IFLA WLIC 2019 Athens

Submitted on: 19.08.2019

From Acquisition to Access to Archiving - Creating Library Data Services that Provide End-to-End Support for Library Acquired Data

Kris Kasianovitz
Stanford Libraries, Stanford University
Stanford, CA, United States
krisk11@stanford.edu

Julie Williamsen
Stanford Graduate School of Business,
Knight Management Center, Stanford University
Stanford, CA, United States.
jwilliamsen@stanford.edu

Abstract:

This case study outlines the challenges and opportunities that we are currently facing with data management of acquired datasets at the Stanford Graduate School of Business (GSB) Library and the Stanford Libraries (SUL) that impact data preservation. We will compare and contrast our current approaches for managing data in our respective environments; discuss what is working and what needs to be changed or approached in a different way. We will also discuss how we envision a library data service that seeks to employ standard approaches, like the FAIR principles, i.e. findable, accessible, interoperable, and reusable. We will identify ways to unify our approaches across our libraries so that we are not working in isolation of each other. We aim to provide considerations that will aid others facing these same data curation and preservation issues.

Keywords: Data Management Lifecycle, Data Curation, Research Data, Data Services, Academic Libraries

Introduction

The Stanford Libraries (SL) and the Stanford Graduate School of Business (GSB) Library are facing growing demands to better manage, preserve, and provide long-term access to our university's research data. We are also being asked to acquire data from government,

commercial, and private entities, which involves contracts and licenses and requires additional preparation for the data to be discoverable and usable by our faculty and students. In addition to being stewards of campus research-generated data and researcher models, analytical scripts and tools for the long-term, we are becoming more involved in creating and managing the mechanisms for data acquisition and licensing, quality checks, documentation, discovery, computation, and preservation.

Data preservation that follows the FAIR principles[1], i.e. findable, accessible, interoperable, and reusable, requires more work than just placing data files into GitHub or a repository with minimal description. We are working to establish appropriate library staffing, infrastructure, workflows, and systems that embrace and address the unmet needs of our growing dataset collections, just as we have successfully done for decades with traditional, physical and electronic collections. It is imperative that we manage, and appropriately curate these data throughout the entirety of their lifecycle.

The datasets, the staffing needs, and the systems are complex, diffuse, highly specialized, and require us to build collaborative relationships between internal library and university departments, and with researchers across campus. In order to address pressing research needs at our two libraries, we have scoped out a set of procedures, workflows, and specifications for systems that will enable the type of data curation that is necessary to ensure long-lived FAIR data, what we endeavour to call a "Library Data Service"[2].

Both libraries are involved with collecting and providing access to a wide variety of data types and formats.[3] The sources of data range from public to private, survey to administrative, and more. These data are acquired for students, faculty, and research teams who also produce "research data"[4] that we selectively collect. Our broad array of variant data presents us with a number of acquisition, curation, and preservation challenges that we are endeavouring to solve, and define best practices to manage our "data stacks as we move forward.

Context for Understanding our Library Environments

The GSB Library and Stanford University Libraries are not administratively under the same university organization, but clientele and services overlap between the two libraries. The GSB Library is part of one of three professional schools on campus, with Law and Medical being the other two. The GSB Library primarily exists to support the research and teaching needs for all Stanford Graduate School of Business affiliates. It also strives to serve the greater

[1] https://www.go-fair.org/fair-principles/

[2] Tenopir et al., 2017, p. 25

[3] Our working definition of data: content in digital form that can be subject to computation and analysis – can be numeric, textual, image, sound.

[4] Pryor, 2012. P Research data can be defined as: "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings." (OMB Circular 110). Research data covers a broad range of types of information, and digital data can be structured and stored a variety of file formats. (University of Oregon Libraries, n.d.)

Stanford community within its subject matter scope and serve the local community. The Stanford Libraries (SL) are not under a particular school or department and support the research and teaching mission across the entire Stanford community.

There is a growing, if not pent up, demand from our faculty and students to access data that typically has not been collected by libraries due to various factors –microdata, licensing issues, personally identifiable information (PII) issues, size is in terabytes or petabytes, cost, etc. from commercial and private entities that in the past typically did not work with libraries. Even though our libraries support and report to different parts of the university, we are collecting data that can be used by all of our constituencies on campus. Thus, it is to our advantage to work in conjunction with each other in order to take advantage of economies of scale and systems that can bolster discoverability and interoperability of our data. Striving to standardize our methods and approaches for providing data services and managing data across our libraries is also desirable and aligns with best practices for data management. In addition to forging a partnership between our libraries, we have also sought partnerships with key campus units where we can align services and infrastructure; specifically Stanford Research Computing Center, University IT, University Industrial Contracts Office, and University Procurement.[5]

Each of our Libraries engaged in their own needs assessment for managing data. By sharing information that we learned from these assessments, it became clear that both the GSB Library and Stanford Libraries are in the beginning stages of providing efficient management for acquired datasets and employing FAIR practices. We are working on the creation of workflows to properly curate, accession, and make our acquired data available for discovery, access, use, and preservation. We are striving to reduce staff time spent on various dataset management tasks, optimize delivery of services to users, and minimize risk such as data damage, corruption, or loss. Identifying our primary challenges and goals are the first steps to building out our Library Data Services.

Each of the following case studies focus on the steps we are taking to solve the challenges related to appropriate data stewardship[6] of our collections. We are in the beginning stages of providing efficient management for acquired datasets and employing FAIR practices. We are working on the creation of workflows to properly curate, accession, and make our acquired data available for discovery, access, use, and preservation. We are striving to reduce staff time spent on various dataset management tasks, optimize delivery of services to users, and minimize risk such as data damage, corruption, or loss. Identifying our primary challenges and goals are the first steps to building out our Library Data Services.

[5] See: https://srcc.stanford.edu/systems-services-overview, https://uit.stanford.edu/services, https://doresearch.stanford.edu/research-administration

[6] *"Data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data."* (Wilkinson et al., 2016, p.1).

Case Study: GSB Research Hub

Research needs at Stanford's Graduate School of Business continue to evolve with a growing focus on empirical research. Traditionally, faculty worked individually or on small teams, today, our researchers are working with larger teams. Where journal articles, smaller datasets, publicly available data and individually collected data used to meet GSB research needs, we have moved to an era where the need is for massive, complex data, proprietary datasets from corporate partners, and government administrative data. All of these data require variant types of storage, security, and computing environments.

In light of this new environment, the Stanford Graduate School of Business conducted an assessment on the current and future trends in faculty research. The goal was to craft a vision of how faculty research at the GSB could expand the frontiers of knowledge, and have an impact on their educational programs, outside organizations, and on society. Through this process it became clear that data needs should be a key focus including; the establishment of systems to house and manage datasets, improve transparency of data holdings across the school, facilitate data acquisitions from outside organizations, and expand field experiments within organizations.

Stanford's Graduate School of Business created an ecosystem to support the research and data needs of the school, with an emphasis on data, by centralizing its research support services into one unit, the GSB Research Hub. The Research Hub's mission is to enable pioneering research across all management disciplines at the school. It provides support for data-intensive projects, human behavioral experiments, library support for teaching and research, support and outreach for faculty research initiatives, as well as conferences and programming. The new structure allows for a deeper integration of services, and a strategic focus to support research needs.

The incorporation of the GSB Library into the Research Hub provides the organization with a deep understanding of traditional library functions that can be directly applied to the school's growing and evolving data needs. Librarians work closely with researchers to identify data sources and acquire datasets and other resources. The data acquisition process is a team-based service that includes; research librarians to support the discovery and evaluation of data sources and providers; contract librarians with a law background to draft and negotiate usage agreements and licenses to ensure the data can be used in the desired way; acquisition librarians to shepherd the agreement through the procurement process; and data access and data curation librarians to coordinate and collaborate with members of the Data, Analytics & Research Computing (DARC) team on data storage, security, access, analysis, and research computing.

The GSB Research Hub contains the following interconnected units:

- The Behavioural Research Lab. The "B-Lab" conducts laboratory-based human subject research, facilitates online research, provides and maintains human subject pools, and provides consultation on experiment and survey design/implementation.

- GSB Library. The Library supports research and teaching through access to databases and datasets, scholarly journals and books, individual and group research consultations from librarians with deep content knowledge. They also provide researchers with data acquisition support including contract negotiation, licensing, and procurement.
- Data, Analytics & Research Computing (DARC). The DARC team supplements the research capabilities and productivity of GSB faculty by providing; technical, administrative, and project management support of research projects, cutting edge research computing environments, secure storage and access to large data sets.
- Research Initiatives. The faculty driven research initiatives include Corporate Governance, Energy Business Innovations, Shared Prosperity and Innovation, Sports Management, Value Chain Innovation, and Venture Capital.
- Convening, Programs, and Operations. The "CPO" provides logistics as well as Marketing and Communications support for all of the teams in the Research Hub.

The GSB's research initiatives each supports focused scholarship in a particular discipline that align with the research of their faculty directors. One of the supported Research Initiatives is the Venture Capital Initiative. As part of this initiative, our faculty want to understand valuations of venture backed companies, how their boards are structured and the impact they have on the economy at large. Of the ten biggest companies in the US, the five largest are backed by venture capital, and the sixth makes venture capital investments.[7]

To build an understanding of venture capital backed firms, researchers need data about these firms. Subject librarians are experts on the data that are crucial for doing business research, and these data need an expert just to pull out basic information. One of the services that the GSB Library offers is acquiring datasets. One aspect of this is negotiating with venture capital data vendors to acquire specific variables about which venture capital firms invested how much, and in what companies. The Library collections staff review corporate data provider contracts like this one. It took several months of negotiation to acquire a venture capital research platform for single company lookups for MBA students and course use, and the same data in a feed format for faculty driven research projects. When the agreement was finalized, library staff set up the authentication to the platform on the library website. This was a challenge because, like many of the companies with whom we work, they did not have a model for providing academics login authentication access.

There was also the challenge of on-boarding the data feed at the GSB. The data engineers on the DARC team, are responsible for making sure the latest data from the venture capital data feed is available and ready to be analyzed by faculty. Maintaining large datasets that need to be updated regularly requires continual maintenance and technical expertise.

[7] Gornall, Will and Strebulaev, Ilya A., Squaring Venture Capital Valuations with Reality (April 20, 2019). Journal of Financial Economics (JFE), Forthcoming. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2955455

The data in the feed is just one piece of the story for the Venture Capital Initiative. Other data points researchers need is understanding the share types and the rules that govern them, this is important for understanding a company's true valuation. If a "preferred" shareholder gets paid first, then those shares are more valuable than a "common" share. The DARC team project managed work to extract data from Certificates of Incorporation documents so our researchers could start with a spreadsheet instead of a pile of legal documents.

With this full data, GSB faculty and their collaborators were able to compute updated valuations for so-called "Unicorn" companies worth over $1 billion, and determined that half were overvalued. Results like this are surprising and important for understanding the US economy.

The Research Hub also facilitates the sharing of this information. In November of 2018, the GSB hosted the Stanford Financing of Innovation Summit[8], bringing together investors and researchers to share insights in the venture capital market. Events like this one provide tremendous value and help disseminate research into the world.

The work at the Venture Capital Initiative has been widely covered in academic journals, trade publications and business news[9] around the world. Recently, research on the value of unicorns as on the cover of the *Economist*[10], providing key insights into IPOs of companies like Lyft and Uber.

Work in the Research Hub at the Stanford Graduate School of Business enables and accelerates research, helping to achieve the mission of the school which is to "Change Lives, Change Organizations, Change the World". As a result of the Research Hub organizational structures to support data needs, GSB students are able to access venture capital data in an online platform[11] for their research, and faculty are able to use the same data in feed in their research and teaching. The Research Hub's organizational structure allows for a collaborative, interconnected approach to support data is a network of interconnected units supporting research at Stanford's Graduate School of Business, and beyond.

Case Study: Stanford Libraries Data Workflow Task Force

Given the growing demand for data and services to support analysis and publication, the Stanford Libraries are reviewing how we manage our data. SL conducted a SWOT Analysis[12] and an environmental scan to help us understand our current strengths and limitations in the

[8] Stanford Financing Innovation Summit (November 2018).

[9] Sorkin, Andrew, How Valuable is a Unicorn? Maybe Not as Much as It Claims to Be, New York Times (April 16, 2017), https://www.nytimes.com/2017/10/16/business/how-valuable-is-a-unicorn-maybe-not-as-much-as-it-claims-to-be.html

[10] A Stampede of Mythical Proportions, the Wave of Unicorn IPOs Reveals Silicon Valley's Groupthink, The Economist, (April 17, 2019)

[11] Richardson, Nora, Venture Capital and Private Equity Research Guide, Stanford GSB Library.

[12] SWOT Analysis for Libraries (Libfocus, 2016)

data lifecycle management for both acquired data and research generated data; identify unmet needs and areas for improvement to focus on over the next one to three years; and lastly to identify partnership opportunities to either solve problems with our campus stakeholders or form allies to advocate for SLs central role in the data acquisition, data discovery, data preservation, and data services.[13]

In order to ensure long-term access and re-use, we have turned our focus towards developing the workflows and infrastructure to curate our acquired data.[14] We formed a Data Workflow Task Force to address these issues. The main work of the Task Force has been to map out the data acquisition workflow specifically to identify the gaps and pain points in acquiring, providing access, and archiving our acquired data, paying special attention to what metadata is important to capture along the way. This group consists of librarians and staff from collections, public service, data services, acquisitions, metadata, and our digital library program.

A "data inventory" or listing of data categories was created to map out the various types of data that we have, are, or anticipate acquiring. Here's a summary of the inventory:[15]

Data hosted by a vendor or third party	**Data the Library hosts**
Memberships, Non-profit organizations	Data on media (CD-ROM or DVD)
Subscription or purchase from a vendor or publisher	Subscription or purchased microdata/structured data from vendor or commercial entity
OpenData	Subscription or purchased unstructured data from vendor or commercial entity
	Digitized library collections for computation
	Stanford Generated Research Data
	Unstructured data from the web

This inventory was a useful starting point to help us evaluate and inventory the content, formats, license/access/sharing conditions. The list is further broken down into subcategories based on the source and publisher of the data, cost/licensing. Key characteristics of the data

[13] Take from Stanford Libraries Data Matrix, internal document created as part of the environmental scan.
[14] Data that are acquired with a variety of contractual issues that typically restrict access to our campus users only and require users to agree to a set of usage terms, e.g. an end user license agreements (EULAs). These data can be free, or cost anywhere from $5,000 to $200,000 USD, with file sizes ranging from 200 gigabytes to several terabytes; they are numeric, text, and audio/visual. Often faculty have specifically requested that our libraries acquire this data or subject librarians have identified datasets that would support research efforts on campus and acquired them as part of their overall collection development work. We want to ensure that we manage these assets in a way that ensures they are discoverable, usable, terms of use are clear, and when allowed, the data is properly documented and archived.
[15] Internal planning document: Categories of Data We Manage at the Libraries. January 2019.

(e.g. structured or unstructured, compute ready, End User License Agreement required) that impact overall data management, storage, and access were included. Common data characteristics SL hosts which present us with challenges are the licenses, lack of documentation and metadata provided by the vendor requiring library staff to do data processing and curation, which includes cleaning, data quality checks, creating technical and descriptive metadata, data documentation, (i.e. readme files, codebooks); often delivered on external hard drives sometimes via FTP; which then necessitates discovery and access mechanisms for the data. Like the GSB Library, we are working with companies who currently do not have good model licenses for the academic institutions, nor do they yet have processes in place to facilitate data transfer to us. We find ourselves working with these vendors and publishers to lay new ground for data acquisition and access in libraries.

To outline the data workflow process, each member of the team charted out the entire process of acquisition to access to archiving of a particular data set from the inventory. We evaluated numeric, textual, geospatial data types as well as data that had been acquired years ago and exists on hard drives, CD-ROMs, DVD's. We identified challenging, unclear, or areas we felt needed improvement and documentation. After each data set was evaluated, a subset of the team collapsed all of the information into a data workflow consisting of the following stages[16]:

Stage	**Key Components**
Identifying and evaluate data to acquire	assessment of content; does it align with overall collections in support of research and instruction?
License negotiation	terms of use including restrictions on sharing; End User License Agreement required; seeking to create a model license across the libraries for data to ensure key components like sharing for peer review are allowed; well-documented terms for both library staff and users.
Ordering and payment processing	Datasets pose challenges for tracking in ILS acquisitions module; how best to integrate acquisition and license terms and ensure this metadata is associated with data through all stages.
Data receipt; Quality Assurance; Structural analysis	DATA CURATION! Develop this workflow; data curator position to be filled.

[16] Internal document from the Data Workflows Taskforce, Workflow Stages Outline

Cataloging	Creating content specific metadata templates to ensure consistent description of our data; ties directly into previous. Can we incorporate DDI standards[17] into our metadata schemas; modelling this in a similar way that we have used the ISO 19115 Geographic Metadata Standard for our Earthworks geospatial metadata[18] .
Access	Data made accessible to patrons in a trusted manner in accordance with licensing terms; determine most appropriate access control system for patron access; documentation- document these controls and any technical development that was undertaken.
Management and permissions	Would ideally use SSO/Shibboleth when available, but not all external vendors can; integrate End User Agreements/Terms of Use into systems.
Use and portability	data are in accessible and usable formats that support interoperability; research teams have data physically stored in or accessible in their preferred compute environment within the limits of licenses.
Stanford Digital Repository (SDR) Accessioning (publishing and preservation)	data deposited in the preservation core (archival storage); publishing dataset file(s) with associated metadata have persistent URL where a user can access the material; license/sharing terms articulated

With each of these stages, we are developing a roadmap for creating a more FAIR-ified data discovery and access environment. The Task Force has been collecting information about workflow stages, actors, processes, and systems as well as the pain points associated with each. Our final effort will be to process and synthesize the information for presenting to our library directors and other collaborators so that these groups have a clear understanding of our recommendations and decisions that need to be made in order to advance the work, address the

[17] DDI Alliance Standards https://www.ddialliance.org/explore-documentation

[18] See https://www.iso.org/standard/26020.html and https://earthworks.stanford.edu/

pain points, and build out workflows and infrastructure. The team will identify and propose ways to address these issues by outlining the possible solutions and how best to test them out; what is needed to take these steps (money, time, staff, more research, more technical resources); identify the individuals and/or teams to advance the work; and offer an assessment of the risks and consequences are if no actions are taken.

Considerations and Conclusions

Libraries have historically established effective infrastructures to acquire, provide access, and preserve information sources in variant, ever changing format types. Today, we are evolving to meet the needs of our researchers that require a wide range of data sources. Whenever embarking on a significant shift in workflow and processes, especially when collaborating across organizations, anticipate that it will take more time than originally expected to accomplish your goals. Additionally, goals and priorities will shift as new learnings occur along the way. Both the SL and GSB Libraries are still identifying areas where we can combine our individual services and form a more collaborative Library Data Service.

Thus far we have found common ground for collaboration and efficiencies by developing a shared licensing model, jointly creating a working prototype for a dataset index to aid in data discovery and access and lastly to explore data storage spaces that can be utilized before data are ready to be published and preserved[19]. The GSB Research Hub has partnered with the Danish company, Consortia Manager to build an electronic resource management system (ERM) designed to support the data acquisition process as well as traditional library resources. Through an API we will be able to pull key information from the platform to incorporate it into a dataset index prototype to create efficiencies in the process. The GSB is improving features in their cloud computing environment, Cloud Forest, and in our on-premise servers that are provided in collaboration with Stanford Research Computing (SRC), and are also administering select datasets through a third-party, Redivis.

Both Libraries have prioritized building and collaborating on a set of model licenses for data acquisitions and piloting the use of ERM in a shared environment. The licensing metadata is highly important for determining access and archiving polices going forward. Our next point of collaboration is for SL to work with the GSB Library to determine if their technical infrastructure can be either implemented or replicated in our environment. Currently SL is evaluating platforms for data curation and access beyond the current set of tools; it is not clear that the third-party solution GSB Library is testing will work for the SL needs. As was noted earlier in the paper, we are attempting to utilize and build upon technical capabilities and workflows in the hopes of creating a more unified Library Data Service. Through these unified

[19] Both Libraries rely upon a well-established digital repository, the Stanford Digital Repository (SDR) which is file format agnostic and enables long-term preservation of all objects deposited. Research data that is ingested and managed in the SDR generally follows the FAIR principles.

efforts we will continue to improve and grow our data management services and data preservation between our organizations.

Acknowledgments

A special thank you to Alex Storer for his assistance with the storyline for the Research Hub case study. Thank you to Ilya Strebulaev, Faculty Director of the Venture Capital Initiative for linking the Research Hub nodes to engaging and collaborative projects. Deep gratitude and appreciation to Maureen McNichols, GSB Senior Associate Dean for Academic Affairs, for the support and guidance she provides to the Research Hub organization.

Thank you to Stanford Libraries Directors and members of the Data Workflows Task Force for their work and use of the internal documents in this paper and presentation.

References

Anonymous. (2015, May 11). FORCE11 Manifesto. Retrieved June 6, 2019, from FORCE11 website: https://www.force11.org/about/manifesto

Association of Public and Land Grant Universities. Public Access. Retrieved March 8, 2019, from https://www.aplu.org/projects-and-initiatives/research-science-and-technology/public-access/index.html

Carlson, J., & Johnston, L. (Eds.). (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. West Lafayette, Indiana: Purdue University Press.

Carnegie Mellon University Libraries. (n.d.). Preparing Your Data. Retrieved May 7, 2019, from https://library.cmu.edu/kilthub/prepare-data

Castiglione, J. (2008a). Environmental scanning: An essential tool for twenty-first century librarianship. *Library Review; Glasgow*, *57*(7), 528–536. http://dx.doi.org.stanford.idm.oclc.org/10.1108/00242530810894040

CLIR. (n.d.). Data Curation. Retrieved March 11, 2019, from CLIR website: https://www.clir.org/initiatives-partnerships/data-curation/

Committee, A. R. P. and R. (n.d.). *2018 top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education.* Research Planning and Review Committee, College & Research Libraries News. https://doi.org/10.5860/crln.79.6.286

Gornall, Will and Strebulaev, Ilya A., *Squaring Venture Capital Valuations with Reality* (April 20, 2019). Journal of Financial Economics (JFE), Forthcoming. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2955455

ICPSR. (n.d.). Data Management & Curation. Retrieved April 12, 2019, from https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/index.html

Indiana University Network Science Institute. (n.d.). CADRE: Collaborative Archive Data Research Environment. Retrieved May 24, 2019, from Indiana University Network Science Institute website: https://iuni.iu.edu/resources/cadre

Johnston, L. (2017). *Curating research data*. Retrieved March 11, 2019 from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596_crd_v1_OA.pdf

Kellam, L. M., & Thompson, K. (Eds.). (2016). *Databrarianship: The academic data librarian in theory and practice*. Chicago, Illinois: Association of College and Research Libraries, a division of the American Library Association.

Knight, M. (2017). What is Data Curation? Retrieved March 11, 2019 from https://www.dataversity.net/what-is-data-curation/

Libfocus. (2016). SWOT analysis for libraries - a compilation piece. Retrieved July 28, 2019, from https://www.libfocus.com/2016/11/swot-analysis-for-libraries-compilation.html

National Science Board, & National Science Foundation. (2005). Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. Retrieved March 12, 2019, from https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf

Open Knowledge Network. (n.d.). The Frictionless Data Field Guide. Retrieved May 21, 2019, from https://frictionlessdata.io/field-guide/

Pryor, G. (2012). *Managing Research Data.* London: Facet Publishing.

Ray, J. M. (Ed.). (2014). *Research data management: Practical strategies for information professionals*. West Lafayette, Indiana: Purdue University Press.

Ribeiro, C., Silva, J. R. da, Castro, J. A., Amorim, R. C., Lopes, J. C., & David, G. (2018). Research Data Management Tools and Workflows: Experimental Work at the University of Porto. *IASSIST Quarterly*, *42*(2), 1–16. https://doi.org/10.29173/iq925

Rice, R. (2016). *The data librarian's handbook*. London: Facet Publishing.

Sorkin, Andrew, *How Valuable is a Unicorn? Maybe Not as Much as It Claims to Be*, New York Times (April 16, 2017), https://www.nytimes.com/2017/10/16/business/how-valuable-is-a-unicorn-maybe-not-as-much-as-it-claims-to-be.html

Stanford Research Administration. DoResearch. Retrieved July 28, 2019, from https://doresearch.stanford.edu/research-administration

Stanford Research Administration. Uniform Guidance: Concepts That Have Changed. (n.d.). Retrieved July 28, 2019, from https://doresearch.stanford.edu/research-administration/major-topics/uniform-guidance-concepts-are-changing

Stanford Research Computing Center. Systems & Services Overview Retrieved May 28, 2019, from https://srcc.stanford.edu/systems-services-overview

Stanford University I.T. Services. Retrieved July 28, 2019, from https://uit.stanford.edu/services

SWOT analysis. (2019). *Wikipedia*. Retrieved from May 21, 2019 from https://en.wikipedia.org/w/index.php?title=SWOT_analysis&oldid=906110045

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, *27*(1), 23–44. https://doi.org/10.18352/lq.10180

University of Minnesota Libraries. (2015). The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/171761.

University of Oregon Libraries. (n.d.). Data Management Best Practices Guide. Retrieved July 28, 2019, from https://library.uoregon.edu/research-data-management/best-practices

Virginia Tech. (n.d.). Data Management & Curation. Retrieved April 12, 2019, from https://lib.vt.edu/content/lib_vt_edu/en/research-learning/research-data-management-curation.html

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.

Wilkinson, M. D., Sansone, S.-A., Schultes, E., Doorn, P., Bonino da Silva Santos, L. O., & Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. *Scientific Data*, *5*, 180118. https://doi.org/10.1038/sdata.2018.118

Young, A. (2019, April 24). Introducing the Contractual Wheel of Data Collaboration. Retrieved May 15, 2019, from Medium website: https://medium.com/data-stewards-network/introducing-the-contractual-wheel-of-data-collaboration-ca4c55938e7a