

UC Davis

E-Research

Title

The Question of the Question: Research Data Inquiries in Relation to Library Services

Permalink

<https://escholarship.org/uc/item/0389x43m>

Author

Michalski, David

Publication Date

2021-10-26

The Question of the Question: Research Data Inquiries in Relation to Library Services

David Michalski

5/27/2020

Abstract

New attention to the power of data in research has brought new kinds of data questions to the university research library. This study seeks to understand the character of research data questions in order to help research libraries develop the structures, technologies, collections, and professional skills needed to meet the needs of the research communities. It employs two original metrics to analyze 42 cases: The Data Question Typology, which allows for the organization of data questions into categories based on researcher objectives, and the Modified READ Scale for Data Questions (MRSQ), which is used to record the magnitude of difficulty presented by each case. It finds that data questions differ significantly across academic fields and that successful research assistance often requires partnerships between subject specialists librarians and technological or computational experts. It concludes with a recommendation on how research libraries can facilitate a collaborative process and workflow for handling a diversity of data questions from across the university.

The Problem

Data questions are not new to the library. Along with journals, monographs, media, and archives, research libraries have been collecting, providing access to, and fielding questions about datasets and data sources for decades. New attention to the power of data in research, however, has brought new kinds of questions to the library from across the university. New methods of data collection, data analytics, and data visualization, coupled with new mandates to share and preserve data, generate inquiries and demand for library support. Finding and evaluating data sources is only one category for the questions the research library now encounters. Researchers are looking to libraries for strategies and assistance as they seek to best share, preserve, manipulate, and analyze the data sources they examine and create during their projects. Researchers and students are also looking for training opportunities that can improve their data literacy and tech skills as they seek to best position themselves and their work in a data-rich research environment and work force.

To handle the diversity of data questions, the library draws on a variety of experts tasked with providing support across the research lifecycle, including data curators, database programmers, data scientists, subject specialist librarians, and other professionals. While this matrix of support

for data questions allows the library to respond to calls for a wide range of services, it can lead to organizational problems, such as the confusion about the division of labor between different library units, and how human resources can best develop the needed expertise for data questions.

Both library patrons and library staff may be confused about what services are offered in response to data questions, and about where the expertise for particular questions can be found. The pace of technology may make some data questions appear novel, but Library staff can also become confused about how responsibilities and expectations are distributed among the library personnel and units. The increased focus of management on data questions aims to alleviate this organization problem. Still, the multifaceted character of the contemporary data question can test a library's ability to give advice, make referrals, or create interdepartmental support teams to keep up with the new demands of scholars.

An understanding of research data questions and their relation to the library's suite of skills, tools, and services can lead to better support, and better referrals and communication between subject and functional experts. Setting clear expectations for both the user and the library staff can also reduce the number of instances wherein a researcher either misses opportunities offered by the library or comes away disappointed by overpromised and under-delivered data support. A better understanding of the data questions coming into the library, and how they are handled by the library's researcher support matrix, can also help the library administration assess whether or not the library structures, technologies, collections, and professional skills are meeting the needs of the research communities it seeks to support.

The Investigation

With these goals in mind, we have developed a study to better understand three questions related to today's data queries:

1. What type of data questions do we receive from the campus community? And which departments/disciplines are asking the different kinds of data questions we receive?
2. Where (that is, by which library unit) are questions being addressed, and what kind of expertise is being brought to bear on different types of questions?
3. How successful is the library in meeting the needs of the researchers asking data questions, that is, does the library have the right tools, collections, services, and levels of expertise in place? How proficient is the library in making referrals and assembling effective interdepartmental teams to respond to data questions?

By tracking and analyzing data questions, this study seeks to provide the information that can help the library better understand our communication and workflow processes in general, and

aid in administrative efforts to make interdepartmental communication and support services more effective and efficient in the area of data questions.

Methodology

Case Collection

We analyzed data questions and responses from across the research university. These are called **cases**. Questions were received through the Library website. They were entered as cases (n=42) either by patrons (**researchers**) onto two web forms found on two different library webpages. One page communicated the library's data management services, and the other communicated the library's data services for the social sciences. Both sites promoted a wide range of data services, including assistance with data discovery, extraction, analysis and visualization, data management, and data sharing and preservation. The web intake forms on these pages routed questions to two corresponding request ticket queues (**Queues**) in a request ticket system we call a **case database**. The data management services website was routed to a data services queue, and questions entered on the data services for the social sciences website routed to a social sciences data queue. The case database was monitored by library professionals making up a **data services group**. The data services group, was an in-part assigned, and in-part a voluntary association of **library professional staff**. The membership of this group was somewhat differentiated in-line with the two request ticket queues that were created to mark the different web-intake forms. A breakdown of the data services group (**Appendix E**) shows the distribution of library professional staff by queue, library administrative position and unit. Each queue was assigned a **case monitor**, who helped to shepherd, with contributions from the data services group, each case from intake to resolution. This included helping, along with the data services group, to assign a **case lead** who was made primarily and centrally responsible for the case and communicating directly with researcher. The case monitor also tracked the progress on cases, served as the case lead when appropriate and by default, and updated the case database with correspondence the case. The data service group members were also responsible for commenting on cases and forwarding case leads or leading cases.

By using the case database to gather, track and discuss questions and responses, all members of the data services group were able to comment, update and contribute to case resolution. More information on the workflow can be found in **Appendix C**. 42 cases were in the case databas were recorded over a period of roughly four academic quarters between July of 2019 and May of 2020. These 42 cases provide the data for this study.

The library received many more data related questions than the 42 that inform these cases. Questions reach the library through a variety of paths. Librarians and other professional staff receive direct emails, encounter questions in research consultations, and through other formal and informal contacts with researchers. The 42 cases tracked through the case database studied

here do not represent the breadth of data questions received either through direct correspondence with research librarians, or through other data-centric library and university support units, which advertise data services and attract research questions and collaborations. Further research on the library's data questions and responses from other intake streams is needed to assess whether the findings about these 42 cases might be indicative of wider trends. Difficulties in obtaining question and response data from across the library personnel and services units qualifies the representative character of these 42 to cases. This study, however, was able to conduct a thorough and multifaceted analysis from this limited set. It also provides a model for how question typology, demography, and question difficulty can be studied by use of a qualitative evaluation of case correspondence.

Coding

For each case, we analyzed the question and responses by applying three instruments. 1. The **Research Data Question Form (RSDQ) (See appendix A)** This form helped us collect the metadata for the cases in the case database, with fields to record the case lead, the affiliation of the researcher, the type of data question involved in the case, elements of collaboration among those assisting in the case, the relative case difficulty, and other attributes. 2. The **Data Question Typology (See Appendix B)** The typology allows for the organization of data questions into categories based on researcher objectives. 3. The **Modified READ Scale for Data Questions (MRSQ) (See Appendix D)** This scale allowed us to record the magnitude of difficulty presented by each case. (See Appendix D)

The information we gathered on these 42 cases information was then divided into the following areas of analysis:

Section 1. Information about the researcher who put forth the data question. This information includes the affiliation or status of the researcher, and the department or academic program to which the research belongs.

Section 2. Information about the question. This includes the actual question, a designation of its 'Type', based on how it fits into our Data Question Typology, and the level of difficulty it presented as measured on a Modified READ Scale for Data Questions.

Section 3. Information on how the ticket was resolved. This information includes who was consulted or brought into a support team, and why. It also includes information about any external referrals that were made, and comments on the library's resolution or response to the question.

Results

Cases by Type (n=42) (See Data Question Typology, Appendix A)

(Percentage (rounded up)) (Number of Questions) (secondary classification)

A. General Research Project Consultation

0% 0 0

B. Data Collection

0% 0 0

C. Data Discovery, Evaluation and Selection.

21% 9 (9 Academic/Subject, 1 also Technical/Computer)

D. Data Access and Extraction

12% 5 (4 Academic/Subject, 4 also Legal, 2 Technical/Computer)

E. Data Formatting, Cleaning and Sequencing

0% 0 (0)

F. Data Visualization and Analysis

5% 2 (2 Technical Computer)

G. Access to Tools and Training in Data Science, Programing, and Software/Tools

5% 2 (2 Technical/Computer, 1 Academic/Subject)

H. Data Publishing, Sharing and Preserving

19% 8 (6 Academic/Subject, 2 Technical/Computer)

I. Copyright, Licensing, and Privacy

2% 1 (1 Legal, also 1 Library Budget Purchasing)

J. Data Management Plans

19% 8 (3 Academic/Subject)

K. Database Construction and Development

5% 2 (2 Technical/Computer)

L. Other

2% 1

M. Not a Data Question

10% 4

Cases (42) by Disciplinary Sector

Social Sciences (Includes 1 Design)
22 (52%)

STEM
17 (37%)

Law
1 (3%)

Health Sciences
1 (3%)

No Subject/Discipline
2 (5%)

Cases (42) by University Affiliation (* recount may be needed)

Professor
37 (88%)

Graduate Student
3 (7%)

Outside University Affiliation
2 (5%)

Cases by Average Difficulty (*See Modified READ Scale for Data Questions, Appendix D)

C. Data Discovery, Evaluation and Selection.
4.9

D. Data Access and Extraction
4.2

F. Data Visualization and Analysis
6

G. Access to Tools and Training in Data Science, Programming, and Software/Tools
3

H. Data Publishing, Sharing and Preserving

4

I. Copyright, Licensing, and Privacy

6

J. Data Management Plans

3.4

K. Database Construction and Development

6

L. Other

5

M. Not a Data Question

1.5

Cases (42) by Resolution Type

Answered
28 (67%)

Referred
11 (26%)

Stalled
3 (7%)

Referred Cases (11) by the Department to which Case Was Referred Includes breakdowns of the case types referred

Datacenter
5 Total

- 1 Data Visualization and Analysis,
- 2 Data Access and Extraction,
- 2 Access to Tools and Training in Data Science, Programming, and Software/Tools

Blaisdell Medical Library
1 Total

- 1 Data Publishing, Sharing and Preserving

Scholarly Communication Office
1 Total

- 1 Copyright, Licensing, and Privacy

Other (Researcher Referred to Textual Information)
4 Total

Discussion

1. What type of data questions do we receive from the campus community? And which departments/disciplines are asking the different kinds of data questions we receive?

It is a common understanding among reference librarians that the questions received from patrons can often change and develop during the reference interaction. One goal of the reference interview is to help researchers develop more complex and detailed questions. As such, through the course of this study, we see the web intake form being used as a means to record the

development of the question between the researcher and the library's professional staff, as well as, a conversation about the question between professional staff with different expertise. At the resolution of a case, when the trail of conversations ended, and the research moved on to the next steps in their projects, it was possible to look back and classify the interaction. While our coding system allowed cases to be tagged with multiple question types, we did not find cases being tagged with more than one type.

The character of the questions we received in our two queues, Data Management Support, and Library Services for Social Science Data, fell across our Data Question Typology, with three exceptions: "General Project Consultations", "Data Formatting, Cleaning and Sequencing questions", and "Data Collection" questions, which were noticeably absent from the support tickets. The absence of these question categories is addressed below.

The two most prominent categories of cases in the support tickets were questions about "Data Management Plans" (19%), and questions about "Data Discovery and Selection" (21%), with the "Data Management Plans", dominating the Data Management Support web intake, and "Data Discovery and Selection" dominating the Library Services for the Social Sciences web form.

The questions were nearly equally divided by the disciplinary groups of the researchers. (Social sciences 52%, STEM 40%) It can also be observed that the STEM questions came exclusively through Data Management Support web intake, and that many, but not all, the Social Science questions came through the Library Services for the Social Sciences web form. While much of this breakdown corresponds to the way the Library has historically structured its outreach and web presence, we did find social scientists were more interested in "Data Discovery and Selection" (9/9), while STEM Researchers are more interested in "Data Management Plans" (7/8), an observation, which may reflect that higher use of shared public datasets in the social sciences, and the rise of data management plans as a requirement in science grant founding. Cases categorized as "Data Publishing, Preservation and Sharing" were evenly split between the social sciences and STEM disciplinary groups, showing that while formal data management plans may not be prevalent in the social sciences, data publishing, preservation and sharing is a concern.

Questions about "Data Discovery and Selection" (21%) have long been supported by libraries, so it is not surprising that we find those questions making their way into the web forms. In this study we see them primarily being asked by researchers in the social sciences. We also see a significant number of cases coded as "Data Access and Extraction" (12%) being asked by social science research. These questions were often resolved by a familiarity with a mechanics of a publicly available dataset, and, at other times, required the manufacture of a customized user interface or API.

It is noteworthy that questions asking directly about "Data Management Plans" featured prominently. If we add those cases to the category, "Data Publishing, Preservation and Sharing" (19%), this study seems to validate the findings in the library literature that show researchers are

turning to the library for assistance in meeting the demands for greater transparency regarding data created and assembled by researchers.

On the questions that did not feature prominently via the web forms.

The Library website data question intake forms only capture a small fraction of the questions presented to the Library. Public service desks, direct emails, online chat, telephone calls and office hours consultations take in data related questions. The two web forms, which have only been available for a few years, are less established, and may be viewed as a less direct connection to service areas or service professionals. From anecdotal information from the library professional staff, we do know that “General Project Consultations”, “Data Formatting, Cleaning and Sequencing”, and “Data Collection” questions do reach the research library. It is expected that email records and other documentation can attest to substantial amount of support in these and other areas. “General Project Consultations” are a staple of the library's Research Services unit and the library's Datacenter, the library unit originally created to assist in the application data science in all research areas. Both units have instituted regular consultation office hours and other mechanisms for scheduling appointments. In fact, Datacenter and Research Services have engaged in a robust collaboration on “General Project Consultation” questions in recent years.

The support for “Data Formatting, Cleaning and Sequencing” and “Data Visualization” (5%) are core functions of Datacenter, and feature prominently on their outreach material and website, and as such, these inquiries are most likely made directly to Datacenter. We also see a dearth of data questions from Health Science researchers, even though we know questions about data, and data management protocols are important in health science research. These questions exist. They reach the library directly through the medical library. A deeper dive into research data questions will need to look more broadly across library intakes.

Similarly, “Data Collection” questions did not make it into our sample. These questions, which are about the support and training available to collect new data, including the evaluation and use of collection methodologies, survey instruments, and assays, are often disciplinary-specific. When they do reach the library, they have been fielded by subject librarians, who commonly refer patrons to the methodological literature in a particular field.

2. Where (that is, by which library unit) are questions being addressed, and what kind of expertise is being brought to bear on different types of questions?

The data questions that came through the webforms were handled in a collaborative manner. Each case was assigned a case lead to make sure it met its eventual resolution: either by providing a solution to a problem, or a referral, which placed the researcher in a better position to meet the researcher's data challenges. Queue monitors were responsible for making sure incoming

questions were addressed in a timely manner with input from the most appropriate library staff. Many times, queue monitors also served as the case lead. The **Flowchart (Appendix C)** illustrates the collaborative process by which addressed. Our review of the cases shows, some cases were relatively routine and straight forward, while others required more internal consultation among the library's professional staff, in-person (or remote) discussions with members of one or more library units and/or the researcher, who posed the question.

The primary library units involved in the set of cases in this study were: Research Services, the Data Management Program, with substantial support from Datacenter, and some additional consultation provided by Collections, The Scholarly Communication Office, and Digital Strategies. It must be noted that, through the course of this study, the Library began transitioning the services provided by its Data Management Program, initially housed in Datacenter to Research Services Department. For this reason, cases which presented as "Data Management Plans" or "Data Publishing, Preservation and Sharing" were increasingly led by professional staff in the Research Services Department, with considerable consultation from the former Data Management Program lead, who has been transferred to Online Strategies.

One way of understanding the collaborative process used by the library professionals who handled the questions coming through these queues is to look at the expertise that was brought to bear on the cases in this study. This study utilized the following categories to score the cases by the kind of expertise required for their resolution: "Technical/Computer", "Legal/Rights", "Academic/Subject", and "Purchase/Library Budgets". (See Appendix B for details.) While data questions often require a base-level of understanding in all these areas, we used these categories to mark only the cases where specialized expertise was necessary for their resolution. One will note, for instance, that the "Data Management Plan" questions were only seldomly scored with an additional level of expertise. The cases categorized as "Data Management Plans" did require a learned familiarity with best practices in data management and some knowledge of the emergent policies associated with different funding agencies from different disciplines, but the cases presented in this study, which were relatively straight forward questions about how to satisfy plan requirements or achieve the FAIR data principles (of findability, accessibility, interoperability, and reusability) did not require additional specialized expertise matching our expertise categories. On the other hand, questions about "Data Publishing, Preservation and Sharing," a broader aspect of data management, which can encompass the selection of disciplinary repositories, metadata decisions, or file exchange, did require additional expertise in the "Technical/Computer" and "Academic/Subject" knowledge areas.

Notably, this study recorded a high number of questions that required "Academic/Subject knowledge (22/42), including all of the "Data Discovery and Selection" questions (9/9). This validates a position that data questions, like most other informational queries that reach the library, are largely embedded in social or disciplinary practices, especially when it comes to the discovery, evaluation and selection of the appropriate data sources for an academic researcher. The other prominent type of expertise required on these cases was "Technical/Computer"

knowledge, which was required on 11 out of 42 questions. As to be expected, we see that data questions often require a combination of “Academic/Subject” knowledge and “Technical/Computer” knowledge. Addressing this challenge is a central concern of libraries providing research data services.

Digging more deeply into the cases, one can read that when “Technical/Computer” expertise is required, the cases demand expertise in machine programming to manipulate data flows, execute computations, or to insure data stability and integrity across data platforms. In the cases where “Academic/Subject” knowledge expertise is required the cases demanded a knowledge of the scope, coverage, or provenance of data sources in relation to the situated goals of the researcher. Understanding the context in which data was collected is most important for “Data Discovery and Selection”, but it is also important for “Data Access and Extraction” (4/5 Academic/Subject, also 2 Technical/Computer, Legal, 1/5) and “Data Publishing, Sharing and Preserving” (6/8 Academic/Subject, 2 Technical/Computer), where a clear understanding of social context of data flows and future data-use can positively influence the selection of preservation sites, and metadata strategizes, as well as the technical and computational choices necessary to address data sharing and preservation problems.

This study shows less “Academic/Subject” expertise was necessary to address cases that we categorized as “Database Construction and Development”, “Access to Tools and Training in Data Science, Programing, and Software/Tools” or “Data Visualization and Analysis”. In reading the transcripts of these cases, one sees that the researchers who put forward their questions had developed a strong understanding of the academic or social context of their data sets and their relation to other datasets in their field before the presentation of their question to the library. As a result, the expertise required in these cases was largely “Technical/Computer” expertise. When “Academic/Subject” knowledge was required on these cases it was to assist “Technical/Computer” experts in their efforts to understand and meet the aims of the discipline. Although this does not show significantly I this study, it is expected that a deeper look into questions presented to Datacenter in the area of “Data Visualization and Analysis” will show considerable “Academic/Subject” knowledge applied, either by the researcher, by Datacenter’s professional staff, or in consultation with Research Services librarians.

The study recorded less of a need for specialized “Legal” or “Purchasing/ Library Budget” expertise, but one does see these categories arising most in the “Copyright, Licensing, and Privacy” and “Data Access and Extraction” cases. A reading of these cases finds that researchers looking to obtain and analyze data can be confronted with difficult access questions, some of which require an expert interpretation of licenses, while others require strategies for acquiring licensed access and use rights (such as Text and Data Mining rights).

Again, this study does not reflect the quantity of questions that arise in these areas. We know management and interpretation of copyright and licenses, and library budgets make up a core function of the research library. Current knowledge of library practices among the data services

group helped to resolve many of the data questions without calling on specialized expertise. Information access and rights issues are, however, some of the most dynamic issues facing the contemporary research library, and continual dissemination of expertise in these areas is crucial to the library's ability to respond to data questions. High-level, innovative research can also push libraries to revise their strategies and responses. This has been the case with the most novel or avant-garde data questions, which can live on after their resolution in our case database to inform future library policy and organizational positions.

3. How successful is the library in meeting the needs of the researchers asking data questions, that is, does the library have the right tools, collections, services, and levels of expertise in place? How proficient is the library in making referrals and assembling effective interdepartmental teams to respond to data questions?

This study did not include a user-satisfaction survey, but efforts were made to include all the correspondence between researchers and the library staff. Reading these exchanges gave us some insight into the library's ability to meet the needs of the researchers who presented data questions. Nearly all the questions met a resolution. The resolutions were grouped into three categories, those that were "answered", those that were "referred" and those that "stalled". In most cases, the problems presented by the case were 'answered' through the interaction with the case leader. (28/42, 67%) In other cases, the researchers were 'referred' by the case leader for additional consultation and services to another group, individual, or to more information that could be used help the researcher better overcome the problems they presented to the library. (11/42, 26%) There were also a few cases, which we marked as 'stalled' (3/42, 7%) In these, the case leads correspondence with the researcher dropped off without a definitive resolution. Again, because this study did not perform a user-satisfaction survey, "answered" does not necessarily mean the researcher was happy with the answer provided. Similarly, "referred" does not mean that the library could not satisfy the researcher, or that the researcher took up the referral. That is, as with other academic library reference assistance, the 'answer' provided may not be the one a researcher wanted to hear, and a 'referral' may be exactly what a researcher desired. Outside of the few disconnected exchanges marked as stalled, reading the cases makes us confident that the researchers came away from their interaction with the library better informed about our services and abilities, and better prepared to engage in the next step of their research process. The case transcripts also gave us a sense of the difficulty involved in each question. It is true that even a simple case can be difficult to a person unfamiliar with a problem presented, and a complex case can seem routine to those with familiarity. To measure the difficulty or complexity of data questions, we modified a scale that librarians have used to measure the difficulty of reference questions. The magnitude scale for reference questions is called the Reference Effort Assessment Data Scale (READ scale). It was developed at Carnegie Mellon University to add a relative assessment of question complexity to library reference/research assistance statistics. Our modified version of this scale, the Modified READ Scale for Data Questions (MRSD) was use to assess the complexity of the 42 cases in our study. **See Appendix D.** Both scales rank the

complexity of interchange between 1 and 6, with 1 designating the easiest and 6 the most complex. As one moves up these scales, there is a need for more and more consultation with the researcher, more trial and error, more strategic decision-making, more nuance and technical or craft skills, and more collaboration with local and distant colleagues. It is important to note that context matters. Even basic questions in any research environment require a learned ability to understand the researcher as well as a familiarity with the setting in which they work. Context differentiates the READ scale, which was designed for a public service reference desk at a large research university, and the MRSD, which handles a very particular set of research questions, those which tend to arise only the course of complex research projects. For instance, level 5 or 6 on the READ scale is an appropriate designation for questions that are commonly presented by graduate student and faculty in the course of prolonged research. In this study, nearly all of the questions in our study were presented by graduate students or faculty members (40/42) who were conducting sustained research projects. As such, the data cases studied here, which are categorized as 2 or 3 on the MRSD, might be scored as 4 or 5 on the READ scale. With an understanding of this range difference, one can get a sense of how much time and collaboration is necessary on the variety of cases that were presented.

Our study found that two of the most difficult question categories were: “Database Construction and Development” and “Data Visualization”. These cases often brought-in expertise from Digital Strategies or Datacenter. Although we did not receive a lot of these from the web forms, the cases in these areas evolved into sustained consultations and collaborations, where computational programming, knowledge of the research area, and questions about library policies and service models were brought into play. “Copyright, Licensing, and Privacy” also scored at the higher level of difficulty as the one question we received in this area was difficult to untangle given current library and university policy information.

Looking across the data question type categories one may be struck by the high average: 4.7. It is hard to make too much out of this. These high scores on the MRSD scale may be a result of the low number of cases in these categories. A larger pool of cases can show whether the protracted collaboration, or consultation needed for these data cases will be sustained with a higher case load. The lowest difficulty rating can be found in “Data Management” 3.4 and “Access to Tools and Training in Data Science, Programming, and Software/Tools” 3. Are these categories simply easier? When one looks at the transcripts, one sees that the library was better prepared in these areas to refer researchers to the appropriate pre-prepared guides. It may be, instead, that anticipation and prep-work can make some areas of research data assistance more efficient.

The numbers here are also the result of our subjective application of the scale. It is undoubtedly based on our position and frame of reference, and most likely reflects both the library’s relative inexperience in handling data questions, as well as the faculty and graduates students abilities to keep pace with fast-changing change and diverse research data practices. The assignment of a score most concretely reflects the level of consultation and collaboration required on data questions at the present time, within our current structure. As similar questions arise over time, the library’s efficiency can improve, and lower average difficulty. Currently, however, our study

shows that data questions continue to challenge the library's ability to assemble consultation teams, and its ability to revise and adapt service models to meet the data problems researchers present. Some steps for improvement to the case database system and data services group are made below.

Still, looking at the case transcripts, one can see that the library was proficient at creating the teams necessary to address data questions. As indicated above many cases required a combination of expertise areas, especially a combination of "Technical/Computer" expertise and "Academic /Subject knowledge". University libraries benefit from a collaborative culture and this is especially true as research problems become more complex. There is a tradition and practice of informal meetings, asking questions, seeking advice, sharing what one knows, and helping one another in the service of researchers. This collaborative spirit also exists among technologists. Like traditional librarians, programmers and data scientists are known for adopting collaborative platforms. One sees this in the collaborative development of software and computational tools. This 'open' environment has proved crucial to research assistance when it comes to data questions.

The development of the case database (See Appendix C) used throughout this study enabled this collaborative culture to shine. Most importantly was its use in designation of a case lead, and the assembly of ad-hoc groups to support the case lead. The data services group was made up of a combination of different subject librarians, data scientists, and the administrators of different data services units. It served both as a consultancy and a reserve of potential case leads. The transcripts show a considerable amount of back and forth, not only between case leads and the researchers, but also between members of the data services group.

Despite this collaborative culture and useful communication system there is evidence that improvements can be made. The obstacles to better communication derive mainly from a lack of a shared understanding of the library unit service models and uneven policy development. The main cause of these problems (as mentioned above) is the relative newness of data services in the library structure, and the rapid advance of data research practices, which are acquired, mastered, and deployed differently by different researchers. The cases show that even simple referrals and answers can sometimes require substantial conversation, either between the case lead and the researcher, in the attempt to develop a shared understanding of the problem, or between members of the data services group, as they sought to develop a shared understanding of responsibilities and library services. While the cases in this study mostly achieved a timely resolution, the inefficiencies caused by the diverse and sometimes inarticulated service models may tax the system, should the case load increase.

The data services group involved in the cases presented in this study was composed of two somewhat overlapping membership groups, (**Appendix E**). The composition evolved somewhat organically and historically to address questions which came into two different intake queues: the Data Management queue, and the Library Services for Social Science Data queue. While it may

be necessary to replicate multiple intake forms on the web to maximize outreach to different communities, it does not appear necessary or efficient to maintain divisions in the case database or in the data services group. Given the number of questions coming in, in-relation to the number of staff acting on the cases, a re-composition of the data services group responsible for monitoring a universal case database seems in order. Results of this study suggest a higher level of collaboration and efficiency can be achieved by reducing the current two queues to a single queue, called the Research Data Services queue, and by composing a single data services group, with proper representation from the units and experts required to field the cases presented to the library. The suggested composition of a new data services group can be found in **Appendix F**. Important additions are: increased representation from Research Services (to enhance subject expertise) and new representation from Medical Library, Library Collections and Budgets and Scholarly Communication units to attend to an anticipated increase in data publishing or data extraction cases that either require an expertise in data privacy, access rights, or the economics of public use licenses.

The development of a central web portal for Research Data Services on the Library webpage can also improve our outreach for data services to the campus, and help researchers (and library staff as well) better understand the various services offered by the library and its units. This site would employ a central intake form for data questions that would feed into a single case database, that would be monitored by a united data services group, and it would also provide descriptions of the different library service units and their specialties and contact information.

A better understanding of what members of the data services group and their units do is important, but it is not the only problem research data services face. The misunderstandings observed may also be made worse by the divergent skill sets among library professionals. The same diversity that powers successful team composition, can inhibit timely team formation. Skill development is a common, ongoing problem in a progressive technological environment. In research data services, however, this problem is compounded by the rapid development of research data practices. Cases in this study show a need for both bespoke data solutions and boutique consultative models, as well as an increased demand for novel areas of support. Preparing for future demands by gauging which of the services demanded today by the most innovative or experimental scholars will become widespread is problematic. Scaled-up models (and prep-work) may not be responsive enough to unique and changing researcher demands. In addition, history shows that the most useful and widespread technical and computational services may become eclipsed by market solutions, and technological fixes. Libraries which offer both innovative and cutting-edge services, while expanding and scaling-up baseline services, will need to communicate and plan effectively, while avoiding the risk of creating a divide that inhibits the smooth transfer of skills from the cutting edge to mainstream.

Library service units, and especially research data service units, cannot afford to be ridged or static. Research data services must grow parallel to the needs of the research community, keeping up with changes in technology, changes in research interests, new understanding of data theory

and research methods, changes in availability of data sources and repositories, and changes to data policies and standards. Planning and nurturing this growth is dependent on not only our ability to track the questions we receive and their resolutions, but our ability to understand how such questions unfold within the library's organizational structure, so we can better anticipate, coordinate, and centrally direct our response to changes in the research community. The objective of this study was to provide useful information for this work, and to develop a model to study research questions more broadly.

Conclusion

In this limited study we find that data needs differ across different academic fields. Broadly, social scientists require assistance in data discovery, selection, evaluation, and in data extraction, while researchers in the STEM fields need support with data management plans, and are concerned about the best practices for data sharing, data publishing and preservation. Noteworthy is the number of cases coded as "Data Access and Extraction" filed by social sciences. This lends support for increased assistance with extracting and sharing previously published data, a task that seems at the core of the library's traditional ethos. We also learned that library is approached to address difficult questions about data visualization, and database construction. A broader study of data questions across the library (and across the university) can test these observations. What is known is that data questions require a mix of expertise and a substantial dialogue between technical and computational specialists and those with subject knowledge.

As the library prepares to meet the future of research data services questions, it must look for ways to bring together subject expertise and computational and technical skills, without hampering the professional growth of these important knowledge areas. Active and horizontal communication systems need to be developed to share problems, insights, questions, and solutions that arise as research data questions are presented. These questions or cases ought to be recorded as a means toward better understanding of what researchers need, and how well the library can respond to researcher needs. This study shows research data questions are dynamic and varied, making collaboration necessary for resolving data cases, and communication crucial for planning and nurturing professional services and skills. Differentiated library research services data units must continue to interface directly with researchers, and to cater and advance their unique services to the specialized needs of researchers, who remain embedded in distinct research environments. The application of a diversity of expertise, coordinated and lead by a chosen case lead, proved to be an effective way of marshaling the power of the library to help solve researcher data questions. The inchoate unit service models, however, can lead to prolonged consultations and confusion, the duplication of services, inappropriate referrals, or a misunderstanding of services and roles. A central research data services case database, with appropriate representative data services group overseeing it, is recommended as a means to link various library units. The version we used on the cases in this study served as a practical means to communicate the differentiated centers of expertise and unit service priorities, with benefits

to the researcher and to the data services group members as each involved learned more about what each other was doing and why. This study also provided some improvements to both the case database and the composition of the data services group, which we hope will accentuate the benefits it can provide.

In constructing and establishing this case database system, the library must be careful not to conceive it as a way to centrally manage or regulate service roles of various library units. Library data service units must be free to evolve their areas of expertise and respond to their studied observations of research trends and data innovations. Instead, the case database system best serves as a platform for active informal collaboration, an efficient and personalized way for a single case lead to coordinate shared library expertise in exchange with a researcher, as well as a means for gathering information for the library's future analysis of data question trends, one that can inform both individual data service units, and library-wide decision-making.

Appendix A

RDSQ: Researcher Data Services Question Form

David Michalski
10/24/2019

---To be entered by the system as ticket is generated---

1. Time and Date: [Stamp]
2. Question Number: [Accession number]

---To be entered by patron through a web form or by a librarian/moderator---

3. Name (Patron): [Text]
4. Email (Patron): [Text]
5. Campus Dept or Program. (including NA):
6. the university Affiliation: [Multiple Choice]

- Undergraduate Student
- Graduate Student
- Post-Doctoral Student
- Faculty
- Staff
- Other Affiliate
- Not Affiliated with the university

7. The Question: [text]

--- Office: To be entered by librarian/investigator---

8. Working Title: [optional title for project/case]
9. How did question come in?

- Web Form from Patron
- Email to _____
- Phone to _____
- In-Person to _____

_Other _____

10. Question Categories (See Appendix B for explanation of types)

- A. General Research Project Consultation
- B. Data Collection
- C. Data Discovery, Evaluation and Selection.
- D. Data Access and Extraction
- E. Data Formatting, Cleaning and Sequencing
- F. Data Visualization and Analysis
- G. Access to Tools and Training in Data Science, Programing, and Software/Tools
- H. Data Publishing, Sharing and Preserving
- I. Copyright, Licensing, and Privacy
- J. Data Management Plans
- K. Database Construction and Development
- L. Other
- M. Not a Data Question

Secondary Categories: Free Floating Sub-categories. Use these when appropriate to further describe question.

- a. Technical/Computer
- b. Legal/Rights
- c. Academic/Subject
- d. Purchase/Library Budgets

11. Lead assigned/Question Coordinator

Name, (from Library Staff Directory)
Library Department, (from Staff Directory)
Expertise (text)

12. Library Support Team Consulted: (Can include multiple people)

Name, (from Staff Directory)
Library Department, (from Staff Directory)
Expertise (text)

13. Internal Library Referrals: Library Department to which patron was referred for further assistance

Multiple Choices Acceptable:

- a. Datacenter
- b. Student Services
- c. Archives
- d. Medical Library
- e. Scholarly Communication Office

- f. Collection
- g. Digital Strategies
- h. Other (text)
- i. To individual: Name, Email, Expertise.
- j. Other

14. If an external referral was made, to who or what organization was the patron directed for further help: _____ (text)

15. Notes on Interactions: [text]

16. Status:
_Ongoing
_Resolved
_Suspended

17. Notes on Status: [text]

18. Difficulty/Complexity

Refer to Modified READ Scale for Data Questions : Multiple Choice : 1-6.

19. Notes on the length of interactions with patrons: [text]

20. Additional Notes for Research Services survey team: [text]

--- Question trail: Cut and Paste interaction for context. To be entered by librarian/moderator---

21. Collected email correspondence. [cut and paste appropriate email]

22. Internal comments related to inquiry. [cut and paste appropriate comments]

Appendix B

Categories for Data Question Types

David Michalski
10/24/2019

This typology was developed to help us categorize the data-related questions that come to the Library. It contains Primary Question Categories, and Secondary Question Categories.

Questions may fall into multiple categories, so please mark all that apply.

Primary Question Categories

A. General Research Project Consultation

These inquiries focus on the researcher's project in general, rather than on a specific aspect of the project. In these consultations a librarian learns about the aim and scope of the project and to introduce and discuss the various options for support the Library can provide across the course of the project. These consultations require good reference interviewing skills, knowledge of the researcher's field, as well as, a clear understanding of various services the Library provides across its organization.

B..Data Collection

These questions are about the support and training available to collect new data, including the evaluation and use of collection methodologies, survey instruments, and assays. They require knowledge of the field of study, and the data collection practices of researchers in the field.

C. Data Discovery, Evaluation and Selection.

These inquiries are centrally about finding and evaluating potential data sources to match the needs of the researcher's project. They often require a disciplinary knowledge of the researcher's academic field, the sources used in that field, and the options available for access to data sources in the field (including those provided by the Library/University. Librarians fielding these questions will often need a familiarity with government information, and other public repositories and data clearinghouses, as well as the information marketplace. The role of librarians is to both help the researcher make informed choices and communicate the needs of the researcher to the Library. Questions about the purchase and licensing of data sources often require consultation of Collection Department.

D. Data Access and Extraction

These inquiries may be as simple as providing the correct URL to the downloading page a data repository, or they may be more complex, in the case of web-scraping data sources, or setting up file transfers or API data transfers. Some knowledge of data formats and file structure may be required, as well as the different interfaces available for popular data sets. Access questions may also merge into questions about the licensing and purchasing of data sets and require consultation of Collection Strategies Department.

E. Data Formatting, Cleaning and Sequencing

These inquiries focus on the arrangement of data, so that it can be studied. It may entail regularizing or harmonizing data sets for the purposes of comparison, or modifying the formats, so that the data can be analyzed with various software, including GIS. These inquiries often require expertise in data science specialist in Datacenter.

F. Data Visualization and Analysis

These inquiries are centered on working with existing data sources, so that researchers can interpret and illustrate findings evident in their data sets. These inquiries require expertise in various data manipulation software, and coding/programming languages, such as R or Python, or cartography/GIS visualization tools. These inquiries often require expertise in data science specialist in Datacenter.

G.. Access to Tools and Training in Data Science, Programing, and Software/Tools

These inquiries consist of consultations about the available tools, skill-sets, and training options necessary to work with data sets. They require knowledge of the popular software/tools within or across disciplines, as well as awareness of the services offered by the data science community at the university, such as Datacenter and beyond.

H. Data Publishing, Sharing and Preserving

These inquiries are about sharing data sets. They require knowledge about both disciplinary and general data repositories, including the university's Dryad repository, as well as the general principles and best practices for making data sets discoverable and usable by researchers (in the present and in the future), including the use of metadata, data description files (readme texts), and best practices for preservation.

I. Copyright, Licensing, and Privacy

These questions may be coextensive with data access or data publishing and preservation. They center on the legal, licensing, and copyright issues, and, the conventions around data privacy. They may require the expertise of the Scholarly Communication Office.

J. Data Management Plans

Data Management Plan questions, are related Data Publishing, Sharing and Preserving but focus more centrally on the writing of Data Management Plans, often required by research funding sources. They can be general, leading to referrals to templates such as the DMPT, or more advanced in that they may include questions about formatting, data access rights, preservation practices, and the policies of various repositories.

K. Database Construction and Development

These inquiries lead to consultations about the models, policies, and standards that govern how data is stored, arranged, integrated, and put to use in data systems. They may include questions about data architecture, the available support and tools for database construction, and options for database hosting at the University and beyond. They may require expertise from Datacenter, or Digital Strategies departments.

L. Other

Use this tag for data question that do not properly fit into the existing categories.

M. Not a Data Question

Surely there will be questions that come in that are not principally about data. They can be marked with this tag.

Secondary Question Categories: Free Floating Sub categories. Use these when appropriate to further describe question.

a. Technical/Computer

Requires coding, programming or other computer knowledge.

b. Legal/Rights

Requires knowledge of license and legal rules and jargon.

c. Academic/Subject

Requires a familiarity with the practices, discourse and aims of the researcher's academic community.

d. Purchase/Library Budgets

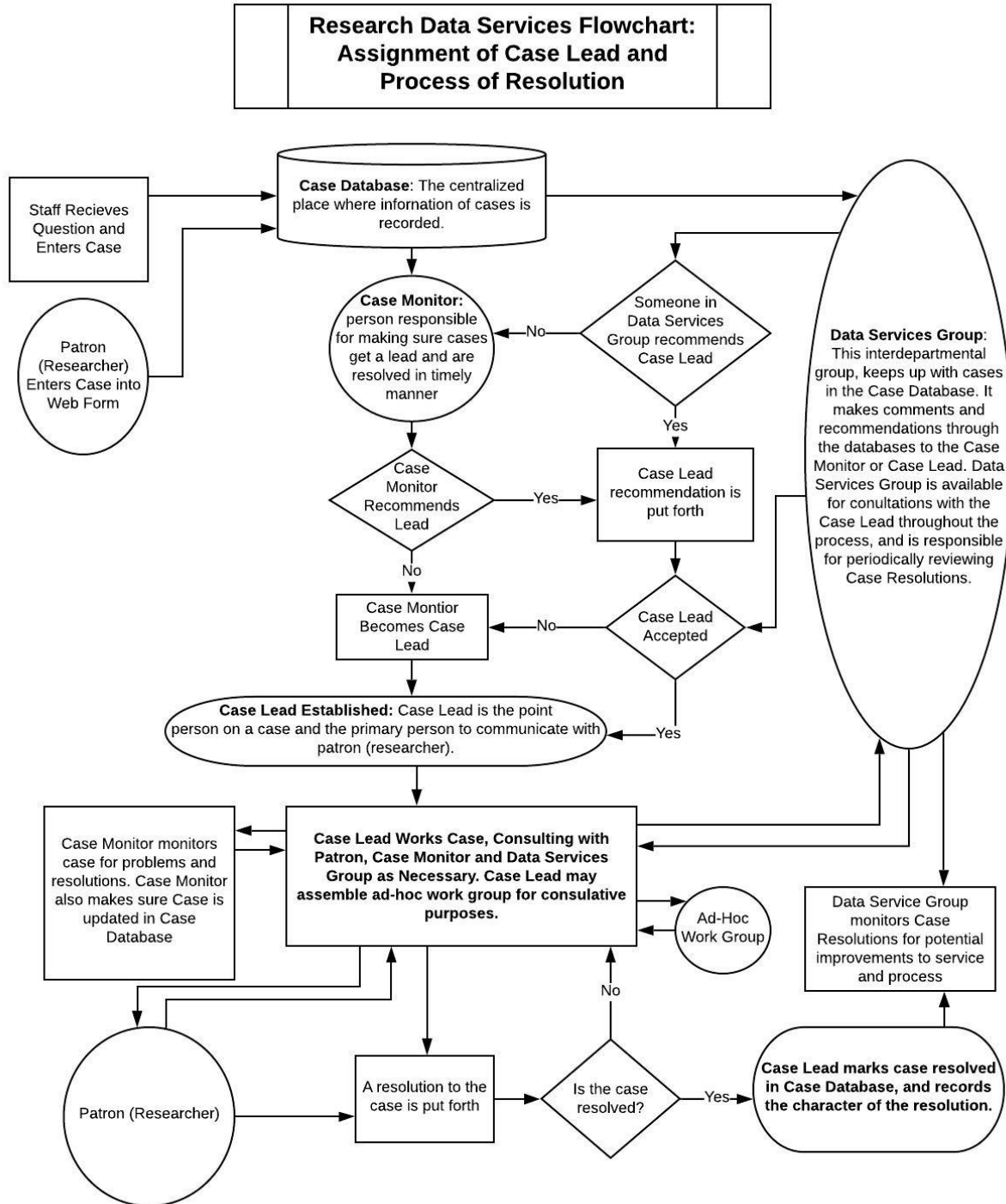
Questions are related to purchasing or licensing of data set access.

Appendix C

Research Data Services Flowchart.

David Michalski

May 22, 2020



Appendix D

MRSD: Modified READ Scale for Data Questions
 David Michalski
 5/22/2020

	Tradition READ Scale	Modified READ Scale for Data Questions
Level One	<p>See: Gerlich, B., & Berard, G. L. (2007). Introducing the READ Scale: Qualitative Statistics for Academic Reference Services. <i>Georgia Library Quarterly</i>, 43(4). Retrieved from https://digitalcommons.kennesaw.edu/glq/vol43/iss4/4</p> <ul style="list-style-type: none"> •Answers that require the least amount of effort; •No specialized knowledge skills or expertise; •No consultation of resources; •Less than 5 minutes; <p>-Examples</p> <ul style="list-style-type: none"> •Directional inquiries; •Library or service hours; •Service point locations; •Rudimentary machine assistance (locating/using copiers, how to print or supplying paper). 	<ul style="list-style-type: none"> •Answers that require the least amount of effort; •No specialized knowledge skills or expertise; •Little consultation of resources; •Less than 5 minutes; <p>-Examples</p> <ul style="list-style-type: none"> •Location of websites for known items questions •Common referrals to known tools (ORCID registration site) •Service point locations;
Level Two	<ul style="list-style-type: none"> •Answers given which require more effort; •Require only minimal specific knowledge skills or expertise; •Answers may need nominal resource consultation. <p>Examples</p> <ul style="list-style-type: none"> •Call number inquiries; •Item location; •Minor machine & computer equipment assistance; •General library or policy information; •More complex machine assistance (how to save to a disk or email records, launching programs or re-booting). 	<ul style="list-style-type: none"> •Answers given which require more effort; •Require only minimal specific knowledge skills or expertise; •Answers may need nominal resource consultation. <p>Examples: Providing basic information about data tools, resources or services so that patron can make an informed choice between alternatives.</p>

		Basic information on scope and coverage of data sources, and repositories, FAIR data management practices,
Level three	<ul style="list-style-type: none"> •Answers in this category require some effort and time; •Consultation of ready reference resource materials is needed; •Minimal instruction of the user may be required; •Reference knowledge and skills come into play. <p>Examples</p> <ul style="list-style-type: none"> •Answers that require specific reference resources (encyclopedias or databases); •Basic instruction on searching the online catalog; •Direction to relevant subject databases; •Introduction to web searching for a certain item; •How to scan and save images; •Increasingly complex technical problems (assistance with remote use). 	<ul style="list-style-type: none"> •Answers in this category require some effort and time; •Consultation with the researcher is necessary to provide the best assistance. •Minimal instruction or problem solving with the user may be required •Deeper knowledge and skills of data downloads, data management, and institutional data services come into play, including informing the researcher on how to meet the best practices for preservation, and data sharing. •Information about standard APIs or Data visualization tools may be presented.
Level four	<ul style="list-style-type: none"> •Answers or research requests require the consultation of multiple resources •Subject specialists may need to be consulted and more thorough instruction and assistance occurs. •Reference knowledge and skills needed. •Efforts can be more supportive in nature for the user, or if searching for a finite answer, difficult to find. •Exchanges can be more instruction based as staffs teach users more in-depth research skills. Examples •Instructing users how to utilize complex search techniques for the online catalog, databases and the web; •How to cross-reference resources and track related supporting materials; •Services outside of reference become utilized (ILL, Tech services, etc), collegial consultation; •Assisting users in focusing or broadening searches (helping to re-define or clarify a topic). 	<ul style="list-style-type: none"> •Answers or research requests require the consultation of multiple resources or colleagues. •Subject specialists may need to be consulted to understand the context of data sources in relation to the research project. More thorough instruction and assistance extracting and downloading files. •More complex knowledge about inter-institutional data services is necessary to make productive referrals. •Efforts can be more supportive in nature for the user, or if searching for a finite answer, more difficult to find.

		<ul style="list-style-type: none"> •Exchanges can be more instruction based as professional staff teach users more in-depth data management or sharing skills. •Instructing users how to utilize complex search techniques for data sources and repositories, or data collection or data manipulation methods; •Interlibrary data initiatives and services become utilized, more collegial consultation is needed to strategize data problem solutions. •Provision of some programming or computational instructions may be necessary
<p>Level ive</p>	<ul style="list-style-type: none"> •More substantial effort and time spent assisting with research and finding information. •On the high end of the scale, subject specialists need to be consulted. •Consultation appointments with individuals might be scheduled. •Efforts are cooperative in nature, between the user and librarian and or working with colleagues. •Multiple resources used. •Research, reference knowledge and skills needed. •Dialogue between the user and librarian may take on a 'back and forth question' dimension. <p>Examples</p> <ul style="list-style-type: none"> •False leads •Interdisciplinary consultations / research; •Question evolution; •Expanding searches / resources beyond those locally available; •Graduate research; •Difficult outreach problems (access issues that need to be investigated). 	<ul style="list-style-type: none"> •More substantial effort and time spent assisting with research and finding information about unique research data questions . •Subject and computational/data science specialists need to be consulted to strategize solutions. •Consultation appointments are required between researchers and library professional staff to strategize different stages of a problem. •Efforts are cooperative in nature, between the researcher and library professional staff and or working with colleagues. •The data problem has multiple steps or stages. •Deeper understanding of the research question is needed to formulate options or referrals. •Dialogue between the user and library staff may take on a

		<p>'back and forth question' dimension.</p> <p>Examples</p> <ul style="list-style-type: none"> •A number of tests or trials may be needed to come to a resolution. •Interdisciplinary consultations / research methods may need to be used; •Questions will evolve substantially; •Solutions often extend beyond those locally available; • Programming and computational support is necessary. •Unique programming or computational solutions may be needed (for extraction or visualization of data).
<p>Level ix</p>	<ul style="list-style-type: none"> •The most effort and time expended. •Inquiries or requests for information can't be answered on the spot. •At this level, staff may be providing in-depth research and services for specific needs of the clients. •This category covers some 'special library' type research services. •Primary (original documents) and secondary resource materials may be used. <p>Examples</p> <ul style="list-style-type: none"> •Creating bibliographies and bibliographic education; •In-depth faculty and PhD student research; •Relaying specific answers and supplying supporting materials for publication, exhibits etc; working with outside vendors; •Collaboration and on-going research. 	<ul style="list-style-type: none"> •The most effort and time expended. •Inquiries can lead to innovations in library services and policies. •Strategies may draw on innovations in data science and disciplinary theory. •At this level, staff may be providing in-depth computational services for specific needs of the clients. •Data architecture questions may come into play. <p>Examples</p> <ul style="list-style-type: none"> •Creating unique or innovative forms of data collection, analysis, or sharing techniques. •Library professionals develop sustained collaborations with faculty and PhD student research; •Work may necessitate partnerships with outside vendors;

		•Collaboration and on-going research.
--	--	---------------------------------------

Appendix E

Breakdown of the current Data Services Group by the Queue
David Michalski
5/22/2020

Data Services Group by the Case Database Queue Monitored

(Note: Any data services group member, irrespective of queue, could be brought-in as a case lead or contributor to case assistance)

Current

Data Management Queue

Instruction and Research, Director (Associate University Librarian)
Digital Strategies, Data Architecture Queue Monitor
Digital Strategies
Collection Development, Head
Digital Strategies (Programmer)
Research Services, Soc Sci and Hum Queue Monitor
Datacenter, Associate Director
Datacenter, Director
Datacenter, GIS Specialist
Research Services, Health Sciences Queue Monitor

Library Services for Social Sciences Data

Instruction and Research, Director (Associate University Librarian)
Online Strategies, Data Architecture Queue Monitor
Collection Development, Head
Technical Services, Head
Research Services, Soc Sci and Hum. Queue Monitor
Research Services, Soc Sci and Hum
Acquisitions, Licensing Specialist
Datacenter, Director
Datacenter, GIS Specialist

Appendix F

Recommended Membership for **New Research Data Services Group**

(One queue, rather than two)

David Michalski

5/27/2020

Data Services Group

Research Services, Head

Research Services, Soc Sci and Hum Queue Monitor

Research Services, Health Sciences Queue Monitor

Research Services, STEM

Research Services, Soc Sci and Hum

Medical Library, Health Informaticist

Digital Strategies, Data Architecture

Digital Strategies (programmer)

Digital Strategies (programmer)

Datacenter, Director

Datacenter, Associate Director

Datacenter, GIS Specialist

Scholarly Communication Office, Head

Collection Development, Head

Collection Development, Social Sciences