UKSG

Bigger on the inside: building Research Data Services at the University of Virginia

Every story has a beginning, where the narrator chooses to start, though this is rarely the genesis. This story begins with the launch of the University of Virginia (U.Va.) Library's new Research Data Services (RDS) unit in October 2013. Born from the conjoining of a data management team and a data analysis team, RDS expanded to encompass data discovery and acquisitions, research software support and new expertise in the use of restricted data. Our purpose is to respond to the challenges created by the growing ubiquity and scale of data by helping researchers acquire, analyze, manage and archive these resources. We have made serious strides towards becoming 'the face of data services at U.Va.'. This article tells a bit of our story so far, relays some early challenges and how we have responded to them, outlines several initial successes and summarizes a few lessons going forward.

A timeline in brief

'People assume that time is a strict progression of cause to effect. But actually from a non-linear, non-subjective viewpoint it's more like a big ball of wibbly-wobbly, timey-wimey … stuff.'

The Doctor (Doctor Who) [1]

MICHELE P CLAIBOURN

Director, Research Data Services University of Virginia Library USA

In 2010 the University of Virginia (U.Va.) Library formed the Scientific Data Consulting group (SciDac) to respond to new imperatives by funding agencies for researchers to share the data gathered as a product of sponsored research. Building on the work of prior library collaborations[2], this group began to research how scientists in our community were handling data, to create tools for assessing their data practices, and to work with colleagues at other institutions to build solutions for data-intensive researchers encountering new and evolving requirements.[3] The SciDac team laid considerable groundwork within the Library, advancing the notion of the research lifecycle and the Library's role throughout, as well as with analogous organizations across the globe. One notable early achievement was SciDac's involvement in developing and advancing the DMPTool, a collaboratively built tool to guide researchers writing a data management plan.[4] In early 2013 the four-person team rebranded as the Data Management Consulting Group, to underscore the University-wide need for data management, documentation and archiving support.

'advancing the notion of the research lifecycle and the Library's role throughout'

In the meantime, the U.Va. Library committed to expanding the Library's data services by supporting additional positions and graduate student assistantships. In Spring 2013 the Library hired a statistical consultant (the author) to build a data analysis consultancy service. The following August the StatLab launched with a second full-time consultant and three doctoral students, offering workshops and individual consultations on the use of statistical software, advanced data analytic techniques, and statistical methods. In October the StatLab and the Data Management Consulting Group merged into Research Data Services (RDS).[5]

Two months later, in December 2013, the Library assumed responsibility for a research computing support team under RDS. Previously reporting to U.Va.'s central IT, and residing in the Library's Scholars' Lab, this team provided support for the distribution and use of University-licensed research software (titles like SPSS for statistical analysis, Matlab for numerical computing and Ansys for engineering simulations). In February 2014 the current configuration of services was realized as a recently hired Data Librarian moved to the RDS team. The current RDS staffing is shown in Table 1.

Service Area	FTE
Director	1.0
StatLab: Data Analytics and Statistics	
Statistical Consultant	1.0
Graduate Students (3)	
Data Discovery and Acquisitions	
Data Librarian	1.0
Graduate Students (1)	
Data Management	
Data Librarian	2.0
Astronomer	0.5
Research Software Support	
Software Specialist	1.0
Undergraduate Students (2)	
Other, Cross-Service	
Data Scientist (Physicist)	1.0
CLIR Post-doc fellow (Sociologist)	1.0
Education Librarians	2.0
Coordinator	1.0
Total full-time staffing	11.5

Table 1. Research Data Services Staffing 2014–2015

More than four years ago, the Library began to envision what it would mean to support the increasing reliance on data across the research spectrum. Building on prior efforts, here and elsewhere[6], the Library made early investments in this vision, reassigning existing staff and hiring new experts. By most measures, it has proved successful. For instance, this past academic year, RDS offered over 35 workshops with, on average, 20 registered learners; and we engaged in nearly 500 consultations with researchers from every school within the University. But this is not the whole story.

> 'the Library made early investments in this vision, reassigning existing staff and hiring new experts'

Thickening the plot

'We're all stories, in the end. Just make it a good one, eh?'
The Doctor (Doctor Who) [7]

The preceding summary, while accurate, elides a few hiccups, dead ends and course corrections along the way; for instance, gaining the awareness of our local community, drawing the ire of some key stakeholders and integrating multiple service points.

One of the early barriers to progress was the inability to draw the attention of internal researchers. Many in the University community were simply unaware of the Library's efforts. The early SciDac team engaged in multiple outreach efforts – conducting nearly 20 interviews with researchers on data management practices from 2010–2012, building a data management web portal for graduate students in 2012, and developing and hosting a multi-institutional data management boot

camp for graduate students in Spring 2013. Yet in the first three years, outreach efforts generated only between 10–20 consultations a year, primarily to review data management plans for grant submissions. Raising awareness of these services was an uphill battle. We gained some leverage by reorienting efforts more strongly towards graduate students, a population not yet entrenched in disciplinary habits. But this was not a challenge marketing alone could solve; many researchers were not that interested in what we were offering. Consolidation of a fuller array of services – better mirroring the full data lifecycle of discovery and collection, wrangling and analysis, sharing and archiving – has helped.

Of course, a fuller array of services first needed to be built. After succeeding in winning resources from the state to hire two data experts, a new obstacle arose. Faculty attention to the Library spiked among a set of key stakeholders. Leaders in multiple departments with some claim on data analysis took umbrage at the Library's efforts to hire experts in statistics and data mining: statistics and data analytics were in their domains; these resources should reside in their school or department; and what did a library know about data analysis? The Library had, in fact, sought input from faculty across multiple departments when creating the two initial position descriptions. The absence of a strong prior network with this data-intensive constituency in the sciences and social sciences – among faculty who, rightly or wrongly, are accustomed to viewing the Library as the place that acquires, manages and preserves research material, not the source of data expertise – meant these initial requests did not rouse these still relatively inattentive parties. So we did not see it coming. In retrospect, more time and effort was needed to build a supportive coalition among these departments. In the absence of those conversations, it was too easy for faculty to see the Library's efforts as part of a larger narrative about the waning of faculty governance and control, to view this as another instance of the administration allocating zero-sum resources in someone else's favor.

'Leaders in multiple departments with some claim on data analysis took umbrage at the Library's efforts to hire experts'

But Library leadership did what enterprising librarians do: invited these faculty onto the search committees, used the sudden attentiveness as an opportunity to communicate the value of the Library to these data-intensive research efforts, and encouraged them to see the alliance of several departments and schools, albeit against the Library, as evidence of the need for pan-University resources. While the agreement of the interested faculty was initially tentative, it was enough to proceed. One search failed; the position description was too amorphous and the faculty found the candidates wanting. The other search, however, ended successfully, with the Library hiring the first statistical consultant (the author) – a social scientist with ten years of faculty experience, seven of them at U.Va. – who joined the Library in February 2013.

Bringing in an academic with existing ties to the internal community proved vital. My prior experience helped to lessen some faculty's sense of the Library as 'the other' and to enhance the Library's credibility among several key departments. Nonetheless, I spent the first few months in the Library on a goodwill tour, talking with faculty about plans and soliciting their feedback. The attention during the search, while not always positive, meant many faculty were willing to meet when I began in the Library. Even so, some early conversations were strained; a few faculty heard workshop proposals as competition for their semester-long classes; some assumed the consultation services were aimed at tutoring students in their classes, rather than the higher-end expertise we had in mind. With misconceptions countered and expectations clarified, the StatLab enjoyed a soft launch in April: researchers sought consultations largely through word of mouth, and I designed an inaugural workshop series for the summer. The previously failed search granted us time to reshape a second position with input from more attentive researchers. By Fall 2013, we had filled the second position, with faculty involvement on the search committee, hired three doctoral students as statistical consulting associates, and fully launched the StatLab.

'With misconceptions countered and expectations clarified, the StatLab enjoyed a soft launch'

The same fall, as mentioned, the Library elected to combine the newly developed StatLab and the Data Management Consulting Group into a single unit: Research Data Services. I began directing this effort. This combination created a center of gravity for drawing in additional data-relevant services and expertise, including a recently hired data librarian and the central IT research-computing specialists. It gave the Library a united front to engage broader conversations occurring at the University, in particular, around the newly formed Data Science Institute – a University-wide initiative emanating from the Provost's office. But it also posed a new challenge: integrating multiple services, originating within different organizational cultures, into a coherent vision under the leadership of a professional still new to libraries. [8] By this time, the group included six library professionals (some with a focus on data discovery and management, some with subject expertise in education research), two IT professionals and three academic professionals: a statistician, an astronomer and a recovering professor (me). In short, we had a blend of experts, a scenario with potential for both exciting emergent properties and persistent divergent perspectives.

> 'a center of gravity for drawing in additional data-relevant services and expertise'

There are no shortcuts to integrating such disparate staff. It takes time and motivation to understand one another, and a willingness to learn and adapt from all sides. We are still engaged in this process. Two years in, we have begun to see useful adaptations, ones I could not have predicted at the outset. RDS began as five separate services (data management, data discovery, data analysis, software support and subject expertise in education); now we have created bridges. Data management experts and data analysis experts contribute to research software support, education librarians are more involved in data discovery, data discovery and acquisitions experts help read data into software and do some basic wrangling. This cross-training gives the whole team a fuller understanding of typical researcher workflows and roadblocks. The integration helps us cross-train our research community as well. We can use the services and partnerships that draw more active research interest (e.g. data discovery and data analysis) to engage a conversation with scholars about issues they are, frankly, if shortsightedly, less excited about (e.g. data documentation and archiving). It turns out, not surprisingly, that helping researchers visualize data, or estimate a causal model, or implement and interpret a classification algorithm, can bring them through the doors. Once here, we have their attention, thus overcoming that first obstacle.

> 'cross-training gives the whole team a fuller understanding of typical researcher workflows and roadblocks'

A plan comes together

'Just do what I do. Hold tight and pretend it's a plan!'
The Doctor (Doctor Who) [9]

Our RDS initiatives have been well received within the University. In our second year of operation, we have offered a successful workshop series, with over 35 hands-on instruction offerings in topics like Introduction to R, Creating a Data Management Plan, Visualizing Data, Querying SQL Databases, Using LaTeX, Social Network Analysis, Version Control with Git and Matching Methods for Causal Inference. Twenty interested researchers registered for each workshop, on average, significantly enhancing our visibility.

Our consultation services have grown, with about 495 consulting interactions in the 2014–2015 academic year, ranging from 30 minutes to more than 30 hours of collaboration. While the majority of these have been centered on data analytic and statistical methods (340), our data discovery and access and data management services have been increasingly active as well. We have begun to expand into more technical and computational areas. RDS co-sponsored Software Carpentry Bootcamps[10] in 2013 and 2014, bringing in instructors to offer the two-day training on software skills like Unix and Bash scripting, version control, and programming in Python, both to help scientists be more productive and to help the science they engage in be more reproducible. But in 2015, after multiple staff became certified instructors, we sponsored our first in-house program. This academic year, we

organized and hosted an R User Meetup for our community, with gatherings ranging in size from 20 to 80.

Library leadership has leveraged and expanded the growing attention toward the Library's data efforts, engaging an ongoing conversation around data, big and small, on campus. As a measure of this success, the Library is prominently mentioned in the University's strategic plan. Under a strategy focused on building research infrastructure and services, the Library is thus charged: 'The Library will provide data services for acquisition, management, and preservation of massive amounts of data, complemented by growing staff and faculty expertise in digital research across disciplines and increased access to digital content in all formats.'[11] In talks around the University about the newly founded Data Science Institute in Fall 2013, the leaders of this initiative regularly referenced the work of the Library around data services, further legitimizing and advertising our efforts. Recently I was recruited by the Data Science Institute for a 25% appointment as the Associate Director of Data Infrastructure and Services, providing a second platform for the work the Library was already doing: developing partnerships with related services to promote a more coordinated system of support for data around the University.

'the Library is prominently mentioned in the University's strategic plan'

The Library's RDS has become a key part of a growing network on our campus. In a new effort this year, we have collaborated with our Education School to launch a restricted data service. One of our staff manages two restricted data rooms, in spaces provided by the Education School, and has deepened her expertise in accessing and securing sensitive data. Working with the Vice President for Research's office, we have developed restricted data procedures to help scholars navigate the often unfamiliar labyrinth of restricted data agreements and compliance. These partnerships have spawned others. We have continued to work with the Vice President of Research's office to formalize a data sharing policy, and have been deeply engaged with the Data Science Institute's efforts to build a secure data computing environment. The position of RDS within the Library, along with key partnerships with other departments at the University, are shown in Figures 1 and 2.

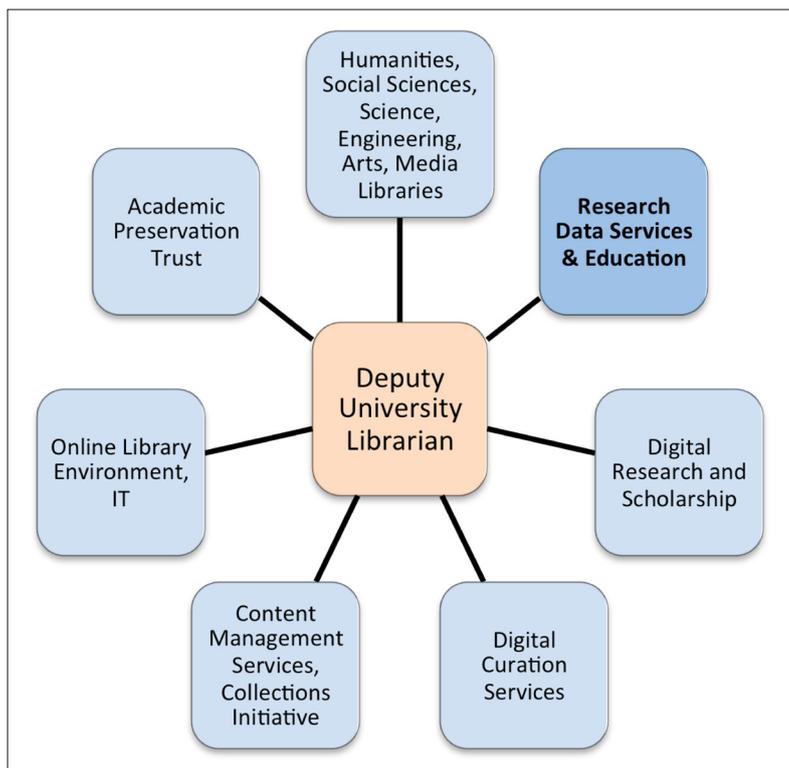'deeply engaged with the Data Science Institute's efforts to build a secure data computing environment'

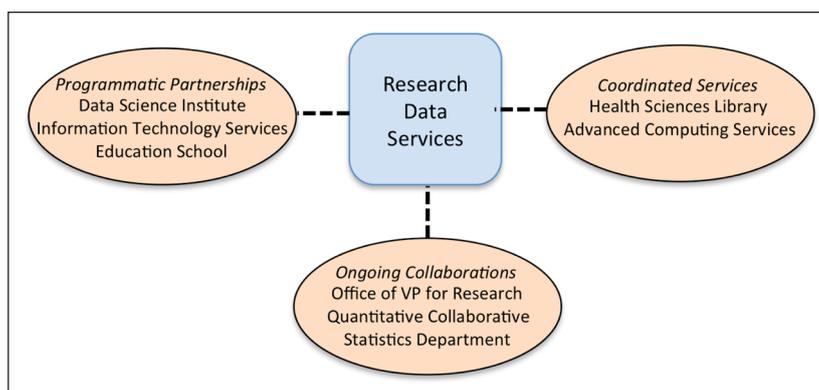Figure 1. The position of Research Data Services within the Library

Figure 2. Research Data Services partnerships in the University

The library in a data-driven age

'You want weapons? We're in a library! Books! The best weapons in the world!'
The Doctor (Doctor Who) [12]

Books are powerful for the information they contain and the knowledge they convey. The same is true of data. Our Library leadership recognized this essential similarity early on and has consistently advanced efforts to expand support for data. They have encouraged the reorientation of existing Library staff, committed new resources to hiring different kinds of librarians, and advocated strongly on behalf of the Library as the face of data services to the larger University. These efforts and investments have begun to pay off. Not only are researchers across the University reaching out to the Library for help with data, the Library has been firmly positioned as a central part of ongoing conversations around data, included in plans for new grants, new programs and new infrastructure.

> 'the Library has been firmly positioned as a central part of ongoing conversations around data'

Our experience, as conveyed here, is open to some interpretation. From my perspective, three lessons stand out. First, beginning with the creation of RDS, we made a conscious choice to prioritize collaborations within our University over networks external to our institution. We doubled down in our efforts to build a reputation among our on-campus constituency, on the theory that internal credibility is a necessary condition for external credibility, but an external reputation is not sufficient for ensuring support among our own community and administration. Second, though we moved swiftly, changes were sequenced and incremental. After developing the data management services internally for a few years, we brought on new staff to build data analytics; an expansion of more focused data discovery followed; we absorbed software support; and finally added additional staff. At multiple points we have deliberately slowed the momentum. New services have been given a soft launch, with time to work out some kinks. Third, we have experimented with the concept of blended librarianship, creating teams of scientists, social scientists, data scientists and library science experts. Library leadership has set clear expectations that librarians with new kinds of expertise should not just adapt to library culture, but should help shape a new culture for the library. An open question at our own Library is whether individual librarians should try to embody this 'blendedness' or if it is sufficient to widen the tent of librarianship to incorporate new disciplinary and data-intensive expertise. Whatever the answer, the inclusion of data services in the Library, incorporating new data analytic specializations our users are not accustomed to finding in the Library, combined with the partnerships RDS has begun to establish with related centers at U.Va., are helping to make our Libraries 'bigger on the inside'.

> 'we have experimented with the concept of blended librarianship'

References and notes

1. Doctor Who, with apologies to Doctor Who (main character of a British science fiction TV show of the same name broadcast by the BBC). The title of this article alludes to the Tardis, the Doctor's time-travelling machine, which is much bigger on the inside than it appears from the outside. Season3, Episode 10: Blink (2007).

2. Hunter, C, Lake, S, Lee, C and Sallans, A, A Case Study in the Evolution of Digital Services for Science and Engineering Libraries, *Journal of Library Administration*, 2010, 50(5), 335–347. DOI: http://dx.doi.org/10.1080/01930821003667005 (accessed 25 April 2015).

3. Fearon, D, Gunia, B, Pralle, B E, Lake, S and Sallans, A L, *ARL Spec Kit 334: Research data management services*, 2013, Association of Research Libraries.

4. DMPTool: https://dmptool.org/ (accessed 25 April 2015).

5. Research Data Services: http://data.library.virginia.edu/ (accessed 10 June 2015).

6. Tenopir, C, Birch, B and Allard, S, Academic libraries and research data services: Current practices and plans for the future (An ACRL white paper), Association of College and Research Libraries, 2012, Chicago, IL. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf (accessed 10 June 2015).

7. Doctor Who, Season 5, Episode 13: The Big Bang (2010), ref. 1.

8. Hunter, C, Lake, S, Lee, C and Sallans, A, ref. 2.

9. Doctor Who, Season 7, 7th Christmas Special Episode: The Doctor, The Widow and the Wardrobe (2011), ref. 1.

10. Software Carpentry: http://software-carpentry.org/

11. The University of Virginia, *The Cornerstone Plan: A Strategic Plan for the Academic Division of the University of Virginia*, 2013.

12. Doctor Who, Season 2, Episode 2: Tooth and Claw (2006), ref. 1.

Michele P Claibourn

Director, Research Data Services
University of Virginia Library, USA
E-mail: mpc8t@eservices.virginia.edu

ORCID ID: http://orcid.org/0000-0002-5969-441X