

What is RDS

SLIDE 1: The Library established Research Data Services (RDS) in October 2013, merging four areas of developing expertise:

- the three-year old **Data Management Consulting Group**, advocating for and advising on data management planning, sharing, archiving, and curation;
- the newly created **StatLab**, providing consultations, training, and support for data analysis, visualization, and statistical computation;
- the **Data Discovery and Collections** expertise of the data librarian, working with researchers to locate existing data sources for analysis and building the Library's data collections; and
- the **Research Software Support** services formerly housed in the information technology center, helping researchers locate, access, and install University-licensed research software titles.

Our work in the first three years has focused on a few key activities and goals: **consultations and collaborations** with researchers; **education and training** for data-intensive research approaches; and building **partnerships and programs** within the University to promote data, computational, and quantitative research across campus.

Consulations and Collaborations

SLIDE 2: RDS revolves around direct engagement with researchers through consultations – efforts to advise, instruct, or support researchers over a hurdle – which often become enduring collaborations – or repeated conversations and interactions around an ongoing research project. The advent of RDS has helped to situate the Library more centrally in the research process of quantitatively and computationally oriented scholars.

- In our first three years, we hosted nearly 1,800 consultations with researchers. The majority of these center around data analysis, wrangling, and statistics (61%). so far, this fall, weve had another 375 interactions.
- Our interactions reach across the entire University. Through June 2016, we'd met with 800 different researchers across all 11 schools and within multiple centers (e.g., Institutional Assessment, Data Science Institute).
- About half of our users engage with our services and staff repeatedly. We have worked with researchers at all levels – from novice undergraduate researchers to expert faculty researchers. The bulk of our engagement is with graduate students (45%), a community we have explicitly targeted.

In our collaborations, we've

- Helped a Commerce faculty reshape experimental data and implement a multilevel model to estimate treatment effects for a repeated measures design;
- Worked with an interdisciplinary team of faculty to draft a Data Management Plan for the NSF Sustainability Research Network proposal;
- Guided a Politics faculty member in acquiring key international time series data for a book project;
- Helped the Office of Institutional Assessment automate survey reports through reproducible scripts.

Workshops and Training

SLIDE 3: Education is a second core mission of RDS. We've been developing our workshop series to meet widespread knowledge needs, providing instruction on the data and computational skills and processes researchers are often expected to know but which individual departments seldom have the resources to provide formally. *RDS educational efforts are not meant to substitute for departmental curricula, but to complement them, by providing a pool of collective resources from which researchers in any field can draw.*

In the first three years, we've offered 90 workshops, open to all, for which nearly 1,900 learners registered¹ The workshop offerings range from introductions to common data analysis environments (R, Stata, SPSS) and advanced statistical methods (Mixed-Effects Modeling, Multiple Imputation, Machine Learning), to computation and programming (Linux, Python, Using APIs) and research data management (Databases, Data Management Planning). We've been working to build efficiencies, as well, by recruiting experts from across Grounds and providing them a ready-made platform to share their knowledge. A taste of our workshops is provided in the wordcloud, with the size and color of the words scaled by workshop registration.

These, too, have proven a University-wide resource, drawing interest from across campus, with researchers in all 11 schools and several additional centers signing up. About a third of participants come from the College of Arts & Sciences, a fifth from the Engineering School, and a fifth from the Education School. Over 60% are graduate students, while faculty, University staff, and research staff (research scientists and post-docs) each compose another 10%.

We've also experiment with multiple bootcamps, including 1- and 2-day bootcamps (Software Carpentry, Data Management).

Meanwhile on grounds...

SLIDE 4: A lot was changing

¹Many learners attend multiple workshops; we have 800 unique registrants. On average, about 60% of those registered attend the workshop.

- The DSI was established, the first of several planned pan-university institutes;
- we hired a new VP for IT with an emphasis on academic computing; our IT department had become increasingly focused on enterprise needs, and there had been growing complaints among researchers that academic computing was neglected;
- the School of Medicine hired their first director of research computing, marking a growing interest on our large health system side;
- and UVA invested in two new high-end computing platforms - Rivanna, a traditional HPC cluster, and Ivy, a secure compute environment for protected data.

New opportunities

These changes opened up new opportunities. At the onset of the DSI, the new director approached the library about bringing me on board in a partial appointment as the first Associate Director, responsible for Data Infrastructure and Services – a position I maintained for 2 years before stepping aside. In that role, we were able to form some great bridges between the Library and the new institute – we’ve been working with them to deliver bootcamps to their students, to resource workshops and bootcamps by outside experts (SWC, COS), and have just completed a 2-year pilot with library staff offering 1-credit short courses through the DSI, but open to all learners (unlike the DSI’s graduate curriculum for the MSDS) – Data Wrangling in R, Data Wrangling in Python, Text as Data, and Applied Causal Inference.

SLIDE 5: Another opportunity was created by the new VPIT. RDS had been working to build relationships with colleagues in the Advanced Research Computing Services (ARCS) for some time. This new environment gave those efforts a boost. In particular, the new VPIT took on the long-recognized need to bring together related research support efforts around campus to help our community discover disparate experts. Recently, these networks have been formalized into CADRE, the Computational and Data Resource Exchange: a partnership of ARCS, RDS, HSL, DSI, SOM, and the VPR. While RDS has had productive partnerships with our colleagues in the Health Sciences and Law Libraries for some time, and with the DSI and VPR, CADRE has been a vehicle for bringing many of the dyadic relationships into a larger network..

A key goal is to improve coordination and exposure of data- and computationally-oriented infrastructure, training, and expertise; and just this fall, we launched a new joint webpage – no small feat! The site is intended as a portal to a wide variety (hopefully all, eventually) of resources and experts for computational and data-relevant research, and is discoverable on all of the partners webpages as well. We’re hoping that users can find it no matter where they start. This has really helped everyone do a better job of advertising for one another, for jointly increasing awareness of institutional resources, and more.

And I hope we can expand it further. Recently I’ve started talking with some colleagues in CADRE about joining in the short course program. And we’ve been inviting colleagues from other labs and efforts to join our monthly meetings and be represented on the webpage

– empirical data lab in the law school, visualization lab in the Arch school, digital humanities developers in the Library’s Scholars’ Lab).

SLIDE 6: And new partnerships and programs have begun to develop as a result. For example, RDS began working with the VPR’s office this fall on funding discovery, providing training and consultations on the use of the new funding discovery tools they’ve licensed; and we’ve been working with the VPRs office and the HSL on UVAs foray into the OSF for institutions. RDS and HSL have been working closely with our research computing team around Ivy. And we’ve started connecting more with the already vibrant DH community within our library and at UVA. RDS, with the support of several partners, founded and hosts the first R meetup in Virginia at UVA to help build community and self-training. And the network has made it easier to kick off new efforts, like hosting the inaugural DataFest event at UVA – a 72-hour data hackathon-style challenge for undergraduates created by the ASA (while RDS spearheaded it, we were able to recruit multiple partners: the DSL, and VPIT, the quantitative collaborative, the School of Engineering, the College of Art and Sciences, and the Statistics department).

Challenges

SLIDE 7: All of this lovely integration and partnering has presented challenges, of course; some expected, and a few that seem predictable in retrospect but I hadn’t anticipated in advance. Among the expected are:

- Time constraints: Maintaining relationships takes time – time that’s also needed for strategizing, managing staff, meeting user needs, and the ongoing onslaught of decisions and demands to communicate out to others the status of programs and coalitions (only now there’s more).
- Staff constraints: And as we expand into new areas, we have to be vigilant that we effectively shift the nature of the work done by our staff rather than just add to it – which also takes time (in prioritizing, in communicating shifting priorities to leadership, to patrons), and an effort to resist our generally admirable instincts to want to support all requests.
- Cultural collisions: not all of our partners have the same expectation or practice of sharing information, and of sharing credit, as our library. This is part of maintaining relationships, of course, but really involves everyone. Some of our partners use ticket systems and want everything to go through there, where our library places a higher value on the personal; some have gotten a little bent out of shape when that we’ve worked to provide initial education to researchers on resources in their domains. There have been hiccups.

Less expected, but following from these

- Partnerships diffuse responsibility with reasonably predictable consequences – e.g., ball dropping, reduced incentive to prioritize a project – but they also diffuse credit or

recognition. And while our institutions regularly voice the desire for more partnerships and coordinations to leverage dispersed resources, that's not how our budgets work. So different actors in these relationships are better positioned than others to highlight joint efforts in advocating for future resources. Similarly, though less consequentially, users, who unquestionably benefit by these collaborations, aren't always sure who's providing resources – as we're in an era of changing budget models, where schools and chairs are seeing their budgets apportioned out to central services like the library (and some of our other partners), that can be a problem.

- Partnerships obscure resource needs: precisely because multiple partnerships, networks, and collaborative programs make the “team” look bigger, it can make it harder for administrators to see the insufficiencies in support – five people spending 5% of their time on an effort still looks like five people. Less benignly, it gives administrators cover for not sufficiently funding initiatives. This has bitten us a bit in a way I'm happy to elaborate on in QandA.

The call to be dynamic and responsive makes the formal institutions intended to guard against this (e.g., MOU's) less practical. Library leadership will undoubtedly need to be involved for these efforts to be sustainable.

The Library Team

SLIDE 8: Before I end, I want to give you a sense of the team. While much of this was going on, in the Summer of 2015, our library re-organized, and the RDS team became the RDS + SNE team – so we're responsible not just for data services, but for liaison work with the social, natural, and engineering science departments at UVA. One of the things this has allowed us (required us) to do is seriously cross-pollinate – most of the data service experts liaison with a department, and most of the primary liaisons are working with researchers on efforts like data management and data discovery as well. We're a mixed bunch, which has been one of our chief strengths.

What's next?

SLIDE 9: Things continue to change, so I'll just run down some of the initiatives we're working on now (all of which I'd welcome ideas on, or conversations about later today!).

- In the summer of 2015, our library also reorganized so that RDS and the Social, Natural, & Engineering Science teams were integrated into one unit – but at the moment of the reorganization, we had only one social science librarian. We're now up to two social science and two science and engineering librarians, so we've been working on integrating these multiple strands of expertise and services.
- So far, RDS has operated without any identifiable space or technologies. We're just beginning an effort to imagine possible spaces as a storefront for RDS and SNE.

- One of the things the integration of RDS and SNE has facilitated is nascent conversation within our group about efforts to better support qualitative research in addition to quantitative research.
- We're taking some first steps into an internally developed projects: Data for Democracy Lab – with an initial project that seeks to pull mostly textual information about the attention and activities of the president from multiple and divergent streams (newspapers, tv news, daily compilations, relevant social media, etc.) and extracting key summary features to represent in an ongoing and dynamic way (e.g., issue attention, ideological positioning, tone, named entities, or the like).