashtags at conferences. Twitter and Instagram accounts were described as a means for interacting with other researchers, but mainly for promoting their findings or individual and lab accolades. Along with researchers maintaining individual Twitter and Instagram accounts, some labs also had accounts. Twitter’s ability to generate a feed was important because it can easily integrate into lab websites and update dynamically. Instagram was described as a way for labs and individuals to share exciting aspects of their research. One trainee described it saying, “Scientists are getting big with their Instagram accounts too, because they’re taking pretty pictures of cells in lab and posting them.” Despite a general appreciation of Twitter and Instagram, some researchers indicated more awareness of the importance of these tools in biomedical science, than consistent use. When it came to using Twitter, researchers made statements like, “I’m not that good at it,” “I don’t use that much,” and “I try to use them. I really do, but I can’t do it religiously.”

While most researchers described low use of social media for science, there were examples of researchers at extreme ends of the spectrum. One trainee was opposed to using social media in the context of her research. She stated:

> I do not promote my research online. I don’t really like social media. I mean I like it personally, like for my personal life I have a social media, but it’s extremely private. It’s locked down. It’s not even in my real name because I don’t want employers to find – I mean I have nothing bad, but I just – I like separation. So, I’m not one of those types of people that like to share their research.

Another trainee similarly did not want to merge the social media with his research. His reasoning was more modesty than privacy. He saw use of social media as overly self-promoting. In contrast, one established researcher had heavily integrated social media into his research workflow. He used Twitter to interact with his followers to promote interesting scientific papers he had discovered, as well as to discover new ones from the researchers he followed. The value of social media to him was in discovering the unexpected. He stressed:

> I'm following enough people with enough personal prejudices and interests that, and they're all different I should say, that I get stuff out of left-field, but it's all sort of on target at the same time. It's sort of a weird… I don’t really know how to describe it, but I don’t get the stuff that I expect to see. I get stuff that’s still relevant, but I would not have expected it existed, even. It's a recommender system.

His online Twitter presence was also a means to attract the types of students who would fit his lab culture. In his view, Twitter helped potential trainees, dissatisfied with the status quo in data science, to locate his lab and discover his lab’s culture of creating reproducible science.

Summary. Researchers were confident in how to disseminate their research findings. Publishing in peer-reviewed journals was the primary method researchers sought to disseminate their work. The ability to successfully publish their findings help to build their professional reputation and increase their ability to be competitive for future funding opportunities. Researchers selected journals to submit their publication primarily on the audience of the readership, but also considered impact factor, time to publication, and the editorial review process of the journal. Researchers were not concerned with author rights, believing that the library or university purchased the copyright of their publication. Conferences were also an important mechanism for disseminating findings and researchers were confident in knowing the most appropriate conferences to present their research findings. Participants did not state a need for assistance with creating graphics for their presentations. Social media was recognized as an emerging communication tool, but not as a key component of the research lifecycle.

Impact subcycle. Table 8 summarizes the frequency of activities referenced by researchers within the impact cycle. *Citation metrics* was the most occurred the most frequently with 16 of 17 researchers referencing.

Table 8, (N=17)
Frequency of coded themes that occurred within the impact subcycle

Cycle Activity	Total Codes	# of References	# of Participants Referencing
Citation metrics	58	30	16
Altmetrics		5	4

Citation metrics. The journal impact factor produced by Clarivate Analytics and h-index were the most common metrics referenced by researchers in measuring impact. Researchers in all career stages appeared to be influenced by impact factor, but in varying degrees. Most established researchers admitted that the journal impact factor some when selecting journals to publish their research, but was one of many factors. An established researcher stated that he “would never make a decision about where to publish if one journal was an impact factor of 12 and the other was 10.5. But I would if it was an impact factor of one and an impact factor of 20.” As previously stated, early-career researchers and trainees were more focused on publishing in high impact journals, as they were more likely to help them establish their professional reputations. An early-career researcher stated that, “I mean, that’s definitely what it is in terms of academia, like you have to do that with publications and the journal’s matter for promotion and tenure.” In contrast, a research trainee stated that she did not care about impact factor, but she also admitted that she did not plan to pursue a career in academia.

Researchers did not mention using citation metrics as a way to measure the impact of their research. In some cases, they connected the use of impact factor as a metric for promotion and tenure, but were far more likely to speak of journal impact factor as the means for obtaining more citations of their work. Publishing in higher impact journals was more likely to generate more exposure to their work, and thus more citations. Most described the process as finding a field-relevant journal and then choosing a journal with the highest impact number. A research trainee stated:

> I definitely always check the impact factor of the journal when I’m considering submitting something to it. So, I would say that does factor into –. If we’ve identified say, two or three possible suitable journals, I would probably submit to the one with the highest impact factor, first.

All researchers expressed awareness of the domain specific journals that would provide the highest exposure to their work. Depending on the research, they may choose different journals, but the goal was always to get the most relevant readers and citations. More citations to their work helped researchers build the case for the importance of their research and obtain future funding. Thus, citation metrics were not perceived as a way to measure their research impact, but as the means to reinforce the research lifecycle.

Citation metrics were not the only factor in choosing publications. The length of time to publication was also a factor, especially for early-career and research trainees. They often balanced the need to publish in high impact journals with the amount of time it took to have their research published. One early-career researcher stated that she needed 5 to 6 publications a year to adequately build her professional reputation.   A research trainee described it as a “tradeoff”, noting that if a researcher wanted to publish in the journal Nature, “you’re probably gonna have two years’ worth of revisions if you even are lucky and you can get that far.” The need to publish in journals with high impact factors was also recognized by established researchers. One described it as a double-edge sword. For most researchers, they may have said that they did not use citation metrics to evaluate the impact of their research, but they also acknowledge that impact factor was a consideration when selecting journals.

<u>Altmetrics</u>. *Altmetrics* was not a heavily referenced theme by researchers, but most expressed awareness of it. Although there was not a high frequency of references, interesting were collected about altmetrics, especially as they pertained to creating biosketches for grant submissions. One established researcher saw the use of altmetrics for social media posts and press releases as a way for early-career researchers to measure the impact of their work. These metrics are especially important for researchers who do not have an extensive portfolio or whose publications are recent and had not yet been cited by other researchers. The value of altmetrics was linked to changes in the NIH biosketch section C which asks for researchers to describe their contributions to science. One trainee described the need as, “So, using different info metrics. That's been really helpful. So, number of citations, number of Tweets. For a presentation, what kind of audience it reached. Or with media outlets coverage.”   Among the researchers interviewed, one relied heavily on support from the library. He described the library’s

journal impact service as “revolutionary” and the “secret sauce” of their grant applications. He went on to add:

> I think we’re happily successful of getting funding and I think I would put a good amount of credit to what the work that [our librarians] do here and helping us buff our bio-sketches to really make them sparkle. So that if you’re a senior person, you can tell a story about how your research has been picked up by guidelines or reviews and how it’s spread across the world and all those sorts of things both with sort of traditional publication citation metrics, but also some of the newer altmetric type things.

Despite this one researcher’s enthusiasm for altmetrics, most researchers did not express a need to use them. This finding is not surprising when compared to the comments about social media. As previously described, most researchers did not describe use of social media as a key component of their research lifecycle, so it is understandable that they did not place a high value of metrics connected to social media. Reponses to questions about altmetrics by most researchers were lukewarm at best. One established researcher stated that altmetrics were “okay” but that he did not see them, “ taking off in terms of really making a huge impact. There’s a few people that it really does impact. They love it because they’re superstars. And like everybody else is just kinda like, it’s just one more thing I gotta follow.”

Summary. Sixteen of the 17 participants acknowledge the awareness of traditional metrics such as impact factor and H index. Impact factor was most used for journal selection, which assisted researchers in obtaining promotion and tenure. Altmetrics was not highly valued by most of the participants, but was noted by one researcher as helpful for changes in NIH biosketch section C, which asks researchers to describe their contributions to science.

Research question 2. What aspects of the research lifecycle are evolving, if any, due to emerging practices in biomedical research? Table 9 summarizes the frequency of activities referenced by biomedical researchers as co-occurring with the code *emerging practice*. *Data analysis, open access, open source software, methodology, social media*, and *data collection* accounted for 42% of the codes applied. In most cases, *emerging practice* co-occurred with other codes providing evidence for the connection between multiple elements of the research lifecycle. Individual elements of the research lifecycle, as described by researchers, concentrated on the concepts of openness, team science and cohort studies, and social media.

Table 9, (N=17)
Frequency of coded themes that co-occurred with emerging practice

Co-occurrence	Total co-occurring codes	# of Co-occurring Codes
Data analysis	142	13
Open access		11
Open source software		10
Methodology		9
Social media		8
Data collection		8
Mentor/Mentee		6
Reproducibility and replicability		6
Grant funding		6
Data management		5
Data privacy and security		5
Library support		4
Presentation		3
Collaborating		3
Organizing and storing information		3
Proprietary software		3
Literature searching		3
Compliance		2
Data preservation		2
Journal selection		2

Citation metrics	2
Licensing and Venture	2
Non-library support	2
Bioinformatics tools	2
Citation management	2
Data literacy	2
Grey literature	2
IRB/IACUC	2
Identify collaborators	2
Systematic Review	2
Attending conferences	1
Conference selection	1
Preprint	1
Writing	1
Lab notebook	1
Metadata	1
Pilot experiment	1
Project management	1
Altmetrics	1
Biosketch	1

Openness. The concept of openness includes *open data*, *open source software*, *reproducibility and replicability*, and *preprints*. Many researchers expressed awareness of a push for open science in biomedical research. An established researcher described this push as the biggest change occurring in biomedical research and that her students and postdocs were likely the generation to adopt emerging practices. As previously stated in research question 1, the current level of researcher's access to data was described as an "embarrassment of riches" by an established researcher. He went on to say that in the past researchers were protective of interesting datasets, but now he finds the current climate is, "please analyze, somebody analyze my data."

Open data. Much of what researchers had to say about emerging trends in openness was connected to data. An early-career researcher contends, "I think open access for research, research products, and research results is really important for researching limited sets. And I don't think discounted access is enough when they don't have large and developed institutions that employ thousands and thousands of people like we have here in the United States." For her research, access to open data is critical, especially when combined with methodology as described in the published literature. The importance of open data goes beyond the ability to replicate science and can allow researchers to explore and even create new science. An established researcher argues, "You could go from the raw data to something that was now no longer general across the all the experiments, but specific to your… something you can investigate with your specific questions in mind."

The ability to access and analyze large data sets was also described as an emerging practice affecting biomedical research. Network technology is also having a significant impact. An established researcher stated that his institution engaged with multiple networks that allowed him to access data from 23 cohort studies that included data from 630,000 patients. In describing harmonizing and analyzing that data he said, "Let me tell you what, in 2006 that really sucked. I mean, it was awful." However, he also stated that they learned a lot from that study and that with modern technology he sees his current ability to analyze this type of data as incredibly fruitful. In terms of the impact on future science he states:

> So whether it's at the kind of population level or the clinical level or the more systems biology level, there's no doubt in my mind that the ability to handle massive amounts of data to harmonize it and link it to other sources of data and analyze and interpret what comes out the back ends while still keeping a tight hold on the limitations of the garbage in I think is clearly the future of clinical and translational science.

For trainees and early-career researchers, access to large cohort studies could represent challenges as they begin to establish their career. This raises an interesting contradiction considering other emerging trends, which seek to open access to data. An early-career researcher believes, “They want you accessing the data from the existing cohorts to try and address your research questions, which leaves not a whole lot of space because these cohorts have now been entrenched at large research institutions for 20 years, 15 years, and they hold onto their data really tightly.”

Access to open data did present certain challenges for researchers. Cleaning the data so that it can be analyzed is a time-consuming task that may deter some researchers from using open datasets. Data science has emerged as a new discipline and can also be considered an emerging practice impacting the research lifecycle. The researcher who described using open data to answer his own research questions also remarked, “I think that there's simply no way to actually do effective data science without having clean data. That’s something that I see from everybody that works… Anytime you talk to a data scientist, they will tell you that the first 80 percent of the job is cleaning the data to the point where you can now import it.” Access to open data is the foundation of his science, but the concept of openness extends along the path of his entire research lifecycle. This include the development of methodologies and software, along with his data, so that others can replicate and build upon his research. He states:

> So, instead I focused myself on methods and software development where I say, Look, there's 15 different people that all have the same problems. This is indicative of a much larger problem. We have access to all this data. We have access to collaborators that all have this problem and we're good at developing methods and good at developing software and we can slay this dragon. So, if you want to go with the dragon metaphor, I'm not slaying the dragon to protect one particular town. It's that there's a dozen towns and maybe 100s that need the dragon slain, and I can argue that I'm the person to do that.

The creation of reproducible workflows not only assisted in solving the current research problem, but also built in a way that others could benefit and build on his work.

Preprints. Other aspects of the research lifecycle being impacted by openness include the emerging trend of preprints. Depositing preprints of scholarly literature into open archives has long been a practice in the domain of physics, and is beginning to garner acceptance in some areas of biomedical research. One established researcher illustrated the value of preprints to research:

> …preprints, like archive in physics, there's a really, I think, growing movement in biology to make use of preprints. We've been posting on bioRxiv, which has been fantastic. And I think I've had a couple students in my lab who, the one thing that would change is – this may be related to the same thing but – is to just go to the preprint literature and see if there's anybody out there doing something like that that maybe hasn't been vetted and peer reviewed yet, but is in the works, in the pipeline. I think preprints are gonna have a really big impact. I only see benefits from preprints. I don't see any negatives yet in my personal experience with them.

Additional aspects of the emerging impact of openness on published literature are further illustrated with the integration of software code into the published work. An established researcher who was also a reviewer for a peer-reviewed journal stated, "if I review a paper I would wanna see that this researcher has published their model in a – it's called SPML, so it's a particular markup language that's kind of community accepted. So sometimes journals will have expectations that files be available in that format, but if not then the community that's reviewing it will hold author accountable for that."

Open source software. *Open source software* was also referenced as an emerging trend. Some researchers acknowledged that the emergence of open source tools was having a greater impact on researchers earlier in their career and that this change is notable over the last decade. A trainee offered, "Once we graduate, we're not going to pay for these SAS licenses, so why don't you teach us something that we can continue to use?" Several early-career and established researchers stated that they remained comfortable with proprietary software, but if they were new to the field they would likely use an open source tool.

The impact of open access and open source software is closely connected to reproducible science. An established researcher described the importance of reproducibility even within his lab, "So I think reproducibility is big. It's a big thing in my mind…I'll have people do experiments, two or three different pairs of hands do the same experiment. So I can see if it's really real." Some researchers have developed their research lifecycle processes to create reproducible results. An established researcher explained:

> I think the main thing that we do is, that is important, is that we try this repeatability issue. So, we put a lot of effort into developing our workflows and our pipelines for each paper so that they're executable. So that there's no manual steps involved or a minimal number of steps, and everything is scripted or automated. And this naturally lends itself to pushing the data out in a form that can be… You know, we prepare our… We write methods that take raw data and then prepare it the way that our pipelines need it to be prepared and we script all of that. I don't know if that makes sense. So, then it's all there and then we can push either the raw form or the processed form to public locations because we have both and we can tell people exactly what was done to get from point A to point B.

This sentiment was echoed by another established researcher who said, "I think sharing the data, including reproducible research workflows such as computational workflows, is really the way to go." However, he went on to add that creating reproducible workflows is not possible if researchers do not include patient privacy needs in the planning stage. He noted that if proper consent forms are not included in the study design then researchers cannot share the data they collect. He saw this as an area where libraries might play a role stating, "I think that's an education issue in which the library system can have a crucial role in teaching people such as clinicians or researchers, what their responsibilities are in order to do research in the genomic era."

Generational differences. Despite acknowledging emerging trends of openness in biomedical research, many established researchers, and even early-career researchers, noted a difference between the generations. An early-career researcher argues that when it came to using open source software and sharing code, "I'm not necessarily savvy enough to do it. I'm half a generation too late for programming. There's a guy I did my PhD with who's ten years younger than I am who started coding in third grade. So, he hates SAS, he writes everything in Python, and that's how he does his stuff. And the difference in perspective on – or even knowledge on how to share code and why to share code and all that kinda stuff – it was just kind of generationally removed." An established researcher whose focus is data science also acknowledged generational differences and the importance of mentorship. He explained:

> But most faculty, most senior faculty, anybody's who tenured, hasn't had the experience of working with bulk data in quite that same way just because, just the latency. So, I think that natural succession of junior faculty into the senior ranks will do a great deal to fix that. What I worry about is the people that get away with doing a half-assed job and then think that's what they teach their students, but that's again why we invest in training.

In this way early career faculty transitioning into experienced faculty will have an impact of the evolution of research practices. Greater access to data and new analytic techniques will bring about change in biomedical research, but only if the mentors themselves evolve with emerging trends.

Team science and cohort studies. Team science and cohort studies were also described as emerging trends in biomedical science. Almost all of the participants described funding trends as the key factor driving this change. Funders' interest in team science may be linked to addressing grand challenges facing society today. One early-career researcher acknowledged this trend by stating, "I think one of the biggest differences for me entering as a hopeful, independent investigator now, at least in my field – 20 years ago we didn't have large, large team-built cohorts. So, everybody got their funding by setting up large, large team-built cohorts. Now, they don't wanna see new large, large team-built cohorts." An established researcher specifically described what he has heard from the director of NIH, "(Francis) Collins basically said that they envision at the NIH, similar approach that the NSF has made, the concept of team science to tackle some of these big problems is really where it's at." The challenge as he described it was that biomedical researchers have evolved to be experts in specific niches of science. This evolution towards specificity requires that researchers assemble teams of experts to create a competitive grant application and conduct their research. Interestingly, he saw a similarity in how some health sciences libraries are trying to diversify their workforce.

Social media. Although the use of social media received mixed reviews from researchers, most expressed knowledge of its use as an emerging practice for biomedical science. Even if a researcher stated that they did not use social media, they recognized its use by their peers. When specifically asked about social media as an emerging practice, researchers described use of social media as a means for interacting with other researchers, communicating results to the general public, and managing their online professional identity. Use of social media by researchers at different stages of their careers varied widely and no evidence was found that there were generational differences dependent on the stage of a researcher’s career.

Many researchers described following the social media accounts of labs and their peers as a means of staying connected. Use of Twitter at conferences was also mentioned as a way to connect and communicate while at a conference. A research trainee offered, “I won't say that I'm a fairly active user on Twitter. I don't know if this is what you're getting at, but I saw a lot of epidemiologists on Twitter and all the conferences on Twitter. A lot of people have blogs now. I have a blog and I think this is an area that people are going into more to disseminate not just their science but ideas.” That individuals and labs were using Instagram was an interesting finding that provides evidence that researchers find value in connecting with peer researchers through social media.

Use of social media was described as a way to communicate finding to the general public. A research trainee remarked, “I think, social media is definitely – it’s changed the way we do all things. And so, to get science and messaging out, there’s probably a stronger tool to the general public.” In contrast, another researcher saw negatives in communicating research results to the general public. She described the use of Facebook by major news outlets as over-simplifying results and creating misinformation for the public.

Social media was also described as a way to manage and extend an online identity. A research trainee described a faculty colleague as having a blog with a huge readership that served as a way for him to “share ideas, and share study findings, and comment on things.” She saw this as an area that she needed to improve in order to get the message out about her research. An established researcher with an influential social media presence found that his identity helped him recruit trainees who shared his philosophy of research. He provided evidence for the importance of social media for in his domain, stating, “In a bioinformatics, most new tools and stuff come through Twitter anyway, well before they had a paper.” However, he admitted to not having found a way to mentor his trainees in the effective use of social media. Instead, he found that his trainees were either predisposed to use social media or not. He explained, “I would say that I've been less than successful at teaching them the self-confidence and, for lack of a better word, sort of, fearlessness that I seem to have with respect to social media.” Still, use of social media was an important aspect of his online identity and has helped further his career. He argues:

> It's (social media) also been very, very good for my career in that I get a lot of invitations between that and my teaching. I get a lot of invitations. These people know that I have things to say and that I'm not a like, like a total idiot even if they disagree with me. And it's also a way, increasingly, and this is a tricky game to play, but increasingly it's been a way to try and shape the future of bioinformatics and data science to some extent by making sure that things that I've personally experienced that I think generalize are things that people are aware of, even if they haven’t experienced them personally. So, I get quoted a lot in nature and news articles and stuff because I have that social media presence, and so people give me calls, basically.

In contrast to this perspective, a research trainee expressed his trepidation about using social media while acknowledging the impact it was having as an emerging practice, “I know a lot of people do a lot of it, social media thing. I don’t right now see how that helps me get published in good journals, get good grants or further my collaborations and in some ways, I think there is a lot of egotism and self-promotion there so I avoid it but it’s probably going to be more and more important, I’ll probably have to adjust.”

Summary. The most frequently referenced emerging practices by biomedical researchers were *data analysis*, *open access*, and *open source software.* Concepts of openness were strongly associated with publications and data, but are evolving to include *reproducibility and replicability*. Some researchers noted that generational differences are emerging between researchers trained in previous eras to those trained today. Data science is a new domain which assists biomedical researchers in cleaning and analyzing data. Funding trends are incentivizing team science and large cohort studies. Social media is impacting how research findings are communicated to fellow researchers and the public.

Research question 3: In what ways do health sciences libraries support the research lifecycle and emerging practices of biomedical research? Health Sciences library leaders were asked to report on library services they provided in support of the research lifecycle. The survey instrument sought to determine what services libraries were providing, planning to provide, or have no plans to provide. To help reveal trends in health sciences libraries' service models, participants were asked if they had provided individual services for more than four years, more than two years, added services in the last two years, planned to provide in the next two years, planned to provide in the next four years, or if they had no plans to add. Of the 51 participants (Table 10), 44 were Directors (or comparable title), one Deputy Director, two Associate Directors, and four Other, which was primarily consisted of Assistant Directors or department heads.

Table 10, (N=51)
Library leader position

Title	N
Director	44
Deputy Director	1
Associate Director	2
Other (Please specify)	4

Since position titles are specific to institutional organization structure, some respondents selected the Other category if their title was University Librarian, Executive Director, Associate Dean, or a similar title. If it was determined that the respondent was the primary person responsible for all library operations, they were included in the Director category displayed in Table 9. Data from the AAHSL membership directory, along with individual library websites, was used to determine if the respondents met the criteria to be counted as a director.

The mean number of years participants served at the institutions was 14 years. Years serving at their current position level, including multiple institutions, is provided in Table 11. The distribution of respondent's institutions by BRMIR ranking is provided in Table 12. Due to the small sample size. BRIMR rankings were not used to determine if there were statistical differences by quintile.

Table 11, (N=51)
Years served at current position *level including multiple institutions*

	N	Minimum	Maximum	Mean
Director	44	0	41	8
Deputy	1	4	4	4
Associate	2	1	6	8
Other Assistant	4	2	10	7

Table 12, (N=51)

Blue Ridge Institute for Medical Research Ranking by quintile

	N	Mean
1st	12	24
2nd	13	26
3rd	11	22
4th	7	14
5th	8	16

To analyze the data and expose patterns, results of the survey were first aggregated into categories of services currently provided by libraries, services they plan to add, and services they do not plan to add. Within each subcycle, aggregate results are analyzed, as well as specific service trends that were identified in the complete dataset. The entire data for all subcycles can be found in Appendix B.

Research planning subcycle. Results from the survey illustrate a strong presence of library services in the research planning subcycle. As shown in Table 13, at least 50% of the respondents indicated that they currently support the research lifecycle with services for background literature searching, citation management, systematic reviews, grey literature searching, locating data sources, methods for organizing and storing information, data management plans, IACUC protocols, NCBI tools (Blast, GenBank, dbSNP, etc), seeking grant funding, and identifying collaborators.

Table 13, (N=51)

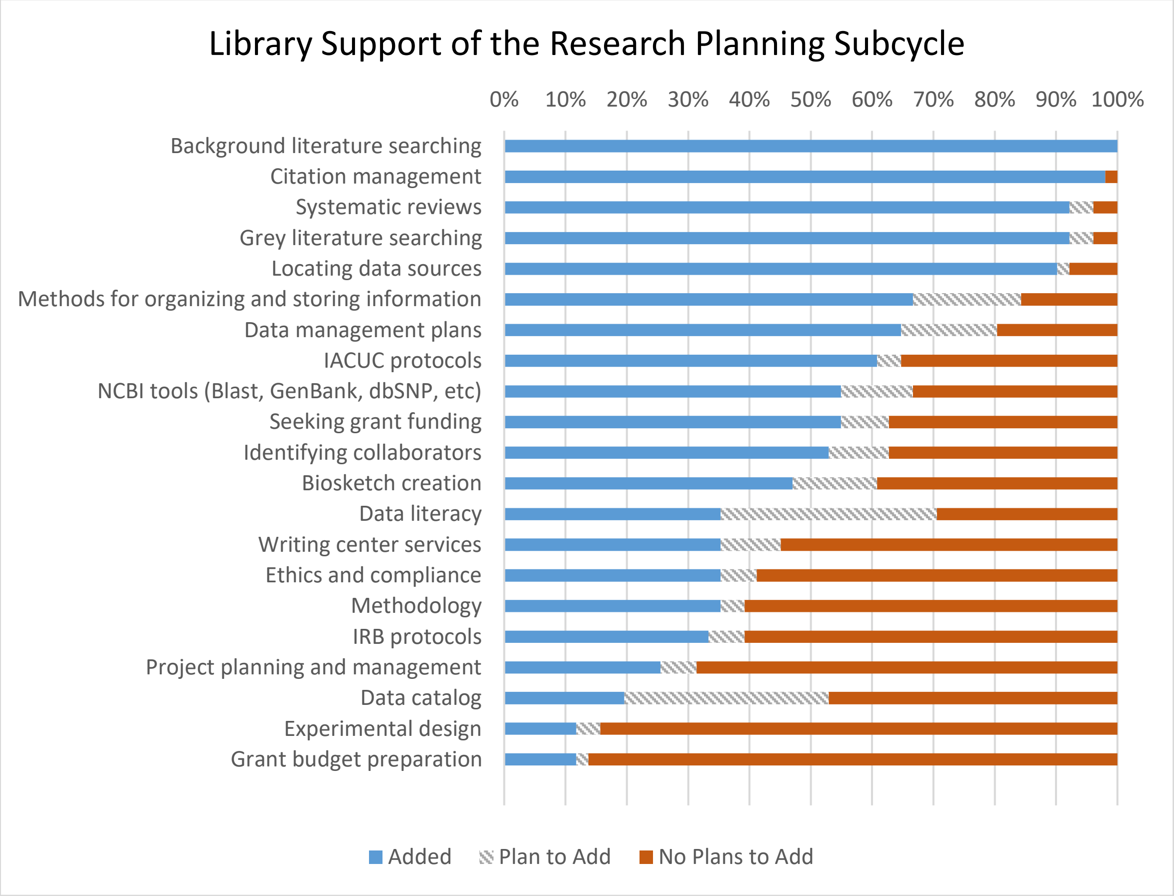
Research planning subcycle: services provided or plan to add

Cycle Activity	Provide		Plan to add		No plans to add	
	N	Mean	N	Mean	N	Mean
Background literature searching	51	100	0	0	0	0
Citation management	50	98	0	0	1	2
Systematic reviews	47	92	2	4	2	4
Grey literature searching	47	92	2	4	2	4
Locating data sources	46	90	1	2	4	8
Methods for organizing and storing information	34	67	9	18	8	16

Data management plans	33	65	8	16	10	20
IACUC protocols	31	61	2	4	18	35
NCBI tools (Blast, GenBank, dbSNP, etc)	28	55	6	12	17	33
Seeking grant funding	28	55	4	8	19	37
Identifying collaborators	27	53	5	10	19	37
Biosketch creation	24	47	7	14	20	39
Data literacy	18	35	18	35	15	29
Writing center services	18	35	5	10	28	55
Ethics and compliance	18	35	3	6	30	59
Methodology	18	35	2	4	31	61
IRB protocols	17	33	3	6	31	61
Project planning and management	13	25	3	6	35	69
Data catalog	10	20	17	33	24	47
Experimental design	6	12	2	4	43	84
Grant budget preparation	6	12	1	2	44	86

Library leaders identified data literacy (35%) and a data catalog (33%) as services they plan to add. Libraries stating that they are already providing data literacy services include 35% of the respondents, with 20% already supporting data catalogs. These numbers may highlight a division among health sciences libraries, as 29% of libraries do not plan to add data literacy services and 47% do not plan to support a data catalog. The services least likely to be supported by health sciences libraries include writing center services, ethics and compliance, methodology, IRB protocols, project planning and management, experimental design, and grant budget preparation. Figure 5 reveals the distribution of services that libraries have added, plan to add, or have no plans to add.

Figure 5. Library support of the research planning subcycle



Services provided by libraries. Table 14 provides evidence that background literature searching, citation management, locating data sources, grey literature searching are established services provided by libraries 4 or more years.

Table 14, (N=51)

Research planning subcycle: services provided more than four years

	N	Mean

Background literature searching	49	96
Citation management	49	96
Locating data sources	37	73
Grey literature searching	36	71
Systematic reviews	29	57

Data management plans, biosketch creation, and systematic reviews are the services identified as transitional services being implemented by library leaders. As Table 15 illustrates, adoption of these services have been implemented by libraries for more than two years.

Table 15, (N=51)
Research planning subcycle: services provided more than two years

	N	Mean
Data management plans	14	28
Biosketch creation	14	28
Systematic reviews	12	24

Emerging trends. Data will likely continue to be an emergent trend, with libraries having provided more data related services in the last two years or planning to add them in the near future. Table 16 shows that services libraries plan to add for data literacy (45%), data management plans (31%), and data catalogs (31%) over the next two years. Data literacy and the ability to discover appropriate sources are evolutionary services where libraries expand on their role in supporting information literacy to include concepts of literacy that include data. Emerging trends in supporting data align with emergent trends identified by biomedical researcher interviews, which include open access since data literacy and data catalogs have strong connections to data access and literacy.

Table 16, (N=51)

Research planning subcycle: library data services

	Added in last 2 years		Plan to add in the next 2 years		Sum mean
	N	Mean	N	Mean	Total
Data literacy	8	16	15	29	45
Data management plans	11	22	5	10	31
Data catalog	5	10	11	22	31

The data supporting trends for library services that extend beyond two years is less clear. Data continues to be an emergent theme, but noted with less frequency. As Table 17 depicts, data catalogs were expressed most heavily at 12%. all other services that libraries plan to add in the next four years were referenced by less than 10% of library leaders.

Table 17, (N=51)

Research planning subcycle: services planning to add in the next four years

	N	Mean
Data catalog	6	12
Methods for organizing and storing information	4	8
Data literacy	3	6
Data management plans	3	6
NCBI tools (Blast, GenBank, dbSNP, etc)	3	6
Biosketch creation	2	4
Project planning and management	2	4
Methodology	1	2
Ethics and compliance	1	2

It is important to consider what services libraries did not indicate that they plan to add along with services provided or that libraries plan to add. Table 18 shows that libraries do not see a strong future role in supporting grant preparation, project management, methodology, and a variety of other functions not directly connected to the creation or management of knowledge.

Table 18, (N=51)

Research planning subcycle: no plans to add

	N	%
Grant budget preparation	44	86
Experimental design	43	84
Project planning and management	35	69
Methodology	31	61
IRB protocols	31	61
Ethics and compliance	30	59
Writing center services	28	55
Data catalog	24	47
Biosketch creation	20	39
Identifying collaborators	19	37
Seeking grant funding	19	37
IACUC protocols	18	35
NCBI tools (Blast, GenBank, dbSNP, etc)	17	33
Data literacy	15	29
Data management plans	10	20

Given the strong presence of data catalog and data literacy services in shown Table 18, it may be that there is a division in the priorities assigned by library leaders.

Summary. Library leaders responses to the survey indicated a significant level of support for the research planning subcycle. More than half of all respondents noted that their libraries currently provide services for background literature searching, citation management, systematic reviews, grey literature searching, locating data sources, methods for organizing and storing information, data management plans, IACUC protocols, NCBI tools (Blast, GenBank, dbSNP, etc), seeking grant funding, and identifying collaborators. Support for data appears to be increasing, with leaders indicating an increase in data related services. Library leaders who have added data services in the last two years or plan to add in the next two: data literacy (45%), data catalog (31%), data management plans (31%), and methods for organizing and storing information (27%). However, 29% of libraries do not plan to add data literacy services and 47% do not plan to support a data catalog, potentially highlighting a division among health sciences leaders. Services libraries do not plan to add in support of the research planning cycle include grant budget preparation, experimental design, project planning and management, methodology, IRB protocols, ethics and compliance, and writing center services.

Experimental subcycle. The results from the survey illustrate that health sciences library services are less integrated during the experimental subcycle. Table 19 reveals that support for managing research data and metadata standards is the only component noted by more than 50% of respondents, although data documentation (49%) and prototyping (43%) are also strongly represented.

Table 19, (N=51)

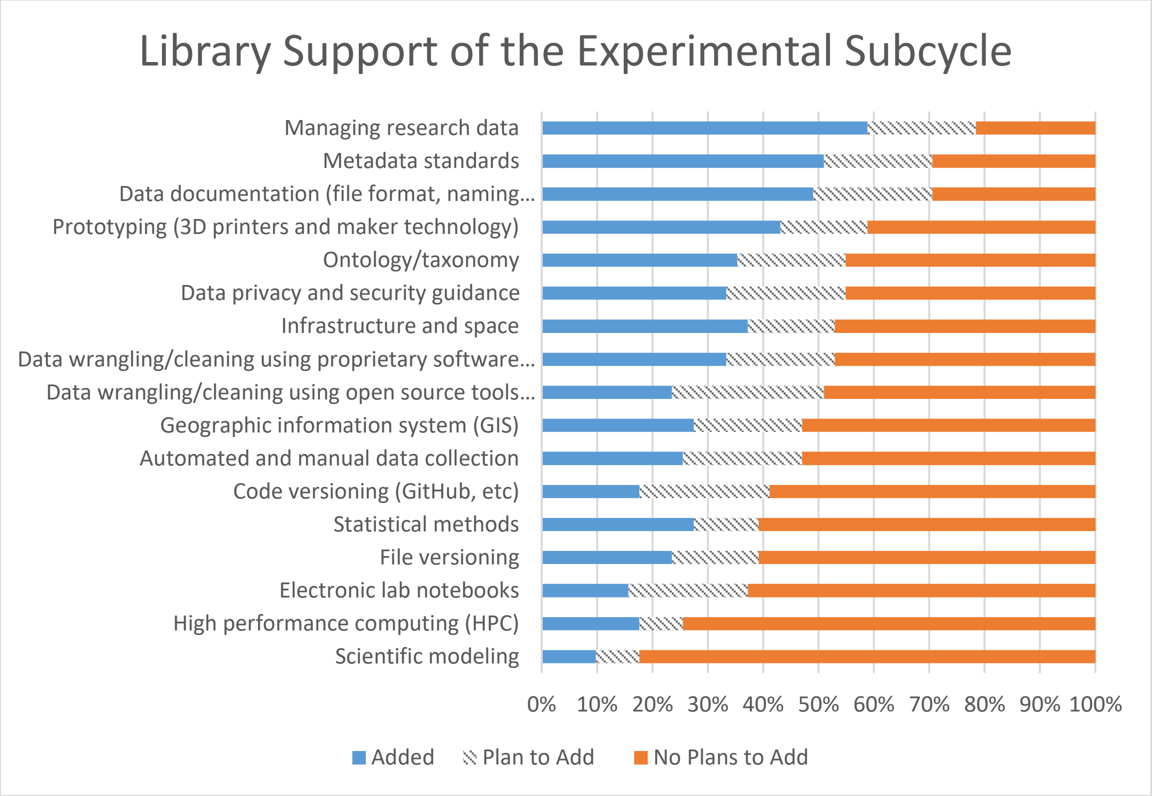
Experimental subcycle: services provided or plan to add

Cycle Activity	Provide		Plan to add		No plans to provide	
	N	%	N	%	N	%
Managing research data	30	59	10	20	11	22
Metadata standards	26	51	10	20	15	29
Data documentation (file format, naming conventions, file organization)	25	49	11	22	15	29
Prototyping (3D printers and maker technology)	22	43	8	16	21	41
Infrastructure and space	19	37	8	16	24	47
Ontology/taxonomy	18	35	10	20	23	45
Data privacy and security guidance	17	33	11	22	23	45
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	17	33	10	20	24	47
Geographic information system (GIS)	14	27	10	20	27	53
Statistical methods	14	27	6	12	31	61
Automated and manual data collection	13	25	11	22	27	53

Data wrangling/cleaning using open source tools such as R, Python, OpenRefine	12	24	14	27	25	49
File versioning	12	24	8	16	31	61
Code versioning (GitHub, etc)	9	18	12	24	30	59
High performance computing (HPC)	9	18	4	8	38	75
Electronic lab notebooks	8	16	11	22	32	63

Some libraries appear to be considering providing services connected to data. At least 20% of library leaders are planning to add services related to collecting and managing data, using software to process data, and techniques for describing data. Figure 6 shows the distribution of services that libraries have added, plan to add, or have no plans to add.

Figure 6. Library support of the experimental subcycle



Services provided by libraries. The most established service provided to biomedical researchers by libraries for more than four years during the experimental subcycle was infrastructure and space at 28%. Other services worth noting include data documentation, GIS support, managing research data, data privacy and security, data wrangling with proprietary software, and data collection (Table 20). It is interesting to note that no health sciences libraries have provided support for code versioning, electronic lab notebooks, data wrangling with open source software, or scientific modeling during the timeframe (Appendix B).

Table 20, (N=51)
Experimental subcycle: services provided more than four years

	N	Mean
Infrastructure and space	14	28
Metadata standards	9	18
Managing research data	9	18
Geographic information system (GIS)	8	16
Data documentation (file format, naming conventions, file organization)	8	16
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	7	14
Data privacy and security guidance	7	14
Automated and manual data collection	6	12
File versioning	5	10
Statistical methods	5	10
Ontology/taxonomy	5	10

Emerging trends. Responses show that trends for supporting data and prototyping for more than two years are emerging, but with an adoption rate of 20% or less. Table 21 displays library provision of these services.

Table 21, (N=51)
Experimental subcycle: services provided more than two years

	N	Mean
Metadata standards	10	20

Data documentation (file format, naming conventions, file organization)	8	16
Ontology/taxonomy	7	14
Prototyping (3D printers and maker technology)	7	14
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	6	12

As seen in Table 22, trends for supporting data and prototyping were also listed as library services that have emerged in the last two years, along with some libraries developing services for statistical support.

Table 22, (N=51)
Experimental subcycle: services added in last 2 years

	N	Mean
Managing research data	11	22
Prototyping (3D printers and maker technology)	11	22
Data documentation (file format, naming conventions, file organization)	9	18
Metadata standards	7	14
Data wrangling/cleaning using open source tools such as R, Python, OpenRefine	7	14
Code versioning (GitHub, etc)	7	14
Ontology/taxonomy	6	12
Statistical methods	6	12

Data support continues to be a consistent theme for the development of library services. Table 23 provides insight into the services libraries plan to add over the next two years. While these numbers remain consistently low, when combined with trends over the last four years, the responses suggest that libraries are thinking about data and how to provide services to biomedical researchers.

Table 23, (N=51)

Experimental subcycle: services planning to add in the next two years

	N	Mean
Data documentation (file format, naming conventions, file organization)	9	18
Automated and manual data collection	9	18
Ontology/taxonomy	8	16
Geographic information system (GIS)	8	16
Data wrangling/cleaning using open source tools such as R, Python, OpenRefine	7	14
Code versioning (GitHub, etc)	7	14
Prototyping (3D printers and maker technology)	6	12
Data privacy and security guidance	6	12
Electronic lab notebooks	6	12
Infrastructure and space	6	12
File versioning	5	10
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	5	10

In the long term, this trend is expected to continue, with libraries indicating that they plan to add data related services. Table 24 summarizes the services libraries plan to add over the next four years.

Table 24, (N=51)

Experimental subcycle: services planning to add in the next four years

	N	Mean
Data wrangling/cleaning using open source tools such as R, Python, OpenRefine	7	14
Managing research data	6	12
Metadata standards	6	12
Code versioning (GitHub, etc)	5	10
Data privacy and security guidance	5	10
Electronic lab notebooks	5	10
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	5	10

More than 75% of participants responded that they had no plans to add services for scientific modeling and high performance computing. Despite a growth in the number of libraries providing data services, many library leaders are not planning to support several data related services. Close to 50% of library leaders do not plan to support software tools for data wrangling, collection, and management as a library service. Table 25 indicates that many of the emerging service trends in libraries have not fully been embraced by library leaders. Aggregate data demonstrate library trends towards establishing data related services over the next four years, but also reveal that 40% or higher of libraries do not plan to add these services.

Table 25, (N=51)
Experimental subcycle: no plans to add

	N	Mean
Scientific modeling	42	82
High performance computing (HPC)	38	75
Electronic lab notebooks	32	63
File versioning	31	61
Statistical methods	31	61
Code versioning (GitHub, etc)	30	59
Automated and manual data collection	27	53
Geographic information system (GIS)	27	53
Data wrangling/cleaning using open source tools such as R, Python, OpenRefine	25	49
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	24	47
Infrastructure and space	24	47
Data privacy and security guidance	23	45
Ontology/taxonomy	23	45
Prototyping (3D printers and maker technology)	21	41
Metadata standards	15	29
Data documentation (file format, naming conventions, file organization)	15	29
Managing research data	11	22

Summary. Responses from library leaders illustrate weaker integration of library services within the experimental subcycle. Managing research data and metadata standards were the only services identified by more than half of the participants. Increased trends for data services were noted by library leaders who have added data services in the last two years or plan to add in the next two. These include data documentation (35%), prototyping (33%), managing research data (29%), code versioning (27%), data wrangling/cleaning using open source tools (27%), ontology/taxonomy (27%), and automated and manual data collection (25%). Others leaders did not indicate that they will increase data services with 49% of respondents stating they had no plans to add services for data cleaning using open source software tools and 47% data cleaning using proprietary software tools. Most library do not plan to add services for scientific modeling, high performance computing (HPC), electronic lab notebooks, statistical methods, file versioning, code versioning (GitHub, etc), geographic information system (GIS), and automated and manual data collection.

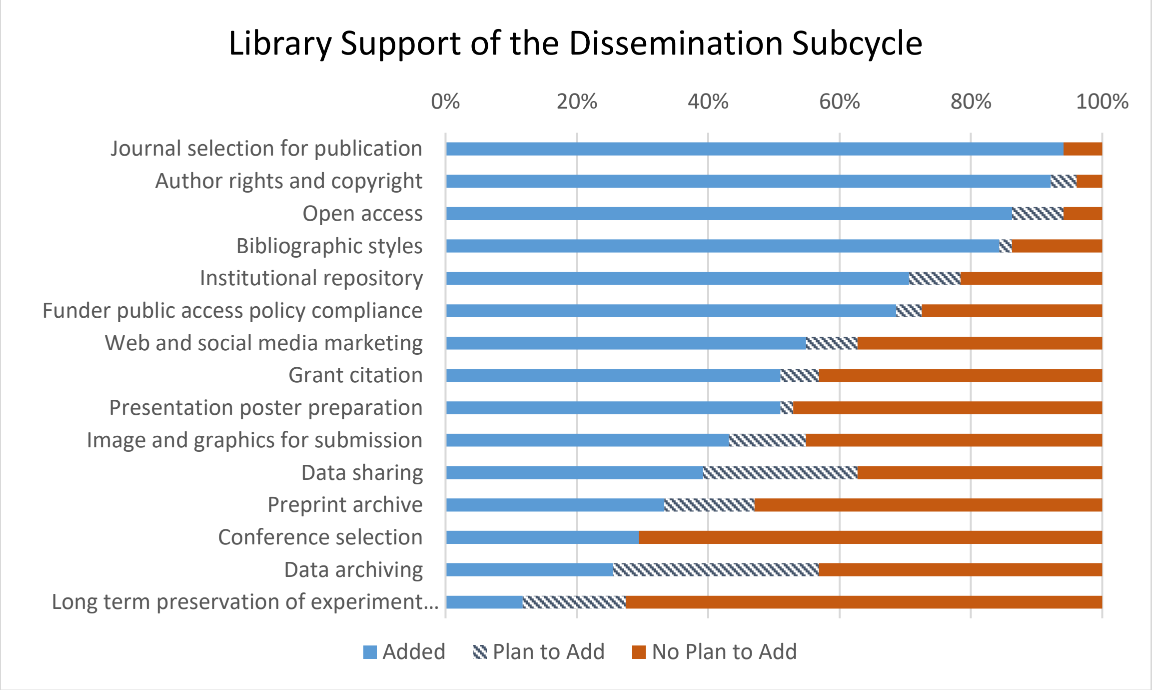
Dissemination subcycle. Results from the survey indicate a strong presence of library services in support of the dissemination subcycle. Table 26 shows that more than half of libraries provide services for journal selection for publication, author rights and copyright, open access, bibliographic styles, institutional repository, funder public access policy compliance, web and social media marketing, grant citation, and presentation poster preparation. Data archiving and data sharing were the services most likely provided by libraries. Figure 7 reveals the distribution of services that libraries have provided, plan to add, or have no plans to provide.

Table 26, (N=51)

Dissemination subcycle: services provided or plan to add

Cycle Activity	Provide		Plan to add		No plans to provide	
	N	Mean	N	Mean	N	Mean
Journal selection for publication	48	94	0	0	3	6
Author rights and copyright	47	92	2	4	2	4
Open access	44	86	4	8	3	6
Bibliographic styles	43	84	1	2	7	14
Institutional repository	36	71	4	8	11	22
Funder public access policy compliance	35	69	2	4	14	27
Web and social media marketing	28	55	4	8	19	37
Grant citation	26	51	3	6	22	43
Presentation poster preparation	26	51	1	2	24	47
Image and graphics for submission	22	43	6	12	23	45
Data sharing	20	39	12	24	19	37
Preprint archive	17	33	7	14	27	53
Conference selection	15	29	0	0	36	71
Data archiving	13	25	16	31	22	43
Long term preservation of experiment materials	6	12	8	16	37	73

Figure 7. Library support of the dissemination subcycle



Services provided by libraries. Services established by libraries to support the dissemination subcycle include bibliographic styles, author rights and copyright, open access, journal selection for publication, funder public access policy compliance. As Table 27 illustrates, more than half of all respondents stated that they have been providing these services for more than four years. These results are not surprising, as health sciences libraries have traditionally provided services related to journal publications.

Table 27, (N=51)
Dissemination subcycle: services provided more than four years

	N	Mean
Bibliographic styles	39	77
Author rights and copyright	38	75
Open access	36	71
Journal selection for publication	32	63
Funder public access policy compliance	26	51

Emerging trends. The responses to question concerning services provided by libraries for more than 2 years reveal some emerging trends (Table 28). Services related to journal publications are still strongly represented, as well as newer evolutionary services such as institutional repositories, grant citation, image and graphics for submission, social media, and data sharing.

Table 28, (N=51)
Dissemination subcycle: services provided more than 2 years

	N	Mean
Journal selection for publication	12	24
Institutional repository	7	14
Grant citation	6	12
Image and graphics for submission	6	12
Author rights and copyright	5	10
Open access	5	10
Funder public access policy compliance	5	10

Web and social media marketing	5	10
Data sharing	5	10

Trends supporting data sharing, data archiving, and institutional repositories are also evident in services added by libraries in the last 2 years as shown in Table 29.

Table 29, (N=51)
Dissemination subcycle: services added in last 2 years

	N	Mean
Data sharing	9	18
Institutional repository	6	12
Data archiving	6	12

Those services libraries are planning to add in the next two years include data archiving, data sharing, preprint archive, image and graphics for submission, and long-term preservation of experiment materials. Table 30 further illustrates that libraries are considering providing data support services. The intent to provide services for image and graphics for publication indicates that libraries are exploring some services to help researchers create and represent knowledge. The creation of preprint archives is an emerging trend in some biomedical domains and 10% of libraries plan to add services that may help support researchers' use of preprint archives. Many health sciences libraries have departments that collect and curate historical materials, and interest in long-term preservation of experiment materials shows that libraries may be expanding their role in preservation to include other aspects of biomedical research.

Table 30, (N=51)
Dissemination subcycle: services planning to add in the next 2 years

	N	Mean
Data archiving	8	16
Data sharing	6	12
Preprint archive	5	10
Image and graphics for submission	5	10

Long term preservation of experiment materials	5	10

Data related services provided by libraries might be a long-term trend as 12-16% of respondents indicated they were planning to develop data archiving and sharing services in the next four years (Table 31).

Table 31, (N=51)
Dissemination subcycle: services planning to add in the next four years

	N	Mean
Data archiving	8	16
Data sharing	6	12

Despite emerging trends for long-term preservation of experiment materials, preprint archive, image and graphics for submission, data archiving, and grant citation, many library leaders do not believe that these are roles are necessary for their libraries. Table 31 shows that more than half of libraries do not plan to add services for long-term preservation of experiment materials, support for conference selection, or support for preprint archives. The preparation of posters, image and graphics for submission, data archiving, and grant citation were also noted as service not being considered by a large percentage of libraries (Table 32).

Table 32, (N=51)
Dissemination subcycle: no plans to add

	N	Mean
Long term preservation of experiment materials	37	72.5
Conference selection	36	70.6
Preprint archive	27	52.9
Presentation poster preparation	24	47.1
Image and graphics for submission	23	45.1
Data archiving	22	43.1
Grant citation	22	43.1

Summary. Library leaders indicted providing a significant number of services supporting the dissemination subcycle. More than half of the respondents stated they provided services for journal selection for publication, author rights and copyright, open access, bibliographic styles, institutional repository, funder public access policy compliance, web and social media marketing, grant citation, and presentation poster preparation. Trends in supporting data sharing, data archiving, and institutional repositories are increasing for libraries. Despite 44% of leaders indicating the provide or plan to add services for data sharing, 37% have no plans to add. Similarly, 56% of library leaders provide or plan to add data archiving services, yet 43% have no plans to provide. This division represents another potential division among library leaders. Library leadership indicated that they are least likely to add services for long term preservation of experiment materials (73%), conference selection (71%), and preprint archive (53%).

Impact subcycle. Survey data reveals that health science libraries have a strong presence in supporting the impact subcycle. Services related to impact metrics were substantially represented, with 86% of libraries supporting metrics such as Impact Factor and H Indices for researchers. Table 33 and Figure 8 also reveals that more than 50% of libraries are supporting the use of altmetrics and online profile management.

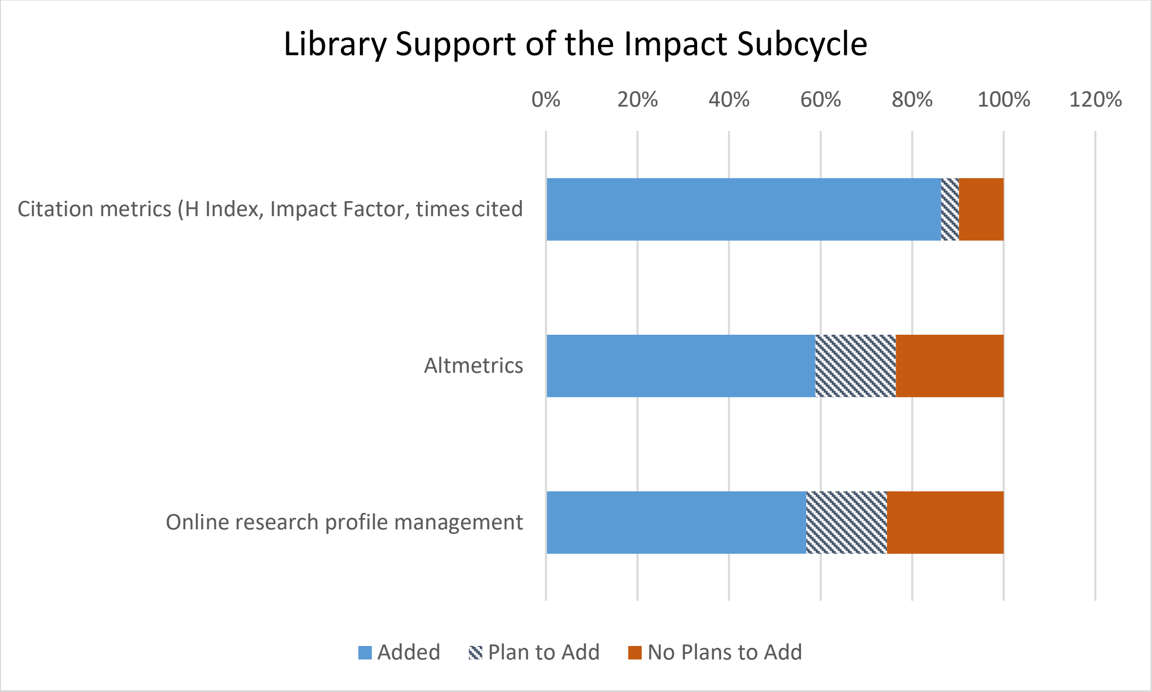
Table 33, (N=51)

Impact subcycle: services provided or plan to add

Cycle Activity	Provide		Plan to add		No plans to provide	
	N	Mean	N	Mean	N	Mean
Citation metrics (H Index, Impact Factor, times cited	44	86	2	4	5	10

Altmetrics	30	59	9	18	12	24
Online research profile management	29	57	9	18	13	25

Figure 8. Library support of the impact subcycle.



<u>Services provided by libraries</u>. More than 60% of libraries have been providing citation metric services for more than four years, indicating that this is a well-established service for libraries. Table 34 reveals that 29% of libraries have also been providing support for online profile management and 18% of libraries supporting altmetrics during the same timeframe.

Table 34, (N=51)

Impact subcycle: services provided more than four years

	N	Mean
Citation metrics (H Index, Impact Factor, times cited	31	61
Online research profile management	15	29
Altmetrics	9	18

Twenty-nine percent of libraries indicated that they have provided altmetrics support services, and 16% providing citation metric support for more than two years. Online research profile management support also included 14% of libraries support for researcher profiles (Table 35).

Table 35, (N=51)

Have provided more than 2 years

	N	Mean
Citation metrics (H Index, Impact Factor, times cited	8	16
Altmetrics	15	29
Online research profile management	7	14

Emerging trends. A significant number of libraries have provided impact services for more than two years with 10% or more of libraries adding these services in the last two years (Table 36).

Table 36, (N=51)

Impact subcycle: services added in last 2 years

	N	Mean
Online research profile management	7	14
Altmetrics	6	12
Citation metrics (H Index, Impact Factor, times cited	5	10

With the exception of citation metrics, this trend is likely to continue for online profile management and altmetric support. Table 37 shows that more than 10% of libraries plan to add these services within the next 2 years. The low percentage of libraries planning to provide citation metric services is likely due to already high rates of adoption for these services, as these appear to already be established services in libraries.

Table 37, (N=51)

Impact subcycle: services planning to add in the next two years

	N	Mean

Online research profile management	6	12
Altmetrics	8	16
Citation metrics (H Index, Impact Factor, times cited	2	4

Few libraries plan to expand these services over the next four years with 6% or less of libraries planning to add online profile management and altmetric services and no additional libraries planning to add citation metric support (Table 38). The reduction in the number is understandable given that almost 75% of libraries are expected to be providing these services within the next four years.

Table 38, (N=51)
Impact subcycle: services planning to add in the next four years

	N	Mean
Online research profile management	3	6
Altmetrics	1	2
Citation metrics (H Index, Impact Factor, times cited	0	0

Table 39 reveals that almost a quarter of libraries do not plan to add online profile management and altmetric services. 10% of libraries do not see themselves as providing citation metrics services.

Table 39, (N=51)
Impact subcycle: no plans to add

	N	Mean
Online research profile management	13	26
Altmetrics	12	24
Citation metrics (H Index, Impact Factor, times cited	5	10

Summary. Responses from library leaders demonstrate a significant amount of services for traditional citation metrics and mixed results for altmetrics and online profile management. Sixty percent of libraries have been providing citation metrics for more than four years. Twenty-nine percent of library leaders indicated that they provide support for online profile management and eighteen percent provide support for altmetrics. However, 26% of libraries do not plan to add services for online profile management and 24% do not plan to add services for altmetrics.

Research question 4. What skills and practices might health sciences libraries cultivate in prompting new roles in the research lifecycle? Health sciences library leaders were asked to respond to four open-ended questions that would help identify potential new roles for libraries. Questions included:

- Please list examples of other actions (not previously mentioned in the survey) that your library has taken to accommodate the changing needs of biomedical research.
- Based on your experience, what new skills should health sciences libraries cultivate in order to more fully support biomedical research?
- What additional steps should health sciences libraries be taking in order to meet the changing needs of biomedical research?
- Over the course of your career, how has the role of health sciences libraries changed when it comes to supporting biomedical research?

The questions were designed to elicit leadership's thinking on not only what new roles should be, but also to reveal trends and actions taken by leaders to address changing practices in biomedical research. Codes were established for the open-ended questions from themes that emerged from the data and were then consolidated. Appendix C provides labels to characterize each theme that emerged. As Table 40 reveals, a variety of themes emerged from the data. Of the 50 library leaders who responded to the open-ended questions, *collaborating* with researchers was the most prevalent theme. *Data science and bioinformatics*, *data management*, and increased *knowledge of the research lifecycle* also appeared frequently. Several other themes emerged, including concepts like *open science* and knowledge of *Awareness of organizational context*. Expert searching was also identified as an important role for libraries.

Table 40, *N=50*
Distribution of coded themes from library leader open-ended questions

New Roles	Sum
Collaboration	45
Data science and bioinformatics	36
Data management	32
Knowledge of research lifecycle	25
Statistics and Methodology	22
Open science	20
Expert searching	19
Awareness of organizational context	18
Domain expertise	15
Librarian training and continuing education	15
Outreach	14
Scholarly impact and online profile management	14
Scholarly communication	10
Grants	8
Knowledge dissemination	7
Information Architecture	6
Space	4
Writing	4
Author Rights and Copyright	3
Collections	3
Compliance	3
IRB/IACUC	3
Virtual and augmented reality	3
Entrepreneurship and innovation	2

Collaboration. Library leaders' description of collaboration included partnering with individual researchers, research teams, offices of the vice president for research, Clinical Translational Science Institutes, and offices of innovation. As one library director saw it, "Health sciences librarians can now be deeply engaged with biomedical researchers as essential partners. Librarians can provide resources and tools to inform and augment the work being done. This shift expands the role of the library: it is both part of the information infrastructure and part of the collaboration framework that is at the heart of the research enterprise." Leaders described collaborative activities as engaging researchers and research support offices to develop, build, and expand library services that supported the research lifecycle.

When speaking of library collaborations many leaders spoke of the need to develop partnerships. "We have always supported biomedical research, previously by the provision of research content and assistance with literature searching. Now, this has evolved to more of a partnership." A different library director went further, stating outright that libraries, "should be working in partnership, not service with faculty, and be seen as integral parts of the research team." Partnerships with researchers was seen by one director as a way to extend the role of her library in creating new knowledge. She stated, " I also believe that libraries have an opportunity more than ever to be part of the 'knowledge creation' process through strategic partnerships, new service models, and our ever-evolving resources."

Many library leaders noted that libraries should actively engage with members of the research community. A theme to emerge was one of evolution through engagement and integration with the research community, but some leaders indicated that it might make some uncomfortable. One director stated that for libraries to create meaningful relationships with researchers they need to look for ways to stretch themselves, even if it required some risk. She went on to add that, "In a world that is going to rapidly transform into machine learning, the library has to position itself downstream and upstream in the knowledge creation process." Another stated simply that "collaboration is key". Library leaders felt responsible for not only managing their organizations, but ensuring their relevancy for years to come.

Data. When combined, data science, bioinformatics, and data management were mentioned sixty-eight times as roles libraries might wish to cultivate in supporting the research lifecycle. Desirable skills for libraries to cultivate included data visualization, wrangling, curation, and discovery. Many respondents listed data management as a general concept with no additional details on the role they viewed libraries providing. However, several participants outlined specific services for data management that included helping researchers to describe, organize, and deposit their data. In many cases, electronic lab notebooks were mentioned as an area to develop library support. One library director believed that the future of libraries entailed helping researchers archive and tag data, as well as teaching them about standards. Others saw libraries helping to make data discoverable, either through data catalogs or by assisting them in depositing their data into institutional repositories.

Several library leaders referred to roles for libraries in assisting researchers with the data management plans required for grant applications. One director summed up the value of creating a position to support data management:

> Just over five years ago, we added a position for research data management. While the RDM person developed resources and tools for researchers to use, consulted with researchers, and taught some basics, she also worked closely with the Office of Research and Innovation on developing and update policies on research data. I think the work in university policy development has been one of her greatest accomplishments.

Other libraries had also established roles or imagined roles in assisting their researchers and institutions in establishing data management policies. Data management services were sometimes seen as a way for libraries to expand and utilize traditional library skills in managing information.

Statistics and methodology. The knowledge of statistics and methodology emerged as important skills for librarians. None of the participants expressly stated that librarians should become the primary provider of these services, though several mentioned providing some level of service or collaborating with other departments to deliver statistical and methodological services. One library contracted for 25% of a bioinformaticist time to provide consultation, education, and training services on behalf of the library. The same library also established a partnership with its Public Health Sciences department to deliver on demand services for statistical and research inquires. This library also hired two full-time data specialists to bolster its ability to support the research lifecycle.

REDCap was mentioned as a tool supported in collaboration with other departments. One director stated, "Our institution collaborates with (a) few other institutions as part of the CTSA. The library supports REDCap (Research Electronic Data Capture) survey builder. We also promote the REDCap Shared Library, a repository for REDCap data collection instruments and forms that can be downloaded and used by researchers at REDCap partner institutions."

Generally, library leaders reported that greater knowledge and awareness of statistics and methodologies would enhance librarian's ability to support the research lifecycle. Statistical literacy was seen as a way for libraries to connect with other aspects of the research lifecycle. A library director stated simply that librarians with a research background that included statistics and study design were better equipped to help biomedical researchers. She explained:

> Librarians who have a research background (published) and who have an MPH degree understand epidemiology - a course that delves in study design and levels of evidence. This qualification makes it easier to teach Systematic Reviews, Study Design, and support research in-depth. Data management is also important as it requires one to understand statistics, which is a required course for the MPH and helps one to better synthesize and evaluate literature.

Another library director connected knowledge and skills with statistics and methodologies as important factors for libraries as they adjust to other emerging practices in science. In his view:

> With changes to federal grant requirements, which trickle down to other types of grants, the need to share biomedical research in a more open and accessible way is growing. We also are being trusted more to provide research services, and be a part of the work being done. This requires us to grow our own knowledge sets, and learn a vocabulary that is often unfamiliar to us at the outset. We are also being asked to write and provide methodologies for research more, which we haven't seen as much in the past. I hope to see librarians being included as authors on more biomedical research, given the amount of work they provide to the research process, with hope that we can see librarians being integrated and embedded, and encouraged to pursue additional education for roles within biomedical research departments outside of the library.

Thus, librarians with skills in statistics and methodologies might demonstrate value to biomedical researchers and lead to deeper collaborations.

Knowledge of the research lifecycle. The open-ended questions support that to gain knowledge of the research lifecycle, libraries and librarians will need to develop a better understanding of the needs of biomedical researchers. Several library leaders relayed that libraries should not, “be afraid to step out of their comfort zone; Continue to listen to user needs and learn where researchers are encountering problems.” An associate director stated that seeing the world from the researcher’s perspective was crucial and that “Half the battle is getting to the table and offering a tangible, useful service to them.” Summing up this dramatic shift in the profession one director stated:

> Having entered the field when it was highly clerical in nature, it has been extremely exciting to see how the contributions have changed and grown. From a time when it was paramount that you had the journals shelved as quickly as possible after photocopying to this current world of information without boundaries has truly been a wonder. But, it has also been a disruptive period and I expect that this will continue with things library IBM Watson and other technologies. For academic health sciences libraries to do more than at a minimum exist, and more importantly to thrive, they have to embrace knowledge in something more than traditional containers of books and journals.

Librarian training and continuing education. Training and education for librarians was suggested as a key factor in positioning libraries to meet emerging and future needs of the research lifecycle. Library directors listed training in data science, metadata/ontology, and the research lifecycle in general as important. A few participants believed that library schools did not adequately prepare new professionals entering the field to fully support research needs. One library director believed that libraries needed to find their purpose in the research lifecycle stating the needs as, “Understanding research workflow, data science skills, ontology knowledge, understanding of team science and where the librarian fits in! I feel like lots of education is needed. Many librarians lack the knowledge, skills, and confidence to step into these areas and offer assistance.” The concept of understanding where libraries “fit in” was expressed by several participants indicating that understanding the needs of researchers and organizations is crucial to developing relevant services at individual institutions.

Awareness of organizational context and outreach. Library leaders communicated consistently the importance for libraries in understanding the needs of the institutions they serve. Most participants who described *Awareness of organizational* *context* mentioned the need to identify support gaps and to develop services in those areas. Many expressed that these gaps can only be discovered though understanding of the research lifecycle. One director articulated his view:

> Health sciences libraries are not the lead entity in almost any aspect of the biomedical research enterprise. But they can be an integral partner and facilitator. This requires a deep understanding of the needs of partner communities and being adept at collaborating. The library should look to add value to the research enterprise; finding the sweet spot in the collaboration matrix where it can make the research process more efficient and effective. This alchemy is going to look different and play out differently at each institution. Because of resource constraints and allocation issues, the library's role in this is generally going to skew heavily toward the consulting, teaching, and training side, rather than the actual doing. This heightens the need for highly effective collaboration.

Health sciences libraries will never be able to provide the depth of services that all labs require, but they can play an important role on building foundational skills that create efficiencies in research workflows.

Several of the participants stated by understanding the resources provided by other units, libraries could develop innovative services that helped the entire organization. Some libraries identified partnering with other departments as a way of delivering needed services without creating redundant services. Areas individual library leaders had identified at their institutions included data management, bioinformatics software, and bibliometrics. Innovation was highlighted as an important factor that could extend libraries ability to support researchers. A library leader declared, “The role of health sciences libraries changes as library leaders are able to illustrate new services of value to their organizations. This is critical. New innovative services need the support of a library leader champion to take the risks and navigate politically to put the services in place and to engage other leaders who value these services to make them effective in the organization.” Several participants indicated that the role of library leaders was to gain acceptance from researchers and administrators of the concept that libraries were the organization to provide innovative services.

Coupled with the need to understand awareness of organizational context, library leaders spoke of the importance of marketing library services, especially around those that support emerging practices in research. An associate director asserted:

> A large part of the effort will depend on serious outreach. So often researchers (and clinicians) do not realize all of the services and expertise the librarians can offer. We are often dependent on word of mouth from one researcher to another. All the data skills in the world are useless if we don't effectively get the word out to our faculty. Of course, librarians need to continue to learn, train, develop new skills that will be of use to the research goals of the institution.

A library director reiterated this sentiment, and saw the value of integrating the library within strategic conversations that occur at her institution. She declared that libraries needed to get in on the “ground floor” on institutional conversations about data, while training library personal in data science skills. Only then, could libraries market the value of the library when it came to data services.

Domain expertise. Library leaders identified recruitment of library personal with domain knowledge as a way to extend libraries’ ability to support the research lifecycle. As one library director stated, libraries need to, “look beyond the MLS degree and cultivate a diverse group of professionals to meet research support demands.” Specific skills mentioned by participants included research methodologies, statistics, science background, information technology, computing, informatics, data management. Background experience or training in public health and epidemiology were seen as desirable skills to compliment traditional library skills. The value of introducing domain expertise was described by a library director:

> Hiring staff with the skills to provide meaningful assistance to biomedical researchers has been critical for making the leap. Hiring a Biomedical Research Support Librarian who understands the work of this community has fostered excellent and successful collaborations, and has identified opportunities for further growth in this area. For example, we have learned that even though there are research cores on campus, a backlog of researchers needing data analysis has arisen. Libraries are poised to help fill this gap by providing tools and services that enable researchers to analyze their data themselves.

A library director whose primary background was in science declared that, "Throughout my career I considered myself a scientist first and a librarian second, and would engage scientists in better understanding their work. This enabled me to introduce information solutions that made sense, and earned great respect for the library." Her experience and knowledge as a scientist not only allowed her to establish deeper connections with researchers, but to coach her staff on how to connect with scientist more effectively. She went on to add that the combination of domain expertise and coaching of her librarians has, "opened countless doors and earned trust and respect across campus." In her view, "Librarians should be prepared to change and library leaders to lead change."

Scholarly impact and online profile management. Several of the participants mentioned that they had established or saw future roles in supporting scholarly impact and researcher profile management. Many leaders felt that library's traditional knowledge of literature and databases uniquely positions them to support other areas that use publication data. Bibliometrics and impact, in particular, were suggested as important areas for libraries to extend services that support the research lifecycle. One director stated that her library's emerging role in supporting impact has allowed the library to, "harness institutional publication data to capture institutional strengths and potential collaboration opportunities."

Expert searching. Despite emerging needs in biomedical research driving libraries to develop new roles, the need for expert searching was often cited by leaders as an important service provided by libraries. Many leaders expressed that the role of expert searching was evolving along with other aspects or research. An associate director explained that she believes expert searching is changing along with its connection to the research lifecycle:

> Library support has become much more sophisticated. It used to be that librarians spent a great deal of time sitting at a reference desk doing simple literature searches. Now, the role has evolved into that of supporting the entire research life-cycle. Literature searches that come to the librarian are not simple searches any longer; they are for systematic reviews and clinical guidelines and grants. Librarians are seen as co-investigators on these types of research teams and are as such co-authors in professional medical journals.

Several leaders noted they have expanded their capability to conduct systematic reviews. Leaders also explained that the nature of search requests have become more complicated over time, allowing librarians to demonstrate value to researchers. Many leaders explained that the expansion of literature search services has lead them to be seen as partners and colleague with researchers. These partnerships have allowed them to co-author papers, led to librarians being written into grants, and allowed them to connect researchers with other services they provide that support the research lifecycle. These services include data management plans, data management, citation management, and publication selection.

Open science and scholarly communication. The utility of educating researchers and creating awareness of open science and scholarly communication were frequently referenced by survey participants. These concepts included building practices that supported open access, institutional repositories, reproducibility, data archiving and discovery, and public access policies. Educational programming and assisting individual researchers in disseminating the results of their work were the most frequently referenced practices that libraries could cultivate.

***Summary*.** Library leaders strongly believed that *awareness of organizational context* and *collaboration* were the keys to libraries developing and cultivating new roles that support biomedical research. *Collaboration* was seen as a way for libraries to increase their knowledge of the research lifecycle and for researchers to become familiar with the services libraries provide. Data was identified as a continued trend and leaders indicated that libraries should develop skills for data visualization, wrangling, curation, and discovery. Increased knowledge of *statistical methods* and *methodology* was noted helpful for developing a better knowledge of the biomedical research process. Participants suggested that librarian training and continuing education was crucial in helping libraries evolve their services. Additional roles and practices leaders identified included scholarly impact, online profile management, scholarly communication, and open science. Leaders suggested that expert searching will remain a continued need and that search requests are becoming increasingly complicated. Leaders also believer that libraries should have roles in supporting open access, institutional repositories, reproducibility, data archiving and discovery, and public access policies.

CHAPTER 5

DISCUSSION

This study explored the changing demands on health sciences libraries created through evolutions in biomedical research workflows and investigated avenues for libraries to expand their role in the research lifecycle. Biomedical researchers were asked in interviews to describe the key activities needed to conduct research and to identify emerging practices in science. Library leaders were surveyed to identify services their libraries provided in support of the research lifecycle and to describe the skills and practices libraries should cultivate for the development of new service models designed to support research. The research questions explored potential gaps in research support services provided by libraries. The study investigated the following research questions:

1. What are the key activities in the research lifecycle for biomedical researchers?
2. What aspects of the research lifecycle are evolving, if any, due to emerging practices in biomedical research?
3. In what ways do health sciences libraries support the research lifecycle and emerging practices of biomedical research?
4. What skills and practices might health sciences libraries cultivate in prompting new roles in the research lifecycle?

Summary of Findings

Phase one of this study reviewed research lifecycle models discovered through online searches and querying of scholarly literature databases. Only one graphical model was found in the peer-reviewed biomedical literature. Specifically, Vaughan et al. (2013) used concept mapping to identify the major areas and activities of the research lifecycle.

Document analysis of models draws from various representations, which sought to visualize the individual research activities as part of a cycle, usually in the form of a linear sequence or circular graph. For the purposes of this study, individual research activities were reviewed and similar terms consolidated into a single list of which was used in the interview protocol. Terminology was categorized under major themes and grouped into five categories: general themes, research planning subcycle, experimental subcycle, dissemination subcycle, and impact subcycle. The categories represent the basic components of the research lifecycle based on the document analysis and are not meant to imply a linear process. The five categories and research activities were used to inform the way the interviews were conducted and in the development of the survey instrument.

Interview data: biomedical researchers. Of the 17 biomedical researchers interviewed for this study, 41% were female and 59% male. Participant institutions were divided into five groups according to their level of NIH funding to create 5 quintiles. Polarity among the quintiles was sought but not achieved. Twenty-four percent of participants were from the 1st quintile, 59% from the 2nd, 6% from the third, 0% percent from the 4th quintile, and 12% from the 5th quintile. To obtain a contrast of perspective on the research lifecycle, participation was solicited from researchers at different stages of their career. Distribution of participants included 47% established researchers, 12% early-career, and 41% research trainees. Seventy-one percent of the participants' primary research was dry lab and 29% wet lab.

Survey data: library leaders. Forty-four of the 51 health sciences library leaders who responded to the survey were directors. Other leaders who responded to the survey included one deputy director, two associate directors, and four other. Distribution of respondents was relatively balanced when grouping by BRIMR institutional rankings by quintile.

Central findings of key activities in the research lifecycle for biomedical researchers. Biomedical researchers shared different perspectives on the key activities of the research lifecycle. Findings were dependent on the comments gathered from researchers and based on their individual processes when conducting research. Central findings for research question included research funding, support, and the interconnected nature of biomedical research.

Funding primes the research enterprise. Biomedical researchers identified the ability to secure grant funding as a primary driver in sustaining their research enterprise. Established researchers noted that they were comfortable identifying and securing funding opportunities. Early-career researchers and trainees were less confident in their ability to locate and obtain funding, often expressing frustration in navigating funder websites and identifying funding opportunities. Researchers noted the importance of professional reputation in submitting competitive grant applications. Publication in scholarly literature was viewed as the primary activity in establishing or maintaining a researcher's professional reputation. Citation metrics, such as H index and journal impact factor, were noted as important metrics in illustrating professional reputation.

Research requires support. For researchers, managing a project was analogous to operating a business. Participants described the substantial amount of support needed to manage their operations. Support was provided by trainees, grant administrators, biostatisticians, librarians, and others. All 17 respondents identified support as a critical aspect of the research lifecycle. All participants also described the significance of the mentor-mentee relationship. Mentors relied on trainees to complete the activities of biomedical research. Mentees depended on mentors to train them to become successful researchers and to help them establish their professional reputation.

Non-library support assisted established and early-career researchers in managing research operations. Participants cited grant administrators as the most likely resource for project management support. *Non-library support* was frequently connected to proposal development, collection and analysis of data, and statistics. Established researchers described their own understanding of statistics as adequate or more than adequate, but also described the value of biostatisticians for establishing and understanding the result of complex designs. Early-career researchers and trainees described a higher frequency of need for statistical support. Trainees relied on *non-library support* to refine their research ideas, enhance their professional reputation, and leverage services they need to complete their research.

Library support was most frequently connected to the themes *literature searching*, *systematic reviews*, *biosketch creation*, *citation metrics*, *data analysis*, and *collaboration.* When asked to describe their comfort level in searching the publication literature, all of the researchers interviewed described comfort in performing standard searches themselves. Support for systematic reviews and other complex searches were recognized as valuable services provided by health sciences libraries. Researchers indicated that they used library literature support services for three primary reasons; librarian expertise, grant application support, and education for research trainees.

There are no subcycles. For the purpose of this study, it was convenient to arrange activities into categorical subcycles. However, researcher depiction of the research lifecycle did not fit within these artificial descriptions. Researchers often described working in multiple subcycles at the same time. The experimental subcycle contained most of the research activity, but researchers were often focused on how their current research project might assist them in obtaining future funding. Many researchers referred to pilot experiments they conducted during a research project as a way to generate preliminary data they could use in the future grant applications.

In some cases, researchers conducted an activity for a purpose outside of the predefined subcycle. For example, attending conferences was categorized under the dissemination subcycle, but 12 of the 17 participants stated that attending conferences helped them to discover collaborators. Other aspects of the research lifecycle were difficult to distinguish one activity from another. *Data collection, data management,* and *data analysis* were characterized as a continuous process and not described as mutually exclusive activities.

Central finding of the key aspects in the research lifecycle evolving due to emerging practices in biomedical research. Changing demands from funders, rapid changes in technology, and trends for open access to the products of research are altering practices in biomedical research. Concepts of openness and data science techniques used to analyze data are affecting the practice of biomedical research. Social media and methods to quantify its value are beginning to emerge as considerations for researchers who seek to describe their contributions to science.

Openness and data science are changing culture. In summary, when asked about emerging practices in biomedical research, *data analysis*, *open access*, and *open source software* were the most frequently referenced themes in the research lifecycle. These emerging trends were also described as generational differences by some established and early-career researchers who noted differences in the practice of biomedical research today from when they trained.

Most researchers acknowledged that trends for openness in biomedical research were rising. The concept of openness was most frequently connected to open access for publications and data, but also included *reproducibility and replicability*. *Reproducibility and replicability* was an emergent theme as researchers begun to consider how to unify data, software analytic code, and findings online. Some researchers noted that data science is a new discipline that can assist with the use of open datasets. Open data was described as lacking adequately metadata and documentation, and may need cleaning prior to analysis. Data science techniques were described as a method to reduce the need for human intervention and assist with the automating the processes of cleaning, mining, and analysis.

Social media and altmetrics are challenging norms. Researchers viewed *social media* as an emerging platform that allowed them to communicate their findings; yet most did not consider use of *social media* as a critical aspect of the research lifecycle. Many study participants acknowledge the use of *social media* in their work at some level, but only one researcher had embraced it as a significant communication medium. The most frequently mentioned social media platforms were Twitter and Instagram. The primary way these platforms were used included monitoring conference hashtags, sharing interesting results with other researchers, or as a way to stay connected with other lab team members.

Changes in NIH biosketch requirements section C, which asks researchers to describe their contributions to science, was also noted as beginning to affect practices in grant application submissions. Although altmetrics was not highly referenced when it came to funding opportunities, four out of the 17 researchers noted the value of altmetrics in describing the impact of recent publications that had yet to receive a significant number of citations by other researchers.

Central findings of health sciences libraries support for the research lifecycle and emerging practices of biomedical research

Survey results revealed a wide array of services and perspectives on the research lifecycle from library leaders. Central findings illustrates strong support in the research planning, dissemination, and impact subcycles. Weaker integration of library services was uncovered for support of the experimental phase of research. Results from the survey illustrate how data services are increasing in libraries, but also reveal a division within the profession's support or planned support of research data.

Strong support for planning, dissemination, and impact. Library leaders noted a significant level of support for the research planning subcycle with at least 50% of the respondents stating that they currently provide services for *background literature searching*, *citation management*, *systematic reviews, grey literature searching, locating data sources, methods for organizing and storing information, data management plans, IACUC protocols, NCBI tools (Blast, GenBank, dbSNP, etc), seeking grant funding,* and *identifying collaborators*. Library leaders also reported a significantly number of services supporting the dissemination subcycle, with at least 50% of respondents specifying that services for *journal selection for publication*, *citation metrics*, *author rights and copyright*, *open access*, *bibliographic styles*, *institutional repository*, *funder public access policy compliance*, *web and social media marketing*, *grant citation*, and *presentation poster preparation*.

In some cases, health sciences libraries appeared fragmented on their plans to support some aspects of the research lifecycle. For example, 29% of libraries are providing support for *online profile management*, and 18% of libraries are supporting *altmetrics*. In contrast, 26% percent of library leaders do not plan to provide services for *online research profile management* and 24% do not plan to support *altmetrics*. Despite this data, trends for libraries supporting these services are likely to increase. The data indicated that 18% of libraries plan to add services for *altmetrics and online research profile management*.

Weak support for experimental subcycle. Responses by library leaders illustrate a weak relationship between library services and the experimental subcycle.   Within the experimental subcycle, only *managing research data* and *metadata standards* are provided by more than 50% of libraries. Biomedical researchers did not express needs in managing research data or metadata standards. More than 50% of libraries do not plan to add services for *scientific modeling*, *high performance computing (HPC)*, *electronic lab notebooks*, *statistical methods*, *file versioning*, *code versioning (GitHub, etc)*, *geographic information system (GIS)*, and *automated and manual data collection*.

Division over data services. Survey data illustrates how library support trends for data are increasing; however, many of the services mirror traditional roles of libraries in supporting storage, discovery, and sharing of information. Data services added by libraries in the last two years or services they plan to add in the next two years include *data literacy* (45%), *data catalog* (31%), *data management plans* (31%), *managing research data* (29%), *data sharing (29%)*, *data archiving (27%), , code versioning (GitHub, etc)* (27%), *data wrangling/cleaning using open source tools* (27%), *ontology/taxonomy* (27%), and *automated and manual data collection* (25%).

Survey results also reveal a division within health sciences libraries when it comes to providing data services. Thirty-seven percent of leaders have no plans to provide *data sharing* services and 42% have no plans to provide services for *data archiving*. Additionally, 49% of respondents stated that they had no plans to add services for *data cleaning using open source software tools* and 47% *data cleaning using proprietary software tools*.

Central findings of skills and practices health sciences libraries might cultivate in prompting new roles in the research lifecycle

Analysis of open-ended responses from library leaders on research question four help to illustrate their view on health sciences librarianship. Awareness of the organizational context was identified as a critical component when cultivating new roles for libraries. Several leaders used the open-ended questions as a platform to stress their view on the importance of library continued support for expert searching. Training and education of librarians was viewed as inadequate and in need of improvement.

Organizational context matters. Open-ended responses by library leaders suggested that *awareness of organizational context* was an important consideration for libraries in identifying new roles for libraries. Study participants described collaboration with researchers and other research support departments as the best mechanism for identifying and creating new roles supporting biomedical research. Leaders also stated that librarians needed to increase their *knowledge of the research lifecycle* if libraries are to successfully develop new roles supporting research.

Library leaders believe in expert searching. Library leaders expressed that *expert searching* remains an important role in the way libraries support biomedical research. Seventeen of the 51 respondents affirmed the need for librarian skills in *expert searching*. Many leaders stated that library search services provide value to researchers as service requests become increasingly complicated. Support for systematic reviews was frequently cited as an area where libraries can demonstrate value.

Training and education are needed. Library leaders noted *librarian training and continuing education* as necessary to cultivate new skills and roles for libraries in supporting the research lifecycle. Leaders identified *data science and bioinformatics*, *data management*, *statistics and methodology*, *expert searching*, and *domain expertise* as areas for developing new roles. *Expert searching* is not a new role for libraries, but some leaders felt strongly about maintaining traditional roles for libraries. Desirable data skills for libraries to cultivate included visualization, wrangling, curation, and discovery. Many leaders described roles for libraries in supporting open access, institutional repositories, reproducibility and replicability, data archiving and data discovery, and compliance with public access policies. Leaders identified increased knowledge in *statistics and methodology* as needed skills by librarians. *Scholarly impact and online profile management, scholarly communication,* and *open science* were emergent themes in biomedical research that library leaders also noted as important roles for libraries.

.

Organizational Change

Organizational change theory provides a framework for understanding how health science libraries can continue to support the needs of academic health sciences centers. Organizations, including libraries, need to adapt to the contextual demands placed on them from their organizational environment in order to maintain professional and institutional legitimacy. Lawrence and Lorsch (1967) argue that organizations can achieve differentiation by developing attributes based on requirements posed by external forces. Libraries derive legitimacy from the services they provide to academic health sciences centers, and thus acquire their provision of resources from their ability to align with organizational priorities. According to Salancik and Pfeffer (1974), in times of great change, “subunits will possess relatively more power to the extent they provide resources for the organization and to the extent that the resources provided are critical, important, or valued by the organization.”

Technology is transforming how information is produced and consumed in the research lifecycle. At the same time, rapid shifts, new entrants, and exogenous shocks in information technologies are having a dramatic impact on health sciences libraries. Scott (2007) offers that from a contingency theory perspective, there is no one best organizational structure and that, “suitability is determined by the goodness of fit between the organizational form and the diverse environments to which they relate.”

Libraries need to remain open to changes in their environments, which include shifts in information technologies. Organizations can be more flexible and evolve more rapidly than their scientific disciplines. Similarly, research labs can make faster adjustments that increase their competitiveness. However, it is problematic if research labs and organizations are changing faster than libraries. As Weick (1979, p. 179) contends, successful evolution can occur without any necessary increase in “productivity or viability of the system”. In essence, rather than focusing on trying to make current and traditional services more efficient, libraries should instead consider a shift in the kinds of services that they offer to ones that consider the demands of the external forces being placed on the organization’s biomedical research enterprise.

Interpretation of the Combined Results

Evaluating the data on the role of libraries in the research lifecycle collected from biomedical researchers compared to survey data from library leaders provides insight into how researchers and libraries view the research lifecycle. Biomedical researcher’s motivation was driven by the ability to obtain and sustain funding. This process relied on their ability to generate significant findings, disseminate those findings in high impact journals, and describe how their previous research makes them competitive for additional funding. For libraries, supporting the research lifecycle was largely connected to traditional services of collecting and accessing knowledge in the form of publication literature and, in some cases, new forms of knowledge including data.

Navigating the research lifecycle enabled researchers to successfully accomplish the tasks of their current research projects and remain competitive in the acquisition of future finding opportunities. This study found that almost all activities within the research lifecycle were connected to other research activities. Support was identified as a critical element of the research lifecycle for all participants in this study. Support mechanisms, which include research trainees, libraries, biostatisticians, and administrative support staff, are woven throughout the entire research lifecycle. Researchers depend on this support to sustain the research enterprise while they prepare for and write the next round of grant applications.

Researchers participating in the study described using the library primarily by accessing its collections. As might be expected, researchers’ most frequent use of assistance from librarians was in the form of literature searching and systematic reviews. However, only nine of the 17 researchers interviewed stated that they used some form of literature searching support. Five of the 17 researchers used the library for assistance with systematic reviews. In contrast, all 51 library leaders stated providing literature search services to researchers, with 47 providing systematic review services and 2 more planning to provide.

This data suggests that libraries are heavily invested in literature search services that researchers believe they can do themselves. Researchers expressed comfort in searching online databases and accessing journal content without the need for assistance. However, those researchers who did utilize library search services were heavy users of the service to support their research. In particular, they valued the service when it came to complex searching and systematic reviews, finding librarian assistance helped them to save time and produce better results. This aligns with the view of some library leaders who stated that the number of searches has gone down, but that the complexity of the request received has increased substantially.

Other library services connected to published literature did not receive significant levels of use from researchers. Forty-seven library leaders reported providing grey literature search services, but no researchers claimed to have used the library for this service. Fifty of the 51 libraries provided services for citation management, but only two researchers reported using the library for citation management support. Three researchers stated that they had used the library for citation metrics with 44 of the libraries providing this service. Researchers also did not state using other traditional library services including *citation management*, *journal selection*, *author rights and copyright*, and *citation styles*. Table 41 provides the number of researchers interviewed who reported using library support for activities within the research lifecycle. In sum, these data suggest that library service models are inadequately supporting biomedical research.

Table 41
Frequency of researchers who reported using library support

	Researchers references of library support
Literature searching	9
Mentor/Mentee	7
Systematic Review	5
Collaborating	4
Data analysis	4
Citation metrics	3
Data management	3
Emerging practice	3
Journal selection	3
Open source software	3
Proprietary software	3
Biosketch	2
Citation management	2
Data preservation	2
Grant funding	2
IRB/IACUC	2
Open access	2
Organizing and storing information	2
Statistical Methods	2
Writing	2
Altmetrics	1
Author rights and copyright	1
Data collection	1
Data literacy	1

Influencing science	1
Preprint	1
Presentation	1
Reproducibility and replicability	1

There were no references to library support: *Attending conferences, bioinformatics tools, citation styles, competition, compliance, conference selection, data privacy and security, ethics, grey literature, identify collaborators, lab notebook, licensing and venture, metadata, methodology, non-library support, pilot experiment, project management, social media*.

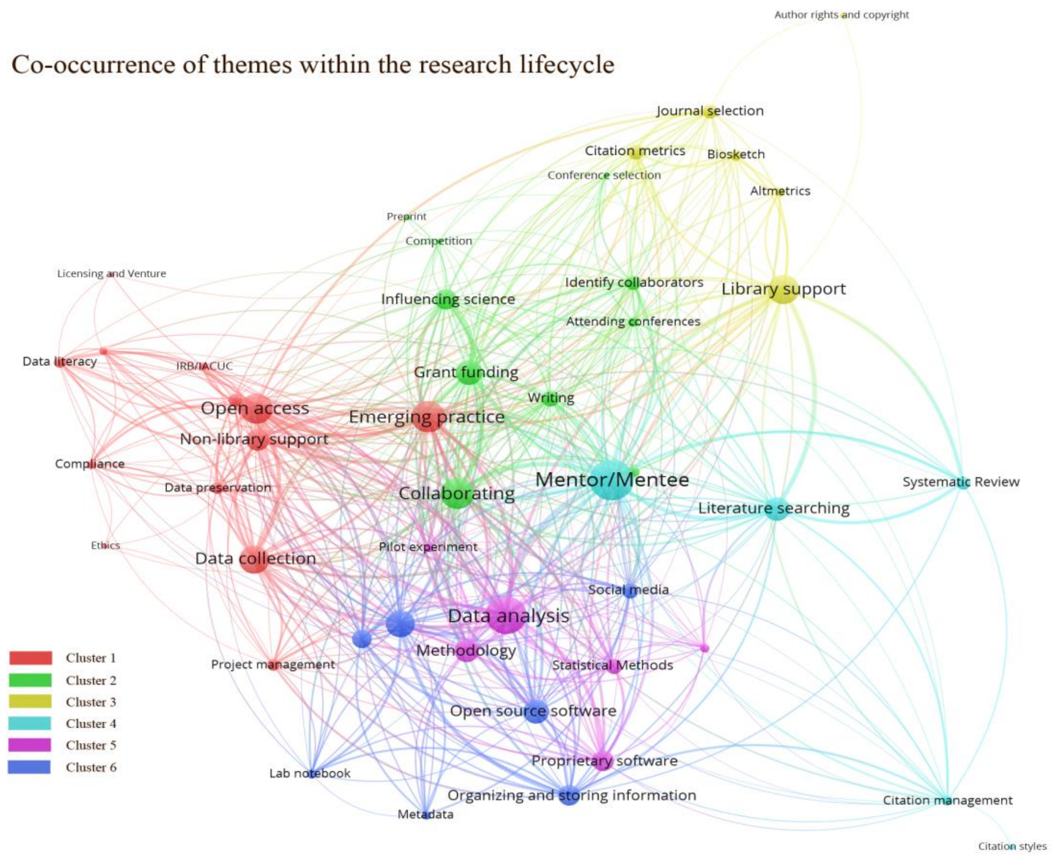
Relationship of Activities within the Research Lifecycle

Data collected from the biomedical researchers interviewed for this study did not reveal a linear or circular process as depicted in most research lifecycle models. Rather, researchers described an interconnected system of activities that helped sustain their research enterprise and enabled the acquisition of future funding. Library support of biomedical researchers did not align with the key activities in the research lifecycle described by researchers, illustrating the gap between library services and the needs of the biomedical researchers.

Co-occurrence of themes gathered from biomedical researchers provides evidence for the connection between activities within the research lifecycle and highlights the placement of library services within the lifecycle. Co-occurrence data were analyzed through the creation of a graph database to better understand the relationship between activities. Graph databases help to visualize patterns and relationships in data that are hard distinguish in numeric or string values. The software VOSviewer, originally developed to visualize bibliometric networks, was used to explore relationships among the themes that occurred within co-occurrence data collected from biomedical researchers.

VOSviewer calculates the associated strength between data objects by using a similarity measure (Van Eck & Waltman, 2010). Also referred to as probabilistic affinity, similarity measures display how alike two or more data objects are to each other. Figure 9 shows the relationship of co-occurrence data from the coded themes as described by biomedical researchers. In the visualization, the coded themes are represented by nodes, the circular disc attached to each theme. The size of the node indicates the number of times the theme was coded in the interview data. Edges are the lines representing the co-occurrence that occurred between two themes. The size of the edges indicates the strength of the connection between the two themes based on frequency data. The colors indicate clusters of alike themes that emerged from the data.

Figure 9. Co-occurrence of themes within the research lifecycle



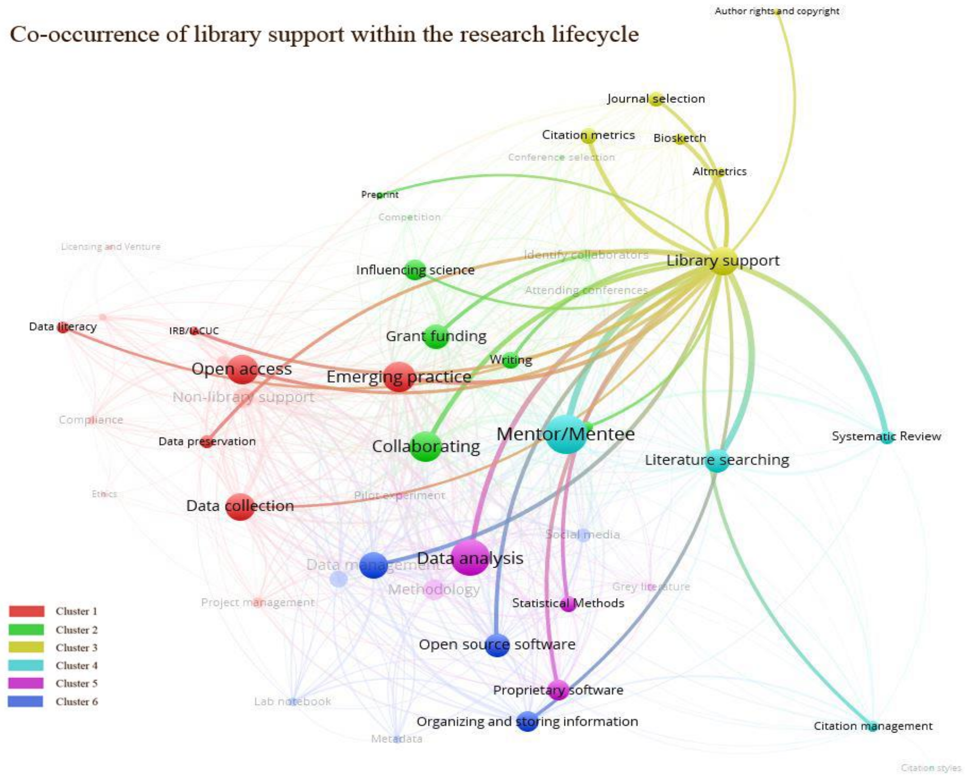
The co-occurrence data displayed in Figure 9 reveals a number of interesting relationships when considering library support and the research lifecycle. Clusters 1, 2, 5, and 6 demonstrate a close relationship among the themes that can be seen by the number of edges between nodes and by their close proximity. Cluster 3 contains the library support node and its large size suggests that it is one of the major themes expressed by researchers; however, the location of the node is on the outside border of the visualization and far away from the high density of edges located in the center of the visualization. Similarly, in cluster 4 only the mentor/mentee theme is located in the high density of edges in the center of the visualization. Most of cluster 4, which contains the traditional library services of literature searching, systematic reviews, and citation management, resides primarily on the outside border of the visualization.

Edge strength between the library support nodes and traditional library services of literature searching and systematic reviews is significant. This is due to the high number of references of library support in the data. However, VOSviewer's association strength measure normalizes the data and corrects for differences in the total number of occurrences and co-occurrences of items. This helps to explain why *library support* does not cluster with *literature searching* or *systematic reviews*, despite having a high frequency of references.

Concerning for libraries, *library support* is clustered with activities not highly used by the researchers. The visualization suggests that the centrality of library services is connected to *citation metrics*, *biosketch*, *journal selection*, and *altmetrics*. Cluster 3 and the *library support* node is the furthest from the center of the graph and remains on the fringe of the high-density activities of the research lifecycle. The central node for cluster 4 is the *mentor/mentee* theme and is clustered with *literature searching*, *systematic reviews*, and *citation management*. This is likely because trainees mentioned searching the literature as part of the research lifecycle and did not significantly connect these themes with library support.

However, Figure 9 does not tell the complete story about library support of the research lifecycle. Data provided from researchers about the library services they use in support of the research lifecycle reveals that they use a variety of services. The interview data also suggests that some library support services are used heavily by some researchers more than others. For example, while nine out of the 17 researchers referenced library support for literature searching and systematic review, there still was a significant number of researchers (8) who did not connect library support with literature searching. Further, one researcher's interview accounted for 43% of the 28 co-occurring *literature searching* codes. This helps to explain why some researchers expressed a high degree of value for library services that did not emerge in data collected from other researchers. Yet, library services are well represented throughout the research lifecycle. Isolating data for just *library support,* Figure 10 provides a view of the breadth services provided by libraries. The lack of centrality of library services among the other themes suggest that researchers use libraries for a variety of independent activities, but that they do not connect *library services* with other activities in the research lifecycle.

Figure 10. Co-occurrence of library support within the research lifecycle



Implications

This study presents two professions in the midst of evolutionary change. New technologies not only allow knowledge to be created digitally, but also influence knowledge creation. David and Foray (2002) state that there have always been entities efficient at creating new knowledge, but that the, “crux of the issue lies in the accelerating (and unprecedented) speed at which knowledge is created, accumulated and, most probably, will depreciate.” Interviews of biomedical researchers provided a diverse range of opinions on emerging practices in the research lifecycle. Researchers were acutely aware of how advances in technology and concepts of openness are affecting research, regardless of career status, BRMIR rankings, or type of science. Some of the researchers interviewed, are witnessing changes as they occur. One established researcher stated:

> No, it's not remotely the same as it was 10 years ago. No way. I think now – in fact, I think it's kind of cool in the sense that the rules – many of the former rules don't exist and so you can kinda make up the rules, a new set of rules for the technology that exists. But I think one of the things that's a big challenge is that it's moved so fast and it's so quick that unless you're really dedicated and more importantly, disciplined to staying on point, it's hard. And I think that it's difficult for people to find their place in the community.

Another established researcher saw a segmentation within the biomedical workforce. In the past, he stated that scientists were more, "jack of all trades" but that today they need to be more focused. As a result, researchers' skills are more specialized than in the past, requiring them to work more with others to accomplish their tasks. Rapid changes are disruptive and can create uncertainty within organizations, but they can also be opportunities for organizational alignment. Libraries should consider how rapid shifts, new entrants, and exogenous shocks in information technologies are affecting not just libraries, but the biomedical researchers they support.

Knowledge creation and production. Tyler Walters (2013), Dean of University Libraries at Virginia Tech, argues that the process for knowledge creation and production (KCP) will increasingly impact how researchers interact with each other and with information scientists. Further, Walters adds that, "the external forces on and subsequent changes to KCP at leading research universities are forcing institutions of higher education to adapt in order to maintain and advance their positions as major knowledge producers. In turn, the services supporting KCP, such as producing, managing, disseminating, and preserving knowledge, must adjust as well (p. 5)." Health science libraries should view unprecedented evolutions in biomedical research as unprecedented opportunities to more deeply integrate support within the research lifecycle.

Validation from the NIH strategic vision for NLM. Findings from this study are validated by the 2015 NIH strategic vision for the NLM. The long-term vision, "calls for NIH to position the NLM as a unifying force in biomedicine that promotes and accelerates knowledge generation, dissemination and understanding in the United States and internationally." The vision also calls on NLM to become the epicenter for biomedical data science research enterprise ("NIH approves strategic vision to transform National Library of Medicine | National Institutes of Health (NIH)," 2015)[5]. Health sciences libraries should investigate ways to become the epicenter for biomedical data at their institutions. Funding opportunities will increasingly contain principles of open access and reproducibility when it comes to data. Libraries are uniquely positioned to develop expertise, services, and educational workshops in these areas. If libraries can position themselves as active partners supporting principles in funding announcements, they can help researchers at their institutions strengthen their grant applications and strengthen their relevancy.

To accomplish these changes, libraries may be able to repurpose or train staff into new roles supporting knowledge, but they need to be open to other possibilities. Many health sciences librarians lack the disciplinary background or computational skills needed to create impactful services supporting emerging practices of biomedical science.

---

[5] The NIH strategic vison calls on NLM to, "Lead efforts to support and catalyze open science, data sharing, and research reproducibility, striving to promote the concept that biomedical information and its transparent analysis are public goods." and "Be the intellectual and programmatic epicenter for data science at NIH, including becoming the center of intellectual and programmatic activities in biomedical data science, and stimulate its advancement throughout biomedical research and application."

Additionally, library leaders indicated that library schools were inadequately preparing health sciences librarians emerging into the filed.

Revisiting the research lifecycle. This study found that library services align with activities that occur at the beginning and end of the research lifecycle. The five areas of support identified by Vaughan et al. (2013) were Idea Development, Funding, Proposal, Conducting, and Disseminating include a variety of support activities based fundamentally on traditional library services. As identified through the biomedical researcher interviews, biomedical research is increasingly complex and many traditional library services will continue to support key aspects of the research lifecycle. The model provided by Vaughan et al. provides a primer for how these services align with the research lifecycle. However, when considering the rapid evolving environment of biomedical research and the consumerization of information technology this study proposes that traditional-based service models are insufficient to meet the current and emerging needs of researchers.

To gain greater awareness of the research lifecycle, libraries might further investigate adopting the UCF research lifecycle model identified in Chapter 2. The model uncovered the processes and activities of the research lifecycle and the institutional departments who support it. The model also identified several gaps in support services, including data curation, data sharing, data visualization, analysis support, and long-term preservation. Health sciences libraries could look to this model and develop strategies to increase their awareness of organizational needs and value they might demonstrate in supporting the research lifecycle. This would not only allow libraries to identify gaps in library support, but could help them establish new partnerships within their organizations.

Reimagining the skills, characteristics, and qualifications of librarians. Health sciences libraries need to reimage the skillset, characteristics, and qualifications needed to support new forms of knowledge creation. Traditional academic tracks for research faculty are becoming increasingly competitive at universities, potentially creating a scientifically trained workforce interested in research support roles at health sciences libraries. Libraries should think strategically about their long-term goals and utilize staff attrition as opportunities to question the need for traditional roles that will no longer be needed in the future. Instead, libraries should recruit professionals with skills needed to support knowledge creation and production to form a new organizational and strategic vision for research support.

In the survey, a number of library leaders viewed library school training as inadequate in preparing health sciences librarians to support biomedical research. However, it is worth noting the value of professionally trained and credentialed librarians. Biomedical research training tends to be domain specific and training can focus on highly specialize techniques. Some libraries have unbundled the librarian skillset by recruiting and hiring biomedical researchers to assist in the development of library services. However, they should also understand that doing so will not enable libraries to provide the breadth of current library services. One value of library training in its current form is in its generality. Masters of library science prepare librarians with a foundation to provide many types of services and allow them to be flexible in support institutional needs. Unbundling the librarian skillset and allowing other actors to fill the general role librarians creates a potential risk to the legitimacy of libraries. Instead, libraries should consider how recruiting non-librarian staff members to address specific needs and combined with services provided by their librarians.

The training needs of librarians will continue to evolve over time, but will also need to consider obligations to other missions of academic health sciences centers such as clinical and education missions. Library support for other forms of research should also be considered since academic health sciences centers have many unfunded research activities. These research activities support quality improvement projects and evidence based practice. Additionally, clinical residents and medical students may have research requirements or electives.

A final aspect of biomedical research that libraries should consider is the mentor-mentee relationship. This study identified importance of the bi-directional relationship between mentor and mentees. This relationship was important for sustaining the operations of biomedical research, as well as producing the next generation of biomedical researchers. Further, generational differences were described by researchers, even among early career researchers who emerged from training within the last 10 years. The generational differences illustrate how the needs of biomedical research training are evolving and the skills and training needed to make that shift. Libraries were identified as an important contributor to the needs of biomedical research training, but are limited by their ability to integrate within the experimental subcycle. Libraries are a common good within academic health sciences centers and need to develop and provide relevant services that demonstrate value to the organization to maintain legitimacy. Emerging roles for libraries in supporting data science could increase the legitimacy of libraries within their organizations and might help promote the financial case for libraries

The future of the profession. Despite the evidence that rapid shifts in technology and biomedical research are altering norms, it is unclear if libraries are well positioned to meet changing needs. Not all library leaders or biomedical researchers agree on the best course of current and future actions for the profession. When asked what additional steps libraries should be taking in order to meet the changing needs of biomedical research one director stated that, “This question assumes that health sciences libraries should be ‘taking steps’ to meet these ‘changing needs of biomedical research’ I'm not sure I agree with the premise.” She went on to add, “To be perfectly frank, I have many reservations about librarians cultivating new skills for the sole purpose of ‘supporting biomedical research’ and questioned which services supporting research belonged to libraries” Additionally she stated:

> Are librarians administrative support or experts offering consultation services? Many researchers still view librarians as glorified secretaries. I think we walk a very fine line here. Providing advice (consultation) on metadata schemas or file naming conventions is one thing. Actually managing their information files is another altogether. Cleaning their data? Not a librarian job.[6]

An established researcher was also hesitant on the role of libraries in supporting fundamental skills in data management and analysis. In his view:

> Most librarians seem to be one significant step, one large step away from active research of any kind. And in a field where the techniques and data-types are evolving so fast, which is true of most data intensive fields, but it's especially true in biology, I worry that any type of training that… it doesn’t rest primarily on the shoulders of research active people, will be out of date and irrelevant quite quickly.

---

[6] Quotes from library leaders taken from open-ended questions in the survey protocol.

Both of points are valid, dramatic shifts in technology and research practices create ambiguity over the appropriate roles for libraries in biomedical research. However, health sciences libraries should look to their recent past to understand the implications for failing to evolve. In 2008, the economic down turn in the United States led to a large number of hospitals laying off librarians. A survey of hospital administrators concluded that the perceived value of hospital libraries was limited and accreditation were revised focus on the information resources available to physicians rather than access to a degreed librarian (Goldstein & Coletti, 2011).

The director's statement on libraries and administrative support is reminiscent of the struggle the medical librarian profession has had over the legitimacy of librarians as peer professionals. As discussed in Chapter 2, the Medical Library Association was originally a physician led organization from 1898 to 1946. Since the inception of medical libraries, librarians have struggled to establish themselves as professional peers with physicians and medical researchers. However, emerging practices in biomedical research provides a remarkable opportunity for libraries to redefine the profession. One director saw swift changes in biomedical research as a chance for libraries to develop new roles in knowledge creation:

> The library has always played a role in "knowledge dissemination/application/preservation/organization/etc." This will continue to be true, but the way we address it will need to expand, given our technology-driven world and the rapid acceleration of research processes. I also believe that libraries have an opportunity more than ever to be part of the "knowledge creation" process through strategic partnerships, new service models, and our ever-evolving resources.

Another director went further, describing then flexibility librarians will need in the future. She stated:

> The "hard" skills sets such as data software, programming etc will change as the technology and science changes. I am more concerned about the "soft" skill sets needed to be a health sciences librarian. By these I mean things like flexibility, communication, willingness to consider new ways of things, curiosity and openness to challenge traditional ways of doing things, the ability to understand the difference between having a "role" in research support versus the idea of doing "tasks" and working outside individual comfort zones. All too often I see health sciences librarians locked into a mindset of being really good at something that does not need to be done or will soon reach its inevitable sunset (read systematic reviews). I believe that librarians need to hold true to their mission of knowledge support but have the ability to evolve along whatever path that will take us.

Some libraries fixated on preserving their traditional expertise could damage their legitimacy if their service models do not align with organizational needs.

Researcher motivation and organizational context. It is clear that the ambiguity created by emergent trends in biomedical research and evolutions in information technologies is challenging the concept of health sciences libraries. Should libraries do more to market traditional services? What is an appropriate role in providing data services? What training and skills are needed for libraries maintain legitimacy?

Library leaders stressed the importance of becoming more aware of their organizational context, with 15 of the 51 respondents noting the need to align with institutional strategies and services provided by other departments. One director stated that libraries need to “understand your institutions priorities, look for opportunities to create services where they are lacking and/or partner with other units where your joint expertise meets an identified need.” Library leaders saw themselves as the ones responsible for instituting change. *Awareness of organizational context* and knowledge of emerging trends can help generate ideas for evolutionary services, but require a champion to develop a vision. As one director saw it:

> The role of health sciences libraries changes as library leaders are able to illustrate new services of value to their organizations. This is critical. New innovative services need the support of a library leader champion to take the risks and navigate politically to put the services in place and to engage other leaders who value these services to make them effective in the organization.

As a result, libraries can increase their institutional legitimacy by promoting new and innovative service models that engage academic health science center leaders in the value of libraries now and in the future.

Libraries should be cognizant of researcher motivation to obtain more funding and desire to continue their research. If not, they could find their organizations providing services researchers neither want nor need. Libraries would be wise to develop and promote services in ways that align with researchers' need to sustain or establish their research enterprise. To do so, libraries need to alter significantly their organizations in ways that embed their skills and practices deeply within the experimental part of the research lifecycle. Given the rapid rate of change they should act quickly and with urgency. Promoting literature searching and systematic reviews because they leverage librarian's traditional strengths is a flawed approach. This perspective speaks more to the desire of librarians to promote their skills, rather than the motivation and needs of biomedical researchers. Libraries need to understand that they are no longer the gatekeepers of knowledge. Knowledge is pervasive, ubiquitous, and omnipresent.

Limitations

The design of the study and conceptual framework led to several limitations. The first phase of the study consisted of a document analysis of research lifecycle models to generate a framework of questions used in the qualitative interviews with biomedical researchers.   In many cases, research lifecycle models were produced from the perspective of an organizational entity such as a library, research support office, or sponsored programs. It is unknown if biomedical researchers were involved with the construction of the research lifecycle models reviewed or to what extent the models were reflective of the actual research lifecycle of biomedical research.

During the second phase of the study qualitative interviews were conducted with biomedical researchers using a semistructured interview protocol developed from the document analysis. Pezalla, Pettigrew, and Miller-Day describe the qualitative interview process as an exchange between the two parties and note that the, “researcher is the instrument”, and as such, has “the potential to influence the collection of empirical materials (2012).” Since the participants were allowed to deviate from the semistructured interview protocol, and because of time constraints on participant’s time, not all concepts were fully explored.

Marshall and Rossman (1989) state that qualitative approaches to inquiry are uniquely suited to uncovering unexpected and new avenues, and that the research questions need to be general enough to permit exploration, but focused enough to delimit the purpose of the study. The design of the qualitative interviews were intentionally exploratory to capture phenomena of interest from the participants’ perspective. Generalizability of the findings was not the aim of this study, and limits the application of findings. Additionally, selection criteria was designed to incorporate a diverse range of participants and this study sought to achieve parity among the types of researchers who met the selection criteria. However, low participation rates among the categories prevented an even distribution of participants, limiting comparison among the categories.

In the third phase of this project, a quantitative survey was administered to leaders at health sciences libraries. Although similar terminology was used in the survey as in phase two, it is possible that interpretation of the definition of terms might be different between biomedical researchers and library leaders, inadvertently creating bias in the responses from being "correct, honest, or accurate" (Furnham, 1986). It is also possible that selection bias occurred since leaders at health sciences libraries were given a choice to participate in the survey. According to the *Encyclopedia of Survey Research Methods* "To the extent that respondents' propensity for participating in the study is correlated with the substantive topic the researchers are trying to study, there will be self-selection bias in the resulting data." (Lavrakas, 2008, pg. 810)

Data collection, analysis, and interpretation of findings was reliant on the principle investigator of this study. Attempts to mitigate bias and limitations of the study in several ways. During the qualitative interviews, participants were encouraged to respond honestly and use the language relevant to their field. They were also encouraged to describe their actual process navigating the research lifecycle, rather than how they saw libraries supporting it. During the interview, participants were asked to clarify responses heard by the investigator to increase then validity and interpretation of the data collected. The library leader survey was sent to all health sciences library directors who met the selection criteria in an attempt to obtain the largest potential sample. When possible, standard terminology concerning the research lifecycle and familiar to health sciences libraries was used. Leaders were also encouraged to devote a significant amount of time to the open-ended questions to capture as much data as possible about the research lifecycle and changing roles and practices for libraries.

Along with the general limitations of mixed methods research several limitations were identified during the research process. Lack of parity among biomedical researcher participants, time constraints, anomalies in the data, and lack of a common language between researchers and library leaders contributed to this studies limitations.

Although parity among the groups of researchers was sought, participation rates varied especially using the criteria of NIH funding level by institution. Participation within the 3rd – 5th quintiles was significantly lower than the 1st and 2nd quintiles. In retrospect this might be expected as there are fewer research projects funded by NIH in the lower quintiles. Lower participation rates prevented comparisons among the quintiles that might have illustrated differences in the practices of biomedical research, or the support they receive, for researchers by institutional NIH funding levels. An additional gap was also noted in the participation of early career researchers. The selection criteria used awardees of T and F grants, which also have lower funding rates and contributed to the limitations of this study.

Time constraints also contributed to the limitations of this study. The interviewed protocol contained a number of concepts related to the research lifecycle and was designed to last an hour. However, the participants were allowed to deviate from the protocol and encouraged to describe their research process so that emergent themes could be identified. As a result, most interviews did not allow for all probes in the protocol to be fully explored, creating potential gaps in the data collected.

Several anomalies in the data were identified that contributed to the limitations of this study. For example, licensing and venture, ethics, and prototyping are known aspects of the research lifecycle, but not heavily referenced by the researchers who participated in this study. Additionally, several library directors felt strongly about expert searching and voiced their opinion in the open ended questions even though the questions focused on emerging needs and identifying new skills for librarians in supporting the research lifecycle.

Suggestions for Future Research

This study explored the key aspects of the research lifecycle and investigated library's current and anticipated support of biomedical research. Future research with a larger sample that included a diversity of participants by gender, NIH funding levels, career stage, and type of science could expand on what is known about biomedical researchers and the research lifecycle. A larger and more diverse sample would allow for comparisons among groups, and help determine if there are differences among biomedical researcher populations and among the libraries who support them.

Future research could also investigate the value of libraries and librarians supporting the research lifecycle. A previous multisite study focused on the value of library and information services for patient care, but did not address biomedical research (J. Marshall, 2013). Evidence for the value of libraries within the research lifecycle may assist library leaders in making arguments for the allocation of resources or in promoting the legitimacy of libraries within the organization.

Selection criteria for this study did not allow for investigation of other types of research support provided by health sciences libraries. Within academic health sciences centers unfunded research often occurs related to clinical and academic missions. For example, research helps to support evidence-based practice and quality improvement. Additionally, medical students and residents often have research requirements or electives. Investigation of the needs and services provided for these forms of research would be beneficial to understanding the complex needs or research in academic health sciences centers.

Further investigation needs to turn attention to the ways in which one activity in the research lifecycle relates to other activities. The development of an interview protocol that illustrates primary and secondary activities could be used to enhance what is known about the research lifecycle and library support.

More researcher is needed to address what skills and training are best for librarian support of biomedical research. Deeper understanding of the continuing education needs of current staff and library students might highlight potential gaps for libraries to address in supporting the knowledge creation process. Additionally, research might investigate the efficacy of library staff with non-MLIS advanced degrees in supporting biomedical research.

Conclusion

Health sciences libraries have traditionally curated knowledge generated by biomedical research and made it available to the next generation of researchers. For hundreds of years this knowledge was in the form of journals and books. Network technologies are shifting the norms of access to knowledge and challenging the traditional roles of libraries as information gatekeepers. Knowledge is increasingly structured and therefore increasingly accessible with less need for human intervention. In the past, health sciences libraries derived their legitimacy from the stewardship of knowledge resources needed to support academic health sciences centers. Today's networked environments have eliminated the need for a physical place for researchers to access knowledge. However, the same network technologies are creating new challenges by exponentially increasing the ability to create new knowledge digitally. Instead of legitimacy, health sciences libraries should focus on service models that established their relevancy to the organizations they support. To maintain both legitimacy and relevancy within the academic health sciences centers they support, libraries need to adapt quickly to the shifting landscape of academic biomedical research. Failure to do so could have long-term negative implications for the profession.

This research strengthens what is known about the research lifecycle and the support provided by libraries. All health systems are different, and there is no one best system for libraries to develop. Results from this study, combined with greater awareness of organizational needs can be used to identify strategic opportunities for libraries. Libraries should be bold in their strategic planning; they should seek out gaps in the knowledge creation process and identify ways their organizations can support the creation and discovery of new knowledge. Libraries should recruit new staff capable of supporting emerging practices in science and assist in the transition of existing staff through training. The results of this research can provide a useful framework for health sciences library leaders to guide their organizations in discovering new roles for libraries in addressing the emerging needs of biomedical research.

Bibliography

A Brief History of NLM. (n.d.). Retrieved October 13, 2016, from https://www.nlm.nih.gov/about/briefhistory.html

About AAHSL. (n.d.). Retrieved January 2, 2017, from http://www.aahsl.org/about-us

American Library Association. (n.d.). Libraries and the Internet Toolkit. Retrieved from http://www.ala.org/advocacy/intfreedom/iftoolkits/litoolkit/openaccess

Anderson, R. K. (1989). Reinventing the medical librarian. *Bulletin of the Medical Library Association*, *77*(4), 323–331.

Association of Research Libraries. (n.d.). Open Access | Association of Research Libraries® | ARL®. Retrieved from http://www.arl.org/focus-areas/open-scholarship/open-access

Ballard, J. (1925). Training for medical librarianship. *Bulletin of the Medical Library Association*, *15*(2), 30–31.

Ballard, J. F. (1998). The president’s address. 1938. *Bulletin of the Medical Library Association*, *86*(2), 229–232.

Birchette, K. P. (1973). The history of medical libraries from 2000 B.C. to 1900 A.D. *Bulletin of the Medical Library Association*, *61*(3), 302–308.

Björk, B.-C., Shen, C., & Laakso, M. (2016). A longitudinal study of independent scholar-published open access journals. *PeerJ*, *4*, e1990. https://doi.org/10.7717/peerj.1990

Bloom, F. E. (1999). Just a Minute, Please. *Science*, *285*(5425), 197–197. https://doi.org/10.1126/science.285.5425.197

Blue Ridge Institute for Medical Research. (n.d.). Retrieved March 18, 2017, from http://www.brimr.org/

Bonner, T. N. (1963). *American doctors and German universities;a chapter in international intellectual relations, 1870-1914.* Linciln,. Retrieved from http://hdl.handle.net/2027/mdp.39015009044200

Braude, R. M. (1997). On the origin of a species: evolution of health sciences librarianship. *Bulletin of the Medical Library Association*, *85*(1), 1–10.

Brower, V. (2001). Public library of science shifts gears. As scientific publishing boycott deadline approached, advocates of free scientific publishing announce that they will create their own online, free-access archive. *EMBO Reports*, *2*(11), 972–973. https://doi.org/10.1093/embo-reports/kve239

Brown, C. M. (1999). Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists. *Journal of the American Society for Information Science*, *50*(10), 929.

Bunting, A. (1998). Excellence, promise, vision, and values: reflections on the Janet Doe Lectures, 1967-1997. *Bulletin of the Medical Library Association*, *86*(2), 258–266.

Calhoun, C. (2006). The University and the Public Good. *Thesis Eleven*, *84*(1), 7–43. https://doi.org/10.1177/0725513606060516

Cataldo, T. T., Tennant, M. R., Sherwill-Navarro, P., & Jesano, R. (2006). Subject specialization in a liaison librarian program. *Journal of the Medical Library Association : JMLA*, *94*(4), 446–448.

Connor, J. J. (2011). Only for “purely scientific” institutions: the Medical Library Association’s Exchange, 1898–1950s. *Journal of the Medical Library Association : JMLA*, *99*(2), 118–126. https://doi.org/10.3163/1536-5050.99.2.003

Consumerization - Gartner IT Glossary. (n.d.). Retrieved October 25, 2016, from http://www.gartner.com/it-glossary/consumerization/

Cooper, I., & Crum, J. (2013). New activities and changing roles of health sciences librarians: a systematic review, 1990–2012. *Journal of the Medical Library Association : JMLA*, *101*(4), 268.

Corrado, E. (2005). The importance of open access, open source, and open standards for libraries. *Issues in Science and Technology Librarianship*, (42).

Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research. *Library Trends*, *61*(3), 636–674. https://doi.org/10.1353/lib.2013.0005

Crawford, S. (1983). Health science libraries in the United States: I. Overview of the post-World War II years. *Bulletin of the Medical Library Association*, *71*(1), 16–20.

Creswell, J. W. (2013). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks: SAGE Publications,.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks: SAGE Publications.

Cummings, M. M., & Corning, M. E. (1971). The Medical Library Assistance Act: an analysis of the NLM extramural programs, 1965-1970. *Bulletin of the Medical Library Association*, *59*(3), 375–391.

D'Alessandro, M. P., Galvin, J. R., D'Alessandro, D. M., Erkonen, W. E., & Choi, T. A. (1999). The virtual hospital: The digital library moves from dream to reality. *Academic Radiology*, *6*(1), 78–80. https://doi.org/10.1016/S1076-6332(99)80069-5

Data Access, Management & Sharing | Association of Research Libraries. (n.d.). Retrieved December 28, 2016, from http://www.arl.org/focus-areas/e-research/data-access-management-and-sharing#.WGO2PRsrKUl

Data Lifecycle | e-Science Portal for Librarians. (n.d.). Retrieved December 28, 2016, from http://esciencelibrary.umassmed.edu/professional-educ/escience-thesaurus/data-lifecycle

David, P. A., & Foray, D. (2002). An introduction to the economy of the knowledge society. *International Social Science Journal*, *54*(171), 9–23. https://doi.org/10.1111/1468-2451.00355

Davidoff, F., & Florance, V. (2000). The informationist: a new health profession? *Annals of Internal Medicine*, *132*(12), 996–998.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*(4), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Dry Lab Medical Definition | Merriam-Webster Medical Dictionary. (n.d.). Retrieved

January 9, 2018, from https://www.merriam-webster.com/medical/dry lab

Eden, B. L. (2012). *The associate university librarian handbook : a resource guide*. Lanham, MD: Scarecrow Press.

Federer, L. (2013). The librarian as research informationist: a case study. *Journal of the Medical Library Association : JMLA*, *101*(4), 298–302. https://doi.org/10.3163/1536-5050.101.4.011

Flexner, A. (1910). Medical education in the United States and Canada bulletin number four (The Flexner Report). *The Carnegie Foundation for the Advancement of Teaching*.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385–400. https://doi.org/10.1016/0191-8869(86)90014-0

Geer, R. C. (2006). Broad issues to consider for library involvement in bioinformatics. *Journal of the Medical Library Association : JMLA*, *94*(3), 286–98, E152-5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16888662

Gold, A. (2007). Cyberinfrastructure, data, and libraries, part 2: Libraries and the data challenge: Roles and actions for libraries. *D-Lib Magazine*, *13*(9/10).

Goldstein, H. M., & Coletti, M. H. (2011). Developing a Strategic Plan for Transitioning to Healthcare Knowledge Services Centers (HKSCs). *Journal of Hospital Librarianship, 11*(4), 379–387. https://doi.org/10.1080/15323269.2011.609076

Goode, V., & Anton, B. (2013). Welch Informationist Collaboration With The Johns

Hopkins Medicine Department of Radiology. *Journal of eScience Librarianship*, *2*(1), 16–19. https://doi.org/10.7191/jeslib.2013.1033

Gore, S. (2013). A Librarian by Any Other Name: The Role of the Informationist on a Clinical Research Team. *Journal of eScience Librarianship*, *2*(1), 20–24. https://doi.org/10.7191/jeslib.2013.1041

Greenbaum, J. (1943). Work analysis of functions and duties of the medical library staff. *Bulletin of the Medical Library Association*, *31*(4), 339.

Greene, J., Kreider, H., & Mayer, E. (2005). Combining qualitative and quantitative methods in social inquiry. *Research Methods in the Social*.

Guston, D., & Keniston, K. (1994). *The fragile contract: University science and the federal government*. Cambridge, MA: The MIT Press.

Haines, L. L., Light, J., O'Malley, D., & Delwiche, F. A. (2010). Information-seeking behavior of basic science researchers: implications for library services. *Journal of the Medical Library Association : JMLA*, *98*(1), 73–81. https://doi.org/10.3163/1536-5050.98.1.019

Hannan, M., & Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology*. Retrieved from http://www.journals.uchicago.edu/doi/abs/10.1086/226424

Hannay, J., & MacLeod, C. (2009). How do scientists develop and use scientific software? In *Proceedings of the 2009 ICSE workshop on Software Engineering for Computational Science and Engineering* (pp. 1–8).

Hansn, K., Bakker, T., Svirsky, M., Neuman, A., & Rambo, N. (2013). Informationist role: clinical data management in auditory research. *Journal of eScience Librarianship*, *2*(1), 25–29. https://doi.org/10.7191/jeslib.2013.1030

Hasman, L., Berryman, D., & McIntosh, S. (2013). NLM Informationist Grant – Web Assisted Tobacco Intervention for Community College Students. *Journal of eScience Librarianship*, *2*(1), 30–34. https://doi.org/10.7191/jeslib.2013.1034

Henderson, M. (2014). Health sciences librarianship. In *Health sciences librarianship* (1st ed., pp. 403–415). Lanham, Maryland: Rowman & Littlefield Publishers.

Hersh, W. (2002). Medical informatics education: an alternative pathway for training informationists. *Journal of the Medical Library Association : JMLA*, *90*(1), 76–9.

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery* (2nd ed.). Redmond, WA: Microsoft Corporation.

Hickson, D. J., Hinings, C. R., Lee, C. A., Schneck, R. E., & Pennings, J. M. (1971). A Strategic Contingencies' Theory of Intraorganizational Power. *Administrative Science Quarterly*, *16*(2), 216. https://doi.org/10.2307/2391831

Huerta, M., Downing, G., Haseltine, F., Seto, B., & Liu, Y. (2000). *NIH working definition of bioinformatics and computational biology*. Retrieved from http://2digitstechcom.ipage.com/uploads/2/9/0/1/2901227/compubiodef.pdf

Internet Use Over Time | Pew Research Center. (n.d.). Retrieved November 1, 2016, from http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/

Jayne, W. A. (1916). Medical literature and medical libraries. *Bulletin of the Medical*

Library Association, *6*(1), 1–12.

Journal of eScience Librarianship: Final Manuscript Preparation Guidelines. (n.d.). Retrieved January 7, 2017, from http://escholarship.umassmed.edu/jeslib/styleguide.html

Kaufman, M. (1976). *American medical education: The formative years, 1765-1910*. Westport, CT: Greenwood Publishing Group.

Kohane, I. S. (2000). Bioinformatics and Clinical Informatics. *Journal of the American Medical Informatics Association*, *7*(5), 512–516.

Kronenfeld, M. (2000). The informationist: a new health profession?" So what are we? Chopped liver? *National Network (Dallas, Tex.)*, *25*(2), 1, 15.

Kronick, D. A., Rees, A. M., & Rothenberg, L. (1972). An Investigation of the Educational Needs of Health Sciences Library Manpower: Part VII: Summary and Conclusions. *Bulletin of the Medical Library Association*, *60*(2), 292–300.

Laakso, M., Welling, P., Bukvova, H., Nyman, L., & Björk, B.-C. (2011). The Development of Open Access Journal Publishing from 1993 to 2009. *PLoS ONE*, *6*(6), e20961. https://doi.org/10.1371/journal.pone.0020961

Lavrakas, P. (2008). *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States of America : Sage Publications, Inc. https://doi.org/10.4135/9781412963947

Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and Environment: Managing Differentiation and Integration*. Boston: Division of Research, Graduate School of

Business Administration, Harvard University.

Lawrence, P. R., & Lorsch, J. W. (1967). Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*, *12*(1), 1. https://doi.org/10.2307/2391211

Li, M., Chen, Y.-B., & Clintworth, W. A. (2013). Expanding roles in a library-based bioinformatics service program: a case study. *Journal of the Medical Library Association : JMLA*, *101*(4), 303–9. https://doi.org/10.3163/1536-5050.101.4.012

Librarian Outreach Kit | DataONE. (n.d.). Retrieved December 28, 2016, from https://www.dataone.org/for-librarians

Liesegang, T. J., Schachat, A. P., & Albert, D. M. (2005). The Open Access initiative in scientific and biomedical publishing: Fourth in the series on editorship. *American Journal of Ophthalmology*, *139*(1), 156–167. https://doi.org/10.1016/j.ajo.2004.10.010

Lindberg, D., & Humphreys, B. (2005). 2015-the future of medical libraries. *The New England Journal of Medicine*, *352*(11), 1067–1070.

Lohr, S. (2014). For Big-Data Scientists, “Janitor Work” Is Key Hurdle to Insights. *The New York Times*, p. 17.

Love, E. (1987). The science of medical librarianship: investing in the future. *Bulletin of the Medical Library Association*, *75*(4), 302–9.

Lucier, R. E. (1995). Building a digital library for the health sciences: information space complementing information place. *Bulletin of the Medical Library Association*, *83*(3), 346–50.

Ludmerer, K. (1999). *Time to heal: American medical education from the turn of the century to the era of managed care*. Oxford, England: Oxford University Press.

Lyon, J. (2003). Beyond the literature: bioinformatics training for medical librarians. *Medical Reference Services Quarterly*, *22*(1), 67–74. https://doi.org/10.1300/J115v22n01_07

Lyon, J. A., Tennant, M. R., Messner, K. R., & Osterbur, D. L. (2006). Carving a niche: establishing bioinformatics collaborations. *Journal of the Medical Library Association : JMLA*, *94*(3), 330–5.

Marshall, C., & Rossman, G. B. (1989). *Designing qualitative research*. Newbury Park Calif.: Sage Publications.

Marshall, J. (2013). The value of library and information services in patient care: results of a multisite study. *Journal of the Medical Library Association*. Retrieved from http://europepmc.org/articles/PMC3543128

McGowan, J. (2012). Agony & Ecstasy: The Transformation of the Academic Health Sciences Library. *Indiana Libraries*, *31*(1), 7–12.

National Institute of Dental and Craniofacial Research: Requests for Applications. (n.d.). Retrieved December 30, 2016, from https://www.nidcr.nih.gov/GrantsAndFunding/See_Funding_Opportunities_Sorted_By/RFA/

Nicoll, L. H. (1992). Grateful Med: an easy to use information tool. *Reflections*, *18*(3), 40–1.

NIH approves strategic vision to transform National Library of Medicine | National Institutes of Health (NIH). (2015). Retrieved December 28, 2017, from https://www.nih.gov/news-events/news-releases/nih-approves-strategic-vision-transform-national-library-medicine

NLM Technical Bulletin. May-Jun 1997. (n.d.). Retrieved October 17, 2016, from https://www.nlm.nih.gov/pubs/techbull/mj97/mj97_web.html

NOT-OD-05-022: Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research. (n.d.). Retrieved March 3, 2013, from http://grants.nih.gov/grants/guide/notice-files/not-od-05-022.html

NOT-OD-08-033: Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research. (n.d.). Retrieved March 3, 2013, from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html

NSF Grant Proposal Guide. (n.d.). Retrieved December 30, 2016, from https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#IIC2f

Odlyzko, A. (1999). Competition and Cooperation: Libraries and Publishers in the Transition to Electronic Scholarly Journals. *Journal of Scholarly Publishing*, *30*(4), 163–185. https://doi.org/10.3138/JSP-030-04-163

OpenWetWare:FAQ - OpenWetWare. (n.d.). Retrieved January 9, 2018, from https://openwetware.org/wiki/OpenWetWare:FAQ

OpenWetWare:Headquarters/Research Pathway - OpenWetWare. (n.d.). Retrieved

January 9, 2018, from https://openwetware.org/wiki/OpenWetWare:Headquarters/Research_Pathway

Overview: Research Lifecycle - UCF Libraries. (n.d.). Retrieved January 9, 2018, from https://library.ucf.edu/about/departments/scholarly-communication/overview-research-lifecycle/

Petrinic, T., & Urquhart, C. (2007). The education and training needs of health librarians - the generalist versus specialist dilemma. *Health Information and Libraries Journal*, *24*(3), 167–76. https://doi.org/10.1111/j.1471-1842.2007.00717.x

Pezalla, A. E., Pettigrew, J., & Miller-Day, M. (2012). Researching the researcher-as-instrument: an exercise in interviewer self-reflexivity. *Qualitative Research : QR*, *12*(2), 165–185. https://doi.org/10.1177/1487941111422107

Ragon, B. (2013). The Political Economy of Federally Sponsored Data. *Journal of eScience Librarianship*, *2*(2), 34–42. https://doi.org/10.7191/jeslib.2013.1050

Rankin, J. A., Grefsheim, S. F., & Canto, C. C. (2008). The emerging informationist specialty: a systematic review of the literature. *Journal of the Medical Library Association : JMLA*, *96*(3), 194–206. https://doi.org/10.3163/1536-5050.96.3.005

Research Lifecycle | e-Science Portal for Librarians. (n.d.). Retrieved December 28, 2016, from http://esciencelibrary.umassmed.edu/professional-educ/escience-thesaurus/research-lifecycle

Robison, R. R. (2008). Informationist education. *Medical Reference Services Quarterly*, *27*(3), 339–347. https://doi.org/10.1080/02763860802199034

Rossman, G., & Wilson, B. (1994). Numbers and words revisited: Being “shamelessly eclectic.” *Quality and Quantity*. Retrieved from http://link.springer.com/article/10.1007/BF01098947

Salancik, G., & Pfeffer, J. (1974). The bases and use of power in organizational decision making: The case of a university. *Administrative Science Quarterly*. Retrieved from http://www.jstor.org/stable/2391803

Shipman, J. P., Cunningham, D. J., Holst, R., & Watson, L. A. (2002). The informationist conference: report. *Journal of the Medical Library Association : JMLA*, *90*(4), 458–464.

Simpson, B. (2013). Hiring Non-MLS Librarians: Trends and Training Implications. *Library Leadership & Management*, *28*(1).

Software Carpentry: Frequently Asked Questions. (n.d.). Retrieved September 12, 2015, from http://software-carpentry.org/faq.html#what-is-swc

Stebbins, M. (n.d.). Expanding Public Access to the Results of Federally Funded Research | The White House. Retrieved March 3, 2013, from http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Surkis, A., McCrillis, A., McGowan, R., Williams, J., Schmidt, B., Hardt, M., & Rambo, N. (2013). Informationist support for a study of the role of proteases and peptides in cancer pain. *Journal of eScience Librarianship*, *2*(1), 35–40. https://doi.org/10.7191/jeslib.2013.1029

Tennant, M. (2005). Bioinformatics librarian: meeting the information needs of genetics and bioinformatics researchers. *Reference Services Review*, *33*(1), 12–19.

The Data Science Venn Diagram — Drew Conway. (n.d.). Retrieved October 3, 2016, from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Vaughan, K. T. L., Hayes, B. E., Lerner, R. C., McElfresh, K. R., Pavlech, L., Romito, D., … Morris, E. N. (2013). Development of the research lifecycle model for library services. *Journal of the Medical Library Association : JMLA*, *101*(4), 310–4. https://doi.org/10.3163/1536-5050.101.4.013

Walters, T. (2013). The Future of Knowledge Creation and Production in University Research Programs and Their Effect on University Libraries. Retrieved from https://vtechworks.lib.vt.edu/handle/10919/47352

Weick, K. E. (1979). *The social psychology of organizing* (2d ed.). Reading Mass.: Addison-Wesley Pub. Co. Retrieved from http://uva.worldcat.org/title/social-psychology-of-organizing/oclc/4775332&referer=brief_results

Wet Lab Medical Definition | Merriam-Webster Medical Dictionary. (n.d.). Retrieved January 9, 2018, from https://www.merriam-webster.com/medical/wet lab

Whipple, E., Odell, J., Ralston, R., & Liu, G. (2013). When Informationists Get Involved: the CHICA-GIS Project. *Journal of eScience Librarianship*, *2*(1), 41–45.

https://doi.org/10.7191/jeslib.2013.1035

Wiegand, W. (1999). 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*, *69*(1), 1–32.

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., … Wilson, P. (2014). Best practices for scientific computing. *PLoS Biology*, *12*(1), e1001745. https://doi.org/10.1371/journal.pbio.1001745

Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, *330*(7500), 1128–0. https://doi.org/10.1136/bmj.38422.611736.E0

Yarfitz, S., & Ketchell, D. S. (2000). A library-based bioinformatics services program. *Bulletin of the Medical Library Association*, *88*(1), 36–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10658962

Appendix A. Description of subcycles and code definitions for Biomedical Researchers

General Codes

Influencing science – External forces affecting how researchers obtain funding, conduct science, or disseminate findings.

Emerging practice – A practice in science that is changing due to evolutions in technology, culture, or external forces.

Library support – Support to the researcher or research teams provided by the library.

Non-library support – Support to the researcher or research teams provided by non-library personnel.

Mentor/Mentee – The role and relationship of mentors and mentees in the research lifecycle.

Reproducibility and replicability – Activities related to producing science in a way that another researcher can reproduce results of an experiment using the same data or to replicate the study with a different sample.

Competition – Concern for research ideas, funding, or preliminary research results to be obtained, discovered, or disseminated by another researcher.

Licensing and venture – The process of transferring university intellectual property to a third party for the development of products, services, or applications. Also known as technology transfer.

Research Planning Subcycle

Literature searching – Searching peer-reviewed literature databases such as PubMed, Google Scholar, etc.

Grey literature – Seeking non-peer reviewed information. Examples include conference proceedings, white papers, device manufacturing information, blog, wikis, etc.

Citation management – Methods used to store and reuse scholarly literature. Incudes software such as Endnote, Mendeley, apps, and other online tools.

Bioinformatics tools – NCBI tools or other forms of data such as Blast, OncoMine, dbSNP, etc.

Identify collaborators – The method in which collaborators are identified, includes the use of online tools or in-person networking opportunities.

Biosketch – Creation or management of biosketch information for use in grant applications.

IRB/IACUC – Activities related to obtaining or complying with Institutional Review Board or the Institutional Animal Care and Use Committee.

Methodology – Experimental design and methods used to conduct a scientific experiment

Data Literacy – Understanding of how to read and use data, the credibility of the source, and permissions and restrictions to its use.

Grant funding – The process of seeking grant funding, grant administration, or grant citation.

Systematic Review – Comprehensive review of relevant literature, using a methodology, to provide a complete and exhaustive summary of the current literature relevant to a research question.

Experimental Subcycle

Data collection – Process of gathering data in research for the purpose of analysis.

Data management – The way that data is planned to be managed or managed during the research process either individually or as a team. Includes data documentation.

Data analysis – The methods, techniques, and products employed to use and analyze data. Includes activities such as data munging and data manipulation.

Organizing and storing information - The methods that a researcher or research team organize and store electronic or non-electronic files, samples, or other objects.

Project management – Activities related to how a researcher or research team plans and manages the research process.

Pilot experiment – Execution of an experiment prior to or within another experiment for the purposes of investigating potential future research.

Proprietary software – Use of software in research owned by an individual or company.

Open source software - Use of software in research that open for further development and maintained by a community of users.

Ethics – Knowledge of and adherence of policies and procedures required to conduct ethical research.

Statistical methods – Techniques used in research to analyze and interpret data.

Collaborating – Activities of a researcher working with other researchers or those who support research.

Metadata – Use of data about data including metadata standards, ontologies, or taxonomies.

Lab notebook – Electronic or paper based system for documenting activities connected to research.

Data privacy and security – Activities associated with ensuring the protection of data, confidentiality, and compliance with applicable laws or policies.

Dissemination Subcycle

Presentation – Activities related to presenting scholarly work include the creation of graphics and PowerPoint poster design.

Attending conferences – Value expressed by the researcher by attending a professional conference.

Writing – Process or activities related to writing journal papers and grants.

Data preservation – The act of archiving and managing data post-research to ensure it continued access for as long as deemed necessary.

Social media – Use of social media to promote science, awareness, or reputation management.

Compliance – Adherence to policies and procedures associated with use of grant funding. Includes NIH public access policy compliance, the submission of peer-reviewed articles into PubMed Central as required by NIH.

Author rights and copyright – Concepts related to the use of intellectual property of others or the rights transferred by a researchers when publishing academic manuscripts.

Preprint – Version of a publication made available prior to its publication in a scholarly journal.

Open access – Access, use, or depositing of scholarly material connected to research. Includes open access articles, open access data, open access publishers, and open access repositories.

Journal selection – The process a researcher uses for selecting a peer-reviewed journal to submit an article for publication.

Conference selection - The process a researcher uses for selecting an academic conference to submit a presentation.

Citation styles – Information necessary to cite a reference in a particular format used for publications or other scholarly reference.

Impact Subcycle

Altmetrics - Collection and or use of non-traditional metrics of scholarly work, including articles, data, and other products of research. Can include downloads, views, or mentions in social media.

Citation metrics – Collection and or use of traditional metrics connected to scholarly publications of a researchers work. Metrics include number of citations, H Index, Impact Factor, etc.

Appendix B: Library support of the research lifecycle

	Have provided more than 4 years		**Have provided more than 2 years**		**Added in last 2 years**		**Plan to add in the next 2 years**		**Plan to add in the next 4 years**		**No plans to add**	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Background literature searching	49	96	2	4	0	0	0	0	0	0	0	0
Systematic reviews	29	57	12	24	6	12	2	4	0	0	2	4
Grey literature searching	36	71	7	14	4	8	2	4	0	0	2	4
Citation management	49	96	1	2	0	0	0	0	0	0	1	2
Locating data sources	37	73	5	10	4	8	1	2	0	0	4	8
NCBI tools (Blast, GenBank, dbSNP, etc)	19	37	4	8	5	10	3	6	3	6	17	33
Seeking grant funding	20	39	6	12	2	4	4	8	0	0	19	37
Writing center services	7	14	4	8	7	14	5	10	0	0	28	55
Grant budget preparation	4	8	0	0	2	4	1	2	0	0	44	86
Methodology	12	24	3	6	3	6	1	2	1	2	31	61
Experimental design	4	8	0	0	2	4	2	4	0	0	43	84
Data management plans	8	16	14	28	11	22	5	10	3	6	10	20
Data catalog	3	6	2	4	5	10	11	22	6	12	24	47
Data literacy	6	12	4	8	8	16	15	29	3	6	15	29
Biosketch creation	6	12	14	28	4	8	5	10	2	4	20	39
IRB protocols	12	24	4	8	1	2	3	6	0	0	31	61
IACUC protocols	22	43	4	8	5	10	2	4	0	0	18	35
Ethics and compliance	13	26	3	6	2	4	2	4	1	2	30	59
Identifying collaborators	17	33	7	14	3	6	5	10	0	0	19	37
Project planning and management	9	18	2	4	2	4	1	2	2	4	35	69
Methods for organizing and storing information	17	33	8	16	9	18	5	10	4	8	8	16
Automated and manual data collection	6	12	3	6	4	8	9	18	2	4	27	53
Data documentation (file format, naming conventions, file organization)	8	16	8	16	9	18	9	18	2	4	15	29
Metadata standards	9	18	10	20	7	14	4	8	6	12	15	29

Ontology/taxonomy	5	10	7	14	6	12	8	16	2	4	23	45
Data privacy and security guidance	7	14	5	10	5	10	6	12	5	10	23	45
Code versioning (GitHub, etc)	0	0	2	4	7	14	7	14	5	10	30	59
File versioning	5	10	2	4	5	10	5	10	3	6	31	61
Electronic lab notebooks	0	0	4	8	4	8	6	12	5	10	32	63
Managing research data	9	18	10	20	11	22	4	8	6	12	11	22
Data wrangling/cleaning using open source tools such as R, Python, OpenRefine	0	0	5	10	7	14	7	14	7	14	25	49
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel	7	14	6	12	4	8	5	10	5	10	24	47
Statistical methods	5	10	3	6	6	12	4	8	2	4	31	61
Geographic information system (GIS)	8	16	3	6	3	6	8	16	2	4	27	53
High performance computing (HPC)	2	4	3	6	4	8	2	4	2	4	38	75
Scientific modeling	0	0	3	6	2	4	0	0	4	8	42	82
Prototyping (3D printers and maker technology)	4	8	7	14	11	22	6	12	2	4	21	41
Infrastructure and space	14	28	2	4	3	6	6	12	2	4	24	47
Journal selection for publication	32	63	12	24	4	8	0	0	0	0	3	6
Open access	36	71	5	10	3	6	2	4	2	4	3	6
Conference selection	13	26	1	2	1	2	0	0	0	0	36	71
Web and social media marketing	20	39	5	10	3	6	3	6	1	2	19	37
Bibliographic styles	39	77	3	6	1	2	1	2	0	0	7	14
Preprint archive	11	22	2	4	4	8	5	10	2	4	27	53
Author rights and copyright	38	75	5	10	4	8	2	4	0	0	2	4
Image and graphics for submission	13	26	6	12	3	6	5	10	1	2	23	45
Presentation poster preparation	20	39	2	4	4	8	1	2	0	0	24	47
Data archiving	3	6	4	8	6	12	8	16	8	16	22	43
Data sharing	6	12	5	10	9	18	6	12	6	12	19	37
Long term preservation of experiment materials	2	4	2	4	2	4	5	10	3	6	37	73

Institutional repository	23	45	7	14	6	12	1	2	3	6	11	22
Funder public access policy compliance	26	51	5	10	4	8	1	2	1	2	14	28
Grant citation	17	33	6	12	3	6	2	4	1	2	22	43
Citation metrics (H Index, Impact Factor, times cited	31	61	8	16	5	10	2	4	0	0	5	10
Altmetrics	9	18	15	29	6	12	8	16	1	2	12	24
Online research profile management	15	29	7	14	7	14	6	12	3	6	13	26

Appendix C. Description and code definitions for library leader open-ended questions

Author rights and copyright – Concepts related to the use of intellectual property of others or the rights transferred by a researchers when publishing academic manuscripts.

Collaboration – Libraries describing roles or potential roles for working with researchers.

Collections – Resources purchased, licensed, or made available to researchers.

Compliance – Assistance provided supporting adherence to institutional, funder, or regulatory policy.

Data management - The way that data is planned to be managed or managed during the research process either individually or as a team. Includes data documentation and data management plans.

Data science and bioinformatics – Concepts related to the methods, processes, and techniques used to analyze biomedical research data.

Domain expertise – Expertise or training in a scientific discipline.

Entrepreneurship and innovation – Activities connected to the way researchers develop novel ideas and manage their labs as small business.

Expert searching – Librarian skills in searching peer-reviewed literature databases.

Grants – Activities related to funded research.

Information architecture – The way in which information is organized and described.

IRB/IACUC – Processes associated with Institutional Review Boards and the Institutional Animal Care and Use Committee.

Knowledge dissemination – Methods and techniques used in the transfer of knowledge.

Librarian training and continuing education – Library school, continuing education, and other forms or education and training.

Open science – Concepts and activities to make the products of science available without cost to the user.

Awareness of organizational context – Library understanding of the needs of their researchers, health systems, and universities.

Outreach – Efforts to market and promote library services.

Research lifecycle knowledge – Awareness of the activities needed to conduct biomedical research.

Scholarly communication – Activities and processes for sharing publish research.

Scholarly impact and online profile management – Assessing the impact of scholarly work and the managing how that information is

Space – Use of library space to support the activities of biomedical research.

Statistics and methodology - Techniques used in research to design, analyze, and interpret data.

Writing - Process or activities related the writing needs of biomedical research.

Appendix D: Biomedical Researcher Interview Protocol

Thank you for taking the time to meet with me today for this interview. Before we begin, I'd like to review the procedure.

The interview is designed to last 60 minutes, but may take shorter depending on your responses.

You have read and signed the consent agreement. Do you have any questions about the agreement?

I am interested in learning about the workflow of a biomedical researcher from identification of the research question to dissemination of knowledge. We will explore standard and emerging scientific practices, as well as the role of the library in supporting biomedical research. If you don't feel like answering a particular question or would like to stop the interview at any point, please let me know. The information that you give in the study will be handled confidentially. You have the right to withdraw from the study at any time without penalty.

Do you have any questions before we begin? May I start the audio recording?

Participant ID: ________________________________

Date: ________________________________

Introductory Questions

1. Tell me a little about yourself. How did you get involved with biomedical research?
2. How long have you been conducting biomedical research?
3. Can you briefly describe the focus of your research area?
4. In your opinion, what are the most significant factors influencing biomedical science today?

Research Workflow Questions

Research Planning Cycle

1. Prior to starting data collection, analysis, and interpretation, can you describe to me the major activities required to generate an idea and obtain funding?
    a. *Probes: Background information (literature searching, systematic reviews, grey literature), citation management, locating data sources, Bioinformatics tools (Blast, OncoMine, dbSNP, others), seeking grant funding, grant planning (writing, budget, application), methodology (experimental design, survey design, etc), data management plans, data literacy, biosketch creation, IRB and IACUC protocols, ethics and compliance, identify collaborators, project planning, organizing and storing information.*
    b. *What are standard practices, what are emerging practices? Discuss education, consultation, and training needs of researcher or their team.*

2. Can you think of an instance where you used the library or asked for the help of a librarian to accomplish the activities in the research planning?
    a. *Probes: Specific ways that they use or do not use library services based on the probes above. Use answers to guide conversation about why they use or do not use library services. Discuss how if the library had additional skills it may or may not demonstrate value. Are there other departments that provide support (who and how)?*

Experimental Cycle

1. Can you describe to me the major activities required to begin collecting and analyzing data?
    a. *Probes: Automated and manual data collection, data documentation (file format, naming conventions, file organization), metadata standards, ontology/taxonomy, data security, versioning code (GitHub), versioning files, electronic lab notebook, data wrangling/cleaning (open source and/or proprietary software), data analysis and visualization (open source and/or proprietary software), statistics, GIS, research computing and technical support, scientific modeling, prototyping (maker technology), Infrastructure and space, Scheduling, project management, grant management.*
    b. *What are standard practices, what are emerging practices? Discuss education, consultation, and training needs of researcher or their team.*

2. Can you think of an instance where you used the library or asked for the help of a librarian to accomplish the activities in the conducting research?
    a. *Probe for specific ways that they use or do not use library services. Are there other departments that provide support (who and how)? Discuss education and training needs of researcher or their team. Use answers to guide conversation about why the use or do not use library services. Discuss how if the library had additional skills it may or may not be useful.*

Disseminating Cycle

1. Can you describe to me the major activities required to disseminate results?
    a. *Probes: Journal selection for publication (for profit publisher, open access), conference selection, web and social media marketing, bibliographic styles, writing center, preprint archive, author rights and copyright, image and graphics for submission, presentation poster preparation, data archiving, data sharing, long term preservation of experiment materials (curation), institutional repository, funder public access policy compliance, grant citation.*
    b. *What are standard practices, what are emerging practices? Discuss education, consultation, and training needs of researcher or their team.*
2. Can you think of an instance where you used the library or asked for the help of a librarian to accomplish the activities in the disseminating research results?
    a. *Probe for specific ways that they use or do not use library services. Are there other departments that provide support (who and how)? Discuss education and training needs or researcher or their team. Use answers to guide conversation*

about why the use or do not use library services. Discuss how if the library had additional skills it may or may not be useful.

Impact Cycle

1. Can you describe to me the major activities required to measure the impact of the research?
    a. Probes: Citation metrics *(H Index, Impact Factor, times cited, others?)*, altmetrics and social media, managing research profile.
    b. *What are standard practices, what are emerging practices? Discuss education, consultation, and training needs of researcher or their team.*
3. Can you think of an instance where you used the library or asked for the help of a librarian to accomplish the activities in measuring the impact of research?
    a. *Probe for specific ways that they use or do not use library services. Are there other departments that provide support (who and how)? Discuss education and training needs or researcher or their team. Use answers to guide conversation about why the use or do not use library services. Discuss how if the library had additional skills it may or may not be useful.*

Appendix E: Health Sciences Library Leadership Survey Protocol

The purpose of the study is to investigate how health sciences libraries can expand their role in supporting the creation of knowledge within the research lifecycle. We are asking for one response per health sciences library from the Director, Deputy Director, Associate Director, or comparable leadership level. Completing the questionnaire will take approximately 25 minutes. To assure confidentiality, I will not refer to your institution or your name in the data analysis and final report. You are heavily encouraged to devote at least 10 minutes to the open-ended questions at the end of the survey.

Introductory Questions

1. Name
2. Institution
3. How many years have you served at your current institution?
4. What is your work title?
   a. Director
   b. Deputy Director
   c. Associate Director
   d. Other (Please specify)

5. Skip logic based on the response for question 4.
   a. How many years in total have you served as a Director including multiple institutions?
   b. How many years in total have you served as a Director including multiple institutions?
   c. How many years in total have you served as an Associate Director including multiple institutions?
   d. How many years in total have you served at this position level including multiple institutions?

Please indicate below any library services that you provide or plan to add supporting <u>biomedical researchers</u>?

	Have provided more than 4 years	Have provided more than 2 years	Added in last 2 years	Plan to add in the next 2 years	Plan to add in the next 4 years	No plans to add
Background literature searching						
Systematic reviews						
Grey literature searching						
Citation management						

Locating data sources						
NCBI tools (Blast, GenBank, dbSNP, etc)						
Seeking grant funding						
Writing center services						
Grant budget preparation						
Methodology						
Experimental design						
Data management plans						
Managing research data						
Data catalog						
Data literacy						
Biosketch creation						
IRB protocols						
IACUC protocols						
Ethics and compliance						
Identifying collaborators						
Project planning and management						
Methods for organizing and storing information						
Automated and manual data collection						
Data documentation (file format, naming conventions, file organization)						
Metadata standards						
Ontology/taxonomy						
Data privacy and security guidance						
Code versioning (GitHub, etc)						
File versioning						
Electronic lab notebooks						

Data wrangling/cleaning using open source tools such as R, Python, OpenRefine						
Data wrangling/cleaning using proprietary software such as SAS, SPSS, Excel						
Data analysis and visualization using open source tools such as R or Python						
Data analysis and visualization using proprietary software such as SAS, SPSS, Excel						
Statistical methods						
Geographic information system (GIS)						
High performance computing (HPC)						
Scientific modeling						
Prototyping (3D printers and maker technology)						
Infrastructure and space						
Journal selection for publication						
Open access						
Conference selection						
Web and social media marketing						
Bibliographic styles						
Preprint archive						
Author rights and copyright						
Image and graphics for submission						
Presentation poster preparation						

Data archiving						
Data sharing						
Long term preservation of experiment materials						
Institutional repository						
Funder public access policy compliance						
Grant citation						
Citation metrics (H Index, Impact Factor, times cited						
Altmetrics						
Online research profile management						

Please list examples of other actions (not previously mentioned in the survey) that your library has taken to accommodate the changing needs of biomedical research.

Based on your experience, what new skills should health sciences libraries cultivate in order to more fully support biomedical research?

What additional steps should health sciences libraries be taking in order to meet the changing needs of biomedical research?

Over the course of your career, how has the role of health sciences libraries changed when it comes to supporting biomedical research?