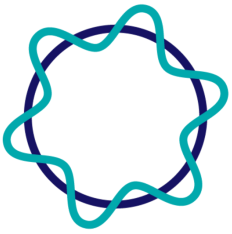
Hiring, Managing, and Retaining Data Scientists and Research Software Engineers in Academia

A Career Guidebook from ADSA and US-RSE



Academic
Data Science
Alliance

August 2023

Chapter Leads and Workshop Organizers:

Editor: Steve Van Tuyl (Academic Data Science Alliance)

David Beck (University of Washington)
Ian Cosden (Princeton University)
Blake L. Joyce (University of Alabama at Birmingham)
Jing Liu (University of Michigan)
Christina Maimone (Northwestern University)
Kenton McHenry (University of Illinois Urbana Champaign)
Micaela Parker (Academic Data Science Alliance)

Workshop Participants and Chapter Contributors:

Nicky Agate (University of Pennsylvania)
Peter Alonzi (University of Virginia)
Arlyn Burgess (University of Virginia)
Sean Cleveland (University of Hawaii System)
Aaron Culich (University of California, Berkeley)
Mike Daniels (Ronin Institute, National Center for Atmospheric Research)
Alex Davis (The Ohio State University)
Sarah Davis (Renaissance Computing Institute, University of North Carolina, Chapel Hill)
Sandra Gesing (Discovery Partners Institute, University of Illinois System, Chicago)
Launa Greer (University of Chicago)
Chris Holdgraf (2i2c)
Daniel S. Katz (University of Illinois Urbana Champaign)
Rachel Levy (North Carolina State University)
R.J. Cody Markelz (University of California, Berkeley)
Nirav Merchant (University of Arizona)
Scott Michael (Indiana University)
Bill Mills (University of Colorado)
Tiffany Oliver (Spelman College)
Alessandro Orso (Georgia Institute of Technology)
Gina-Maria Pomann (Duke University)
Rahim Rasool (University of Chicago)
Erick Ratamero (The Jackson Laboratory)
Kristina Riemer (University of Arizona)
Trevor Spreadbury (University of Chicago)
Sarah Stone (University of Washington)
Jess Sweeney (University of Chicago)
Karen Tomko (Ohio Supercomputer Center)
William J. Tomlinson (Software & Application Innovation Lab, Boston University)
David Uminsky (University of Chicago)
Maryam Vareth (BIDS, University of California, Berkeley; CI2, University of California, San Francisco)
Váleri N. Vásquez (University of California, Berkeley)
Elizabeth Vu (Sloan Foundation)
Bruce E. Wilson (Oak Ridge National Laboratory)
Jeffrey Young (Georgia Institute of Technology)

Hiring, Managing, and Retaining Data Scientists and Research Software Engineers in Academia

Chapter Leads and Workshop Organizers ii
Workshop Participants and Chapter Contributors ii
Table of Contents iii
Foreword vi
Acknowledgments vii
Sponsors viii

Chapter 1 - Introduction and Rationale 1

1.0 Summary 1
1.1 Introduction to the Guidebook: “What is this all about?” 2
1.2 Who is the Audience? 3
1.3 Abstracts for the Other Chapters in the Book: 4

Chapter 2 - Articulating the Need 6

2.0 Summary 6
2.1 Context: Why DS/RSE Professionals? 7
2.2 Identifying Key Stakeholders 8
2.3 Addressing the Need and Getting Buy-In 11
2.4 Positioning 13
2.5 Potential Models for DS/RSE Positions 14
2.6 Funding Sources and Sustainability 14

Chapter 3 - Before Posting: Position Descriptions, HR, and Compensation 15

3.0 Summary 15
3.1 Working with HR 16
3.1.1 A Few Terms 17
3.1.2 HR Process Basics 17
3.1.3 Challenges DS/RSE Positions Pose for HR 18
3.2 Defining a Position 19
3.2.1 Responsibilities and Requirements 19
3.2.2 Position Logistics 21
3.2.3 Find or Write a Job Description 21
3.2.4 Job Title 22
3.3 Compensation Considerations 23
3.3.1 Compensation Basics 23
3.3.2 Compensation Challenges and Strategies 24
3.4 Posting the Job 26
3.4.1 Posting Logistics 27
3.4.2 Posting Components 27
3.4.3 Keep Postings Focused 28
3.4.4 Language Matters 29

Chapter 4 - Recruitment 30

4.0 Summary 30
4.1 Finding Candidates 31

4.1.1 Building a Pool of Candidates Prior to Job Postings 31
4.1.2 Strategies for Advertising Job Postings 33
4.1.3 Guidance for Hiring Managers 36
4.2 Application Review, Interviewing, and Evaluation 38
4.2.1 Preparing for Interviews 38
4.2.2 Structure of Interviews 38
4.2.3 Elements of an Interview 40
4.2.4 Evaluation Rubric 43
4.2.5 Reference Letters/Checking References 44
4.3 Considerations for Offers 45
4.4 Post-hoc Review and Process Improvement 46
Chapter 5 - Expectations, Metrics for Success, and Onboarding 48
5.0 Summary 48
5.1 Determine Group Expectations & Organizational Needs 49
5.2 Onboarding for Individual Success 50
5.3 Determine Metrics for Individuals & Groups 51
5.4 Revisiting Metrics and Expectations 52
5.5 Examples of Setting Expectations 53
Chapter 6 - Retention 56
6.0 Summary 56
6.0.1 Rationale 56
6.1 Professional Development Opportunities in Diverse Academic Environments 57
6.1.1 Professional Profile 59
6.2 Professional Development Approaches and Opportunities 63
6.3 Mentoring 65
6.4 Credit 67
6.5 Learning from Private Sector Engineering 69
Chapter 7 - Managing Data Scientist/Research Software Engineer Teams 71
7.0 Summary 71
7.1 Organization 71
7.1.1 Economies of Scale 72
7.1.2 Specialization in Key Areas 72
7.1.3 The Ability to Amplify Activities on Projects 73
7.1.4 Institutional Knowledge and Sustaining Software 73
7.2 Team Management Examples 73
7.2.1 University of Washington eScience Institute 73
7.2.2 University of Illinois Urbana Champaign 74
7.3 Teams 75
7.3.1 Project Based 75
7.3.2 Thematic Based 75
7.3.3 Expertise Based 75
7.3.4 Other Models 75

7.4 Forecasting 76
7.4.1 Resource Management 76
7.4.2 Soft-Funded Environments 76
7.5 Commercial vs Research: Awareness of Differences in Management and Expectations 77
7.6 Differentiating Team Types and Needs 78
7.6.1 Project Team Needs 78
7.6.2 Data Scientist and Research Software Engineer Team Needs 79
7.6.3 Organizational Needs 80
7.7 Healthy Work Practices 81
References 83
Appendix A - Career Paths 88
Appendix B - Position Description, MOU, and Buyback Program Examples 95

Cite as:
Van Tuyl, S. (Ed.). (2023). *Hiring, Managing, and Retaining Data Scientists and Research Software Engineers in Academia: A Career Guidebook from ADSA and the US-RSE*. Academic Data Science Alliance.
DOI 10.5281/zenodo.8264152

Foreword

Welcome to a first-of-its-kind guidebook traversing the largely uncharted territories of two vital career paths that have emerged to redefine the way we conduct scientific and scholarly discovery. We are honored to present this guide, which contains the collective efforts of dozens of volunteers from our two organizations: the Academic Data Science Alliance (ADSA) and the United States Research Software Engineer Association (US-RSE).

Academic research has undergone a paradigm shift in recent years, propelled by rapid advancements in technology, computing, and data-driven methodologies. This transformation has given rise to a new generation of research professionals: Data Scientists and Research Software Engineers (RSEs). With the increasing complexity of research challenges around data and software, these individuals have become the architects of cutting-edge software tools, the masters of data analysis, and a growing force behind groundbreaking discoveries from physics to digital humanities.

As the leaders of US-RSE and ADSA, we have witnessed the rise of these new career paths and the profound impact they have on the research community. We have seen the extraordinary dedication and passion that RSEs and Data Scientists bring to their roles, supporting and contributing to research projects, and collaborating across disciplines to achieve inventive and transformative results. Their contributions are vital to the academic ecosystem, revolutionizing the research process and fostering a new culture of collaboration and innovation. However, we have also observed unique difficulties encountered by these nascent roles in academia, an institution reticent to change.

Within the academic community, the roles of RSEs and Data Scientists are often underestimated, overlooked, and inadequately defined. Their contributions to projects frequently go unrecognized, or at best appear as a nod in the acknowledgments sections of scholarly publications, the currency of academia. Thankfully, as the need for computational expertise and data-driven insights continues to soar, these roles are increasingly gaining well-deserved recognition and credit for their contributions.

As these roles continue to grow and evolve, it has become evident that RSEs and Data Scientists face some strikingly similar opportunities and challenges. We, therefore, undertook this combined and unified effort to more effectively address the needs of these new professionals. As Maya Angelou once said, "We are more alike, my friends than we are unalike." This collaboration between US-RSE and ADSA signifies our shared commitment to fostering a supportive ecosystem that caters to the aspirations of RSEs and Data Scientists alike, amplifying their impact and propelling academic research forward.

This guidebook is a testament to our desire to support those who dedicate their talents to these increasingly important roles. It is an embodiment of our collective vision to nurture and empower RSEs and Data Scientists throughout their careers. Within this guidebook, you will find a collection of guidance, current best practices, and practical advice from accomplished professionals and group leaders who, over many months of collaborative volunteer efforts, have generously shared their knowledge.

We extend our heartfelt gratitude to all the contributors, volunteers, and our broader communities for their unwavering support in bringing this guidebook to fruition. Your dedication to advancing these nascent career paths and your commitment to fostering a supportive environment for RSEs and Data Scientists have made this endeavor possible. We sincerely thank our financial supporters, in particular the Alfred P. Sloan Foundation and the Gordon & Betty Moore Foundation, whose enduring support for these roles reaches back over a decade through the Moore-Sloan Data Science Environments and their continued support of ADSA and US-RSE. We also acknowledge the generous contributions of the University of Virginia, Princeton University, the University of Chicago, the University of California Berkeley, and Indiana University.

And finally, we owe a special thank you to Steve Van Tuyl, whose dedication and tireless efforts were instrumental in organizing, editing, and shepherding this guidebook to the finish line.

As we embark on this expedition together, we hope that this guidebook serves as a useful reference, resource, and guiding light for individuals forging their paths as RSEs and Data Scientists, emerging leaders preparing to establish new groups, and the institutions seeking to create an inclusive and thriving academic research environment.

With warmest regards,

Ian Cosden
Founding & Current Chair, US-RSE

Micaela Parker
Founder and Executive Director, ADSA

August 2023

Acknowledgements

We gratefully recognize the efforts of all of the workshop participants and reviewers who have provided content, ideas, and comments to this guidebook. A special thank you goes out to the working group members who pulled this workshop together and guided the editorial process throughout. Our colleagues at the University of Chicago helped us coordinate on-site logistics. A number of the content areas in this guidebook were inspired by discussions at the US-RSE Workshop in Princeton, NJ in 2022. Last but not least, the team at ADSA - Megan Atkinson, Stella Min, and Veronica Woodlief - was critical in the organization and management of the workshop and subsequent efforts to complete the guidebook.

The Guidebook was generously supported by the following organizations:

ALFRED P. SLOAN FOUNDATION

GORDON AND BETTY MOORE FOUNDATION

UNIVERSITY of VIRGINIA

Berkeley UNIVERSITY OF CALIFORNIA

INDIANA UNIVERSITY

PRINCETON UNIVERSITY

THE UNIVERSITY OF CHICAGO

renci RESEARCH \ ENGAGEMENT \ INNOVATION

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Chapter 1 - Introduction and Rationale

THIS CHAPTER COVERS:

Introduction to the guidebook: An overview of the content of the guidebook, including a general rationale for why the contents are important and unique

Who is the audience?: A listing of some interested parties and which sections of the guidebook might be of most interest to each party

Abstracts for other chapters: A brief summary of each chapter in the guidebook

KEY TAKEAWAYS:

Data Scientist and Research Software Engineer positions are unique: The changing nature of computing and research offers opportunities to recognize new types of positions in our research teams

This guidebook is useful for academics in many roles: Much of this guidebook is written for hiring managers, though elements are relevant to administrators, human resources employees, funding agencies, or data scientists and research software engineers themselves

1.0 Summary

The importance of data, software, and computation has long been recognized in academia and is reflected in the recent rise of job opportunities for data scientists and research software engineers. Big data, for example, created a wave of novel job descriptions before the term Data Scientist (DS) was widely used. And even though software has become a major driver for research (Nangia and Katz, 2017), Research Software Engineer (RSE) as a formal role has lagged behind in terms of job openings, recognition, and prominence within the community. Despite their importance in the academic research ecosystem, the value of DS and RSE roles is not yet widely understood or appreciated in the academic community, and research data, software, and workflows are, in many domains, still regarded as by-products of research. Data Scientists and Research Software Engineers (DS/RSEs) face similar challenges when it comes to career paths in academia - both are non-traditional academic professions with few incentives and a lack of clear career trajectories. This guidebook presents the challenges and suggestions for solutions to improve the situation and to reach a wide community of stakeholders needed to advance career paths for DS/RSEs.

The modern research environment requires an understanding and application of software engineering, evolving computational and statistical techniques, and application of unique technical solutions to research problems. While, in some ways, this has always been true of the research environment, the increased complexity of questions and an array of techniques for answering those questions has created a need for new types of positions that can focus on the application of advanced technologies and methods to the research endeavor. Academic institutions often want to hire domain scientists in fields such as biology, chemistry, or physics with literacy in DS/RSE, but it is no longer feasible for these domain scientists to be experts in the wide range of skills and techniques that may be critical to research in these fields.

Unfortunately, there is a shortage of faculty and staff who work in the areas of data science and research software engineering, which may be attributed to competitive salaries and benefits offered by non-academic institutions for individuals with relevant skill sets. The US Bureau of Labor Statistics projects a 36% increase in Data Scientist positions nationwide over the next decade[1], and a 21% increase for Software Engineers, generally, over the same time period.[2] This rise in the demand for individuals who work in the field of data science and software engineering will require academic administrators to evaluate the opportunities and address the challenges related to producing, attracting, and retaining students, faculty and staff who work in these areas.

1.1 Introduction to the Guidebook: "What is this all about?"

This guidebook summarizes observations about the current career path challenges encountered by academic data scientists (DS) and research software engineers (RSE). Our goal with this guidebook is to elevate the recognition of academic data scientists and RSEs and elucidate "good enough" practices for recruitment, management, career development, and retention of staff DS/RSEs in academic settings. This guidance was generated by a diverse working group of data scientists, research software engineers, hiring managers, and others in the field.

It defines generally what DS and RSE roles entail and offers suggestions for how to clarify the professional trajectories of these roles. While the focus here is academic data scientists and research software engineers to manage scope, there are a range of professions and job categories that share many characteristics with these two groups. These related professions include cyberinfrastructure professionals, information scientists, and data librarians. In some cases, there are also different names for what we describe as academic data scientists and research software engineers. Similarly, data scientist and research software engineer positions in academia can have a lot of overlap with these same job titles in adjacent sectors such as national labs, non-profit organizations, and even industry. The degree to which our observations and recommendations are applicable to these very related professions is left to the expertise of the reader.

Definitions for positions such as DS and RSE abound and can have significant overlap with related professions such as Information Scientists, Cyberinfrastructure Professionals, and Tenure Track Faculty. In many cases, there isn't a bright line between concepts like data and software, or between scientist and engineer, and any individual may drift among types of positions throughout their career.

While we do not intend to create authoritative definitions for these positions, we will use the below definitions to help the reader navigate the text.

- **Data Scientists** use computational and mathematical tools and create workflows to analyze data to create knowledge for a domain of research (e.g. medicine, wildlife biology, political science)
- **Research Software Engineers** design, develop, maintain, and extend software to support, enable, and accelerate research

[1] https://www.bls.gov/ooh/math/data-scientists.htm

[2] https://www.bls.gov/ooh/computer-and-information-technology/software-developers.htm

> "this guidebook is meant to be a reference for hiring managers and administrators on the motivation, means, and strategies for building and sustaining successful research programs and rewarding career paths for DSs and RSEs"

Given that research is about discovering new knowledge and inherently involves figuring out how to do things for the first time, there are important differences between the research software engineers and academic data scientists who are part of the institution's research enterprise and the information technology professionals and data analysts who support an institution's business functions. The research environment often requires individuals who have a much broader software engineering or data science knowledge base, who are comfortable using more leading-edge (if not bleeding-edge) technologies, who can see unexpected applications for technologies, and who are interested in learning elements of domain science essential to developing tools and pipelines that are fit for use and fit for purpose in an environment with evolving and often ambiguous requirements. These factors are why many successful research software engineers and academic data scientists started out as domain scientists who discovered a passion and a mindset for these roles that are focused on enabling scientific discovery.

1.2 Who is the Audience?

The audience for this guidebook is meant to encompass a variety of individuals who are employed in DS/RSE positions, those who manage and hire DS/RSEs, and those who interact with these individuals in the workplace. Importantly, this guidebook is meant to be a reference for hiring managers and administrators on the motivation, means, and strategies for building and sustaining successful research programs and rewarding career paths for DSs and RSEs. This guidebook also is a reference for data scientists and RSEs as to best practices for how to engage in a productive and fulfilling career in academia.

As a practicing DS/RSE, you may wonder about possible career paths and the best strategies for advancement and growth in your area. These roles are often ill-defined in relation to other faculty and staff roles, which may limit opportunities for growth via a concrete career path. This guidebook aims to provide examples of possible career paths (Appendix A - Career Paths) and an understanding of the challenges and opportunities presented by their position.

As a human resources representative working within a research organization, you may be interested to learn about titles and responsibilities for these positions, as well as how to be competitive in terms of hiring and retention of skilled DS/RSEs. This guidebook will clarify key challenges as well as potential strategies that can be used to hire individuals in roles that provide pathways to long-term career progression (Chapters 3-4).

As a PI of a research project, the lead of a research computing group, or hiring manager, you may look to justify the addition of DS/RSEs to your team. This guidebook will help you make the case that DS/RSEs are a vital part of the research enterprise and that their expertise and full-time focus on research-related tasks can

not easily be replaced simply by hiring more post-docs or students. It will help you identify strategies, including potential key stakeholders, for getting institutional support and buy-in for establishing new positions (Ch. 2,4,5-7), as well as define positions, career tracks, and salaries that will help you hire people into positions once you secure them (Appendix A and Appendix B).

As part of Institutional Leadership, you may be interested in understanding the definition and importance of these roles for your research organization as well as the appropriate payroll classifications for individuals in these roles. The chapters on the need for DS/RSEs and models for administering and funding (Ch. 1-2) can be used to guide the institutional decision-making process and to support career paths for individuals in these roles (Ch. 6 and Appendix A).

As a member of a funding agency, you may want to have a full picture of the skillsets, appropriate roles, and salary ranges for data scientists and RSEs that support funded projects (Ch. 1-3) and to have a deeper understanding of how to support these types of positions with your funding vehicles.

1.3 Abstracts for the Other Chapters in the Guidebook:

Chapter 2. Articulating the Need
This chapter covers how to make a case for hiring DS/RSEs in academia. Why are these positions necessary? How does one go about creating a position that matches the needs of the organization? And who are the key stakeholders in this process? We explore options for where the position could be located in the organization and what will be the components of the job. We also cover how to get stakeholder buy-in and consider the future of the position after the initial work is completed.

Chapter 3. Before Posting: Position Descriptions, HR, and Compensation
This chapter covers how to work with HR before posting the position description. We include how to: define a position, write a job description, determine the appropriate compensation, and craft a job posting. The chapter also includes discussion of some of the challenges that arise with this process specific to DS/RSE roles.

Chapter 4. Recruitment
This chapter covers how and where to recruit for DS/RSE positions. We include information on how to prepare a potential applicant pool for your job posting, including performing informational interviews and leveraging existing communities to attract a diverse pool of applicants. We also cover how to build and prepare a search committee to execute a search that is effective, fair, and equitable, and how to structure interviews and interview review sessions for unbiased review of candidates. Finally, we discuss items to consider when making an offer to a successful candidate and how to review your hiring process when complete to identify lessons learned for future searches.

Chapter 5. Expectations, Metrics for Success, and Onboarding
This chapter covers the groundwork needed to set individuals up for success in DS/RSE positions. We cover how to define and measure success for DS/RSEs and how to set clear metrics and expectations that will enable individuals to grow their skills and have a fulfilling career while balancing their own aspirations with group and university expectations. Readers of this chapter will consider the discussions needed during hiring, onboarding, and regular check-ins about expectations, metrics of success, and pathways to promotion.

Chapter 6. Career Development
This chapter covers how to retain individuals in DS/RSE roles and provide them with rewarding trajectories within academia. The discussion is geared toward directors/supervisors who manage research groups with DS/RSEs and domain scientists. We first discuss career opportunities and satisfaction that the academic research environment can offer. We then describe the diverse range of career paths that are possible for DS/RSEs, and how they and their managers can intentionally design career paths based on their current skills and aspirations. Then we compile examples of professional development approaches and opportunities, as well as perspectives from industry. At the end of the chapter, we discuss strategies for working with HR to support and institutionalize DS/RSE career advancement.

Chapter 7. Organization and Management of Research Software Engineer and Data Scientist Teams
This chapter explores the hierarchy of needs required for individuals to flourish and enable teams to achieve their objectives. Some needs are individual, such as training and compensation. Other needs focus on team dynamics such as accountability and healthy work practices. The final layer of needs is focused on an organizational level such as values and managing relationships with stakeholders, such as PIs. When building a strategy for their team, managers and directors can consult this list and assess how their institution lines up, or does not, with these needs.

Appendix A. Career Paths
This appendix discusses some of the implications for career paths for staff versus faculty positions. Many DS/RSE positions have some responsibilities that would qualify them for faculty roles and others that would qualify them for staff positions. There is no one right answer as to how DS/RSE positions should be classified, but it is important to understand the implications of the choice and to structure the position to be successful in the chosen type of role. The chapter includes examples of career ladders and career paths for a variety of DS/RSE career types (contributor, management, etc.).

Appendix B. Position Description and MOU Examples
This appendix presents example position descriptions from actual job announcements gathered by the workshop attendees. Position descriptions are grouped into Example Sets - groups of position descriptions from a single institution for either data scientist or research software engineer positions. These position descriptions are presented in their original form, and variability in their structure and content gives a sense of how different position descriptions can be.

Chapter 2 - Articulating the Need

THIS CHAPTER COVERS:

Context: Why DS/RSE positions are important and how they differ from other institutional employees

Identifying Stakeholders: Who are the important advocates, administrators, and blockers to creating DS/RSE positions at the institution?

Addressing the Need: How to frame the need for DS/RSE positions at the institution?

Models for DS/RSE Positions: Examples of models for positioning in the institution and focusing DS/RSE work

Funding and Sustainability: Considerations for short and long-term funding for DS/RSE positions

KEY TAKEAWAYS:

Understand Institutional Context: Every institution is its own labyrinth of regulations, politics, and personalities. Understanding this context will help you articulate the need for DS/RSE positions and allow you to talk to the right people at the right time.

Articulate "what" and "why": DS/RSE positions are new enough in academia that it will be helpful to create talking points for the different groups of stakeholders at your institution. Among the most common questions are "what is a DS/RSE?" and "why do we need DS/RSEs?" Being able to articulate what and why will be critical to building support.

Consider Sustainability from the Start: The longevity of the positions you create can have a major impact on the success of the DS/RSE enterprise. Try to avoid short-term or "one-off" contracts, as they can be detrimental to recruitment and retention of DS/RSEs and may have negative impacts on associated research projects.

2.0 Summary

This chapter addresses strategies for articulating the need and justification for Data Scientists (DSs) and Research Software Engineers (RSEs) as well as getting buy-in from higher-level management for the specifics of a new DS/RSE role. It focuses on providing helpful, implementable solutions to the following questions, which are best asked and addressed early in the process:

- **Why?** If you don't currently have DS/RSEs and think they are the right means to address the work to be done, why are DS/RSEs the answer? What problems will DS/RSEs solve?
- **Who?** Who are the key decision-makers and allies at your institution and what do they care about?
- **Where?** Where should these positions fit within the institution?
- **What?** What are the specifics of the DS/RSE role? What questions need to be addressed ahead of time to ensure success?
- **How?** How do you get buy-in and what are some successful strategies?

We assume you are at an academic institution and fall roughly into one of the following categories. Based on the category that aligns with your situation we highlight the sections that may be most relevant to you.

Principal Investigator (PI) with or without funding wanting to hire a DS/RSE for a specific goal. See Sections 2.1, 2.3, 2.5, and 2.6.

Staff member in an IT or research unit with aspirations to create a DS/RSE group (consisting of one or more DS/RSE positions). See all sections.

Staff member in an IT or research unit with aspirations to hire a DS/RSE into an existing group. See Sections 2.5, 2.6, and 2.7.

A university administrator wanting to create DS/RSE roles at your institution. See all sections.

Note that while this guidebook has chapters of interest to individual DS/RSEs, this chapter isn't aimed at current/future DS/RSEs, unless they also fall into one of the above categories.

RESOURCE

RCT Newsletter is a website and newsletter with many resources on hiring and managing Research Software Engineers, Data Scientists, and other related professionals.

https://www.researchcomputingteams.org

2.1 Context: Why DS/RSE Professionals?

If you need to convince your stakeholders why they should be funding or supporting DS/RSE positions, you should be prepared to address the following questions. And even if you don't need to convince anyone else, it's still good to keep these questions and their answers in mind while working on other parts of the process of hiring DS/RSEs.

What is the DS/RSE role? How does it fit within the research process and change the current paradigm of software development or data science?

You can base your description of the role on the templated role descriptions provided in Appendix B - Position Description and MOU Examples, but you should plan to tailor it to the stakeholders with whom you are engaging. You will want to address how the DS/RSE will integrate with the rest of the research team. You should look at how this role will improve upon the status quo in the team and the institution, and what benefits this might bring to both the research groups working with the DS/RSE and to the institution as a whole.

What is the value added by having a DS/RSE role?

Depending on the stakeholders and the research projects you are addressing, you likely will want to emphasize what full-time DS/RSEs can bring to the process that graduate students and postdocs cannot: best practices in software development, breadth and depth of knowledge that accelerates time to project completion, maintainability and robustness of code or workflow processes, or documentation, for example. A professional DS/RSE can also provide continuity in supporting a project over a longer period of time than a single

grant funding cycle, postdoctoral position, or degree. And a DS/RSE can bring time and cost savings, efficiency, and productivity to the research team as a whole.

What are some of the commonly mentioned disadvantages to the DS/RSE role as compared to the current paradigm?

Commonly mentioned disadvantages include high salaries, a potential loss of control over the role (e.g. if it becomes centralized), potential misalignment of DS/RSE skillset with project needs, and a lack of familiarity with the research environment. Preparing answers to these concerns can help in strategic conversations.

What are peer institutions doing to address these types of issues?

You can leverage some of the use cases and group descriptions from this guidebook, as well as pull examples from peer institutions. ADSA maintains an [3] institutional member directory and US-RSE has a list of RSE [4] groups that can be helpful for comparisons with peer institutions. Depending on your audience, present examples from institutions that are addressing similar research problems to yours, have comparable disciplinary strengths, or share a Carnegie classification with your own. Individuals and groups at peer institutions will often be happy to discuss the development and growth of their DS/RSE programs and positions.

2.2 Identifying Key Stakeholders

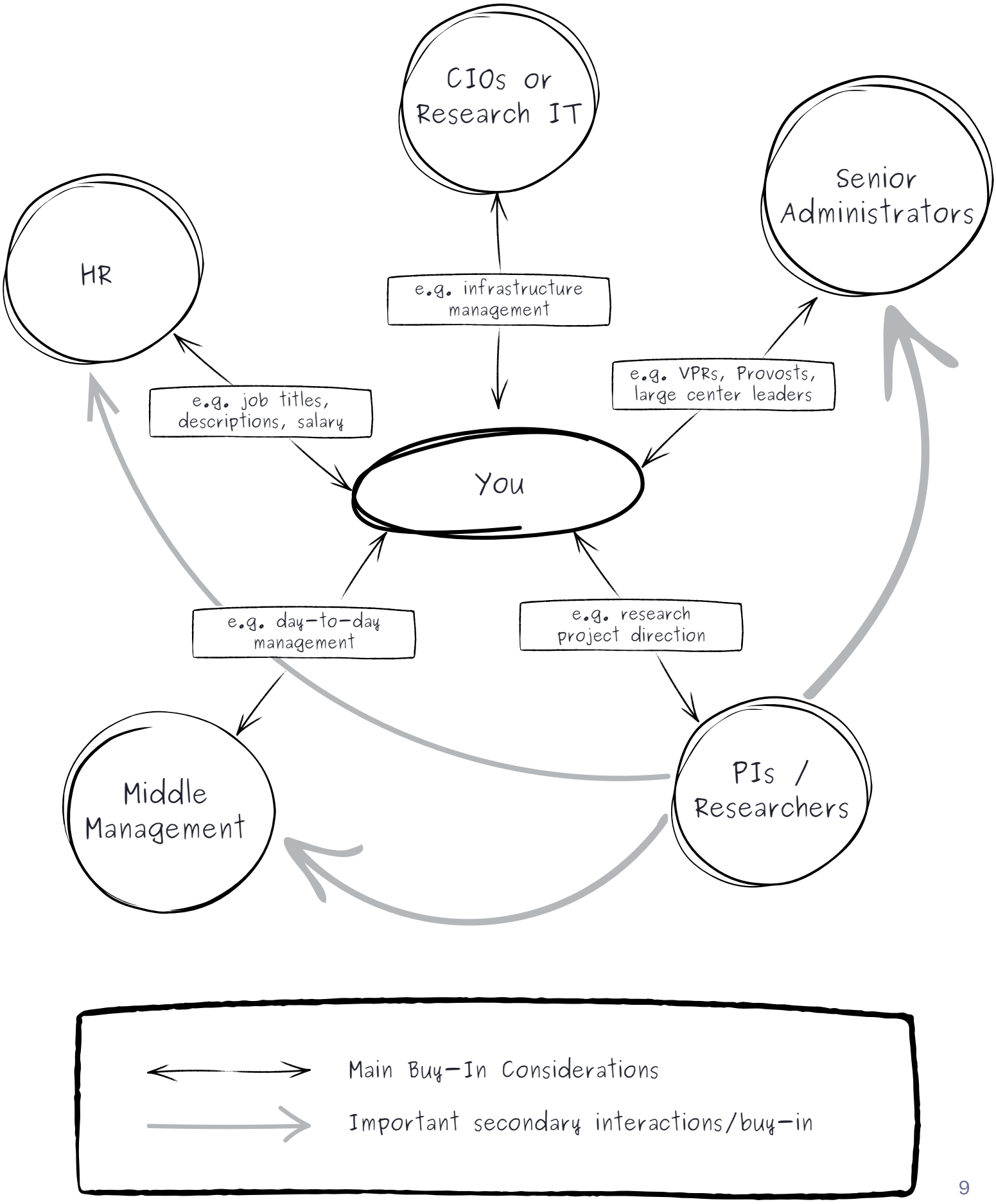
As with any other strategic initiative, it's critical to identify the individuals and groups within your institution that you need to support the investment in DS/RSE positions. This is like any other strategic initiative in that sense, so here we focus on some of the people you should consider for your specific situation and institution. It's unlikely that you'll have to get buy-in from them all every time, but you should at least be intentional about who you contact (and who you don't). Additionally, the order in which you will want to approach these stakeholders will vary depending on your institution and your position.

- **Researchers who will benefit from DS/RSE.** This group includes PIs (primary faculty, research scientists, etc.) and other researchers. You may want to start your outreach with this group. These people are likely to turn into your advocates, or if you are in the envious position of having funding without a clear connection with researchers, you'll want to identify a few initial PIs to help define the scope and role in a manner that will maximize the likelihood of success for the future position(s). PIs and other researchers are typically hyper-focused on their research, are quick to articulate immediate challenges, and are always thinking about future funding and research directions. You'll want to identify what technical skills and expertise a DS/RSE could provide that would have maximum impact on their work. You may need to define what DS/RSEs do, and how such collaborations can work in the context of your institution and the PI's work. Many PIs have never worked with a DS/RSE and would benefit from a clear, possibly

[3] https://academicdatascience.org/

[4] https://us-rse.org/

Potential Stakeholder Buy-In



very detailed, articulation of the role and the way to interact with a DS/RSE. You may need to sell the idea or you may need to temper expectations. Having a few key and influential faculty members advocate on your behalf to champion the need can be very effective. You may need to recruit these advocates, convince them that DS/RSE can have a major impact on their work, and encourage them to present the need to senior administrators. You may need to arm your advocates with specifics to ensure the message is consistent with your needs.

- **Middle management**. This stakeholder group, possibly including department chairs or deans, may not be able to perform the work of a DS/RSE themselves, nor might they be highly familiar with what that work entails, but they may well be responsible for setting the vision and the day-to-day administration of the department in which these roles will sit. They will want to know that their leadership is supportive of the idea and may be concerned with the practical and operational logistics of having a DS/RSE in their department or unit, such as who their manager will be, who will do their performance review, where they will sit, etc. You want to present a similar case of value to the department as you've made to PIs, perhaps with fewer technical details. You'll want to carefully listen to their concerns as they could torpedo your efforts if they don't buy in.

- **HR**. Because HR will approve job titles, descriptions, salaries, levels, etc. and influence where a new position will fit in your organization, you'll want to engage them in early conversations in order to negotiate position definitions, the administrative and physical placement of the position(s) at the university, compensation, and management structures. This may need to be done in parallel with other discussions with other stakeholders. Keep in mind that HR may have trouble understanding the specifics of the DS/RSE role and type of work. You'll need to find a way to articulate the specifics in a way that will ensure the positions are classified and structured appropriately. Read more about Working with HR in Section 3.1 - Working with HR.

- **CIOs and/or Research IT leadership**. This group will include the key administrators who manage the IT infrastructure that the DS/RSE professionals will rely on to do their work. When making the case to research IT leadership that DS/RSEs are the right way to meet the research need and that they will benefit the institution as a whole, you may want to focus on the collaborations that DS/RSEs can have with IT infrastructure providers, and highlight the fact that the DS/RSE group can both champion the value of the research infrastructure and be championed by the research infrastructure team as exemplar users of those systems to facilitate research in a highly efficient manner.

- **Senior Administrators**. This group may include VPRs, Provosts, and/or large center leadership. They often care about attracting and retaining the best faculty and researchers, increasing scholarly output, increasing the success and amount of externally sponsored research, improving or maintaining institutional rankings, comparison to peer institutions, ensuring the quality of research, and financial sustainability and impact. When making the case to Senior Administrators you may need to be quick, data-driven, and prepared with the key items that they care about. You may want to consider summarizing examples of peer institutions, financial sustainability plans, and clearly articulating the potential impact on the

research enterprise. The support of the faculty champions you identified earlier will likely go a long way here.

2.3 Addressing the Need and Getting Buy-In

If you have already secured funding, you may already have a clear remit and specific requirements for the DS/RSE position. However, if you don't, or if you are still considering how to fund the position, answering specifics about the role may be helpful. It is important to get agreement from all stakeholders on what the DS/RSE will do, as this lays the foundation for a future job description, and getting buy-in at this stage helps prevent mismatched expectations in the future. Some of these questions may be simple and straightforward to answer but ensuring all parties are in agreement is critical.

What will the DS/RSE work on initially?

Defining the initial project(s) with the key stakeholders serves three key functions. First, it will bring clarity to the role definition by specifying the type of technical work and the role within the research project(s). Second, we've found that the "if you build it, they will come" is less persuasive than the "here is the work that is already lined up for a future DS/RSE." Clear illustration of projects that will benefit from a DS/RSE position, or alternately, projects that cannot move forward without the DS/RSE makes the impact on research clear and tangible. Note, though, that you may need to hire without full knowledge of someone's future work – effectively, make a bet on future work coming in – when there is a commitment for partial funding in a new area and you need to hire a new person to meet that commitment. Having an accepted initial slate of projects prevents early competition for DS/RSE time, likely to the benefit of the research and, importantly, to the DS/RSE themselves.

What will the outputs of the DS/RSE work be and how will the person be evaluated, promoted, and mentored?

The outputs of the DS/RSE position need to align with how the success of the position will be measured. Outputs of such a position might include software, datasets, algorithms, analyses, answers to research questions, reports, or papers. As an example, there could be a time tradeoff between writing code at a professional level (e.g., documentation, tests, reproducibility, etc.) and primarily writing manuscripts. Qualitative measures such as "researcher happiness" – what is the research group doing now that they could not have done without the DS/RSE working on a project? – may also be valuable. Further, you can clearly define how the DS/RSE will be mentored based on their individual career goals and how they will grow and learn in the role. Additional information can be found later in this guidebook on how to develop expectations (see Chapter 5 - Expectations, Metrics for Success, and Onboarding), career development (see Chapter 6 - Career Development), and manage and quantify outputs (see Chapter 7 - Organization and Management of Research Software Engineer and Data Scientist Teams).

How will the DS/RSE get credit for their work?

How credit is allocated for these positions should align with how the DS/RSE's time is spent and how outputs are measured for success. As a few examples, small one-off automation scripts might just require an acknowledgment in a paper, while writing a key software package or data pipeline should include co-authorship of the software and/or paper about it. A substantial project might require the

DS/RSE to be the lead (author) in writing a software package or paper. You should consider if the different scopes of these projects are weighted appropriately in credit for the time, outputs, and credit. You should also consider whether your DS/RSE's department or unit should have a standing policy or some formal guidance on assigning appropriate credit to the DS/RSE for different types of projects. This policy should be communicated to collaborators as early in the research process as possible to avoid conflict or confusion about assigning credit (see Section 6.4 - Credit).

What happens to the DS/RSE and to their work when the initial project(s) end?

Short-term positions for DS/RSEs are suboptimal from the perspective of both the institution and the DS/RSE. Filling short-term positions without a longer-term plan for the DS/RSE position risks the DS/RSE leaving the projects and institution early in order to find new opportunities. Personnel churn is quite disruptive to the research enterprise, especially with complex and specialized software systems and methodologies at play.

If the output of the DS/RSE's work will be some kind of software, whether a full-fledged package, prototype, or operating application or service, we strongly recommend having a discussion with stakeholders about who will maintain the software/service when the project and/or funding ends. Maintenance and/or operations can be a significant effort. Is it reasonable to assume this new DS/RSE is expected to take a new project but also maintain a previous project? If not, is the project PI prepared to maintain it?

If this project will not cover all the DS/RSE's time, what other responsibilities will this person have, and how will they be supported?

It should be clear why this question needs to be answered, but if you don't get early agreement from stakeholders, you run the risk of having multiple people provide their own answers to the question. Will there be a teaching component? Will the DS/RSE be responsible for seeking their own funding support? This also applies to a DS/RSE when the initial projects on which they have been funded end. Answers to these questions can fundamentally change the classification (e.g., faculty vs. staff) and unit (e.g. IT, research department, center) of the future position. For example, some institutions do not allow staff to act as PI on grants, removing this possibility from the list of DS/RSE responsibilities.

Where does this position fit within the institution?

This may be dictated by the source of the funding or the scope of the work. If not, or if there is some flexibility, there are a number of possibilities for hosting a position. Many of these will be unique to your institution, but it is important that you seek input into the answers and pursue the possibilities. Will this be a central group, embedded in a research lab, or a department or center? How will the position be classified? This is typically a question for HR, but stakeholders may have different expectations. Who will this position report to? Is there a DS/RSE manager?

In many cases, it may make sense to separate the reporting structure from the funding source, e.g. using matrix management. As you discuss this with the key stakeholders, you'll want to consider who has the resources, time, and knowledge to be an effective manager for a DS/RSE. Aspects of this question are also addressed in later sections of this guidebook.

2.4 Positioning

There are many options for where a DS/RSE can fit in a campus research environment. The first question is if there will be a group of DS/RSEs or an individual DS/RSE associated with one or more projects. If there is a single individual, the DS/RSE probably will belong to some existing group that is not DS/RSE-specific where there is some existing management and structure of the group, such as a research lab or department. If there is a DS/RSE group, some administrative structure is needed, including at least a manager, but for a larger DS/RSE organization, there could be multiple levels of management. In either, there is a home for the DS/RSE, which could be non-centralized (e.g. in a faculty member's group, in a department, in a college) or centralized (e.g. in a university IT organization, or in a research center/ institute/organization that has a campus-wide mission).

It is impossible to say what type of positioning works best for any given organization, given differences in administrative practices, vision for the DS/RSE group, and institutional history. Where the impetus for the DS/RSE position(s) comes from can have a major impact on how the position is structured and where it sits in the institution. For example, the details of positioning will differ for hiring a DS/RSE on a major grant in the Chemistry Department versus a Provost-led initiative to stand up centralized DS/RSE services at the institutional level. Consider looking at other new initiatives, research groups, and administrative units at your institution and see how they emerged and what issues their management may have faced.

If the DS/RSEs are not part of a centralized organization, or even if there is both a centralized organization and individual DS/RSEs elsewhere on campus, either the DS/RSEs themselves or the central DS/RSE organization should build a campus-wide (distributed) DS/RSE community. The goal of this community is to create a venue where DS/RSEs can assemble, talk, and have colleagues for troubleshooting, moral support, and discussing potential organizational issues. This helps communicate technical issues and solutions between DS/RSEs. It also helps individual DS/RSEs who might otherwise feel isolated, as the only DS/RSE in their organization.

Additionally, a distributed DS/RSE community can have a collective voice when needed. For example, if the DS/RSEs are not part of the campus IT department, their central organization or distributed community probably needs to have at least some regular informal or formal communication with the IT department.

2.5 Potential Models for DS/RSE Positions

DS/RSE positions can have different focuses, each of which likely has a particular type of funding required.

Service to a Particular Team: All or part of the DS/RSE's time is dedicated to a single research group or team. For example, building or maintaining software for the team or analyzing data for the team. This type of work, dedicated to a team, is likely to be funded by that team, such as through grants, the startup funds of the faculty member who leads that team, or institutional support dedicated to that team.

Expertise in a Particular Technology: Focusing on development and support for the use of a specific technology at the institution (e.g. Python, applied linear algebra), or for computational expertise in a specific domain area (e.g. chemistry, linguistics). This would likely be funded by the institution, or by a project in a specific, related technology area.

Provision of DS/RSE Consulting Services: Usually funded by the institution or a large unit such as a school or well-funded center, a DS/RSE may focus on consulting with researchers through short and medium-term engagements requiring their skills or expertise.

Independent Research: Research to build software or a tool that the DS/RSE thinks will be useful to others either in the institution or in a broader community. This could be funded by grants won by the DS/RSE, grants won by others, or through institutional support.

Contribution to Community Projects: Contributions to extra-institutional community projects, such as community open-source software used by a broader domain area (e.g. Jupyter). This is mostly likely funded by institutional support at a large scale, but it also could be funded by grants won by the RSE or other related projects that use the community projects.

2.6 Funding Sources and Sustainability

Last, it is important to think about how the position you are trying to fill is going to be sustained after the initial project ends, whether it's an external grant, internal funding associated with a faculty member, or some other funding vehicle. No matter which is the case initially, it's extremely likely that the career of the DS/RSE at the institution will endure longer than the initial funding, and it is important for administrators, managers, and the DS/RSE to have a full and transparent understanding of current and expected future funding streams.

Some questions to consider include:

- Will the DS/RSE be partially or fully responsible for finding their own future projects, such as writing grants to lead their own projects, finding internal or external collaborators, etc.? Has an appropriate amount of time been allotted in the DS/RSE's position description to meet these responsibilities?
- Will some of the DS/RSE's time always be institutionally supported with a corresponding obligation to provide a service (e.g., consulting, hands-on DS/RSE work) to the institution? What guarantees do you have to this institutional support and what is the timeframe on which that support is certain (e.g. 5 years? indefinite?)?
- What flexibility and autonomy does the DS/RSE have to move or shift time between externally-funded projects vs internally-funded work?

Chapter 3 - Before Posting: Position Descriptions, HR, and Compensation

THIS CHAPTER COVERS:

Working with HR: why you need to work with HR, a brief overview of the process, and some of the challenges that DS/RSE positions create for HR

Defining a Position: considering the responsibilities and requirements for a position and then mapping those to official job descriptions

Compensation Considerations: identifying and addressing some of the challenges with DS/RSE compensation

Writing the Job Posting: tips for creating a job posting that will support your efforts to attract a diverse set of qualified candidates for your position

KEY TAKEAWAYS:

DS/RSE positions are often challenging for HR: Due to their new and rapidly evolving nature, DS/RSE positions may be challenging for academic HR departments and can take extra effort to define, hire, and appropriately compensate.

Take a long-term view with HR: Working with HR to develop job descriptions, career ladders, and compensation ranges that are appropriate for DS and RSE positions can require a long-term, iterative approach. Accepting good enough solutions in the short term while working toward better solutions in the long term is a workable strategy, and is often the clearest way forward.

You can't get a unicorn for $50,000: If there is a mismatch between your position description and your target salary, either adjust the level of the position or increase the budget.

Job postings matter, so take care with them: The language and structure of job postings will affect who applies. Keep listings concise and use inclusive language to support recruitment and hiring efforts.

3.0 Summary

Before posting a position for DS/RSE, you need to define the expectations of the role; determine the official job title, classification, and description; set the compensation; write a compelling job posting; and determine what materials are needed from candidates for evaluation. To do this effectively, close coordination with Human Resources (HR) is typically needed. This chapter aims to provide guidance on the essential elements of a job description, identifying the right job classification and level, effective job postings, coordinating with HR, and compensation considerations.

There are two common situations when seeking to hire: 1) there is an existing job description defined with HR, with an associated compensation range, job title, and other key elements already in place; 2) you need to create a new job description with HR because an appropriate one does not exist or notable changes are needed to

an existing role, such as a compensation evaluation or a significant change in the job responsibilities to meet your current needs.

If you think you are in the first situation, double-check that the existing position is still a good fit for what is needed, as DS/RSE roles evolve more quickly over time than positions in better-established career paths. In particular, references to specific technologies or the compensation range for the position may need to be updated. If you are fortunate to have an existing position and compensation range that is appropriate for your open role, congratulations, you can likely skip some of the sections in this chapter.

3.1 Working with HR

As a hiring manager (the person responsible for making the hiring decision and running the hiring process), working with HR is critical to structuring and supporting roles for DS/RSEs in academia. Beyond ensuring compliance with employment law and university policies, HR is the unit that has information on:

- The job families and descriptions that are available at your institution
- How to determine equitable compensation levels
- Requirements for the hiring and promotion processes

While HR is often framed as a hurdle to overcome, they are important partners in the hiring process, and developing a productive relationship is key to supporting DS/RSE careers.

HR may have different goals and priorities than the hiring manager for a position. For example, HR may be focused on compliance, while the hiring team is looking for flexibility to meet an exceptional candidate's expectations. This may result in conflict, such as HR telling hiring teams that something is not possible even in situations in which there are ways to achieve what is being requested. But this is important: HR is paying attention to necessary things you might not. Part of HR's role is to keep positions and policies consistent across the institution, and this is important for both equity and ensuring the career paths of DS/RSEs can be supported. Yet HR's focus on consistency and existing policy can be a particular challenge for DS/RSE positions that do not conform to existing models and are rapidly evolving.

Besides keeping this difference of perspective in mind, there are a few strategies for working effectively with HR:

- Become familiar with the rules and regulations at one's institution. This will help you understand HR's constraints and may also lead you to potential solutions
- Identify the individual in your department, center, office, or school who has designated HR responsibilities to assist you
- Engage with others around your institution who may have implemented new or different staffing models to discuss strategies and engage with HR
- Help educate HR staff on the complexities of DS/RSE positions and how they differ from other established positions on campus (use this guide!)
- Changing HR policies, developing new positions, or securing exceptions so that your positions meet the needs of DS/RSEs can require a long-term approach

3.1.1 A Few Terms

Institutions use different terms to refer to similar concepts. In this chapter, we are using terms as described below.

- **Job Description**: a document defining a position's responsibilities and the qualifications for the position. The job description is the basis for assessing an individual's performance in a position and is used in determining the grade of the position and compensation range
- **Job Posting**: the text officially defining and advertising an open position posted on a university's career website where applications are submitted. Those applying for a job see the job posting. At some institutions, the job description is also the job posting
- **Job Family**: a broad set of related staff positions and occupations, consisting of multiple career ladders. Examples of job families include: finance, IT, research technologists, librarians, and administrators
- **Career Ladder**: a set of closely related job descriptions defining progressive levels of role, for example: Associate Data Scientist, Data Scientist, Senior Data Scientist, Lead Data Scientist. Position grades and compensation increase as you move up a career ladder to reflect increasing responsibility and qualifications
- **Position Grade**: also known as a pay band or level, this determines the broad salary range for a specific position. These may be determined institution-wide or may be specific to particular job families or categories

Check your institution's HR website for information on the terminology they use.

3.1.2 HR Process Basics

When working with HR on writing a job description and getting a position posted for hiring, expect to engage in an iterative process. While this is true for posting any position, it is especially applicable to DS/RSE positions that may be new for an institution.

1. Put together initial information on what the job will do and the elements to go in the job description.
2. Work with HR to see where and how the position will fit into the hiring frameworks of the university. This includes determining the class of the position (staff, faculty, research staff, postdoc), and then the job family, level, and compensation grade for the position. Determine if there is an appropriate existing job description/ position, or if a new one will need to be created.
3. Check the actual target salary within the compensation range. This is often the midpoint, any deviations from the midpoint or other target may need to be justified.
4. Make adjustments as needed and finalize the job description. Go back to #2 to make sure everything is in alignment.
5. Determine what, if any, additional information beyond a resume will be required during the application (such as a cover letter or diversity statement).
6. Craft a job posting. At some institutions, the job description and the job posting are identical. In such cases, you may need to revise the job description with its dual role as a job posting in mind. Return to #2 as needed.
7. Discuss with HR what the selection, interview, and process procedure entails to ensure there is clarity on participants' roles and the steps involved. For example, what will be required by HR to select a candidate and make an offer?

Ensure you know the procedure before beginning the interview and recruitment process.

8. Get the job posted on the university website. Use the tools in Chapter 4 - Recruitment to advertise the job elsewhere.
9. Work with HR on the process to screen and interview candidates. The specific requirements vary by institution. This part of the process is covered more in Chapter 4 - Recruitment.
10. Select your candidate and make an offer.

Many institutions have guides for hiring managers that provide an overview of the process, policies, and terminology for that specific institution. For example, see Harvard's Hiring Manager Packet (Hiring Manager Packet 2021) or Stephen F. Austin State University's Guide for Hiring Managers and Search Committees (Hiring Toolkit: A Guide for Hiring Managers and Search Committees 2023).

In discussions about hiring between DS/RSE leaders from different institutions, the primary theme to emerge is that each institution has different policies, practices, and expectations. Some of these differences are determined by an institution's public or private status. Other rules may be determined by statewide employment laws. Still others are the quirks of each institution. Do not assume that how things worked at a previous institution will match how they work at your current institution, or that policies are even necessarily the same across schools at the same institution. **Again, it is imperative to engage with HR early and often to learn the specifics for your institution.**

3.1.3 Challenges DS/RSE Positions Pose for HR

Research software engineering is an emerging profession, and the field of data science continues to evolve, especially in the context of academic research. This can create a few challenges for HR.

No existing job family: Many institutions have established job families into which staff hires must fit, but there may not be a job family or career ladder that exists that fits these new DS/RSE roles. For example, the closest matches may be those for IT or Research Study/Technologist positions, where the former may not fit roles with research responsibilities, and the latter may only have roles that are too junior for DS/RSE positions.

Subtle differences to existing positions: At institutions with highly regulated or formalized job families, career ladders, and positions, HR may push for DS/RSE positions to be mapped to similar existing positions, such as statistical analysts, general software developers, or even roles like systems analysts. Articulating exactly how DS/RSE roles are different in their responsibilities or the requirements for the position is a useful exercise for refining your DS/RSE role and determining whether similar positions are a good fit or not. See Appendix A - Career Paths for benefits of defining DS/RSE-specific career ladders.

Comparable Positions: While there are benefits to defining DS/RSE-specific positions, doing so can make it even more challenging to find the correct comparison positions for compensation evaluations. While the mix of technical skills and research experience/domain knowledge is what makes these roles unique, the inclusion of research responsibilities can pull compensation evaluations downward,

as HR may look to other research support positions, or even non-tenure research faculty positions, as comparison points. These positions generally have salaries below data science and software engineering positions outside of a research context. More on compensation below.

Evolving position definitions: As DS/RSE groups develop, the responsibilities of those hired onto these teams may change quickly. Even for established groups, changes in the job descriptions may be required due to changes in prevailing technologies or the evolving needs of researchers. This may result in more frequent updates and revisions to job descriptions than HR prefers.

Misalignment of job levels or requirements: HR may have strict guidelines and requirements in terms of years of experience or degrees for particular position levels in a career ladder that do not match expectations from DS/RSE fields. For example, there may be a requirement that all "senior" roles have 10 years of related experience, which may not be realistic for DS/RSE roles. Similarly, HR may want to list specific majors or fields for degrees, which would exclude applicants from less common backgrounds (e.g. social science or the humanities) that you may want to hire.

3.2 Defining a Position

If you are hiring someone into an existing position that is working well, or already have someone in a comparable position, you can and should use that existing job description. It is always worth reviewing, however, whether the current description is a good fit for what the new hire will be doing, especially since the DS/RSE fields are evolving rapidly.

If you are creating a new position, start by defining the role's responsibilities and the experience you need someone to have to be successful in the role. This can be done first before engaging with HR. Once you have this information, work with HR to determine whether there is an existing position description that fits what you need for the role. If there isn't, then you will need to work on a new job description. There is wide variance across institutions on how involved the process of creating a new job description is likely to be.

3.2.1 Responsibilities and Requirements

Start by determining what you need from the position separate from an official job description. What work is the person in the role going to be responsible for and what skills do they need to be successful in that role?

Responsibilities:
There are some broad categories of activities in which many DS/RSEs engage and responsibilities that those in DS/RSE roles have:

- Teaching or training
- Collaborative consulting: often short-term, may be free or for a defined fee, often without direct credit in publications
- Software development, coding, writing analytics code
- Research collaborations: usually longer term, driven by a research PI, usually with authorship credit or acknowledgment
- Independent research: research directed by the DS/RSE
- Outreach, communications, and community building
- Management and supervision of others: students or staff
- Management or supervision of a service, such as a training or consulting service

A position *cannot* successfully do all of these well. Some roles are primarily teaching and short-term consulting. Other roles partner a data scientist or RSE on a single project for a longer period of time. Still others may be in a PI role. What work will the person spend most of their day doing? Make a list of the top 3-5 things for which this role will be responsible and on which they will be evaluated.

Beyond the categories of responsibilities, you also need to determine the level of responsibility the role will have. This will help you map the position to the appropriate level of the career ladder (see Appendix A - Career Paths). Levels of responsibility increase as the position level increases. Those in entry-level positions are expected to need guidance and have responsibility only for small projects and their own work, while those in senior and advanced positions are expected to work independently, direct the work of others, and take on larger and more complex tasks and projects. For example, within the area of teaching or training, an entry-level position may be expected to teach workshop materials developed by others, while someone at the lead level may be expected to determine what topics the team will offer training on and develop new workshop materials.

Class of the position (faculty vs. staff):
If the position has responsibilities for conducting independent research or initiating research projects, publishing research as a lead author, or bringing in grant funding, these are strong indications the position should likely be classified as a faculty position or at least a class of position that has PI status at your institution. See Appendix A - Career Paths for more details.

Requirements:
What skills and experiences do candidates need to successfully fulfill the responsibilities of the role at the level required?

- **Degree:** academic environments have a strong bias towards requiring graduate degrees, but is one truly necessary for the role? Could someone with experience working in a research environment be successful without a graduate (or undergraduate) degree?
- **Experience**: Instead of defining years of experience (HR will likely have a say in this later), focus on the skills and experiences you're looking for someone to have.
- **Technologies**: Does the position require knowledge of specific languages or frameworks, or would more general experience with the ability to quickly learn the specifics of the role still allow someone to be successful?
- **Domain knowledge**: is specific domain knowledge required for the position, or are you looking for more of a generalist who learns quickly and can pivot between technologies and projects as needs change?
- **Communication and collaboration**: Does this position work as part of a team? Interface with other researchers? Lead or provide a service? Communicate publicly? What skills will someone need to work well with other humans?

With requirements, it can be useful to distinguish between the true minimal qualifications – what will someone have a difficult time being successful without – and preferred qualifications that would be beneficial for someone in the role but could be learned on the job. However, it is important to carefully differentiate these types of requirements as they can have an impact on your later discussions with HR, the quality of the hiring pool, and the success of the DS/RSE who fills the role.

3.2.2 Position Logistics

In addition to defining what the role will be doing and the skills needed, there are some administrative and logistical details about the position that can affect the job description and classification process.

Information to gather for a productive conversation with HR:

- What is the funding source for this position? Is it a grant, central administration funding, or some other combination?
- Will the position be term-based (a fixed length of at least a year, often with renewal possibilities) or permanent (no set end date or renewal process)?
- Is this a full-time or part-time (or potentially split/partial) position?
- What level of position are you targeting: someone just completing either undergraduate or graduate education, someone with work experience, or someone with the ability to lead or manage others?
- Who will supervise this position? Will this position serve as a supervisor? If so, for what position(s)?
- Will the position allow for remote or hybrid work?

3.2.3 Find or Write a Job Description

Once you have the set of responsibilities and required experience, work with HR to determine whether there are any existing job descriptions at the university that will fit your position or that can at least serve as a basis for a new job description. Institutions vary in terms of openness to new job descriptions. In some cases, proposing a new job description would be a major endeavor. At others, HR may be open to new or revised job descriptions for each role. This is where working with others who have hired people at your institution is highly beneficial.

Look for an existing description:
As you're looking for existing job descriptions, consider roles and job families that may be outside of your primary organization. Job titles may not have any resemblance to "data scientist" or "research software engineer" but the responsibilities might be a good match (see more job titles below). It is more important to check whether the skills and compensation range for any position you find matches your expectations, as choosing an existing job description may determine what salary you're able to offer, the position's career path and promotion possibilities, and the pool of other employees used for equity or performance evaluations.

Your HR representative should be able to help you search for potential existing job descriptions once you share your list of responsibilities. If you cannot find a good match, it will still be important to find the closest matches you can and highlight very specifically what is missing from or wrong with the existing positions you find. This is an important exercise for having a productive conversation with HR about the need for a new position or career ladder.

Writing a new description:
HR likely has a template that you will be required to fill out with the details of the job description and some of the information on the logistics of the position noted above. The similar existing job descriptions you've already identified will help you determine the appropriate level of detail for job descriptions at your institution. Lean towards fewer, more general requirements where possible and appropriate for your institution to help the description be flexible and relevant enough to be used again in the future.

Even where it is not required, writing job descriptions for a full career ladder (see Appendix A - Career Paths) for your new position is a worthwhile and recommended exercise. This will help to ensure that there are clear distinctions between different levels, help you determine what level of your position should really be, and help HR grade the position correctly (see below).

Positions will be distinguished by:

- the complexity of the work someone is expected to do in a role
- the degree to which someone is expected to work independently or lead the work of a team
- the scope of their interactions across an organization and outside of their unit
- supervisory or managerial responsibilities

Your HR department likely has key phrases and terms they use to distinguish position levels or that they look for to classify a position as a senior, lead, or other level. There may also be formal expectations about degrees, years of experience, or other requirements for different levels. Whether there is flexibility in these requirements varies by institution, especially with the public vs. private status of the institution.

Creating a full career ladder will also help if you have the opportunity to post a position at open rank or post multiple positions at different ranks, a strategy that can be advantageous for attracting a range of candidates (more below).

Position grading:
Existing job descriptions will be associated with a grade that will determine the broad salary range and determine where the position is in its career ladder. For new positions, HR will need to review the position and determine the grade. The position grade, level, and compensation may be closely tied. More on compensation considerations can be found below. If HR does not grade the position at the expected level, a revision of the job responsibilities or requirements may be necessary.

3.2.4 Job Title

At some institutions, especially if you are writing a new job description, you can determine the job title. Where possible, using "Research Software Engineer" or "Data Scientist," with the appropriate level (senior, lead, principal) added where relevant, helps send a clear signal to those in these growing communities that the role aligns with the type of work they are looking for. Using these consistent job titles may also make it easier and clearer for folks to move between academia and industry.

In other cases, the official job title may be determined by the job description. Where the official classification is something like “business analyst 3,” technology specialist II," or “statistical scientist,” it is worth discussing whether the job posting (see below) can list a different "working" or "business" job title that may be more descriptive of the position and attractive to candidates; if so, using "data scientist" or "research software engineer" is again recommended. If the job title cannot be altered for the job posting, you may still be able to give the position a working title for day-to-day operations.

3.3 Compensation Considerations

Matching industry salaries within academia and government research labs, where funding is tuned for summer salaries, graduate students, and postdocs, is extremely difficult. In addition, budgets may be inflexible due to institutional or funding agency requirements and proposal teams tend to push compensation elements of the budget lower in order to acquire funding.

Taking these factors into account, it is important to include the full picture with regard to compensation, as well as utilizing the budget available as strategically as possible. In regard to the overall compensation picture, it is important to take into account the very different climate within an academic setting, desirable aspects such as being encouraged to learn, the ability to be creative/architecting solutions vs simply resolving issues in a sprint, the potential for real-life societal impact (as opposed to increasing a profit margin by some small percent), and typically a very generous amount of personal time and sick leave. Universities also tend to have good benefits packages and tuition waivers for individuals and their family members if desired. These benefits can go well into the tens of thousands of dollars if enumerated and as such should be stated as part of an offer.

Compensation includes benefits beyond just salary, but the focus in this section is on salary, as the other forms of compensation offered by academic institutions are generally fixed and cannot be negotiated for individual positions. However, the additional benefits can be useful to advertise in recruiting candidates for academic positions (see Chapter 4 - Recruitment).

3.3.1 Compensation Basics

There are two primary factors in determining the salary for a position:

1. How much money do you have available in your budget and/or approved by your unit to pay someone?
2. What salary range does HR say is appropriate for the position?

At academic institutions, these two factors generally combine to result in a fairly narrow target salary range, often with a hard limit on what compensation can be offered for the position. It is important to know what the compensation targets and limits are before posting, as a mismatch between the job description and compensation levels may necessitate a revision of the job description or budget.

Academic salaries for specialized technical roles are lower on average than industry salaries for equivalent positions. This is true across a range of positions and is a perpetual challenge for hiring. It may not be realistic to try to match market salaries for DS/RSE roles, and when you consider total compensation packages that include stock options and bonuses, academic roles will not compete. Yet, the other benefits of working in an academic environment can make salary differentials acceptable for many candidates. However, this only holds when academic salaries are in the vicinity of market rates. The greater the gap between the salary for your position and the market rates for similar positions, the more challenging recruitment and retention will be.

3.3.2 Compensation Challenges and Strategies

Possible compensation challenges for DS/RSE positions include:

1. HR's compensation evaluation is below your expectations and below market rates
2. Your budget – whether determined by your institution, funding agencies, or other forces – is below market rates for your area
3. Stakeholders are reluctant to pay DS/RSEs equivalent to or more than faculty positions
4. Rising market salaries for DS/RSE positions result in compensation evaluations for open positions that are notably higher than compensation for existing employees, creating equity challenges
5. Compensation evaluations have not been updated recently, resulting in grades and salary targets for existing job descriptions being out of line with the rapidly evolving market rates for DS/RSE positions

Different considerations factor into each situation. Read below for details:

HR Compensation Evaluations: If the HR compensation evaluation is below your expectations, start by inquiring about how equivalent positions were determined for the comparison. HR may not be willing or able to share the full details of the compensation evaluation and the data they used, but they should be able to explain the factors that influenced the evaluation and provide examples of the types of positions they used in the analysis.

Note that the inclusion of significant research responsibilities in the job description may result in HR looking to postdoc or staff scientist positions as equivalents. While the combination of research expertise and technical skills is a hallmark of academic DS/RSEs, and having expertise in both areas does ask more from those in the roles, that does not necessarily translate into higher salary evaluations. While it varies by institution and field, this means you may also find that DS/RSE positions classified as faculty positions have lower compensation ranges than those classified as staff positions.

If there is a mismatch between the positions HR picked as equivalent for compensation evaluations, gather DS/RSE job postings and position descriptions to share with HR. Job titles can play an important role in the compensation analysis HR may perform to determine the appropriate salary range for a position. A "data scientist" may be compared to industry data scientist salaries in your area and result in a higher compensation evaluation (and thus a higher limit on the salary offer you can make) than a "research analytics consultant" or "systems analyst." Similarly, a "research software engineer" is more likely to be compared to a software engineer than a "computational scientist" would be.

Strategies for Limited Budgets: If you have a limited budget, and you do not have the means to increase it, you must adjust the job responsibilities and requirements. Posting a job with responsibilities and requirements that should be associated with a salary well above what you can pay will only be a frustrating experience for candidates and the hiring manager. You are also likely to alienate candidates who might be interested in future positions. Hiring a candidate that is willing to work for a salary that is too low, even if they are truly qualified, will create significant retention challenges down the line.

If you have a limited salary, create a position appropriate for someone who can develop as they gain experience. This is one key advantage of working at academic institutions: there is generally support for professional development, further education, and a culture of learning. For DS/RSE roles, this may mean finding someone with domain knowledge who is still developing their technical skills, or vice versa. Or hiring someone with the baseline skills to learn quickly but no former work experience.

However, you must ensure that the expectations for the role align with the salary and qualifications. Hiring an entry-level person into a higher-level role will result in dissatisfaction and frustration both for the person in the role and the rest of the team. Instead, create a true entry-level position with appropriate expectations. For this to be successful, you will need people who can mentor and supervise such a position. Remember, the expectations for the lower levels of the career ladder are that the person's work is limited in scope and they receive support and guidance from others.

Addressing Hierarchy-Based Expectations: Compensation evaluations for DS/RSE roles may result in salary ranges above those for the faculty with whom the DS/RSEs will be working, especially when the DS/RSE roles are staff positions. To address concerns that may arise around this differential, consider highlighting that DS/RSEs in collaborative or service roles are being asked to provide specialized skills and expertise that is not otherwise available as part of a research team; this expertise has value that is reflected in the salaries. Non-DS/RSE Researchers are leaders in their fields. If they want support staff that are similarly skilled in their areas of expertise, that requires competitive salaries. As in every other area, those who insist on "cheap" DS/RSEs will need to be willing to sacrifice either the quality of the work or the speed with which it gets done.

Addressing salary differentials can be easier when DS/RSEs are hired and managed as part of a centralized team rather than being placed directly in research teams (see Chapter 7 - Organization and Management of Research Software Engineer and Data Scientist Teams).

Compensation Equity: DS/RSE salaries have been increasing in industry more quickly than in academia (Burtchworks 2022, Colby 2022). This means that compensation evaluations for new positions may result in higher recommended salaries than what existing employees receive. Addressing this discrepancy can be challenging, as the process for giving existing employees raises at academic institutions may be complicated. Standard yearly compensation pools for raises and cost of living increases are often extremely limited.

There are no easy answers here, but paying attention to equity and maintaining it across equivalent positions is important for retention. If a new position has a higher salary, and you do not have other ways to ensure salary equity, consider encouraging existing employees to apply for the open position. You may also consider engaging with HR to discuss ways to reduce inequity in pay.

Updating Compensation Evaluations: Another implication of rapidly increasing DS/RSE salaries in industry is that compensation evaluations can quickly become out of date. If you are hiring using an existing job description, it will likely be worth the extra time to ask HR for an updated compensation evaluation or for a position to be regraded if it hasn't been updated in a few years.

SALARY RESOURCES

A few additional resources to help with compensation evaluations:

- Level.fyi includes industry salaries and benefits for positions comparable to DS/RSE and is useful for getting an idea of market salaries
  https://www.levels.fyi
- Bureau of Labor Statistics Wage Data by Geographical Area, Industry, and Occupation. See in particular Software Developers (15-1252) and Data Scientists (15-2051). There are not yet specific occupational categories for research data scientists or software engineers
  https://www.bls.gov/bls/blswage.htm
- Higher Education Salary data; relevant positions are listed under Professional Salaries, but those roles are not necessarily within research domains (they may be on the business side of the university)
  https://www.higheredjobs.com/salary/
- American Statistical Association Work and Salary Survey including salary comparisons by sector and employee satisfaction information
  https://www.amstat.org/asa/files/pdfs/YCR-2020WorkandSalarySurvey.pdf

3.4 Posting the Job

Job postings are the gateway to the role and often the first interaction a job candidate will have with the organization.

The goals of the job postings are to:

- Make your position attractive to qualified candidates so that they want to apply
- Let candidates know what skills and experience are important so they can highlight that in their application
- Encourage a diverse pool of applicants to apply both with explicit language and the usage of unbiased terminology

While you can and should advertise positions in a variety of ways (see Chapter 4 - Recruitment), all positions should be listed on your institution's official job board. This will ensure the position is picked up by job board aggregators, providing exposure to a broad audience, and it is also key for ensuring an equitable hiring process.

Job Descriptions vs. Job Postings: At some institutions, the job description may be the same as the job posting: the job description will be posted directly or with limited modification. In these cases, it is especially important to keep the job description succinct, as there may be few opportunities to edit it before posting. At institutions where you can write a separate job posting, take advantage of this (more details below). If you are stuck with using a less-than-ideal job description as the job posting, consider writing summaries and alternative descriptions as part of your recruitment and advertising efforts (see Chapter 4 - Recruitment).

3.4.1 Posting Logistics

Beyond determining whether the job description will be posted directly or if you can edit it, there are other logistics to check on with HR before posting:

- Will the position be listed publicly/externally for anyone to apply, or will it only (or first) be listed for current employees of the institution or other restricted populations?
- Can you sponsor a visa? Does this have any other implications for the posting and hiring process?
- What are the requirements for working in the office vs. remotely? If there is flexibility around remote work, are there restrictions on where the employee can reside (e.g. in-state, out-of-state)?
- What materials will be submitted as part of the application? In addition to a resume, is a cover letter required? Will you need a teaching or DEI statement? Do you need to add additional questions to the standard application?
- How long will the position be posted? Is there a minimum or maximum? Can you review resumes and interview before the posting period ends?
- Can you have permanently open positions so that interested applicants always have a way to signal their interest? This strategy can be particularly helpful for those who are in more geographically remote areas or smaller labor markets
- Can you list a position at multiple levels or open rank?

It is also a good idea to understand the interview and evaluation process at your institution (see Chapter 4 - Recruitment) before posting the position, as that may have implications for the specifics of the posting.

3.4.2 Posting Components

Your institution likely has a template for job posting, but if there is flexibility, reviewing other job postings to find a format that works well for your position is a good idea.

Job postings should typically include the following components:

- **Position overview**: usually written as a paragraph at the start of the posting, a brief description of the role. For DS/RSE positions, aim to indicate whether this role is part of a service supporting a set of researchers, dedicated to a specific group or project, or an independent research position
- **Team overview**: a brief description of what unit, department, project, or other organization this role will be a part of and what the group does generally. Aim to provide a sense of the team culture, priorities, or work environment. For DS/RSE positions, whether the role is part of a centralized team with colleagues in similar positions or a standalone role that is part of a research group is an important distinction
- **Job responsibilities**: what will someone in this job be expected to do? Start with the job description, but this should not be a comprehensive list. For DS/RSE roles, is the focus on longer-term collaborations? Shorter term services? Teaching (formally or informally)? Managing others? A single role cannot do everything well. What will most of the person's time be spent doing? A bulleted list is expected and helpful
- **Requirements**: what qualifications are you looking for in a candidate? More on this below
- **Application materials**: if an application requires more than a resume and cover letter, what else will the candidate need to put together to apply? How will these additional materials factor into the candidate evaluation process?

- **Benefits**: candidates who first encounter your job posting on a job board will not necessarily have access to information about the benefits your institution offers. Consider highlighting key benefits in the job posting directly
- **Salary**: some states now have salary disclosure laws. We encourage you to include information about the salary in the posting when possible even if you are not required to. An “anticipated hiring range” of $10,000-$15,000 is much more useful than the full salary range for the position grade. This will help attract an appropriate set of candidates
- **Position term**: is this a permanent or term position? If it's a term position, but you expect it to be renewed, be as clear about this as possible. Many good candidates may not be comfortable with term positions with uncertain renewal likelihood
- **Work location**: be as clear as possible about whether the person needs to work on-site and with what frequency. If remote or hybrid work is allowed, provide this information.
- **Visa sponsorship**: if you cannot sponsor a work visa for the position, clearly indicate this in the job posting to avoid applications from those without independent work authorization
- **Equal employment opportunity or diversity statement**: your institution may have standard language that is included on all postings. If not (or sometimes even if there is a standard one), it is worth seeing if you can include such a statement in your specific posting. If not, consider working it into the general position/team description

3.4.3 Keep Postings Focused

When writing a job posting, keep the overall goals in mind: make the position attractive to a diverse set of applicants and provide candidates with the information they need to submit a useful application. Keep the posting focused on the minimum information and details you need to achieve these goals. When in doubt about whether a responsibility or requirement needs to be included, leave it out.

Responsibilities:
Start with the responsibilities listed in the job description; if the job description includes a very detailed list due to institutional requirements, summarize and condense the points so that a candidate can better determine what the job entails. Long lists of responsibilities can discourage people from applying, as they often imply that a position is not well defined or that the person will be asked to take on responsibilities that would be more appropriately handled by multiple people.

Requirements/qualifications:
This is an area where there may be tighter controls on wording, as the listed requirements and qualifications may have implications for the candidate evaluation process, especially at public institutions. Where possible:

- "Required" or "minimum" qualifications should be just that – things that an applicant absolutely has to have to hire them. If you're targeting 8 years of experience, but would hire someone with 4, then it isn't required or a minimum qualification.
- Keep the list of required/minimum qualifications as brief as possible.
- Provide multiple ways a candidate can meet a requirement, for example through education or experience.
- Where degrees are required, avoid limiting to a specific field. Those with

RESOURCES FOR WRITING JOB DESCRIPTIONS

Writing Job Descriptions:

- Academic Data Science Alliance Jobs Archive
  https://academicdatascience.org/resources/jobs-archive/
- Tiernok.com - Writing Better Job Ads
  http://www.tiernok.com/posts/2021/writing-better-job-ads/
- Interviewing.io - How to Write (Actually) Good Job Descriptions
  https://interviewing.io/blog/how-to-write-good-job-descriptions
- re:Work - Guide: Create a Job Description
  https://rework.withgoogle.com/guides/hiring-create-a-job-description/steps/introduction/

Inclusive Language:

- Project Include
  https://projectinclude.org/
- Textio
  https://textio.com/
- Gender Decode
  http://gender-decoder.katmatfield.com/
- Hire More Women in Tech
  https://www.hiremorewomenintech.com/
- Linkedin Talent Blog - 5 Must-Do's for Writing Inclusive Job Descriptions
  https://www.linkedin.com/business/talent/blog/talent-acquisition/must-dos-for-writing-inclusive-job-descriptions
- Glassdoor - 10 Ways to Remove Gender Bias from Job Descriptions
  https://www.glassdoor.com/employers/blog/10-ways-remove-gender-bias-job-listings/

- social science, humanities, art, and other backgrounds may have gained the necessary experience for DS/RSE post-education.
- For preferred or desired qualifications, explicitly note that candidates are not expected to have everything listed.

3.4.4 Language Matters

The language in job postings has been shown to influence who applies to a position (Gaucher et al. 2011, Kang et al. 2016). The resources section includes guides to help you avoid language that will discourage people from applying. Beyond following best practices for using inclusive language in job postings, it is important to share the posting with a diverse set of people before actually posting it to get feedback on whether they understand the job and requirements and whether the posting has any red flags from their perspective. Make sure that you are not inadvertently excluding candidates whose backgrounds are different from yours due to a lack of knowledge about expectations or likely experience levels.

Chapter 4 - Recruitment

THIS CHAPTER COVERS:

Finding Candidates: Building a pool of candidates, strategies for advertising positions, and guidance for hiring managers.

Application Review, Interviewing, and Evaluation: Application and interview evaluation rubrics, structure and timeline of interviews, and checking references.

Considerations for Offers: Pay equity, redirecting a hire, non-compensation perks in academia, and managing failed negotiations.

Post-Hoc Review and Process Improvement: Prompts to reflect on after the search process is over.

KEY TAKEAWAYS:

Build an Equitable and Transparent Process: Structuring the application and interview processes so that they are equitable and transparent benefits the candidates, the search committee, and the institution.

Keep Improving in Hiring:
Continuous improvement of the interview process will mean better interview experiences for candidates and search committees in the future.

4.0 Summary

Identifying and evaluating candidates for Data Scientist and Research Software Engineer positions can take many different forms, depending on institutional regulations, the type of position, and the needs and culture of the hiring organization, among other concerns. In this chapter, we offer advice for finding a high-quality and diverse pool of applicants, a variety of ways to structure interviews and subsequent evaluations, and some considerations for extending offers to candidates.

Early and widespread outreach in advertising positions can increase the size and diversity of your candidate pool. In some cases, identifying quality candidates can begin well before your first application arrives - outreach to communities of interest (especially those representing underrepresented groups) and targeted outreach to qualified candidates can boost the applicant pool. Try to leave plenty of time for potential candidates to find and respond to your job postings.

There are a number of ways to structure application reviews, interviews, and overall evaluation of candidates. How reviews, interviews, and evaluations are conducted depends on a number of factors, which makes it difficult to make prescriptive suggestions about how these should work. That said, in this chapter we offer guidance on these topics and others related to this stage of hiring.

4.1 Finding Candidates

4.1.1 Building a Pool of Candidates Prior to Job Postings

Prior to posting the job and perhaps even prior to building the job descriptions, work on building a pool of applicants. STOP! If you haven't already done this, start now. Curating a pool of potential applicants will inform the job description shaping process by getting input from a potential pool on what features of the job description will appeal to your pool and help you ensure your potential applicants can meet the requirements for the position. Below are a few ways to build and leverage a quality pool of candidates at this stage of the search.

Targeted Outreach to Underrepresented Groups

Targeted outreach to members of underrepresented communities should happen early and often, not just when the position is posted. This outreach can help build awareness and enthusiasm for your upcoming job posting. This is also where prior coalition and partnership-building with leaders from underrepresented communities can be really valuable, allowing a hiring entity to build on a relationship of trust and positive reputation prior to reaching out about a specific recruiting task.

We often (though not always) know when we may have a position opening in the future. Even without a position description or a job posting to pass along, one can still engage with potential candidates about the upcoming positions at your institution. This engagement can take many forms, but informational interviews and general introductions to the DS/RSE group and the types of projects the position might work on can keep your institution front-of-mind for future applicants.

Targeted Outreach to Organizations and Individuals

It can be difficult to identify the right individuals and communities for DS/RSE positions as each job posting has the potential to be highly specialized in certain areas. Creating and curating a list of individuals in adjacent communities who may be able to advise you on and assist with amplifying messaging about your upcoming positions and identifying individuals who may be interested in applying (ADSA and US-RSE can help!). Prior to your search, reach out to the members of this advisory list to let them know about your impending appointments and ask them to amplify messaging and to pass along contact information for any specific individuals they think might be interested in applying.

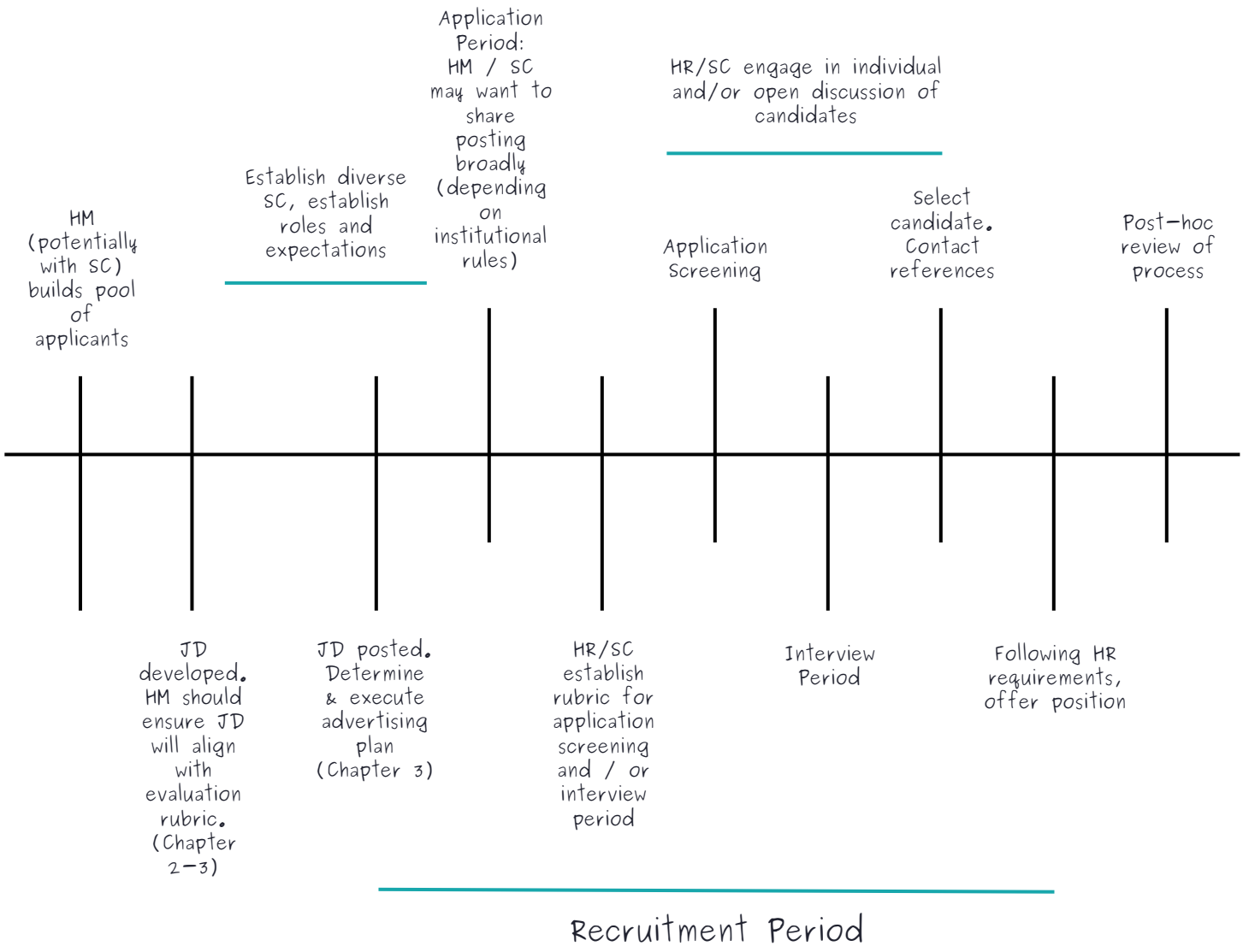
Where appropriate (and allowed by institutional rules), the hiring manager and search committee should actively reach out to their research community to promote the position. This is in addition to the advertising done through job boards, mailing lists, etc. As experts in the field, search committee members may be in the best position to find potential qualified applicants. Committee member outreach should include a significant outreach to a diverse community, including potential candidates from underrepresented minorities (URMs). That is, search committee members should not rely on historically homogeneous networks that will result in a candidate pool that does not reflect DEI values.

Informational Interviews

Conducting individualized informational interviews can be incredibly useful for both the potential candidates and the hiring institution. These one-on-one meetings allow the potential candidate to ask questions about the position and workplace and also allow the manager to

Timeline of Recruitment and Hiring

HM = Hiring Manager
SC = Search Committee
JD = Job Description



understand what potential candidates find interesting and valuable in a job description. However, it can be difficult to scale this process given the time commitment (at least 30 minutes) for each one on one interview. A more sustainable approach may be to run multiple pre-planned “Open Q&A sessions” about a job, class of jobs, or your workplace in general. The institution can choose a time and date and share information with the potential application pool via a telepresence tool (e.g. Zoom) in order to share more detailed information about the position.

Find opportunities to message relevant communities that it is okay to reach out to arrange an informational interview, especially if a candidate is not available during the “Open Q&A sessions”. This is an opportunity to achieve a shared understanding of the goals for the position (or more generally of an organization) and those of the candidate; it also encourages applicants and clears the decks of folks for whom a job opportunity might not be a fit.

Institutions should set aside time for a person to be a general contact for “what it’s like to work for my institution” and make this effort an explicit part of their job duties. This creates an “open door” for informational queries and interviews and can provide your team with an ongoing opportunity to engage in dialog about upcoming positions. In practice, this can take a variety of forms, but open Q&A sessions about the DS/RSE group, discussing the DS/RSE group at conferences and other events, and meeting potential candidates before and during the interview are a few.

4.1.2 Strategies for Advertising Job Postings

The variety of venues for advertising job postings is dizzying, from free-to-use venues like social media and Slack groups to pay-to-post job advertising services (e.g. Dice, Indeed, etc.). Depending on your institution, you may have a budget set aside or be able to leverage existing project or unit budgets to advertise positions in higher-cost venues. However, it pays to be selective with advertising, as there can be a limit to the return on investment for posting jobs in a large number of venues.

Job Descriptions vs. Job Postings
At some institutions, the job description may be the same as the job posting: the job description will be posted directly or with limited modification. In these cases, it is especially important to keep the job description succinct, as there may be few opportunities to edit it before posting. At institutions where you can write a separate job posting, take advantage of this (more details below). If you are stuck with using a less-than-ideal job description as the job posting, consider writing summaries and alternative descriptions as part of your recruitment and advertising efforts.

Below are some strategies for maximizing the reach of your job postings.

Paid Advertising Services:
Paid job advertising services are a good way to get your position out to a large audience quickly, as these services tend to have large distribution channels and user bases. These services, unless otherwise noted, can be very focused on industry jobs, so they may provide a limited number of applicants. Even services that are focused on academic settings can be very broad in terms of types of positions posted - DS/RSE positions can be intermingled with faculty, staff, and administrative roles. This tradeoff between

volume of user base and potential applicants in the user base is worth considering, especially as many of these sites have a cost to post positions.

Social Media:
Social media can be an effective way to amplify messaging around your pending or open positions, but should never be your sole method for advertising. Social media channels can, however, be useful for outreach to specialized communities that may not have their own job posting venues or services. These can often include underrepresented minority groups and groups who do not squarely fit the expected backgrounds for DS/RSE positions. Leverage your candidate pool (see Section 4.1.1 - Building a Pool of Candidates Prior to Job Postings) to advertise in these spaces.

Professional Society and Community Networks:
Professional societies and communities often provide venues for advertising positions, either on formal job boards or via mailing lists, discussion groups, and social media (see above). These venues can provide you with an audience that is likely more specialized in the areas into which you are hiring. For example, the Academic Data Science Alliance and the US Research Software Engineer Association both maintain job pages, active social media channels, and venues for instant or threaded discussions. Likewise, advertising positions with domain-specific professional societies or community networks may help draw applicants coming to DS/RSE positions from a domain research background. Note that some professional society journals also charge a fee for advertising, but the costs tend to be lower than the larger services described above.

Underrepresented Minority Outreach:
If it is allowed by your organization, reaching out to specific communities representing underrepresented minorities in your field can greatly help diversify your pool of potential applicants. It can be challenging to reach communities who are underrepresented in data science and research software engineering, but if you and your organization value diversity in the applicant and employee pools, outreach at this stage (and ideally before - see Section 4.1.1 - Building a Pool of Candidates Prior to Job Postings) is critical for identifying a diverse set of interviewees.

Table 1 shows the relative investment of funds, referral volumes, time to curate the resource, time until candidates are identified, and the diversity of the corresponding applicant pool.

Service / Location	Cost	Referral volume	Time to curate the resource	Time until candidates are identified	Diversity of applicant pool
Job advertising services	High	High	Low	Low	Low
Job advertising services (URM focus)	High	High	Low	Low	Low
General Purpose Social Media (Twitter, Mastodon)	Zero	Varies by network size	Low	Low	Low
LinkedIn (sponsored)		High	Low	Low	Low
LinkedIn (network only)	Zero	Varies by network size	Varies by network size	Varies by network size	Low
University Post-doc email lists	Zero	Varies by network size	High	Medium	Medium
Professional Societies and Associations	Varies	Medium	High	Medium	Low
Professional Societies and Associations (URM focus)	Varies	Low	High	Low	High

4.1.3 Guidance for Hiring Managers

For the purposes of this guidebook, the Hiring Manager is the equivalent of Search Committee Chair, though there may be situations where these roles may be separate (and/or may not be the candidate's direct managers after onboarding). Regardless, the Hiring Manager/Chair should be someone who will work regularly with the candidate once they are appointed to the position and should have a fairly deep understanding of how the DS/RSE group functions and what types of projects they work on.

Building a Search Committee

Search committees may vary in size, and there may be limits on the number of committee members, as defined by the Human Resources office in your organization. That said, the ideal size of a search committee is 4 to 6 people. Consider having the following individuals and knowledge sets on the committee:

- **Search Committee Chair:** Responsible for convening and coordinating the work of the search committee. In some cases, this may also be the Hiring Manager.
- **Technical Expert**: Responsible for helping evaluate the technical skills of the candidate. This may also be a subject matter expert and/or team member, but it does not have to be.
- **Subject Matter Expert**: Responsible for helping evaluate the subject matter knowledge of the candidate. This may also be a technical expert and/or team member, but it does not have to be.
- **Team Members**: Peers of and/or members of the academic community(ies) with whom the candidate will work directly.
- **Logistics Manager**: Responsible for driving the process and logistics for the search committee including meeting and materials review reminders, and organizing applicant materials. This can be the hiring manager, search committee chair, or another individual. This person does not necessarily need to be on the search committee, but if they are not, they should be in close communication with the Search Committee Chair.
- **Search Equity and/or DEI Officer**: Responsible for ensuring that the search is equitably conducted and that the search is grounded in the DEI policy or ethos of your institution. Some institutions have programs for placing Search Equity officers on search committees - contact your Human Resources department to inquire.

Some of these roles may overlap; for example, the hiring manager may also be a subject matter expert, or all members of the search committee have participated in DEI training. If there is a limited pool of DEI-trained folks, don't ask the same DEI representative to be on every search committee. This is a significant burden that can impede their ability to make career progress.

Communicating Roles and Responsibilities

The Search Committee Chair should ensure that all search committee members understand the role of the search committee and what their individual role is on the committee. It is helpful to hold at least one meeting with the search committee before beginning to review applications to clarify the roles of committee members and the committee process. Search committee members should, at a minimum, be cognizant of:

- The position being hired for, the team the position is being hired into, and the rationale for the hire. This includes information beyond what is listed in the position description and includes information about institutional goals for

the position, what day-to-day interactions will look like for the position, and what are the career growth opportunities for the position.
- A detailed understanding of the position description including why certain knowledge, skills, and experience are required versus preferred.
- What are the institution's rules for job searches, including what types of topics are off-limits for evaluation (e.g. marriage status, sexual orientation, religion)?
- Guidelines of the review and selection process both for the institution and for the hiring unit.

Diversity, Equity, and Inclusion and the Search Process

All members of the search committee, and preferably anyone who will interact with candidate materials or the candidates themselves during interviews should be aware of the value of diversity, equity, and inclusion in the workplace and in the hiring process. This includes consideration of members of the search committee and defining with whom the candidates will meet during interviews.

Some areas of bias that may be of particular concern for hiring into DS/RSE positions include:

- **Implicit Bias**: attribution of certain traits to a member of a group (e.g., socio-economic, nationality) that is based on preconceived notions about the group (e.g., people from industry don't understand the academic research environment, people without a computer science degree will not be a good fit).
- **Affinity Bias**: considering how similar in background and experience the candidate is to the current team or search committee members (e.g. this candidate graduated from the same computer science department as me so they must be great!)
- **Affect Heuristics**: using emotional or "gut feeling" responses to candidates in evaluation (e.g. this candidate has an annoying voice so we shouldn't hire them)
- **Halo Effect**: preferring candidates who have done "high profile" work in the past (e.g., they worked at Google or Microsoft or were interviewed by the New York Times about a project)
- **Contrast Bias**: evaluating candidates against one another, rather than considering individual candidates on their merits (e.g., this candidate's previous projects are more relevant than the rest of the candidate pool - though the candidate may still not meet the requirements of the position)

Ideally, all members of the search process should have some formal training in DEI issues, either provided by the institution or acquired through a third party.

RESOURCES

- Handbook on Diversity and the Law: Navigating a Complex Landscape to Foster Greater Faculty and Student Diversity in Higher Education | American Association for the Advancement of Science (AAAS)
  www.aaas.org/programs/diversity-and-law
- Coleman, A., Keith, J. L., & Chubin, D. (2012). Summary and Highlights of the Handbook on Diversity and the Law: Navigating a Complex Landscape to Foster Greater Faculty and Student Diversity in Higher Education. American Association for the Advancement of Science.
  ofew.berkeley.edu/sites/default/files/summary_and_highlights_on_diversity_and_the_law.pdf

4.2 Application Review, Interviewing, and Evaluation

4.2.1 Preparing for Interviews

When preparing for interviews, it is important to understand the rules and requirements for interviewing at your institution. This may also include an initial application review by Human Resources, regulations about how interviews are structured, or other institution-wide considerations that the hiring manager should take into account. Talk to your HR office about what rules are in play at your institution and how much flexibility you have to structure your own interview process and ensure that the entire search committee understands this process.

4.2.2 Structure of Interviews

While the structure of interviews is highly dependent on rules set out by your institution and by norms set within your institution by those hiring into similar positions, there are a few universal practices we find helpful.

Enforce a Script

Enforce a script when interviewing candidates so that all candidates are exposed to the same language (technical and otherwise) and compete on a fair playing field without bias. The opening script for the interview should include an introduction to the participating interviewers, a brief explanation of the interviewer roles (for example, a primary question-asker, primary note-taker, hiring manager, etc.), and the reason behind the standardized process of asking all applicants the same set of key questions.

Longer (e.g. half- or full-day) interviews should have dedicated and structured time slots that make it clear what to expect during that section of the interview (e.g. technical interview, social time with the team, etc.). Each section of the interview should include:

- What to expect from this interview, how long the discussion will take, and approximately how long is allocated for questions and answers.
- Introductions of the search committee members present on the call including why or how the member will interact, relate or manage the position.
- Description of the organization and how this position will fit into the organization
- Description of project(s) the position will cover

Potential Roles and Responsibilities
It is useful to discuss the roles and responsibilities of the position with the candidates in order to align expectations. Often, this is an opportunity to discuss the day-to-day work of the position in much greater detail than the position description, job posting, or screening interviews allow. This is especially true for situations where the types of information in position descriptions and job postings are restricted by administrative requirements (e.g. if it is a general position description used to fill a variety of roles at the institution). This can also be an opportunity to evaluate the candidate's expertise and interest in the position details.

Ask Consistent Questions
Enforce asking the questions in the same way for each interview phase. This can be helpful to avoid imparting biases during an interview by "customizing" questions for each candidate. Where scheduling permits, have the same individuals ask and evaluate the same questions from interview to interview. Not all interviewers have the same insight, passion, and knowledge base from which to evaluate interviewees on specific questions. Being consistent with who asks questions and who evaluates answers will remove differential bias.

Avoid Hidden Evaluations
It can be tempting to evaluate candidates based on every element of their interview, not just structured elements. Evaluation on informal elements of the interview (such as speaking style, clothing, and amount of eye contact) can seem valuable, but the lack of structure and consistency can introduce bias into the evaluation. Examples of hidden evaluations include: clothing choices, hairstyle, marriage status, or whether the candidate speaks a prestige dialect (i.e. white academic vernacular).

Allow for Candidate Questions
Allow for time at the end of each interview for the candidate to ask questions with the overarching goal of enabling the candidate to learn as much about the organization, job role, and culture as possible. This is beneficial to the candidate but also gives the search committee members more ways to understand the candidate's strengths, concerns, and needs for success.

In early interviews such as the initial screening, this can be as short as 10 minutes, but as the hiring process converges on final interviews, consider allowing more time for complex, engaged, and insightful questions. Interviews should be a two-way street, where the candidate learns about the institution and the role as much as the organization learns about the candidate. This will lead to a higher outcome of a mutual fit for the role and increase job satisfaction.

Prepare to Talk About Industry
You may also want to incorporate a discussion about the tradeoffs between academic and industry positions. In almost all cases, academic DS/RSE jobs will pay less than industry positions, but academic institutions tend to offer generous fringe benefits packages along with some less tangible benefits when compared to industry. This conversation might cover: differences in monetary compensation, recognition of the impact of the work as a motivator in the academic setting, potential for better work-life balance in academia, promotion paths at your institution, or the prevalence of research collaborations in academic settings.

Post-Interview Activities

After each interview, consider allowing 15 minutes for the search committee participants to discuss and record their thoughts on the interview. Depending on the committee dynamics, the members may want to record their thoughts separately before the group conversation. Regardless of the order of "record" and "discuss", the hiring manager should ensure that all opinions are heard and considered. You may also use this time to allow for the search committee to express their own biases, either positive or negative (e.g. "I have a bias because the candidate graduated from the same graduate program as me", or "this candidate interrupted me and I'm very sensitive to being interrupted").

Following Up with Candidates

When following up with candidates, use a script for communications, as with other aspects of the interview process (see above). If you plan to send follow-up resources to any of the candidates (e.g. information about benefits, responses to specific questions during the interview), send the same set of resources to all of the candidates in order to avoid inequity in information sharing.

4.2.3 Elements of an Interview

The timeline for your interview process depends on a number of factors such as human resources regulations and organizational norms. Below are some of the major elements of an academic interview, and some options for how to approach each element. Many of these elements should be combined to comprise a complete interview, though the number of elements and their order can vary widely depending on the length of the interview, the type of position, or institutional hiring rules, among other concerns.

Screening: This step may come in several forms, and many of these forms can be employed in the same job search:

- **Search Committee Screening**: Usually conducted by the search committee, this screening usually involves reviewing application materials and ranking candidates on some kind of rubric (see Section 4.2.4 - Evaluation Rubric).
- **Phone/Video Screening**: An initial, short interview with a set of candidates to ask clarifying questions and for the candidate to ask questions of the hiring manager or search committee. This screening should be scripted to ensure consistency across candidates (see Section 4.2.2 - Structure of Interviews).
- **HR Screening**: Some institutions require HR to act as the initial screener for application materials, passing qualified candidate materials along to the search committee after review. HR Screening can be very important to identify candidacy-ending factors upfront (e.g. is remote work allowed, clarification of salary range) and save both parties substantial wasted effort downstream. It is critical to work with your HR department to understand the rubrics they employ for this type of screening and to engage in negotiation about the rubrics as needed.

- **Hiring Manager Screening**: An initial review by the Hiring Manager - the purpose of this screening is, in many ways, very similar to the HR Screening. Because there is only one person reviewing materials in this screening, it is important to have a predefined rubric in place for the Hiring Manager, in order to reduce bias

General Interviews:

- **Information Sharing**: At multiple points during the application and interview process, the search committee should review the details of the position with candidates and offer opportunities for candidates to ask questions.
- **Leadership Interview**: In some cases, leadership of the unit (beyond the hiring manager), department, or other administrative unit may want to meet candidates. This type of interview may also include staff in roles such as program manager, project manager, or product owner - all of whom may interact with the candidate at some point during their tenure
- **Team Interview**: A meeting with the team the candidate will be part of, but without the hiring manager. This can offer opportunities for honest discussion of the work environment and more detailed information about the day-to-day work. Example questions could include:
  - What sizes of teams have you worked on in the past?
  - What level of team collaboration do you prefer? (Note: There isn't a single right answer for this. A candidate might prefer to mostly code by themselves but should express willingness to get feedback and help from team members.)
  - What mechanisms have you used for collaboration in the past? This should include both technologies (Slack, email, VC, bug tracking) and processes (code reviews, design documents, team standups, bug tracking). What do you like/dislike about <specific_productivity_tool> that you used?
  - What mechanisms have you used to document your work?
  - The team lead has made a request that you do not think is technically feasible. How do you express your concern?
  - How do you estimate the amount of time a given task (e.g., addition of a new feature) will take from design to release? They should elaborate on their design, development, documentation, and testing process.
  - Your primary task will take ~3 months end-to-end. How do you break this down into manageable tasks? How do you prioritize? How do you track progress?
  - You disagree with a teammate on a technical decision. How do you resolve it?
  - Have you previously been involved with interviewing candidates for your team? If so, what did you look for and what areas did you cover?
  - Have you previously worked in a hybrid setup? If so, what worked well? What didn't work well?
- **Informal Gatherings:** Some interviews will include less formal gatherings with stakeholders (e.g. coffee with the candidate), offering opportunities for stakeholders and candidates to experience collaboration, curiosity, and communication styles in the workplace. Because of the informal nature, it is difficult to create a standardized evaluation rubric for participants, and search committees often seek informal feedback from participants. Pay attention to potential bias in these settings and mitigate against it (see Section 4.2.3 - Elements of an Interview)

- **HR Interview:** This is often more of an information session for the candidate than an actual interview. An HR representative can use this opportunity to discuss and answer questions about pay, benefits, career paths, and position responsibilities. This is also an opportunity for the candidate to give the HR representative feedback on the interview process, and raise any concerns that might have emerged during the process.

Technical Interviews: Used to evaluate candidate familiarity with technologies and problem-solving skills. These may take many forms, but should cover the range of topics below:

- **Hands-On:** Pick an area, show a code example, and explain your decision
  - Debug a simple snippet of code
  - Write a function to reverse the letters in a string
  - Write a function to convert a string in hexadecimal to an integer
  - Describe the difference between pass by value and pass by reference
- **Concepts and Algorithms:** This should be a discussion that focuses on ideas, rather than examples or live coding. The focus of this interview is on the candidate's ability to elucidate concepts and display their problem-solving skills. Example questions might include:
  - What's the last major system you designed, and what would you change about it?
  - Tell me about a time when you designed a system and it ran out of capacity in some way. Was the problem foreseen, or unforeseen, and what tradeoffs were made before and after the incident?
  - Describe the Service Level Agreements (SLAs) for a system you've designed. How did you monitor them?
  - Do you have any experience in writing, running, or deploying code for cloud-based systems? Which ones? What technologies did you use and why?
  - Imagine you are designing a system to process and prioritize incoming alerts on the order of 10^7 per night. What technologies could you use? What considerations would you need to take into account?
  - Imagine you are designing a system to upgrade a legacy forecasting system to one that needs to process 100X the amount of data on a well-known problem with generally adopted open-source frameworks or tools. What would you need to consider?
  - Imagine that a researcher has developed a reasonable algorithm and tested it on a sample data set on their computer. To handle the new data volumes, the algorithm will need to be distributed over O(100s) of cores. How would you do this?
- **Example Work**: An opportunity for candidates to show examples of relevant previous work. With planning, this can also be an opportunity for the search committee or technical interview group to explore a candidate's previous work and ask technical questions. Note: not all candidates will be able to provide examples of previous work due to intellectual property or security clearance issues - this may be especially true of candidates coming from industry or government positions.

4.2.4 Evaluation Rubric

Creating and using an evaluation rubric for candidates can take some of the subjectivity away from hiring decisions. That said, some elements of evaluation require subjectivity - for example, counting papers as a stand-alone metric can lead to differential outcomes for candidates if one is from social science where the publication quantity is relatively low vs. someone from computational biology where the publication count can be quite high. Some considerations for the rubric may include:

- Listed programming proficiency, ideally in multiple languages
- Past examples of software development
- Past examples of leadership on projects (organizational, team, or other leadership roles)
- Past experience supporting research in an academic setting
- A desire to continuously learn

Portfolio Evaluation
The rubric can include a component called "portfolio evaluation" where the search committee reviews examples of previous projects that the candidate has referenced in their application materials. These may come in the form of code repositories, live or archived projects, publications, and the like. This type of evaluation can be quite subjective, but it offers an opportunity to see work the candidate has released "into the wild" and to explore how those projects unfolded.

It may be helpful to incorporate some portfolio evaluation in your interview sessions with the candidate, as this gives them an opportunity to answer questions about their previous work. If you plan to use a portfolio evaluation, you should ask candidates to provide links to aspects of their work as part of the application package.

Achievement Relative to Opportunity
Evaluating candidates from a breadth of domain backgrounds can make evaluating candidates on equal grounds challenging. Frameworks such as Achievement Relative to Opportunity[5] support a fair and equitable assessment of career progression and achievements over a period of time, given the opportunities available to the candidates. This framework helps to ensure that the overall quality and impact of the achievement are given more weight than their quantity, rate, or breadth relative to personal, professional, and other circumstances.

Implementing the Rubric
Elements of the rubric should be graded on a scale that is clearly communicated to search committee members (e.g. three to five levels of "gradation" in the scale), and members should rate the candidates independently. You may also consider offering or requiring a "notes" section for each element of the rubric. This will allow search committee members to offer a little more nuance in their evaluations or to help explain how they arrived at the evaluation.

For interviews that include stakeholders, collaborators, and peers outside the search committee, the committee may want to solicit anonymous feedback so that negative information is shielded but considered in selection. Note that anonymous feedback related to hiring is not allowed at some institutions, so it is important, once again, to know the rules and regulations of your institution and/or unit.

[5] https://www.monash.edu/academicpromotion/achievement-relative-to-opportunity

After candidates have been evaluated, the search committee should meet and reach a consensus on which candidate(s) to select based on committee evaluations and any additional information about the candidates or the interviews. Depending on how hiring is managed for the position, it may also be valuable to provide a ranked list of candidates, in the case that the committee's top choice does not take the position. Likewise, it may be of value to indicate if any of the candidates are considered unqualified, and should be taken out of the selection process.

4.2.5 Reference Letters/Checking References

Reference letters are a complicated aspect of candidate review and evaluation. It is important to let the candidate know when letters / references will be requested and often this is least intrusive in the final steps of candidate evaluation. It is important to note whether or not the inclusion of a current supervisor as a reference is required and to be sure to provide ample notice and information about how references will be requested so the candidate can inform their references on their own terms about the outreach. Requesting references can be time-consuming for both the committee to review and the reference to write. Furthermore, there are well-documented biases that exist in unstructured letters of reference, e.g. Dutt et al. 2016. These biases are often strongly associated with a candidate's minority status or gender. Therefore, a structured reference request based on a set of quantitative questions should be used to provide as unbiased feedback as possible.

Following are example questions that can be used for a phone or asynchronous reference check.

1. What is your relationship to the candidate and how long have you known them?
2. What were the job duties and responsibilities of the position that this candidate held?
3. How would you describe the candidate's overall work performance?
4. What are some of the candidate's strengths? What key accomplishments or impact did the candidate have on the organization?
5. What area of development could the candidate focus on?
6. In stressful situations, describe how the candidate reacted.
7. Did the candidate mainly work independently or with a group of people?
8. How do you think the candidate's skills and experience will match this position?
9. Would you hire/work with this candidate again?
10. Is there anything else that you can tell me that would be helpful to us in making our decision?

4.3 Considerations for Offers

Within an academic setting, employees that tend to stay longer are those that are less motivated by monetary compensation than by other factors such as impact and recognition. Targeting hiring these kinds of individuals is important, even if additional training is required to get them to the level needed. Leveraging a documented career path is also important here, where staff advances through the ranks by developing the skills needed to better sustain the group as a whole, skills such as serving as the point person with collaborators and leading development teams, or serving as a Co-PI on new proposals, building a portfolio of funded activities around them (see Appendix A - Career Paths).

Before making an offer, the hiring manager should be very clear on the HR rules and requirements for offers, as they will be institution specific. In some settings, a central HR entity may be required to extend the offer, while in other settings the hiring manager will take on this role. Offers extended by HR can give the hiring manager fewer opportunities to include specific considerations during the offer phase, so communicating with HR and understanding the nuances of the process are, again, important to the search. In some environments, the items below need to be considered, though at other institutions these considerations are not relevant or are explicitly not allowed.

Finally, given the impossibility of matching monetary compensation with industry positions, it is also important to recognize that some degree of attrition is inevitable. As a result, institutional sustainability needs must also be considered, including reducing barriers for onboarding, investments in cross-training, and an emphasis on documentation – as discussed later in this chapter.

Pay Equity

Pay equity is the process of reducing the difference in pay between different groups of people (e.g. by sex, race, etc.) and many institutions have specific policies in place to help achieve pay equity. A few ways to move your organization into a more equitable space include: publishing the salary or salary range for the position with the job posting, not asking for salary history from applicants (this helps prevent carrying over inequitable pay from a previous job, and in some states, it is illegal to ask for this information), and creating a uniform structure for salary negotiations in your organization. Salary ranges and expectations should be clear to the hiring manager ahead of the search process, and to the extent that they can be, they should also be included in the job description, as noted in Section 3.3 - Compensation Considerations.

Redirecting a Hire

There may be occasions in which a candidate is not a good fit for the current position but might be a good fit elsewhere in the organization. For example, a unit hiring an RSE may find that a candidate would be a great fit for another unit that is anticipating hiring an RSE. Be sure to discuss this potential outcome with HR as early as possible, so that you have a sense of how much flexibility you might have to shift a candidate into an open or new position elsewhere in the organization, forward a candidate's application materials to another hiring manager, or hire the candidate into a different role within your unit.

Non-Compensation Perks

Below are a few non-compensation perks you might consider offering as part of a broader compensation package. Ensure the availability of these perks at your institution before extending them as part of an offer.

- Quantified benefits package (including vacation, etc)
- Flexible schedules
- Possibility of partially or fully remote
- Access to university courses and matriculated degree programs for employees or educational assistance
- Compute time
- Support staff resources (e.g. ability to hire staff or students at a later date)
- Professional development opportunities including access to formal and informal training programs
- Travel and accommodation support to attend workshops, conferences, and other meetings that build professional networks and yield national exposure
- Childcare, eldercare, and family care support
- Immigration and naturalization support (if the institution sponsors visas)
- Working titles can be used to satisfy the needs of some applicants, e.g. a title that achieves parity with industry. This is situational and can sometimes be difficult to do depending on the setting as some organizations do not allow a change of title.
- Better work-life balance compared to similar jobs in industry
- Longer lead times on projects (not "shipping" deliverables on fast timelines)
- Research autonomy
- Clear recognition of contributions to research
- Opportunities for teaching
- Not being a cog in a machine, making a difference

Failed Negotiations and Salary Realignment

If pay is a consistent issue during negotiations, this may signify that the assigned pay range for a job is misaligned with the expectations of applicants. At that point, you may need to discuss realignment of the job and/or salary range as described in Section 3.3 - Compensation Considerations and consider reorganization or reposting. Be sure to discuss equity with existing positions and other posted positions when considering realignment.

4.4 Post-hoc Review and Process Improvement

Congratulations! Your hiring process has been successful and you have a fantastic new member of your team. The search committee's work is not done, however. Rarely is one hire ever the last hire in a given team, position, or rank. This is an opportunity, while the experience is fresh in the minds of the search committee, to autopsy the process, and identify strengths, weaknesses, and areas of improvement for the process.

Key questions to focus on during the post-hoc review and process improvement phase include:

Prior to Posting

- Did your outreach prior to opening a job posting generate interest from a diverse pool or potential applicants?
- Did your job postings generate a diverse, large enough, and well-aligned pool of applicants?
- What are some key takeaways from interacting with Human Resources during this process?

Interview Process

- What questions were ambiguous and led to confusion on behalf of many or all of the candidates?

- How useable were the rubrics you used? Were any elements poorly defined?
- Did the search committee composition accurately reflect the subject matter expertise and skill evaluation necessary for complete assessment of candidates?
- Were there missed opportunities for efficiency in the organization of the search committee, how the feedback from individual interviews was shared, and how candidates were evaluated?

Making Offers

- What aspects of the compensation negotiation process went well? Were there elements of the compensation package that seemed more or less appealing to the candidate?

After the Hire

- Ask the new hire what they thought of the process. This is an opportunity to get unvarnished feedback about the process including how effective the process was, was time well used from the candidate's perspective, etc. Furthermore, this is an excellent first opportunity to show them their opinion is valued. Keep in mind, however, it's biased by successful hires who clearly performed well in the format.

Chapter 5 - Expectations, Metrics for Success, and Onboarding

THIS CHAPTER COVERS:

Group Expectations & Organizational Needs: setting the goals for your group and its context in the larger organization should come before setting expectations for new hires.

Onboarding for Individual Success: onboarding suggestions to set an individual up for success and suggestions for improving retention.

Metrics for Individuals and Groups: suggested metrics tied to generalized sets of expectations for DS/RSE positions.

Revisiting Metrics and Expectations: where and when to revise metrics and expectations.

Example of Setting Expectations for Effective Evaluation: nurturing growth during the assessment of DS/RSEs.

Example of a Creative Balance of Autonomy and Service: creating a balance between the needs of the individual and the needs of a service-oriented group.

KEY TAKEAWAYS:

Start with your group's and organization's expectations: Before you can craft the expectations for any new or current position, re-visit your group's goals and how they fit into the mission of your larger organization.

Understand that hiring is a two-way street: Expectations are not just about your needs. New hires and current employees have expectations for you as the manager, your group, and the organization. Be sure you understand them before, during, and after hiring, and ensure these fit with the culture and support you, the group, and the organization can provide.

Match the metrics to the expectations: A list of suggested metrics for common categories of DS/RSE job duties.

Be flexible: Be creative to make DS/RSE positions fulfilling and sustainable.

5.0 Summary

The responsibilities of people in roles such as data scientist and research software engineer can vary by position and by context. They may be defined on an individual, group, or institutional level, and can be highly dependent on the priorities and organizational preferences of a manager. In this chapter, we discuss the ways individuals in these roles might be evaluated, both in terms of the duties they might be expected to fulfill and the metrics that are used to assess their performance.

Performance metrics can be used in many ways, including areas that are the focus of other chapters. For example, when you are applying for a job it can be valuable to ask about expectations and metrics for evaluation. During recruitment, applicants may want to know more about metrics for success (see Chapter 4 - Recruitment). During the onboarding process, you will become familiar with expectations for yourself as an individual, as a member of a group/team and as part of an organization.

Managers (see Chapter 7 - Organization and Management of Research Software Engineer and Data Scientist Teams) may identify areas of professional development (see Chapter 6 - Career Development) that can help you meet expectations. You may also serve or support stakeholders external to the organization that have a separate set of expectations and metrics. The key as an employee is to pick up on the explicit description of these expectations, such as in performance appraisals or promotion pathways, and also to be aware that there may be implicit or cultural expectations in an institution that are not codified, but expected such as working on the weekend or participating in social activities. Awareness of these expectations helps an employee identify jobs where the expectations align well with their own values and professional priorities, which can be key to workplace health and success.

We will start with the expectations and responsibilities that an individual might receive from a group in terms of their duties, skills, and competencies. Conversely, an individual might have expectations of the group that they join. The following sections describe metrics that might be used to assess individual performance and ways that these metrics might change over time. Finally, we provide case studies with examples of successful practices and lessons learned.

5.1 Determine Group Expectations & Organizational Needs

Before you can set expectations and goals for the individual DS/RSEs in your group, it is critical to understand what expectations and goals are for the group as a whole. This can take many forms such as a Mission and Vision or a set of work priorities, but are almost always aligned with larger institutional goals or strategic planning. Codifying group expectations gives the group a way to anchor the group's work to institutional goals, and allows the DS/RSE to anchor their goals and expectations to the group's.

It is helpful to set boundaries and scope on activities that can be pursued by the DS/RSE and supported by the organization. Many DS/RSE activities require close working relations and partnerships with institutional resource providers (central IT, libraries, etc.), and ensuring alignment with their goals can help facilitate relationships.

Contributions by individuals towards the goals of the group (often referred to as "key duties") encompass interactions with collaborators and several core competencies. For RSEs, these competencies may broadly include project management skills (scoping, stakeholder engagement, etc.), software development (version control, software testing, etc.), and operations (deployment, performance monitoring, etc.). For DSs, competencies may include a working knowledge of modern applied statistics, machine learning frameworks (e.g. Keras, Tensorflow, etc.), information visualization, domain expertise, data management skills, and so on. As noted elsewhere, many of these competencies and their related expectations can be found in either type of position, e.g. a DS may have stakeholder engagement duties, and RSEs may have domain or visualization expertise expectations.

5.2 Onboarding for Individual Success

According to the Job Openings and Labor Turnover Survey (Job Openings and Labor Turnover - May 2023, 2023) from the Bureau of Labor Statistics, 84,000 individuals voluntarily quit the information industry in August 2022 while another 79,000 quit private educational services. For software engineers in particular, many choices exist for employment given the feasibility of working remotely, and it is common to switch companies every few years (Sharma and Stol, 2020). Consequently, the onboarding process represents a unique opportunity to instill confidence in the new hire, reassuring them that their new work environment will complement their background and support their professional growth. To improve retention, academic institutions should adapt onboarding to directly address individuals' expectations for employers and definitions of success. Rather than simply plugging the individual into the machinery of the university through traditional approaches, it is more effective to "shape the onboarding process around individual identity" and "encourage new employees to [...] use their signature strengths in their jobs" (Cable, Gino, and Staats, 2013).

Organizational Culture, Mission, and Work Environment

Setting up an individual for future success during the onboarding process includes giving the person an understanding of how they fit into the larger work environment. How much ownership and autonomy do they have to decide on the projects they work on? How do their projects fit with the mission of the group and the larger organization? Who should they talk to if there is a problem? Following a review of 92 studies of work motivation and productivity among software engineers, Bass et al. (2008) found that engineers are more motivated when they can "identify with the[ir] tasks"—i.e., having a sense of the tasks' purpose and how they fit into a larger project; a personal interest in the task; and the ability to identify "quality work" upon completion. Because academic institutions struggle to offer data scientists and RSEs salaries comparable to those in industry, offering non-monetary perks that increase autonomy–like flexible work hours and choice of projects–is an effective strategy that should be highlighted during onboarding. Retention is also often correlated with a hire's sense of autonomy and independence.

Communication channels and procedures between different groups should also be discussed, so new hires can quickly direct questions to the right person(s) with minimal frustration. Employee participation with others and good management support and communication were also identified as common motivators across the studies. To become acquainted with their team and the larger academic unit's culture, mission, and purpose during the onboarding process, new hires can rotate through meetings with different groups and team members (directors, PIs, engineering leads, etc.)

Finally, the academic unit should vocalize its commitment to diversity and inclusion and explain how it is implemented. Creating a welcoming, inclusive, and safe work environment is crucial to retaining employees of diverse backgrounds and building models and software applications that are fair and unbiased. Connecting employees from historically under-represented backgrounds with representatives of affinity and employee resource groups (ERGs) also leads to more engagement in the workforce compared to peers who are not involved in such groups (Grillo and Kim, 2015).

Technical Ramp-Up

To produce quality work on a project, hires must develop familiarity with existing code bases, project histories, technology stacks, and IT support options. Training can be administered formally through on-campus coursework or short-format trainings for specific skills as well as informally through pair programming, "knowledge transfer sessions" with senior staff members, and initial meetings with external collaborators. The existence of written documentation is crucial, especially for complex models and applications. The first few tasks should be smaller in scope as the hire builds proficiency. Individuals should be granted access to compute resources as soon as possible, given that system permissions may not immediately propagate.

Mentorship and Professional Growth

During the onboarding process, managers should delineate structures for technical and personal mentorship and help hires schedule recurring meetings with the mentor(s) of their choice. Because turnover is higher in diverse groups where managers are "inconsistent in the quality of the relationships they develop", academic institutions should also train managers to exhibit "diversity role behaviors" (Grillo and Kim, 2015). Hires can take a skills assessment and create a professional development plan with their mentors. Managers should also make clear what percentage of their time is allocated for professional development. In addition, managers should explain how to access free or discounted university resources, such as campus libraries, online learning platforms, tuition and conference reimbursements, etc. Finally, expectations for growth and achievement - including paths to promotions and salary increases - should be transparent and equitable. (For more information on this subject, see Chapter 6 - Career Development)

5.3 Determine Metrics for Individuals & Groups

Once expectations have been set, both for the individual and for the group, consideration needs to be given to how progress on these expectations will be measured. Below is a non-exhaustive set of categories of metrics, and a set of example metrics that can be used by individuals to track their own progress, by managers to assess and evaluate the progress of individuals, and by groups to collectively identify priority areas for the group or organization. Importantly these categories and the associated metrics are not meant to represent a complete list. We recognize that many variations of these positions exist, and thus many additional expectations and metrics could be added. We are assuming that the fraction of time spent by the new hire on each of the following categories has already been established in position descriptions (Chapter 3 - Before Posting: Position Descriptions, HR, and Compensation) and discussed during interviews (Chapter 4 - Recruitment). The broad categories for the distribution of time, which are defined in Chapter 3, include:

- Teaching and Training
- Collaborative Consulting
- Software Development
- Research (independent, as PI)
- Research (as collaborator)
- Outreach, Communications, and Community Building
- Management & Supervision

For each of these categories, we have listed possible metrics below. Note: some metrics are replicated because they apply to more than one category. Importantly, there are additional metrics that could be used in promotion but may not necessarily be included as part of expectations, such as external funds awarded from industry,

donors, or fellowships; recognition within and outside of the university system (best paper award, best staff award); impact on the field through an optional conference or committee participation; mentoring other staff or students; independent development of a new program, workflow, or event; taking on a leadership role in any of the above.

Teaching and Training

- Number, length, and type of workshops taught
- Curriculum development
- Number and types of datasets archived

Collaborative Consulting

- Hours of consulting offered
- Number of consulting engagements
- Number of different departments or units served
- Feedback from external collaborators
- Number and types of datasets archived

Software Development

- Hours spent on project management (identifying milestones, planning meetings, etc.)
- Number of deployed applications
- Site engagement/traffic
- Repository engagement
- Number of projects as technical lead
- Number of software tools maintained and hours spent on maintenance
- Number of software tools developed and hours spent on development
- Number of contributions to other codebases or open-source projects
- Technical certifications earned
- Number of users of a developed software tool
- Evidence of open scholarship

Research (independent, as PI, or as a collaborator)

- Grant dollars (as research lead or collaborator)
- Number of proposals submitted (as lead or collaborator)
- Hours spent on project management (identifying milestones, planning meetings, etc.)
- Number of papers submitted and accepted

Management & Supervision

- Number of developers managed
- Number of mentees
- Hours of mentoring
- Services and products managed

5.4 Revisiting Metrics and Expectations

Managers and their reports should periodically review progress on metrics, and potentially revise metrics and expectations for the work. This feedback can be determined and communicated through multiple different means. For example, group meetings such as retrospectives are a time to evaluate how your metrics are proceeding. Retrospectives occur after a set amount of time, such as two or three weeks, and are a time for all members of a project or team to determine what's going well and what went poorly, and what work to continue or stop doing. Assessing progress on metrics can also happen during individual meetings such as weekly one-on-one meetings. These sorts of summative assessments should happen with regular frequency and be scheduled in advance so they do not get overlooked.

Both metrics and broader expectations can and should be revised over time. These can change as interests, skills, or career goals change. They may also need to be changed based on what other group members or collaborators want to be doing, or as the goals or progress of particular projects change. Whatever platform or system is used to describe and communicate metrics, these should be updated consistently over time (at least annually) and with clear processes for managing changes.

5.5 Examples of Setting Expectations

Setting Expectations for Effective Evaluation

Pomann et. al. (2020) outlined 16 competencies that are key for collaborative biostatisticians to master under three main categories: communication and leadership, clinical and scientific domain, and statistical expertise. These competencies also apply to DS/RSE positions and can be applied to other fields. Your leadership team can build on these to define compe-tencies needed to succeed, which can then be used to define activities that staff will be expected to complete or goals they need to achieve. The authors outline a systematic process to evaluate the skill which they name the "Teach, Implement, Evaluate (TIE) Approach". They outline methods for helping the staff member gain the skill (Teach), then examples to provide oppor-tunities for the staff member to apply the skill (Implement), and a pre-defined process for providing feedback to the staff member about their implementation (Evaluate).

RESOURCE

Pomann et al.'s 16 Competencies for collaborative biostatisticians also apply to Data Scientist and Research Software Engineer positions. ***See image on next page.***

https://doi.org/10.1017/cts.2020.518

These TIEs can be used to monitor the progression of how each person is developing professionally and facilitates a documented, equitable, and personalized evaluation process. For example, one of the key competencies is "understanding the scientific domain" and the staff member may be expected to "have a critical understanding and ability to evaluate scientific papers". In medicine, it is very important for a data scientist to evaluate previous manuscripts for numerous reasons. Before deciding on an analysis plan, a data scientist may need to critically review manuscripts to assess how the analysis of a specific clinical outcome has been evaluated in the past. This can help them determine whether similar methods are appropriate for a study they are working on.

If a supervisor observes that a staff member does not know how to pull the important points out of previous manuscripts to do this, they could develop a TIE to help the staff member learn this. In this example, the "teach" may describe a meeting that will include the supervisor providing guidance on how they read manuscripts to gather important information and key points they typically look for. The "implement" step could be that the supervisor and staff will read the same paper and pull out the important aspects that help them to decide on future design considerations. The "evaluate" step should be measurable and include feedback for further improvement. So this could include comparing which main points the supervisor documented that the staff member did not. This activity could be repeated until the staff member begins to identify all points that the supervisor is identifying.

The Duke Biostatistics, Epidemiology, and Research Design (BERD) Methods Core is a central data science collaboration unit that houses a large group of collaborative biostatisticians who collaborate with clinical and translational investigators throughout the School of Medicine. When a particular skill or competency has been identified as a job duty that the staff member needs to improve on during their performance evaluation, the BERD Core uses the TIE

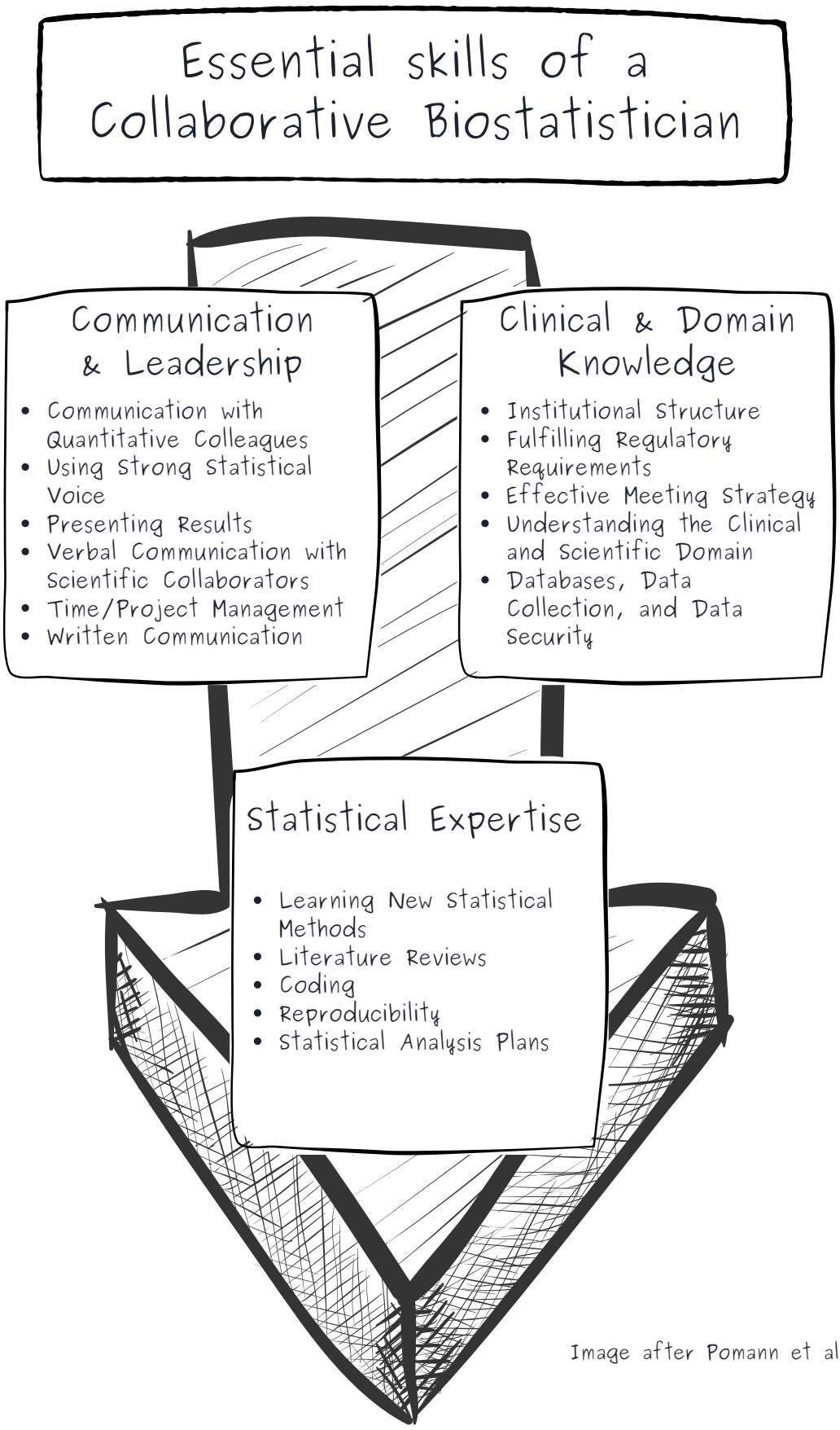


Image after Pomann et al. 2020.

Pomann et. al. (2020) outlined 16 competencies that are key for collaborative biostatisticians to master under three main categories: communication and leadership, clinical and scientific domain, and statistical expertise. These competencies also apply to DS/RSE positions and can be applied to other fields.

method to implement a proactive professional development and evaluation plan (thus hopefully avoiding the need for a performance improvement plan). When necessary, this tool could also be used to guide a formal performance improvement plan.

Setting Expectations for Balance of Autonomy and Service

At the University of Washington, the eScience Institute employs both full-time (100% FTE) Data Scientists and part-time (fraction of FTE) "Research Scientists." Research Scientists have many of the same job duties and service expectations as Data Scientists - albeit at a reduced amount - but they have a home department elsewhere on campus, i.e. they are employed formally by another unit. eScience buys out a portion of the Research Scientist's time from their home department, mediated through an MOU (see Appendix B - Position Description and MOU Examples). In exchange, the Research Scientist contributes to the service mission of the Institute. The percentage of buy-out is negotiated between the individual, eScience, and the home department, and can be changed year to year. The MOU sets the expectations for service, time commitment, and compensation and is signed by all three parties: the individual, the home department, and eScience. The expected service may include: leading one or more projects for their Incubator program or Data Science for Social Good program, offering tutorials or office hours for the campus community, etc.

Despite receiving full salary support, eScience Data Scientists are encouraged to participate in their own independent research projects. To acknowledge and reward successful grant proposals that offset some of their salaries, eScience offers a Buy Back program. Detailed in Appendix B - Position Description and MOU Examples, this program, in essence, returns 50% of the salary cost savings from the grant back to the Data Scientist in the form of a research budget that can be used to pay student helpers, graduate research assistants, attend conferences, purchase software or cloud credits, etc.

Chapter 6 - Career Development

THIS CHAPTER COVERS:

Professional Development: including formal and informal professional development, mentoring and apprenticeship, and principal investigator status

Professional Profiles: discusses the development of professional profiles for individuals and teams to help identify strengths and gaps in skill sets

Mentoring: an in-depth discussion of types of mentorship including one-on-one (traditional), peer and near-peer mentoring, and group mentoring

KEY TAKEAWAYS:

Opportunities exist, but navigating can be difficult: There are many modes and methods of professional development for DS/RSEs, but the abundance of options means that individuals and teams will need to seek out and plan for relevant opportunities

You can't "set it and forget it": Individuals and teams grow over time, and professional development opportunities can lend agility to skillsets, team makeup, and individual satisfaction. Regularly discussing and planning for professional development can help sustain individuals and teams over time

6.0 Summary

This chapter is geared toward directors and managers who supervise the Data Scientists (DSs) and Research Software Engineers (RSEs). We first discuss career opportunities and satisfaction that the academic research environment can offer. We then describe the diverse range of career paths that are possible for DS/RSEs, and how they and their managers can intentionally design career paths based on their current skills and aspirations. Then we compile lists and examples of professional development approaches and opportunities, as well as perspectives from industry. Last, we discuss strategies for working with HR to support and institutionalize DS/RSE career advancements.

6.0.1 Rationale

Professional development is key to the success of DS/RSEs both individually and organizationally. Robust professional development opportunities are necessary to remain competent in data science and software engineering; it is essential for the career advancement and job satisfaction of the staff scientists and for the best interest of their organizations. (Note: Although the availability of professional development opportunities is vital both during recruitment and after hiring, this section will only focus on professional development opportunities after hiring, not on how to incorporate these into job offers or recruitment efforts.)

Professional development is essential for DS/RSE because of the nature of their research fields. Both data science and software engineering are extremely fast-moving fields. Tools and methods that were relevant when a DS/RSE was in school or even when they were hired can quickly become irrelevant. The state of the art is constantly changing in DS/RSE work, as new research creates novel techniques. If a data scientist is not given time to stay up to date, their work will suffer. Similarly, software engineering is constantly evolving as new

software and hardware are released and old software becomes obsolete and unmaintained (e.g. framework modernization like Angular to React, or architectural modernization like monolithic to microservices).

Additionally, professional development is also often necessary for the organizations. Many DS/RSE roles are embedded in smaller teams and less well-defined than in industry. This requires them to wear many distinct and diverse hats. Many of these responsibilities will not be predictable at hiring time and many roles entail a combination of responsibilities that will make it simply impractical to find someone who is already experienced with all of them. Sometimes, some of these new responsibilities emerge after the hiring. Therefore, it is wise to allow the DS/RSE to take time to focus on professional development and acquire new skills in order to ensure that they can maintain the desired level of competency and quality in their work. Here we remind the reader that, importantly, the supervisor should discuss with the DS/RSE the new expectations of the position and provide information on how progress will be measured and evaluated for career progression (see Chapter 5 - Expectations, Metrics for Success, and Onboarding for examples of metrics to consider).

Ample professional development opportunities can be a major feature of working in an academic institution. When considering how to retain DS/RSEs, it is important to consider their options and what makes an academic institution unique. All other things being equal, DS/RSEs can leave for industry positions with the expectation of a significant raise. Therefore it is important for academic institutions to ensure all other things are not equal. The unique position of academic institutions as places of learning and creative, cutting edge research makes them ideal and natural places to offer a DS/RSE the freedom and resources for intellectual pursuit as part of professional development.

6.1 Professional Development Opportunities in Diverse Academic Environments

Before we dive into specific professional development opportunities, we would like to point out that there isn't a stereotypical academic institution's environment for a DS/RSE. The academic environment for one DE/RSE can be drastically different from that of another, and this diversity drives the need for a diverse (and flexible) set of offerings for professional development and career advancement for DS/RSEs in different types of units and environments.

A DS/RSE can be situated mainly in a research or lab group, within a department, or within a larger centralized unit. Their roles within these settings will also differ (e.g., supporting one research project, supporting projects in a specific research area, building capacity across areas, and others.), and are often less standardized when compared to a narrowly defined entity at a university (e.g., IT, HR, etc.). Other factors in academia, besides how a DS/RSE position is embedded within a unit, also impact how opportunities should be designed for their professional development and career advancement.

Below we describe in some detail a number of factors that impact the professional development and career advancement of a DS/RSE.

Decentralized placement of DS/RSE

From the top down, universities and academic institutions can differ in how they are governed and organized based on institution type (public vs private), organization, history, culture and other

factors ("Colleges and Organizational Structure of Universities" n.d.). This notion of governance also applies to the areas within universities where DS/RSEs exist and function. DS/RSE can be found in individual research groups led by faculty; at the department level where they are servicing multiple research teams; at the academic center-level (university and even state-wide), where a conglomerate of research-aligned academic entities co-exist; at the institute-level that is composed of a broader affiliation of researchers (University of Arkansas Office of the Provost n.d., Ohio Supercomputer Center n.d.); and even separate entities designed to service university research in totality (Software & Application Innovation Lab n.d.).

Funding variability across different functional units

It is a widely known fact that there are several funding disparities that exist from institution to institution, based on a plethora of factors (Yi 2019, USA Facts 2023). Interinstitutional funding disparities are out of scope for this portion of the document, and all mentionings of funding variability are presented in the context of intra-university variability, in reference to how an academic institution chooses to allocate funds to various functional units, and the typical level of extramural funding available based on research areas (Nietzel 2021). This phenomenon illustrates that where a DS/RSE is placed can have an impact on their job expectations, who their stakeholders are, and subsequently resources, availability, and how to advocate for professional growth and career progression.

Current position mapping within Human Resources

The existence of a generalized Human Resources (HR) position mapping infrastructure, for any organization, is often instantiated to standardize job duties, build a structure for internal equity (compensation), and provide pathways for promotion and other areas (Herrity 2023). However, the downside of this type of position mapping, particularly in reference to DS/RSE in an academic institution, revolves around the variability and the fast-paced changes of the position duties. Because of the large variabilities of the area/unit of employment for a DS/RSE, job functions, core technical competencies, and beneficiaries of their work outputs, a one-size-fit-all HR classification of DS/RSE positions often does not accurately position some DS/RSEs in the right career path to start with. Because of the fast-paced changes in the roles of DS/RSE, the HR job classification system often does not inform DS/RSE and their employers about the career paths that a DS/RSE can take. An urgent issue for discussion across many academic institutions is how to accurately reflect various types of duties and the contributions of DS/RSE to the research enterprise and how to clearly define progressively larger sets of responsibilities and the recognition and compensation that come with them. But for now, when a DS/RSE seeks professional development opportunities and the next career moves, lack of clarity in the HR classification of their positions can translate into a lack of clarity both for themselves and their supervisors about what goals they should strive for for career advancement, thus difficulty in seeking the most useful opportunities.

An isolated professional environment

Regardless of the placement of a DS/RES (in a research team, under a research center, etc), it is yet rare for an academic institution to have a robust and well-connected DS/RSE population. Oftentimes a DS/RSE can be the sole contributor to a project, work on small dedicated teams with a particular focus, or be a part of a larger

research team in a well-funded research area. DS/RSE who work within a larger team of DS/RSEs may have a more structured understanding of how to advance professionally and gain skills to contribute to personal and team growth, simply because there likely have been many precedents. However, small teams, and teams that may be large but only have a small number of DS/RSE, can fall prey to various constraints that must be accounted for when pursuing professional development and career advancement opportunities, such as lack of clarity from both the DS/RSE and their supervisors, the need to explore opportunities solely by themselves, lack of peer support, and lack of champions on campus for DS/RSE advancement.

Diverse avenues for advancing domain expertise

However, an academic environment also offers substantial benefits and opportunities for DS/RSE. For example, the university community often champions continuing education and learning for everyone and in general provides a variety of opportunities. DS/RSEs also have the opportunities to leverage university benefits either classically (through courses offered via tuition remission) or in an ad hoc manner, pursuing opportunities for exposure to cutting-edge advancements in their own fields, to other fields, and building skills through their peer departments, faculty, staff, centers, initiatives, programs, etc. They may have numerous options for gaining domain expertise in desired areas where a DS/RSE sees their career taking shape or growing, through seminars / workshops and many similar informal training opportunities; trying their hands at new projects; expanding their areas of expertise through collaboration on new projects; and so on. In addition, they can be recognized as authors on research papers, participate in teaching and mentoring activities, and leading research projects and grants, all of which can strongly boost their credentials.

The prevalence of these environmental elements for DS/RSEs at academic institutions provides insight and potentially unique entryways for the discussion about how to bring to bear meaningful practices, tools, policies, and considerations that ensure these staff members have positive experiences when seeking career progression and professional growth.

6.1.1 Professional Profile

DS/RSEs might be very new positions at many institutions, and their roles might be rapidly evolving at other institutions. One immediate issue is that the existing job titles or descriptions do not accurately reflect their responsibilities and roles. As discussed in Chapter 3 - Before Posting: Position Descriptions, HR, and Compensation, one can work with HR to create position descriptions that more closely match the DS/RSE role. This process, however, can be lengthy and a mismatch with hiring timelines. For this reason, DS/RSEs often hold positions for which the HR position description is “the best match within the current system” but is in effect vague or inaccurate.

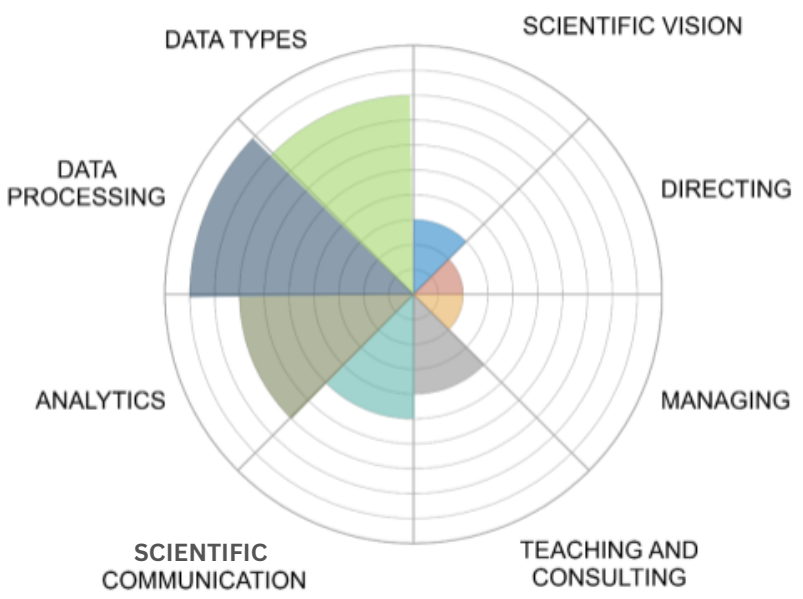
In the absence of an overhaul of HR classification for DS/RSE, we recommend that hiring managers and DS/RSEs focus less on job titles, and more on technical competencies and responsibilities to help define current and desired professional skills. Consider developing a “professional profile” for each DS/RSE that includes technical and managerial (if applicable) expertise and responsibilities. This helps the DS/RSE, their hiring manager, their unit leadership, and HR to think a little more precisely about the role of the DS/RSE within the institution, department, lab, or

research group and what skills they bring to the work. It is also a foundation for designing professional development goals and approaches by comparing current skills with desired skills. The polar plots in the next several pages are one of many ways to depict a professional profile. We choose this way to illustrate our points because it is easy to visualize. What attributes / expertise should be included in a professional profile also varies greatly from position to position.

Using a data scientist position as an example, consider mapping core competencies for the position that span the range of data acquisition, cleaning, wrangling, archiving, analytics, modeling, and visualization, and depict the DS's competencies on the map. This approach has two additional advantages: First, it can help the DS/RSE see their own strengths and gaps, and help them chart an individual path for professional development. Second, it helps show the combined expertise of a team of DS/RSEs and demonstrates the strengths of the team and the gaps that need to be filled.

The remainder of this section uses DSs to illustrate our recommendations, though RSE profiles can be similarly constructed.

Figure 1. Data Scientist #1



Data types
Work with various types of data: text, spatial, video; health record, social media, etc.

Teaching and consulting
For researchers and students

Data processing
Management, wrangling, QC, sharing, etc

Managing
Managing projects, staff members and research groups

Analytics
ML, NLP, Statistical inference, etc

Directing
Directing a team or a program

Scientific communication
Writing, data storytelling

Scientific vision
Identifying important research challenges and envisioning research direction

Figure 1 shows an example of the professional profile of a DS, whose current strength lies in the expertise of dealing with various data types and data processing.

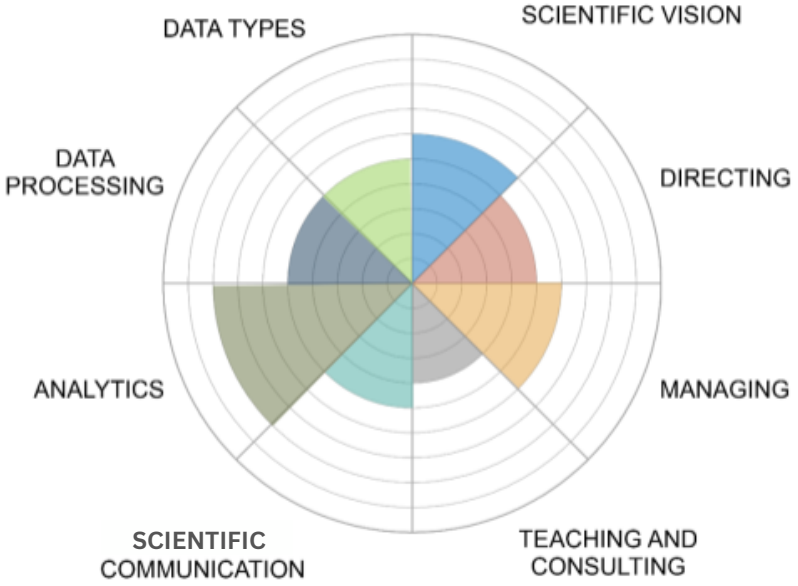
This profile will allow this DS's manager and collaborators to easily see how this DS can fit into a research project. Equally importantly, this DS and their manager can use this profile to discuss career advancement - for example by identifying important areas for growth, or opportunities for mentorship. This DS clearly has strengths in the hands-on components of a data science project - data processing, experience with a variety of data types, and data analytics. Also, clearly, this DS is not yet ready to communicate research as a scientist, project leader or teacher, and is not yet experienced in management and leadership. Depending on the DS's own preferences for career advancement, they might choose to focus professional development efforts on continuing to strengthen their hands-on technical expertise, or on developing scientific leadership skills.

This profile can be easily customized to fit with the needs of a DS or their unit, by adjusting how specific each section is. For example, the data analytics section can be expanded to include many specific types of analytics. Or, if teaching is not at all under the purview of the unit or within the interest of the DS, then it does not need to be included.

A note about public vs. private profiles

Some might opt to display such profiles (or more likely only the strengths) publicly, for example, on the team's website. This could be very similar to many current practices when a researcher lists their research interests and focus areas. This can help potential collaborators with the DS or their unit to easily understand the fit of the DS (or their unit) for a particular research project. However, it may also present a static picture of the DS's expertise and may position them in "a box" in some collaborators' minds. Therefore, whether to use such a profile publicly, only between the DS and their supervisor, or with the entire team needs to be carefully determined.

Figure 2. Data Scientist #2



Data types
Work with various types of data: text, spatial, video; health record, social media, etc.

Teaching and consulting
For researchers and students

Data processing
Management, wrangling, QC, sharing, etc

Managing
Managing projects, staff members and research groups

Analytics
ML, NLP, Statistical inference, etc

Directing
Directing a team or a program

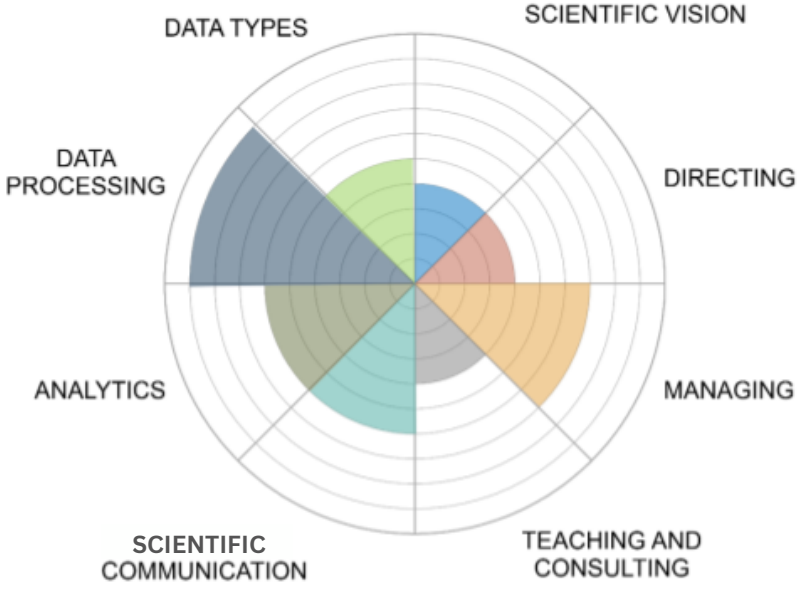
Scientific communication
Writing, data storytelling

Scientific vision
Identifying important research challenges and envisioning research direction

Figure 2 shows the profile of another DS, who clearly has different strengths than DS #1. This person has a much more balanced profile, with strengths not only in some aspects of hands-on technical work (data analysis) but also in research vision, leadership, mentoring, and communications.

When comparing the information in the profiles of DS#1 and DS#2, one consideration beyond their current strengths is that they may or may not have the same career aspirations. For example, both might want to stay on the technical path, or one might want to move to a director role and one might want to be more involved in setting scientific directions. In any case, a professional profile will help chart their path from where they are to where they want to be.

Figure 3. Data Science Team Composite Expertise



Data types
Work with various types of data: text, spatial, video; health record, social media, etc.

Teaching and consulting
For researchers and students

Data processing
Management, wrangling, QC, sharing, etc

Managing
Managing projects, staff members and research groups

Analytics
ML, NLP, Statistical inference, etc

Directing
Directing a team or a program

Scientific communication
Writing, data storytelling

Scientific vision
Identifying important research challenges and envisioning research direction

In addition, a composite team profile **(Figure 3)** can help the manager and team members understand the strengths and weaknesses of the team of data scientists.

This will help them plan new hires and professional development for the team. For the particular team shown in the figure, if they wanted to develop training activities for researchers they could see that they will need to build the team's strength in teaching and scientific communication. To round out the technical skills of the team, the new hires or professional development for existing members will need to focus more on analytics skills, not data processing skills.

The profiles that a team designs for itself will need to incorporate the specific requirements of expertise that their work needs. For example, a team that works with many schools and colleges of a university may need to deal with many structured and unstructured data types (health data, administrative data, research data; text, image, video, spatial, digital record, survey, etc). They may want to explicitly list such data types on the polar plot. In contrast, a team that deals with mostly one type of data (for example, a team that supports research with health record data) may want to focus on building strengths in analytics and hence explicitly list the major types of analytics that they use (ML, federated learning, Bayesian inference, etc).

6.2 Professional Development Approaches and Opportunities

Although the specific environments of academic institutions can vary widely, they all have the opportunity to offer useful and exciting professional development opportunities. These include opportunities for formalized continuing education, the ability to learn and practice technical and soft skills, and receiving recognition and building credentials based on their work.

It is critical that the institution provides time and monetary assistance to employees for a variety of professional development opportunities. This support can also be helpful in the hiring process, as a highlight of non-compensatory benefits; it sometimes may even be make-or-break for DS/RSEs considering leaving for another position.

Academic institutions are natural places to offer formal and informal continuing education opportunities. It is both to the benefit of the institution and the individual to offer and promote courses and learning materials to DS/RSEs at low or no cost. These will stimulate intellectual curiosity, improve the quality of work, and may help increase retention. Furthermore, DS/RSEs should be provided the ability, including time and budget, to attend conferences, trainings, and workshops of interest in order to gain exposure to tools, methods, and colleagues from other institutions.

Types of Professional Development to Consider

Managers and DS/RSEs should work together to identify professional development priorities and create time in the duties of the DS/RSE to engage in professional development.

- **Formal Continuing Education:** Universities and colleges can provide excellent opportunities for formal training in a

variety of areas (technical and soft skills). These opportunities can include courses (for credit or audited, with the tuition partially covered by the institution as part of the employee benefits), workshops, and training.

- **Technical Skills**: Formal training for technical skills, especially at an advanced level, such as a certificate program, may be more difficult to come by in an academic setting, beyond taking graduate-level classes. That said, the flexibility offered in academia means that DS/RSEs may be able to carve out time for informal learning. This can take a number of forms including taking on a challenging project and learning on the job, collaborating or co-learning with others interested in these skills, reading papers or books on the topic, or other forms of exploratory learning.
- **Professional Skills**: Basic competencies in leadership, management, public speaking, and interpersonal communications are critical to DS/RSE roles, and development in these areas can improve team dynamics and interactions with collaborators. Strengths in these areas beyond basic competence can open doors to DS/RSE for more career opportunities including different paths such as leadership positions. Similar to technical skills, formal learning opportunities for these professional skills may not be as available or apparent in an academic setting.
- **Teaching and Mentoring Skills**: Many academic institutions have an office (or offices) dedicated to improving faculty and staff teaching capabilities. While programs from these offices, or in other areas at the institution, may seem focused on faculty-student learning situations, DS/RSEs may also be able to benefit. Also consider informal opportunities for improvement in these areas, such as teaching small workshops.

or training student employees.

- **Mentorship or Apprenticeship**: This refers to a DS/RSE receiving mentorship. Acquiring skills to mentor others is treated together with teaching skills. Oftentimes, senior DS/RSE in one's own institution, close research collaborators who are more advanced in the career path, and those who are in the positions that the DS/RSE would want to be, are natural mentors. Sometimes the supervisor of the DS/RSE is also a valuable mentor. Depending on the skillset of the DS/RSE and the size and specific setups for the DS/RSE, mentorship opportunities may not exist within your institution; in those cases, one can seek mentorship opportunities within a broader community, such as US-RSE. See 6.3 for a more extensive discussion.
- **Principal Investigator Status**: While PI status is not always desired, and often challenging to arrange, by those in DS/RSE positions, this is an area of responsibility that can be appealing and empowering to some. Not all institutions allow non-faculty members (or even some classes of faculty) to act as PIs on grants. Managers should consider approaching your research office/office of sponsored projects to discuss whether and how you can achieve PI status for DS/RSEs in their unit who have shown strong scientific vision and leadership and can benefit from having a PI status.

6.3 Mentoring

Mentorship consists of relationships between people who exchange formal and informal knowledge toward the goal of identifying and overcoming challenges to achieve personal and professional goals. Mentorship is an important part of sustaining the growth and retention of DS/RSE. The professional development space, and specifically mentorship, is nascent and evolving as DS and RSE fields mature as professions, so the following section summarizes the opportunities currently available. Prospective mentors and mentees can reference these resources as they upskill in their new roles.

A few key features of DS and RSE fields include: the fast-paced nature of new tool/methods development; that many's work is interdisciplinary or can hop from one research field to another (applying similar technical methods); that project and job appointments may be short-term (Milewicz and Mundt 2021). This may lead to collegial relationships that can last only as long as a project or job appointment. Mentorship is more effective when mentor-mentee relationships are durable and long-term, even when individual projects are completed, or when colleagues move between disciplines and institutions.

For DS/RSEs, formal knowledge gained through mentorship might include technical expertise and skills. Informal knowledge goes beyond subject matter expertise and may include non-technical professional skill development, psychosocial support, and tips for navigating the 'hidden curriculum' which is the unspoken norms and values which, for better or worse, often determine the success of an individual as they hope to progress in their careers.

The ways mentoring relationships look may vary widely; for the purposes of this section, we distinguish between three types: traditional, peer, and near-peer mentorship.

Traditional mentorship

Traditional mentorship is a mentor-mentee designation where there is a sizable difference in expertise between the two people in the pairing, which can be measured in a number of ways including subject matter expertise or number of years "in the field." From the perspective of a mentee, a mentor serves an aspirational function - someone who may be considered a role model. Mentors may also serve as an accountability "elder" - someone to whom one may feel responsible to answer to, but can be out of the authority line (i.e. not a direct managerial relationship). Although the mentor-mentee relationships may focus on mentee-driven needs, there can be two-way benefits of the relationship; from the mentor's perspective, the act of mentorship can provide personal satisfaction and reinforce professional identity.

Some mechanisms for this type of mentorship may include periodic check-ins in the form of office hours, one-on-one sessions, or structured long-term programs such as the Pilot Mentoring Programme by the Society of Research Software Engineering[6]. Structured programs may also be shorter in scope or toward the purpose of one goal, such as helping early career researchers navigate a particular professional milestone (see one example[7] from RSECon 2022).

[6] https://society-rse.org/events/pilot-mentoring-programme/

[7] https://rsecon2022.society-rse.org/call-for-mentors/

RESOURCES

Mentorship models and evaluation of the effectiveness of different practices within DS and RSE are non-existent. However, organizations and individuals are actively compiling and discussing such practices. Below are a few resources discussing the challenges of mentoring, generally, and how DS/RSE mentorship might differ.

- Academic Mentorship Needs a More Scientific Approach - Montgomery et al. argue for a complete overhaul of mentoring practices, pushing us towards collaborative, evidence-based approaches
  - Montgomery, B. L., Sancheznieto, F., & Dahlberg, M. L. (2022, September 12). Academic Mentorship Needs a More Scientific Approach. Issues in Science and Technology. **https://issues.org/academic-mentorship-scientific-approach-montgomery-sancheznieto-dahlberg**
- An Exploration of the Mentorship Needs of Research Software Engineers - discusses the unique challenges of mentorship for RSEs and how organizations can support RSE mentorship
  - Milewicz, R., & Mundt, M. (2021). An exploration of the mentorship needs of research software engineers. arXiv preprint arXiv:2110.02251.
- Academic Data Science Alliance - Resources for Mentors and Mentees - ADSA has developed a set of living documents with guidance and resources for mentors and mentees
  - Min, S. (2022, October 5). Resources for Mentors and Mentees. Academic Data Science Alliance. **https://academicdatascience.org/resources/2022-10-05-blog/**

Peer and Near-peer Mentorship

Peer and near-peer mentorship contrasts with traditional mentorship in that the mentor-mentee gap in social, professional, or age level is little to none. In the case of peer mentorship, the gap is none; in the case of near-peer mentorship the gap is slight - one of the individuals may have recent experience which would inform a slightly different type of relationship from a strictly peer mentorship relationship. A common example of a near-peer mentor is a first-year college student mentoring a high school senior on college admissions.

Peer and near-peer mentorship gets less attention than traditional mentorship but appears in practice organically. If the demographics of DS/RSEs reveal a critical mass of individuals at the early career phase all professionalizing at the same time, peer and near-peer mentorship mechanisms could be intentionally fostered by the DS/RSE or their supervisors, as a supplement or perhaps a substitution for traditional mentorship depending on the supply and demand of mentors and mentees.

In addition to the organic formation of peer or near-peer mentorship, employees and their supervisors have many formal approaches to developing such mentorship. Example mechanisms for peer mentorship include running programs through a cohort model where there is structured space for individuals to come together, learn as a group, and make connections where they might not otherwise. Together peers can navigate unique activities of DS and RSE like pair programming, code reviews, and collaborating on papers. As opposed to an accountability “elder”, peer mentors can be seen as accountability “buddies.” Peer mentorship can also occur through institutional and professional society

communities of practice. Near-peer mentorship mechanism may include a recent alumnus pairing with new members to a program.

Mentorship Communities for DS/RSEs
Hiring managers and DS/RSEs should seek out mentorship opportunities in a number of settings. The hiring unit or institution may already have a mentorship program in place, which can be a good starting point for finding mentorship opportunities. Hiring managers may want to explore these institutional offerings to determine whether the format and style of mentorship are aligned with DS/RSE needs.

Universities may have pockets of DS/RSEs scattered in many departments that are isolated from larger professional networks. Creating informal peer mentorship communities across these isolated groups can provide a sense of community at the institution and allow DS/RSEs to assist one another beyond mentorship (e.g. informal training, community).

Last, professional societies and less formal communities may also offer opportunities for mentorship. Professional societies such as the Academic Data Science Alliance, the US Research Software Engineer Association, and others may have existing mentorship programs to tap into. There may also be informal communities either on-campus or inter-institutionally that offer mentorship, such as the Staff Collective for Data Science at the University of Michigan. These might include communities focused around a technology (e.g. SciPy), topic area (e.g. High-Performance Computing), or the social state of the individual (e.g. communities for underrepresented minorities in DS/RSE roles).

6.4 Credit

For DS/RSEs to have professional success, whether that takes the form of increased compensation, internal promotion, or employment elsewhere, their contributions have to be both legible and visible. This might take the form of setting the expectation advocating with faculty partners for the inclusion of DS/RSE collaborator names in published research that their work enables (see: Princeton Research Software Engineer (RSE) Partnership Guide 2021), publication of links to public repositories alongside other research outputs, embedding user roles in metadata, providing recommended citations for data products, code, or software or generating DOIs for the same. This might also mean developing internal recognition structures that allow for and encourage the elevation of essential but often “invisible” maintenance and/or consultative work that might traditionally fall under the “service” and “teaching” responsibilities of an academic position. It is also important for DS/RSEs and their managers to understand the need for service within the community, and to be able to articulate the impact of their work, both in terms of advancing research objectives and the impact of those objectives on society at large. Articulating this impact is important both for recognizing the individual’s professional successes and for contributing to the individual’s job satisfaction from the impact of their work and recognition of that impact.

Because the work for DS/RSEs is more focused on enabling research objectives, it can be important to recognize and assign credit for individual contributions as well as the overall group achievements, many of which are enabled by the work of DS/RSEs. Metrics for advancement and personal satisfaction are likely to vary between data scientists and software engineers, particularly if software engineers come from more

traditional Computer Science training pathways and/or from industry. Data scientists working in academia often come from applied science domain backgrounds, and they are likely to value authorship on peer-reviewed papers in disciplinary journals in addition to publications in more methods-focused forums. Software engineers may be more interested in leading or being a primary contributor for high visibility, impactful software products.

RESOURCES

For more discussion of Credit for DS/RSE Positions, see:

- Carver, J. C., Weber, N., Ram, K., Gesing, S., & Katz, D. S. (2022). A survey of the state of the practice for research software in the United States. PeerJ Computer Science, 8, e963. **https://doi.org/10.7717/peerj-cs.963**
- CRediT. (n.d.). CRediT (Contributor Roles Taxonomy). Retrieved July 7, 2023, from **https://credit.niso.org/**
- Katz, D. S., McHenry, K., Reinking, C., & Haines, R. (2019, May). Research software development & management in universities: case studies from Manchester's RSDS group, Illinois' NCSA, and Notre Dame's CRC. In 2019 IEEE/ACM 14th International Workshop on Software Engineering for Science (SE4Science) (pp. 17-24). IEEE. **https://doi.org/10.1109/SE4Science.2019.00009** (Preprint: arXiv 1903.00732)
- Katz, D. S., Chue Hong, N. P., Clark, T., Muench, A., Stall, S., Bouquin, D., Cannon, M., Edmunds, S., Faez, T., Feeney, P., Fenner, M., Friedman, M., Grenier, G., Harrison, M., Heber, J., Leary, A., MacCallum, C., Murray, H., Pastrana, E., ... Yeston, J. (2021). Recognizing the value of software: a software citation guide. F1000Research, 9, 1257. **https://doi.org/10.12688/f1000research.26932.2**
- Jackson, M. (n.d.). How to cite and describe software. Software Sustainability Institute. Retrieved July 7, 2023, from **https://www.software.ac.uk/how-cite-software**
- Journal of Open Source Software. (n.d.). Retrieved July 7, 2023, from https://joss.theoj.org
- Scroggins, M. J., & Pasquetto, I. V. (2020). Labor Out of Place: On the Varieties and Valences of (In)visible Labor in Data-Intensive Science. Engaging Science, Technology, and Society, 6, 111–132. **https://doi.org/10.17351/ests2020.341**
- Smith, A.M., Katz, D.S., Niemeyer, K.E., FORCE11 Software Citation Working Group. (2016) Software Citation Principles. PeerJ Computer Science 2:e86. DOI: 10.7717/peerj-cs.86

6.5 Learning from Private Sector Engineering

So far, our focus on DS/RSE professional development is on the individual. As DS/RSEs form communities either formally or informally, and as the number of DS/RSEs increases in academic institutions, the next natural question is to think about how such communities can promote the work and career advancement of their members at the technical level, beyond mentoring and advocacy.

In this section, we include two examples from industry to illustrate how DS/RSE as a team or a community can advance their technical work and the careers of all members in the community. These practices could accelerate the work of many DS/RSEs in academia and strengthen their technical skills but this can happen only if there is coordination among DS/RSEs and their teams. By nature, academic research projects are oftentimes much smaller in scale compared to industry development projects. As a result, technical solutions (including code packages) are developed often for one person (the person who manipulates the data or carries out the data analysis), or one research group (for multiple similar projects), and only very occasionally for multiple research groups employing similar methods to address similar scientific questions. As such, DS/RSEs often work in isolation and often lack opportunities for coordinated development and deployment of research tools. However, this landscape is changing and we are seeing more coordinated efforts among academic DS/RSE. Therefore, we hope these industry examples will be useful resources.

Both of the best practices from private industry that we feature underwrite software **sustainability and reproducibility**: *architectural convention* and *project management:*

Architectural conventions from private industry consist of a tool kit of proven design patterns (e.g. tiered applications, microservices, relational vs unstructured databases, MVC, etc), and situations in which they can be effectively deployed. By following these common architectures, DS/RSE teams can not only minimize reinvention but build in such a way that new team members and external users can quickly recognize and reproduce software architectures by drawing on past experience. Similar advantages accrue when appropriate frameworks are used for building common components. Furthermore, the typically heavily modular design patterns of modern software engineering lend themselves to the adaptability required by DS/RSEs and their stakeholders in experimental or exploratory situations - which are primarily the mode of operation in academic research, by making feature replacement well-defined and surgical, rather than tightly coupled to monolithic applications.

Below are some considerations for learning about architectural conventions:

- What are some common software architectures, and how to choose an appropriate one?
- What sort of frameworks exist at the sub-architecture level that facilitate code reuse and avoid creating bespoke expertise?
- What are standard instrumentation patterns and tools for software development and design, such as unit and integration testing, metrics instrumentation, and security scanning?

Industrial **project management** priorities center on the division of labor across engineers in a way that makes it practical for a new team member to begin contributing at their skill level to a project, including in cases where they are picking up the portfolio of responsibilities being handed off by a DS/RSE who has left or wishes to change their project focus. Project management best practice also ensures the transparency of the development project to stakeholders adjacent to the engineering team, like managers and PIs. This added transparency increases the longevity and durability of a software project by making it easier for lead developers and managers to transfer responsibilities as needed.

Below are some considerations for project management that could benefit your DS/RSE team and team members:

- What does an agile development cycle look like for the work of RSEs in academia? What does an agile data science project cycle look like in academia?
- How do you build an agile DS/RSE team?
- What does a collaborative software development and data science research cycle look like, that includes feedback and review from team members?
- How can tools like issue trackers and Gantt charts increase the transparency of the projects for DS/RSEs, their managers, and PIs?

Chapter 7 - Organization and Management of Research Software Engineer and Data Scientist Teams

THIS CHAPTER COVERS:

Organization of Teams: including discussions of group size, specialization, and longevity

Team Management Case Studies: two examples of different team organization and management strategies

Team Styles: examples and discussion of different models for organizing and focusing a team

Forecasting: resource management in different environments

Training and Professional Development: from individual to team and organizational needs

KEY TAKEAWAYS:

Team makeup varies: There is no one size fits all model for team makeup - as you build and grow your team, look to colleagues to see how they structure their teams and what you might learn from their experiences

Team size: Team size varies, and teams often start small and specialized. As the makeup and size of the team changes, consider evolving towards other models for organization and focus

Look for efficiencies: As your team grows and takes on more projects, seek out ways to reduce duplication across projects, including identifying and maintaining reusable infrastructure

7.0 Summary

The model of funding projects on finite grant cycles is incompatible with the career path of DS/RSEs and in the short-term presents a major structural obstacle to employee retention. Different organizational structures are required for DS/RSEs to overcome this structural issue. The incentive and management structure should be loosely coupled to the grant cycle. This in turn brings further opportunities in terms of addressing the topics of previous chapters such as recruitment, training, and retention where now economies of scale can be taken advantage of. This chapter explores how such a group might be managed, touching on strategies to organize and effectively manage a team and how to ensure the sustainability of the team from a financial, technical, intellectual, and personal perspective.

7.1 Organization

There is a tendency in academia, due to its underlying reward system, for researchers to “own” every aspect of their work - discoveries, credit, software, and even students and staff. This can result in the creation of duplicate infrastructure and added internal competition across the institution as individual research groups create (nearly) identical capabilities arbitrarily and compete for resources. From the perspective of DS/RSEs at the national level, this proliferation of competing endeavors is the reason for common sayings such as “yet another workflow system,” and the endless data management systems that seem to do the same thing (De Silva et al. 2021, Amstutz et al. 2022). This fractured

landscape, most especially internal to an organization, does not project “stability” to potential DS/RSEs and tends to drive many away, wanting instead a more assured position in industry or government.

An organization can address much of this by standing up a cross-cutting DS/RSE unit that can provide services to the many research domains at the institution. This breadth of stakeholders can also help smooth out the variability in funding by diversifying the stream of income. A centralized unit can also provide opportunities to influence administrative decision-making and impact career paths and technical policy, among other concerns. We describe some of the other windfalls from a more organized institutional approach below.

7.1.1 Economies of Scale

As with startups, initially, when the team is small everyone must be generalists as there is not enough work yet to support any specific specialty. Because of this staff members tend to find themselves not only doing multiple roles, but also putting in extra hours, or perhaps conversely having things fall off the plate as one must juggle priorities, handle context switching, address emergencies, and still meet deadlines. A larger team has the ability to specialize and react to a variety of requests from stakeholders. This is something that is tricky, if not impossible, with a small team, especially one that has an uncertain funding stream. A second key gain of larger more organized teams is the ability to provide meaningful career paths and retain staff. With an organized body of management and staff, one can largely ensure a constant influx of funded projects. Third, is the ability to share the overhead for, and even optimize, needed resources such as hiring and infrastructure. For example, a larger team can share the load of interviewing and onboarding new staff, or holding a specific position for project management. Last, a larger team may hold more sway in administrative decisions at the institution - for example, allowing for PI status for DS/RSEs or managing visa issues for employees. Failure to engage in these types of issues can be incredibly impairing, especially since the market for tech talent is fierce, and incentives for leaving academia can be strong.

7.1.2 Specialization in Key Areas

As described above, the ability to specialize is possible with a larger coordinated body of staff. This goes beyond just management however to also include areas of technical expertise. As an example, user interface/user experience (UI/UX) is being more and more recognized as a crucial need within science as difficult-to-navigate applications hinder uptake and usability. Poor UI/UX can limit the reach of the software which can, in turn, make maintaining the software more difficult (i.e. smaller user community may mean fewer contributions). In many cases, the UI/UX portion of a project is one small part, and a small team cannot employ a UI/UX specialist because they won't have enough work. Larger teams support more projects, creating opportunities to fill full-time roles with specialized skill sets. Other key areas of specialization among DS/RSEs could include dev ops, machine learning, or possibly an early foot in emerging areas such as quantum applications.

7.1.3 The Ability to Amplify Activities on Projects

Larger DS/RSE groups can produce an amplification effect, by which efficiencies are gained across projects through careful coordination. An example of this effect is the support the DS/RSE group can provide to small projects or teams that have relatively limited needs. Many of these smaller projects can be chained together by the DS/RSE group, especially when the needs of these small projects are clustered by technology or workflow types. A handy motto is that big projects anchor small projects and small projects broaden big projects (e.g. helping larger efforts reach even further into new communities). Additionally, bringing staff together allows for information sharing across projects, the ability for others to step in to meet urgent needs, and the crucial ability of senior staff training newer staff.

7.1.4 Institutional Knowledge and Sustaining Software

Institutional knowledge is key to sustaining things such as developed software and larger teams of DS/RSEs can bring to bear the knowledge and resources needed to do that more robustly. The DS/RSE group can approach software sustainability and maintenance strategically by having multiple individuals knowledgeable about a piece of software (i.e. increasing bus factor) and carving out time specifically for maintenance, updates, and other technical debt concerns. A mix of senior and junior staff can also be helpful for transferring institutional knowledge, turning software sustainability into part of the mentoring or training process. In addition, a team with longer institutional memory will be able to preserve architectural and methodological knowledge gathered over the course of a number of projects and technologies.

7.2 Team Management Examples

7.2.1 University of Washington eScience Institute

Since 2008, researchers in non-traditional tenure-track positions have been an essential part of the University of Washington eScience Institute, with titles of either Data Scientist or Research Scientist. Data Scientists at eScience are researchers whose passion lies in developing innovative solutions to real-world problems and tackling challenging data science problems in multiple domains. Individuals in this role have expertise in one or more areas related to data science (machine learning, statistics, databases, large-scale data systems, visualization, interfaces that facilitate large-scale data understanding), demonstrated experience in software engineering, and, ideally, demonstrated ability applying skills to solve problems in a disciplinary field.

Data Scientists both plan and execute data science research, translate research results into solutions with broad applications in the sciences, and produce practical applications delivering lasting impact across domains. Projects are sourced through joint collaborations with eScience affiliate faculty, through eScience leadership, and through the independent research goals of the individual. The Data Scientist's emphasis is on collaborative software development, innovative technical consulting, development and delivery of workshops and training, and other activities designed to advance the research and practice of data-driven and data-intensive discovery across fields.

Programmatically, Data Scientists are funded entirely through the eScience Institute. As well as working on collaborative research projects, Data Scientists support core eScience activities like the Data Science for Social Good summer

program and the Incubator Program. Many individuals in the Data Scientist role have successfully competed for external funding. In such cases, eScience has reduced the commitment to eScience by these individuals to allow them to pursue this research agenda.

Research Scientists represent a new kind of engagement for data-science-savvy folks around campus. Research faculty and staff already in permanent positions can have a portion (typically 25-50%) of their salary covered by eScience in exchange for equivalent hours spent on eScience activities. They have part of their salary covered, lessening the burden on their grants, and have the opportunity to both teach and learn data science skills through engagement with eScience. In return, this greatly expands the domain and methodological expertise eScience can offer the campus community. Individuals in these roles enjoy expanded autonomy, with flexibility over the research projects that cover salary. In most situations, eScience also has flexibility with the split in salary, allowing Research Scientists to be more engaged with eScience when other funding is limited and then focus more intensely on research when funding opportunities arise.

7.2.2 University of Illinois Urbana Champaign

The University of Illinois Urbana-Champaign established the National Center for Supercomputing Applications (NCSA) in 1986 as part of a national effort to establish Higher Performance computing capabilities in support of U.S. scientific research. While created to support computation, NCSA is probably best known for its developments and impacts in regard to software, with the best-known example being that of the web browser. At NCSA the software developers that have supported scientific efforts have been called research programmers, and recently been re-titled to research software engineer. These RSEs come from many different backgrounds, spanning researchers within specific domains that have focused on software development, to industry software developers that prefer the academic lifestyle and rewards, to recent graduates looking to gain additional experience before moving on. Teams are organized such that a senior RSE with 15+ years of experience developing research software oversees newer software engineers in the environment, coaching them in the intricacies of software engineering best practices and becoming part of the research endeavor. These senior RSEs serving as managers have 25% of their time covered for management, serving on efforts with their team akin to a player-coach. Being a soft-funded team the leadership works to maintain a constant influx of new projects with senior and lead RSEs writing proposals themselves and/or with researchers and, over time, developing a rich portfolio of expertise areas, software frameworks, projects, and continuing collaborators. The RSE career path at NCSA aims to foster this ability, i.e. working closely with collaborators and writing proposals to fund developed ideas, helping the team maintain this ability to maintain activities and even grow. Additionally, as NCSA is fairly well known for its supporting role to research, many faculty members also approach the organization with support needs on currently funded efforts. A management committee made up of the lead RSEs collaboratively works to staff incoming efforts, triaging if need be, prioritizing based on alignment with the team's expertise and goals, ensuring all staff members are allocated for one or more years out, and hiring new staff if need be based on foreseen needs.

7.3 Teams

There are numerous ways to organize teams, each with its pros and cons dependent on not only the goals and mission of the overall organization but also its funding makeup. Many of the approaches below tend to best fit large teams, though these examples may carry helpful information for smaller teams to consider.

7.3.1 Project Based

Organizing teams based on projects is probably the most natural form of grouping. Benefits are clear in that the team can be very focused on the project's needs and timelines. There are some drawbacks, however. One drawback is that staff often end up siloed from other efforts, efforts that might benefit from the team's expertise and developments. Project-based teams may also suffer from duplication of effort, as different projects might address similar needs. Probably most importantly, these teams tend to be more susceptible to the uncertainties of soft funding. If you plan to build project-based teams, it is critical to plan for gap funding in order to maintain the team long-term, or, how the team is to be dissolved and merged into other new teams that may arise. Last, this mode of organization does not work for small project teams (e.g. with capacity to support only 0.25 or 0.5 FTE) as this will require the DS/RSE on the team to split their time across many projects, which can create distractions and competing priorities.

7.3.2 Thematic Based

Another option is to organize teams based on domain of expertise (e.g. biology, geoscience, astronomy, etc.). This approach benefits from an ability to focus, in this case thematically, even though staff members might work on different projects. Thematic teams also have an opportunity to focus mentoring and professional development on relevant topics and technologies. One downside of thematic teams includes some of the same siloing we see with project-based teams, including duplication of effort across teams. For example, in the cyberinfrastructure world this has resulted in a great many data management systems which essentially serve the same function as data challenges and needs are by no means unique, or unique enough, to a given domain.

7.3.3 Expertise Based

One might also organize groups based on technical expertise (e.g. front end development, back-end development, DevOps, UI/UX), which has the benefit of allowing the group to focus, technically, and grow expertise in a given technical area. A benefit of this approach is that it can allow for the growing and shrinking groups as demand for skillsets changes over time. Downsides include that many DS/RSEs in academic settings tend to have a diversity of skills and may not have experience in such a technically focused setting. Further, this approach tends to best fit very large teams with sufficient sustainability and a constant flow of projects. Some areas, such as UI/UX, often involve a relatively small percentage of an FTE on individual projects, so a large number of projects is needed in order to support one FTE, let alone a team of people with such skills.

7.3.4 Other Models

There are of course other means of organizing teams, where some goals other than those above are of emphasis. For example, one approach, as described later on, is to build out a body of DS/RSEs specifically as demand grows. In this case, each team might be a mix of domains, skills, and projects, be led by a sort of advisor or coach possessing significant experience in this area (in this case supporting technical needs in an academic setting), and who

works with the team to implement best practices on the technical side, as well as research/collaborative aspects on the academic side. This approach adopted at the University of Illinois right now has leads that are themselves from the technical research ranks, RSEs themselves serving as player-coaches, and in so doing additionally provides leadership opportunities for the RSE career path there. As with every approach, there are downsides, in this case largely around keeping track of all the projects your staff members are working on, some of which the player-coach may not be on themselves.

7.4 Forecasting

A critical aspect of managing a team of DS/RSEs is ensuring they are constantly fully allocated across their duties (projects, teaching, consulting, etc.). How this is done, as well as the criticality of this, varies, especially from the perspective of core-funded versus soft-funded groups (see Section 7.2 - Team Management Examples).

7.4.1 Resource Management

Regardless of the funding environment, it is important to track the capabilities of the team. One will find that, in reality, staff members have multiple areas of expertise at various levels of proficiency. Given the fluctuations in projects over time, as well as often limited resources within academic settings, being able to track and optimally assign these skill sets is key. This can be done in a number of ways, from high-level expertise areas amongst staff to very detailed in-depth surveys that are conducted regularly (so as to capture new skills learned). Skill areas can be domain-specific knowledge, to more technical skills such as machine learning, front end & back end development, dev ops, etc. In addition to allowing the group to support a wider range of efforts and increased opportunities, resource management also allows the group to see weak spots where demand is high but expertise is low and then work to fill that through either training or hiring. Section 6.1 - Professional Development Opportunities in Diverse Academic Environments discusses additional aspects of resource management, including the concept of Professional Profiles.

7.4.2 Soft-Funded Environments

In a soft-funded environment in particular it is critical to have staff commitments forecasted out one, if not two years, in advance, depending on organizational rules around appointment of soft-funded staff. This can be a difficult task, especially if there is not any core funding support.

The first necessity of teams in this situation is establishing a means by which projects come to the team. This can come from connections to campus, where researchers know to approach the group for support, or from staff within the group who work to bring in projects as Co-PIs on grants, or possibly even PIs. With such a system in place, leadership must then align staff with projects and somehow “make it all work” such that all projects are supported and all staff are fully allocated, ideally aligned with interests and skills. This is a challenging task, to say the least. In particular, this makes it very difficult to be nimble and to quickly respond to new projects and needs, as the group is largely in a state of “just enough staff”. In fact, any fluctuations in funding - lulls or peaks - can be very difficult to handle without some sort of buffer to smooth out the changes. The ideal buffer is of course some sort of core funding. That said, other approaches include stretching out efforts, such as delaying staffing on a new project until a current project ends, or if possible leveraging a no-cost extension on a project allowing the funds to last longer. To respond to peaks in needed support, it is

helpful to have an efficient hiring and training setup. Maintaining a continuous search to build out a resume pool so that candidates can be brought on quickly as needs arise, as well as leveraging senior staff to train newer staff so that less experienced staff might be brought on and brought up to speed more quickly.

7.5 Commercial vs Research: Awareness of Differences in Management and Expectations

DS/RSE teams' workflows, scope, and expectations in academic institutions differ from those common in commercial settings. In many commercial environments, there is a clear and common understanding of a standard valuation metric, such as a cash flow-weighted return on investment. There is often a product manager who has profit and loss responsibility for the specific area and training in how to express requirements in ways that are more straightforward to express in software engineering and data science terms. By contrast, the research setting involves the generation of new knowledge, and the path by which that knowledge might be obtained is itself something that requires research. In addition, in academia, a number of external stakeholders have dependencies for your team (i.e. datasets, scientific code, etc). But these individuals do not report to you or your team and thus you have no control over when exactly these entities will provide you with items. This results in dependencies/blockers that prevent other work from proceeding. This is especially true of multi-year research projects.

Due to the nature of research, planning from A to Z for a project often looks different at point C, as the end-product/goal shifts as dictated by what was discovered in steps A and B – this is not a failure, but just the way science/research evolves. As mentioned in the section on scoping below, managers need to be aware and upfront with stakeholders and continue to get updates from the DS/RSEs on additional requests and modifications to scope to ensure the requirements are reasonable and align with the DS/RSEs skill set and Full-Time Equivalent (FTE) allocation. Attempting to get up-front requirements in a waterfall management fashion, disappearing for 6 months to build an implementation, is a nearly guaranteed recipe for disaster in an academic setting. A collaborative agile approach is typically required, working closely with the researcher and students, adapting and changing course continuously.

Cadence is also different. Industry can typically pursue a more tightly coupled roadmap/milestone and development sprint schedule. The research side however can see roadmap/milestones shift significantly from project inception to implementation. Development sprints are still valid and are certainly better than the waterfall approaches since the nature of Agile work lends itself to the shifting requirements. For data scientists, often how long a particular analysis will take might be unknown as datasets and information have to be explored and worked with; so managers should be aware and flexible in how they design sprints or organize work blocks and seek input from the involved DS/RSEs.

Last, there is the notion of a principal investigator (PI) within academia which is completely foreign within an industry setting. Somewhat akin to the founder of a startup, the PI has an intimate stake in the establishment, direction, and outcome of a funded project, the results of which impact their reputation and career. Different PIs will lead differently, sometimes working heavily in the weeds, sometimes leaving much of the effort to others. Regardless, at the end of the day, the PI is considered to have the final say in project activities. This

can be foreign to those outside academia where decision-making and responsibilities are delegated within a command structure, and perhaps project managers have direct control of specific areas. Not appreciating this difference can spell disaster for those new to academia. In light of this, managers coming from industry will very likely need to make some adjustments to their project and team management approaches when they transition into an academic setting. Relatedly, managers should be prepared to support and ease this transition for software engineers and data scientists coming from industry and expect that this may be a significant adjustment and transition. There may be some inherent value in building mixed teams where some members have followed academic career paths while others have spent some time in industry, ideally bringing the relevant best practices of both worlds to bear.

7.6 Differentiating Team Types and Needs

Within a single organization, there may be many different types of DS/RSE teams, each with its own considerations and needs. Below are a few types of DS/RSE teams you might encounter or consider building at your organization, along with guidance for how these teams might be managed and developed.

7.6.1 Project Team Needs

Many DS/RSEs do their work while deployed on a specific project or product team. For example, they may work alongside scientists and principal investigators that define the scientific objective and the needs that they have. This may be a permanent team that also includes an engineer's line manager, or it may be a different and temporary team to which a DS/RSE is deployed during the project. In either case, there are a set of principles to follow that ensure a DS/RSE is positioned to succeed in executing on the project.

Definition of Success

Clear expectations, metrics, and pathways for advancement are important. DS/RSEs are non-traditional roles in academia and, depending on the university, are hired across a breadth of position titles. They might be hired in academic tracks or in professional staff tracks, where benefits and freedoms can be quite divergent. In some cases there may be a limited number of position grades that even exist, thwarting pathways for advancement. This may be exacerbated due to the tendency by some to hire research software engineers/data scientists at the highest grade possible to access more competitive compensation rates. Since position titles and their associated advancement metrics are often not a good fit for these positions, it is important that managers clearly define the expectations and benchmarks for each level and clearly convey these to their reports.

Project goals must deliver value to the scientific team but also be realistic in the scope of the available time and resources. A common pitfall of embedded engineers is that they become too focused on technical problems, and miss the "big picture" of the scientific need they are meant to serve. Additionally, many principal investigators are not trained in how to translate abstract scientific needs into concrete, actionable, and realistic technical goals for an engineer to implement. This results in wasted effort or frustration due to an inability to properly communicate a project's technical needs, and design a plan for meeting them. It is crucial for a project team to have a clear definition of their goals, and how specific technical needs fit into those goals. This is often carried out by a facilitator (e.g., the engineering manager or a product manager) that can ask the right questions to translate science objectives into technical steps. This

should be assumed to be an iterative process, fostered by regular touch-ins between researcher(s) and developer(s) with continued communication support by the manager throughout.

7.6.2 Data Scientist and Research Software Engineer Team Needs

A DS/RSE team is composed of multiple DS/RSEs along with a manager that oversees the distribution of work and projects amongst the team, cultivates each team member on a career pathway, and treats the team as a system that works together. This is sometimes the exact same team as the "Project Team" above, though in many cases can be a centralized DS/RSE team that deploys DS/RSEs to many other project teams over time. The DS/RSE team is concerned with developing and supporting DS/RSEs as individuals and as a team, as opposed to focusing on meeting specific project deliverables. However, a DS/RSE team manager may interact with Principal Investigators or Project Managers to integrate feedback about the performance and needs of engineers.

Mentorship and Community

Access to more senior personnel is important in order to learn from and get assistance/advice. For the centralized team, having staff at different levels of the career ladder makes it relatively easy to set up mentorship relationships.. However, it may be necessary to explicitly make that mentorship an expectation of more senior members of the team and to create structures to recognize and reward it as part of their work (rather than in addition to it). Ensuring documentation of the team's preferred processes and workflows also facilitates a variety of learning styles. For DS/RSEs who are embedded in research groups across the institution, finding and building meaningful community can be more challenging and will require thought and coordination across units.

Additionally, peers/community of practice to exchange ideas and discuss experiences is important. With the embedded/ distributed model, the formation of a peer group or community of practice should be considered to provide additional professional support. Although regional/ national organizations exist there is certainly benefit to a local community that understands the institution and politics that go along with it, since every place has unique challenges to navigate that local peers have likely run up against and can offer the most specific advice and feedback.

Trust

DS/RSEs are likely to be attracted to academia by opportunities to work on intellectually engaging and meaningful projects, as well as a greater sense of creative freedom, as compared to related roles in industry. As appropriate to individual maturity and seniority, managers should provide opportunities and avenues for DS/RSEs to make technical decisions and have increasing levels of autonomy in project work. However, managers should set expectations as to the range (i.e. stability, maintainability, etc.) and ensure there are open lines of communication related to DS/RSE approaches so that there is trust on both sides in the final product. This can help enable investment, productivity, and satisfaction across the team when members feel there is trust and support. Staff should have the sense that they have the freedom to explore, however, must also deliver timely results.

7.6.3 Organizational Needs

Organizational needs reach across the entire team of DS/RSEs and include the boundary where this team interfaces with the broader University or Research Lab. Typically these needs are the responsibility of someone serving in a director role. For example, a PI of the grant, if their grant provides for a team of DS/RSEs, or the explicit director of a unit which could be situated within a larger IT structure, a School at a University, or a stand-alone entity. These needs deal with relationships between this team and the groups with which they work and also the sustainability of the entity. These needs require the director to look both inside and outside of the team.

Mission and Values

Setting a mission statement and an accompanying set of values for your group is an opportunity to build out criteria or guiding principles to undergird your preferred working conditions, create and uphold boundaries, attract new talent looking for an alternative to corporate environments, and even prevent burnout in your staff. If you publicly state that your team values openness, for example, you can refer to that value to champion open-source solutions (and/or refuse proprietary ones) and make space for your team not only to contribute to open-source projects and open infrastructure but to be recognized for that work (in an annual review, for example); if reproducibility is a core value, you can set aside time, space, and kudos for your team to engage in code review, replicable workflows, FAIR data curation, and related practices.

Sustainability and Position Security

Centralized structure and funding provide a sense of job security not realized by soft money positions housed in individual labs where available support follows grant cycles. Ideally, hiring onto a centralized team guarantees full and sustainable position support with administrative staff handling billing and fractional FTE allocations. If the team or individual DS/RSE funding model is different then managers should be upfront and communicative about the status of funding with both the DS/RSE staff and the directors/funders/PIs to ensure transparency and trust in position status and future plans.

Identifying, Consolidating, and Escalating Issues

Given that DS/RSE are non-traditional occupations within most academic environments, a key need is for someone who can understand the particular and systemic needs facing these professions and who has the positional authority to be able to escalate those needs to institutional decision-makers. As one example, the processes associated with visa processing at one institution were substantially impacting the institution's ability to attract and hire research software engineers. Because that institution had a more centralized model for research software engineers, the effect of these visa processing practices was easier to identify, quantify, and escalate to higher levels, resulting in positive changes.

Project Point Person

The manager/point person who engages with potential project PIs needs to employ a structured framework to understand project objectives and engineering needs. These are then clearly conveyed to the rest of the team of software engineers and/or data scientists who will be working on these projects. Engineering needs are set and tracked by this point person. Faculty PIs don't always have the full picture when it comes to all the labor required to facilitate their research project or process. The Project Point Person acts as an advocate supervisor on the team, translating and de-

mystifying the engineering process for the researcher, serving as a point of contact (and triage), and explaining the rationale behind project timelines and potential delays. This role is crucial to giving the research software engineer/data scientist sufficient bandwidth to actually work on the project itself.

Defined Software Stack & Managing Customer Expectations

The overall area of expertise that the unit maintains is a high-level strategic decision that influences every aspect of how the team functions. From hiring through to project execution, it is critical to select the right skill set for the team to possess. This decision is made thinking about the broader community outside of the team, as well as looking toward the future. Lingering on one language too long or continually jumping to every new one is a recipe for disaster. Each effort must decide which programming languages to support. Managers need to set expectations with projects as to what technologies can be applied based on current staff/team knowledge/expertise. If other technologies are necessary then training/ramp-up has to be accounted for in the scope of work otherwise the project will not be successful and the research software engineer/data scientist will be frustrated. Much like the positions on a sports team, the managers of a research software engineers/data scientists team can maintain a roster document that describes the skill sets of the team members, their level of proficiency, and their development path (i.e. the resource management described previously). This document can be used in many contexts from, making assignments, to scoping projects with PI partners, to annual review processes.

7.7 Healthy Work Practices

Academia has recognized work-life balance issues. Protecting your team's time and drawing clear boundaries around working hours is vital for managers of research software engineers and data scientists, many of whom may have chosen to take a lower salary particularly because they value that balance. This may mean setting clear expectations for faculty partners, for whom the work-life boundary is porous, and who may choose to work 60 or 70-hour work weeks on their own research—and expect (implicitly or explicitly) others to do the same. As a manager, you should document iterative deadlines, expectations, and project scope early, and be prepared to reinforce them when necessary. You may also need to pay attention to your own practices, including communicating (or committing) outside of work hours, so that the DS/RSE understands team expectations and boundaries.

Appreciating variety without depletion (context switching) in an academic setting is important. Often there are more projects than people to work on them so there is a tendency to split DS/RSEs across more than one. Current experiences on successful DS/RSE groups indicate that people should really only work on two projects at a time or risk too much overhead related to jumping from one project to another and thus also increasing the incidence of burnout.

One main benefit of academia can be flexible schedules. While many universities are striving to return to pre-Covid expectations with all staff working full-time onsite, this expectation could further challenge hiring efforts for these very competitive positions, especially with industry positions increasingly offering similar flexibility. While laws (particularly tax and employment laws) may substantially limit an institution's ability to employ

remote personnel outside of their particular jurisdiction, seeking the greatest flexibility practical within those laws for remote and hybrid work is particularly important for attracting and retaining these types of professionals. Likewise, managing work schedule flexibility to the extent possible, while maintaining effective work relationships, is essential in attracting and retaining these very sought-after professionals.

References

Amstutz, P., Mikheev, M., Crusoe, M.R., Tijanić, M., Lampa, S., et al. (2022). Existing Workflow systems. Common Workflow Language wiki, GitHub. https://s.apache.org/existing-workflow-systems updated 2023-05-03, accessed 2023-06-29.

Bass, J. M., Beecham, S., Razzak, M. A., & Noll, J. (2018, May). Employee retention and turnover in global software development: Comparing in-house offshoring and offshore outsourcing. In Proceedings of the 13th International Conference on Global Software Engineering (pp. 82-91). https://doi.org/10.1145/3196369.3196375

Brady, H., Geiger, S., Culich, A., Das, D., Howard, T., Mazel-Cabasse, C., Fiore-Gartland, B., Noren, L., Cullens, C., Hench, C. (2019). Developing Desirable and Sustainable Career Paths for Academic Data Scientists. Slide Deck. Retrieved June 13, 2023, from https://docs.google.com/presentation/d/1zJyGuqoYu_vIgsT2716xt4tPIg2N77g-sLcPj_e6uXM/edit#slide=id.g37ae830f00_9_98

Bureau of Labor Statistics (2022). Job Openings and Labor Turnover - August 2022. Bureau of Labor Statistics - Job Openings and Labor Turnover Survey (JOLTS) program https://www.bls.gov/news.release/archives/jolts_10042022.pdf

Cable, D. M., Gino, F., & Staats, B. R. (2013). Reinventing employee onboarding. MIT Sloan Management Review, 54(3), 23. https://www.hbs.edu/ris/Publication%20Files/Reinventing%20the%20onboarding%20process_3b5ac7ce-f71b-40d9-bb8a-93276b979570.pdf

Career Framework - Human Resources at Ohio State. (n.d.). Retrieved July 7, 2023, from https://hr.osu.edu/career-roadmap/career-framework/

Carver, J. C., Weber, N., Ram, K., Gesing, S., & Katz, D. S. (2022). A survey of the state of the practice for research software in the United States. PeerJ Computer Science, 8, e963. https://doi.org/10.7717/peerj-cs.963

Colby, G. T. (2022). The Annual Report on the Economic Status of the Profession, 2021-22. American Association of University Professors. https://www.aaup.org/report/annual-report-economic-status-profession-2021-22

Colleges and Organizational Structure of Universities (n.d.). In StateUniversity.com Education Encyclopedia. https://education.stateuniversity.com/pages/1859/Colleges-Universities-Organizational-Structure.html

Cosden, I. A., McHenry, K., & Katz, D. S. (2022). Research Software Engineers: Career Entry Points and Training Gaps. Computing in Science & Engineering, 24(6), 14–21. https://doi.org/10.1109/MCSE.2023.3258630

CRediT. (n.d.). CRediT (Contributor Roles Taxonomy). Retrieved July 7, 2023, from https://credit.niso.org/

Da Silva, R. F., Casanova, H., Chard, K., Altintas, I., Badia, R. M., Balis, B., ... & Wolf, M. (2021, November). A community roadmap for scientific workflows research and development. In 2021 IEEE Workshop on Workflows in Support of Large-Scale Science (WORKS) (pp. 81-90). IEEE.
https://arxiv.org/pdf/2110.02168.pdf

Data Science & Analytics Salaries Are on the Rise. (2022, February). Retrieved July 7, 2023, from https://www.burtchworks.com/industry-insights/data-science-analytics-salaries-are-on-the-rise

Developing Employee Career Paths and Ladders. (n.d.). Society for Human Resources Management. Retrieved July 7, 2023, from https://sso.shrm.org/IDBUS/SHRM/PORTAL-EE/JOSSO/SSO/REDIR?josso_cmd=login&josso_partnerapp_id=portal-sp

Dutt, K., Pfaff, D., Bernstein, A. et al. (2016). Gender differences in recommendation letters for postdoctoral fellowships in geoscience. Nature Geosci 9, 805–808 (2016). https://doi.org/10.1038/ngeo2819

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. Journal of Personality and Social Psychology, 101(1), 109–128. https://doi.org/10.1037/a0022530

Geiger, R., Mazel-Cabasse, C., Cullens, C. Y., Norén, L., Fiore-Gartland, B., Das, D., & Brady, H. (2018, June 6). Career Paths and Prospects in Academic Data Science: Report of the Moore-Sloan Data Science Environments Survey. https://doi.org/10.31235/osf.io/xe823

Griffith, E. H., Sharp, J. L., Bridges, W. C., Craig, B. A., Hanford, K. J., & Stevens, J. R. (2022). The academic collaborative statistician: Research, training and evaluation. Stat, 11(1), e483. https://doi.org/10.1002/sta4.483

Grillo, M., & Kim, H. K. (2015). A Strategic Approach to Onboarding Design: Surveys, Materials, & Diverse Hires. https://ecommons.cornell.edu/handle/1813/74358

Harvard University (n.d.). IT Research Computing Matrix. Harvard Human Resources. Retrieved June 15, 2023, from https://hr.harvard.edu/sites/hwpi.harvard.edu/files/humanresources/files/it_research_computing_matrix.pdf

Herrity, J. (2023). Job Levels in the Workplace (With Sample Titles and Duties). Indeed.com. https://www.indeed.com/career-advice/career-development/job-level

Jackson, M. (n.d.). How to cite and describe software. Software Sustainability Institute. Retrieved July 7, 2023, from https://www.software.ac.uk/how-cite-software

Job Openings and Labor Turnover - May 2023. (2023). [News Release]. US Department of Labor - Bureau of Labor Statistics.
https://www.bls.gov/news.release/archives/jolts_07062023.pdf

Journal of Open Source Software. (n.d.). Retrieved July 7, 2023, from https://joss.theoj.org

Kang, S., DeCelles, K., Tilcsik, A., & Jun, S. (2016, March 29). The Unintended Consequences of Diversity Statements. Harvard Business Review. https://hbr.org/2016/03/the-unintended-consequences-of-diversity-statements

Katz, D. S., Chue Hong, N. P., Clark, T., Muench, A., Stall, S., Bouquin, D., Cannon, M., Edmunds, S., Faez, T., Feeney, P., Fenner, M., Friedman, M., Grenier, G., Harrison, M., Heber, J., Leary, A., MacCallum, C., Murray, H., Pastrana, E., … Yeston, J. (2021). Recognizing the value of software: a software citation guide. F1000Research, 9, 1257. https://doi.org/10.12688/f1000research.26932.2

Katz, D. S., McHenry, K., & Lee, J. S. (2021). Research software sustainability: Lessons learned at NCSA. In 54th Annual Hawaii International Conference on System Sciences, HICSS 2021 (pp. 7249-7256). IEEE Computer Society. http://hdl.handle.net/10125/71494

Katz, D. S., McHenry, K., & Lee, J. S. (2021). Senior-level RSE career paths (with an s). Daniel S. Katz's blog. Retrieved June 13, 2023, from https://danielskatzblog.wordpress.com/2021/09/27/senior-rse-paths/

Katz, D. S., McHenry, K., Reinking, C., & Haines, R. (2019, May). Research software development & management in universities: case studies from Manchester's RSDS group, Illinois' NCSA, and Notre Dame's CRC. In 2019 IEEE/ACM 14th International Workshop on Software Engineering for Science (SE4Science) (pp. 17-24). IEEE. https://doi.org/10.1109/SE4Science.2019.00009 (Preprint: arXiv 1903.00732)

Lloyd, J. (2022, September 12). Academic Mentorship Needs a More Scientific Approach. Issues in Science and Technology. https://issues.org/academic-mentorship-scientific-approach-montgomery-sancheznieto-dahlberg

Mazumdar, M., Messinger, S., Finkelstein, D. M., Goldberg, J. D., Lindsell, C. J., Morton, S. C., ... & Parker, R. A. (2015). Evaluating academic scientists collaborating in team-based research: A proposed framework. *Academic medicine: journal of the Association of American Medical Colleges*, *90*(10), 1302.

Milewicz, R., & Mundt, M. (2021). An exploration of the mentorship needs of research software engineers. arXiv preprint arXiv:2110.02251.

Min, S. (2022, October 5). Resources for Mentors and Mentees. Academic Data Science Alliance. https://academicdatascience.org/resources/2022-10-05-blog/

Montgomery, B. L., Sancheznieto, F., & Dahlberg, M. L. (2022, September 12). Academic Mentorship Needs a More Scientific Approach. Issues in Science and Technology. https://issues.org/academic-mentorship-scientific-approach-montgomery-sancheznieto-dahlberg

Nangia, U., & Katz, D. S. (2017, August 30). Track 1 Paper: Surveying the U.S. National Postdoctoral Association Regarding Software Use and Training in Research. figshare. Retrieved June 13, 2023, from https://figshare.com/articles/journal_contribution/Track_1_Paper_Surveying_the_U_S_National_Postdoctoral_Association_Regarding_Software_Use_and_Training_in_Research/5328442/3

NCCC 170 and SCC 13 (USSES) Multistate Research Coordinating Committees and Statistical Information Exchange Groups (2018). Overview of Statistics as a Scientific Discipline and Practical Implications for the Evaluation of Faculty Excellence. American Statistical Association. Retrieved June 13, 2023, from https://www.amstat.org/asa/files/pdfs/POL-Statistics-as-a-Scientific-Discipline.pdf

Nietzel, M. T. (2021). The 19 U.S. Universities Spending $1 Billion Or More On R And D. Forbes.com. https://www.forbes.com/sites/michaeltnietzel/2021/02/08/the-19-us-universities-spending-one-billion-or-more-on-r-and-d/?sh=19a0364a183a

Ohio Supercomputer Center (n.d.). Ohio Supercomputer Center: An O H Tech Consortium Member. https://www.osc.edu/

Pomann, G. M., Boulware, L. E., Cayetano, S. M., Desai, M., Enders, F. T., Gallis, J. A., ... & Thomas, S. M. (2020). Methods for training collaborative biostatisticians. Journal of Clinical and Translational Science, 5(1), e26. https://doi.org/10.1017/cts.2020.518

Princeton Research Software Engineer (RSE) Partnership Guide. (2021). Princeton Research Computing. https://researchcomputing.princeton.edu/services/research-software-engineering/rse-partnership-guide

RCD Professionalization Working Group. (2018). Research Computing and Data Professionals Job Elements and Career Guide. Zenodo. https://doi.org/10.5281/zenodo.7108947

Scroggins, M. J., & Pasquetto, I. V. (2020). Labor Out of Place: On the Varieties and Valences of (In)visible Labor in Data-Intensive Science. Engaging Science, Technology, and Society, 6, 111–132. https://doi.org/10.17351/ests2020.341

University of Arizona (n.d.). *Career Streams*. University of Arizona Human Resources. Retrieved June 15, 2023, from https://hr.arizona.edu/content/career-streams

Sharma, G. G., & Stol, K. J. (2020). Exploring onboarding success, organizational fit, and turnover intention of software professionals. Journal of Systems and Software, 159, 110442. https://doi.org/10.1016/j.jss.2019.110442

Sharp, J. L., Griffith, E. H., & Higgs, M. D. (2021). Setting the stage: Statistical collaboration videos for training the next generation of applied statisticians. Journal of Statistics and Data Science Education, 29(2), 165-170. https://doi.org/10.1080/26939169.2021.1934202

Smith, A.M., Katz, D.S., Niemeyer, K.E., FORCE11 Software Citation Working Group. (2016) Software Citation Principles. PeerJ Computer Science 2:e86. DOI: 10.7717/peerj-cs.86

Software & Application Innovation Lab (n.d.). Software & Application Innovation Lab: Where Academia Meets Application. https://sail.bu.edu

Soito, L., & Hwang, L. J. (2017). Citations for Software: Providing Identification, Access and Recognition for Research Software. International Journal of Digital Curation, 11(2), 48–63. https://doi.org/10.2218/ijdc.v11i2.390

Hiring Toolkit: A Guide for Hiring Managers and Search Committees. (2023). Stephen F. Austin State University. https://www.sfasu.edu/docs/human-resources/hiring-manager-toolkit.pdf

University of Arkansas Office of the Provost (n.d.). Academic Policy 1622.30: Centers and Institutes. https://provost.uark.edu/policies/162230.php

The United States Research Software Engineer Association (n.d.). Skills Used by RSEs. https://us-rse.org/resources/training/skills/

USA Facts (2023). What do universities do with the billions they receive from the government? https://usafacts.org/articles/what-do-universities-do-with-the-billions-they-receive-from-the-government/

Yi, A. L. (2019). Lessons Learned: A Case Study of Performance Funding in Higher Education. Third Way. https://www.thirdway.org/report/lessons-learned-a-case-study-of-performance-funding-in-higher-education

Appendix A - Career Paths

A.0 Summary

Having a well-defined and articulated career path for DS and RSE positions in academic units is important, as these roles in industry have well-defined roles and clear career ladders. To ensure that those in the RSE and DS positions in academia see these positions as viable, long-term careers, and therefore keep their expertise and knowledge within their academic units a career path will be necessary. This is not intended to sweep under the rug the various challenges associated with academic positions, which will almost always be smaller in terms of personnel than their industrial counterparts. Still, on the flip side, these positions carry significantly more intellectual freedom and individual contribution. The goal of this section is to highlight some challenges of being a staff DS/RSE, drawing upon some of the established best practices from industry - combining them with the freedom and flexibility associated with academia.

A.05 Building a Formal Career Path

In the sections above, we have focused on professional development. When a DS/RSE takes advantage of these opportunities and builds new skills, the next natural question, then, is: what is the next position in my career path? If they are willing to remain in the same team or the same institution but would like to seek a more advanced position, then having clear paths for career progression is essential for retaining highly skilled and motivated staff. Possible career paths should have transparent and well-defined processes for progression that are backed by corresponding HR processes and appropriate job classifications. Working with HR to understand the job classification system and criteria for progression is essential for supporting DS/RSEs.

There are multiple points in the process of supporting DS/RSEs when working with HR is critical. Engaging with the HR team when creating positions and hiring has been discussed in previous sections. For supporting career progression, understanding the job classification and compensation processes is essential.

Each organization has a framework for classifying staff positions and all staff are mapped into this framework when hired. Understanding how you or your employees are mapped and what progressions are available in the existing framework are the first steps toward progression. Your institution's HR representatives can help you understand and navigate the HR process for career progression. A career framework will likely have specific criteria for each level in a career category, those criteria should be the foundation for DS/RSE-specific career path progression. Understanding your organization's job classification framework (e.g. at The Ohio State University, Career Framework - Human Resources at Ohio State. (n.d.)) and corresponding terminology. Understand that your HR department's goal is to regularize job descriptions and levels across the organization. Criteria for each level will be in general terms that apply broadly across the organization. For example, for positions mapped as individual contributor specialized, there are brief descriptions for progression from level 1 to level 5 with respect to knowledge, complexity, autonomy, etc. For example, for the knowledge aspect of career levels, the criteria for each level are:

1. Application of basic concepts
2. Application of general industry knowledge & experience
3. Applies a thorough understanding of concepts
4. Advanced understanding of concepts
5. Utilize broad & unique knowledge

Justification for career progression should reference these HR guidelines with specifics on why an employee should progress to the next level. Working with your unit HR representative to navigate the process for requests/approvals for level advancement. If the HR framework does not have job classifications that are well matched to DS/RSE positions then DS/RSE managers should develop clear progressions that map to the HR levels and criteria, with HR's consensus, and share them with employees so that expectations are clear. When current job classifications for DS/RSE do not have appropriate progression levels or compensation, a DS/RSE may consider working with your unit HR or manager to reclassify the position to a more appropriate job within your organization's job framework.

An increase in pay usually accompanies an increase in responsibility. Again it is critical to understand HR processes regarding compensation to get a reasonable increase. Discuss with HR what is a customary increase, and if that increase is not sufficient for retention, discuss with HR additional mechanisms for increasing salary. They may include pursuing a process for an equity increase if the staff member is not being paid equitably with other employees at the same job category and level at your institution. Your organization may have a process for market increases, if so, find relevant market data. You may also refer to Section 3.3 - Compensation Considerations for more discussion of compensation strategies.

For managers, do the best you can with the annual raise process and use bonuses (when available) at your organization to reward high-performing staff members. Finally, most institutions will have a process to request matching if a staff member is offered a higher salary for an external job. Don't miss an opportunity to counter an employee's job offer - making counter offers can signal to the employee, and other members of the team, their value to your organization, and may result in the retention of the employee.

A.1 Faculty or Staff?

At academic institutions, different categories of positions have different requirements, growth opportunities, salary ranges, evaluation criteria, and responsibilities. While the main distinctions are between faculty and staff roles, there can also be multiple types of faculty and staff positions at a given institution. It is important to understand the distinctions between the available categories when deciding in which category to situate data scientist (DS) and research software engineer (RSE) positions.

The category of the position is likely to have significant implications for:

- **Hiring Requirements**: what, if any, background experience does the candidate need to meet requirements for experience and education?
- **Interviewing Requirements**: what form the interview takes and whether there are resources available for in-person interviews
- **Compensation**: what is the salary range for this position? What other types of compensation are available?
- **Promotion and Evaluation**: what types of work products or interactions are used for evaluation for promotion? What promotion pathways are available for the position?
- **PI Status**: can the position hold "PI Status" at the institution? In other words, can the position lead grant proposals, interact with the institutional review board as a PI, and lead other initiatives at the institution?

Many DS/RSE positions have some responsibilities that would qualify them for faculty roles and others that would qualify them for staff positions. There is no one right answer as to how DS/RSE positions should be classified, but it is important to understand the implications of the choice and to structure the position to be successful in the chosen type of role. A data scientist who is in a faculty role but lacks the time and support to publish original research will not be able to be successful in their position. A staff DS/RSE who is expected to bring in funding will have a difficult time doing so if they cannot apply for grant funding independently.

A.2 Career Ladders

A career ladder is the progression of positions within an organization from entry-level to higher levels of skill, responsibility, and benefits. Depending on the institution, a career ladder may also be referred to as a job family, although "job family" typically refers to a broader class of jobs comprising multiple career ladders, such as a research computing job family with system administrators, RSEs, research facilitators, etc. or even more broadly, such as an IT job family.

Different institutions may use different names for these positions, and many DS/RSE groups are not yet large enough to support positions at all levels of the career ladder. However, using this shared ladder of career levels can help DS/RSEs understand and plot a career path that may span academic institutions, industry, and government roles. Inflating staff titles, as is common in many university departments, is discouraged. (As an example, note the prevalence of "director" titles at your institution for staff who have no direct reports.)

Below are general templates that often must be modified to conform to the HR policies and expectations at any individual institution. Chapter 3 - Before Posting: Position Descriptions, HR, and Compensation addresses working with HR. See Appendix B - Position Description Examples for job descriptions along a career ladder.

Benefits to having Data Scientist and Research Software Engineer career ladders include:

- Recruiting people to positions where the job title matches their expectation is easier.
- Full career ladders set clear expectations of the requirements and responsibilities at each level to support professional development and expectations about promotions.
- Having career ladders helps establish DS/RSE roles as a career instead of a random position with a research lab or group that may go away in the future.
- Career ladders increase visibility for DS/RSE roles within academic institutions and help establish these roles as part of the research infrastructure and ecosystem.
- They establish the set of roles that HR uses for compensation comparisons and evaluations.

A.3 Example DS/RSE Career Ladders

The levels of a career ladder, especially for staff positions, may be distinguished as Data Scientist I, Data Scientist II, Data Scientist III, or similar ordering schemes, but within technical fields, there are widely used titles that can help denote the level of the position and experience expected.

Individual Contributor Career Ladder: Not all DS/RSEs want to move into a management position. Having a strong Individual Contributor career ladder is critical to those seeking to stay in a contributor role long-term.

- Associate or Assistant Data Scientist/ Research Software Engineer
  - This is an entry-level position that may correspond to a first position for someone with no work experience and a Bachelor's degree.
  - Expect someone in this position to require significant mentoring and development.
  - Not all organizations have this level of position, especially smaller ones without the capacity to provide the necessary mentoring and supervision.
- Data Scientist/Research Software Engineer
  - This may be an entry-level position for someone with a graduate degree, and it is also appropriate for those with several years of RSE/DS experience.
  - Expect someone in this position to be able to perform assigned work independently, and identify and resolve problems, but not take on significant strategic, supervisory, or mentoring responsibilities.
- Senior Data Scientist/Research Software Engineer
  - This is a position for DS/RSEs with several years of work experience.
  - Expect someone in this position to independently handle more complex work and take on planning responsibilities within the scope of their own work. Those in this position are expected to begin mentoring others.
- Lead Data Scientist/Research Software Engineer
  - This is a position for experienced DS/RSEs.
  - Expect someone in this position to be able to work without significant direction and lead project teams. Leads should be aware of the broader context beyond their individual projects and help set strategy.
- Principal Data Scientist/Research Software Engineer
  - This position is for very experienced DS/RSEs who are experts in their area.
  - Expect someone in this position to provide significant mentorship to their team, lead project strategy, take responsibility for large and complex projects and engage with others beyond their individual projects and institution.
  - This level of position is currently rare at academic organizations given the developing status of the DS/RSE fields and groups, however, it is important for retaining and recognizing high-performing individual contributors who do not want to move into management positions.

Management Career Ladder: DS/RSEs typically move into management roles following senior or lead individual contributor positions.

- Manager, Data Science/Research Software Engineering
  - This position has official responsibilities for supervising staff, including hiring and firing, performance reviews, work assignments, and compensation/promotion decisions.
- Associate Director, Data Science/Research Software Engineering
  - In large organizations, this position may oversee multiple teams, each of which has a supervising lead or manager. In smaller organizations, this position may take on many of the responsibilities of a director, but for a smaller group of people.
  - This level of position may not be found in all organizations.
- Director, Data Science/Research Software Engineering
  - This position has responsibility for multiple teams or areas and typically supervises other managers. It sets the overall strategy and direction for the organization.
- Beyond Director
  - After the director level, roles typically gain responsibility for people and services outside of data science and research software engineering, such as covering research computing and data services broadly, or an entire data science institute.

A.4 Data Scientist and RSE Career Paths

All academic DS/RSEs have a career path – the series of positions they hold during their career – that does not necessarily neatly follow a career ladder. In part, this is because DS/RSE career ladders are new, and many academic institutions still do not have these pathways well-defined. However, it is also the result of the varied experiences that DS/RSEs need to be successful in their roles. DS/RSE roles and teams actively interact with domain researchers, and for this reason, many DS/RSEs find themselves traversing DS, RSE, and domain research (among other areas) career ladders through their careers.

However, to recruit and retain DS/RSE talent, organizations may want to develop distinct career pathways within the university setting for these careers. While academia may not be able to offer the rapid career progression offered in industry, it can develop structures that allow DS/RSE professions to understand what growth in their role means and have definable metrics to get to the next level.

A.4.1 Transparency

It is critical that you have transparent career ladders so that employees and candidates know what pathways are available to them and how they can move between career ladders for related roles as a part of their overall career path. These lateral moves from career ladder to career ladder should be normalized and administrative processes should facilitate, rather than hinder, such pathways. It is even important to show very immature DS/RSEs, such as undergraduates or graduate students, where their position sits on a career ladder to help illustrate the variety of options available for their careers.

It is common practice in academia to use levels of the career ladder as a way to provide appropriate compensation for candidates. This represents a miscalibration of the career ladder, which should be addressed with human resources. Such miscalibrations can result in limited career growth opportunities and, when combined with a lack of pathways to traverse career ladders, is likely to negatively impact the retention of quality employees. If you must use the career ladder to increase compensation for candidates, you should be prepared to openly discuss the potential impacts on career progression.

A.4.2 Entry Points and Exit Points

The boundary between academia and industry can be quite porous for DS/RSEs, and in tech more generally, though the direction of flow is more often from academia to industry. It is common knowledge in academic and industry circles that cross-pollination from one sector to another can have benefits for the individual and the institutions. With this in mind, you should not assume a single entry or exit point to or from academia, and certainly, you should not assume that these transitions only happen low on the career ladder. Cosden et al. (2022) discuss entry points for RSEs at length, and much of this discussion can also apply to DS positions.

When constructing career ladders and traversal points among ladders, also consider whether entry points from industry and exit points to industry are likely. Illustrating these entry and exit points as part of your career paths may help employees understand what opportunities they have ahead of them, including those outside of academia. Do note that moving out of academia does not necessarily mean one will not return to academia, but without clear entry and exit points, reintegration into academia may be hindered.

RESOURCES

Brady, H., Geiger, S., Culich, A., Das, D., Howard, T., Mazel-Cabasse, C., Fiore-Gartland, B., Noren, L., Cullens, C., Hench, C. (2019). Developing Desirable and Sustainable Career Paths for Academic Data Scientists. Slide Deck. Retrieved June 13, 2023, from **https://docs.google.com/presentation/d/1zJyGuqoYu_vIgsT2716xt4tPIg2N77g-sLcPj_e6uXM/edit#slide=id.g37ae830f00_9_98**

Geiger, R., Mazel-Cabasse, C., Cullens, C. Y., Norén, L., Fiore-Gartland, B., Das, D., & Brady, H. (2018, June 6). Career Paths and Prospects in Academic Data Science: Report of the Moore-Sloan Data Science Environments Survey. **https://doi.org/10.31235/osf.io/xe823**

Harvard University (n.d.). IT Research Computing Matrix. Harvard Human Resources. Retrieved June 15, 2023, from **https://hr.harvard.edu/sites/hwpi.harvard.edu/files/humanresources/files/it_research_computing_matrix.pdf**

Katz, D. S., McHenry, K., & Lee, J. S. (2021). Research software sustainability: Lessons learned at NCSA. In 54th Annual Hawaii International Conference on System Sciences, HICSS 2021 (pp. 7249-7256). IEEE Computer Society. **http://hdl.handle.net/10125/71494**

Katz, D. S., McHenry, K., Reinking, C., & Haines, R. (2019, May). Research software development & management in universities: case studies from Manchester's RSDS group, Illinois' NCSA, and Notre Dame's CRC. In 2019 IEEE/ACM 14th International Workshop on Software Engineering for Science (SE4Science) (pp. 17-24). IEEE. **https://doi.org/10.1109/SE4Science.2019.00009** (Preprint: arXiv 1903.00732)

Katz, D. S., McHenry, K., & Lee, J. S. (2021). Senior-level RSE career paths (with an s). Daniel S. Katz's blog. Retrieved June 13, 2023, from **https://danielskatzblog.wordpress.com/2021/09/27/senior-rse-paths/**

Mazumdar, M., Messinger, S., Finkelstein, D. M., Goldberg, J. D., Lindsell, C. J., Morton, S. C., ... & Parker, R. A. (2015). Evaluating academic scientists collaborating in team-based research: A proposed framework. Academic medicine: journal of the Association of American Medical Colleges, 90(10), 1302.

RCD Professionalization Working Group. (2018). Research Computing and Data Professionals Job Elements and Career Guide. Zenodo. **https://doi.org/10.5281/zenodo.7108947**

University of Arizona (n.d.). Career Streams. University of Arizona Human Resources. Retrieved June 15, 2023, from **https://hr.arizona.edu/content/career-streams**

Appendix B - Position Description, MOU, and Buyback Program Examples

Introduction

Below are examples of position descriptions, memoranda of understanding for DS/RSEs who are co-appointed with another department, and a "buyback program" contract for DS/RSEs who wish to pursue their own research. These examples were provided by contributors to this guidebook, and are presented in their original form (though anonymized). The variability in structure and content gives a sense of how different these documents can look across institutions, though we consider these examples to be an excellent foundation.

Data Scientist Position Descriptions

Example Set 1

Context: Private R1 university where the position is part of a research computing team funded primarily through central (hard) funding. This is a staff position without an independent research agenda. The job description and job postings are the same.

Data Scientist

As a Data Scientist with [team name], you will support world-class research from across the university by implementing practical solutions to researchers' data science challenges. In this position, you will:

- Consult with and provide advice to researchers facing a wide range of data challenges,
- Develop and deliver training for students, faculty, and staff on data science and computational skills, and
- Collaborate with researchers on projects requiring data science or data visualization skills.

Working with other members of the [team] and partners across the university, you will provide the data services, resources, and support researchers need. [Broader team] consists of specialists who collaborate to support the full scope of computational and data needs of researchers. We facilitate data-enabled research and training across disciplines and fields. We value continual learning, problem-solving, and sharing our diverse backgrounds and skills with each other and researchers.

Responsibilities:

- **Provide data science support**: Support researchers by collecting and organizing data; writing scripts and software programs; analyzing, modeling, and visualizing data; recommending predictive and statistical analysis approaches; and providing other data science services as a part of collaborative research projects.
- **Deliver data science training**: Develop and teach technical workshops on data science, programming, and data visualization topics for researchers.
- **Develop and communicate analysis strategies**: Provide short-term consultation support to researchers at all levels: assess researchers' data and computational challenges or problems and develop strategies to address those problems. Communicate the steps necessary to accomplish a data-centered task to researchers of varying levels of technical background and expertise.
- **Support service operations**: Perform organizational, planning, and logistical tasks required to ensure the efficient operation of data science services, such as organizing workshops and ensuring consultation requests are fulfilled.

- **Student mentorship**: Mentor undergraduate and graduate student data science consultants working with the team.
- **Documentation**: Write both technical and non-technical documentation of projects, code, and consultations.
- **Team collaboration**: Work together with [team] colleagues, school/college partners, other IT groups, the library, and other university data science groups to provide integrated support for the research computing and data needs of [UNIVERSITY] researchers.

Required knowledge, skills, and abilities:

- Successful completion of a full 4-year course of study in an accredited college or university leading to a Bachelor's degree in a related field; or appropriate combination of education and experience.
- 4 or more years combined work experience and/or post-baccalaureate education in a related field, including experience conducting research projects involving statistical and/or predictive data analysis outside of coursework or supporting such projects as part of a team.
- Knowledge of Python and/or R at a sufficient level to fulfill job responsibilities, including teaching others.
- Demonstrated experience with two or more additional computer programs, languages, or skills beyond Python and R, such as, but not limited to: Stata, SAS, MATLAB, JavaScript, SQL, git, bash, high-performance computing (HPC), machine learning, bioinformatics, advanced data visualization, computer vision/image processing, or text analysis.
- Experience using a range of techniques and technologies to clean, analyze, and model complex or messy data in a variety of formats.
- Demonstrated interpersonal and communication skills, with the ability to build relationships and effectively communicate with a wide range of people in the [UNIVERSITY] community.

[UNIVERSITY] is an Equal Opportunity, Affirmative Action Employer of all protected classes, including veterans and individuals with disabilities. Consistent with its obligations under the law, [UNIVERSITY] will provide reasonable accommodation to any employee with a disability who requires accommodation to perform essential job responsibilities.

Senior Data Scientist

As a Senior Data Scientist with [team name], you will support world-class research from across the university by implementing practical solutions to researchers' data science challenges. In this position, you will:

- Consult with and provide advice to researchers facing a wide range of data challenges,
- Develop and deliver training for students, faculty, and staff on data science and computational skills, and
- Collaborate with researchers on projects requiring data science or data visualization skills.
- Documentation: Write both technical and non-technical documentation of projects, code, and consultations.
- Team collaboration: Work together with [team] colleagues, school/college partners, other IT groups, the library, and other university data science groups to provide integrated support for the research computing and data needs of [UNIVERSITY] researchers.

[Same additional introductory context as above.]

Responsibilities:

- **Provide data science support**: With minimal guidance, assess researchers' data challenges or problems, and develop and communicate strategies to address those problems. Lead small to mid-size researcher support projects involving collecting and organizing data, writing scripts and software programs, analyzing and visualizing data, or other data science services.
- **Deliver data science training**: Develop and teach technical workshops on data science, programming, and data visualization topics for researchers.
- **Support service operations and development**: Independently perform organizational, planning, and logistical tasks required to ensure the efficient operation of data science services, such as organizing workshops and ensuring consultation requests are fulfilled. Support the development of service improvements and new services.
- **Mentorship**: Mentor and provide guidance to undergraduate and graduate student data science consultants working with the team. Mentor less experienced colleagues.
- **Engage with the data science community**: Foster relationships with researchers and schools/colleges to understand new research trends, identify skill gaps and data science needs, and raise awareness of services. Engage with data science-relevant communities within and outside of [UNIVERSITY]. Represent [team] in meetings within and outside of [UNIVERSITY].
- **Team collaboration**: Work together with [team] colleagues, school/college partners, other IT groups, the library, and other university data science groups to provide integrated support for the research computing and data needs of [UNIVERSITY] researchers.

Required knowledge, skills, and abilities:

- Successful completion of a full 4-year course of study in an accredited college or university leading to a Bachelor's degree in a related field; or appropriate combination of education and experience.
- 6 or more years combined work experience and/or post-baccalaureate education in a related field, including experience conducting research projects involving statistical and/or predictive data analysis outside of coursework or supporting such projects as part of a team.
- Knowledge of Python and/or R at a sufficient level to fulfill job responsibilities, including teaching others.
- Demonstrated experience with two or more additional computer programs, languages, or skills beyond Python and R, such as, but not limited to: Stata, SAS, MATLAB, JavaScript, SQL, git, bash, high-performance computing (HPC), machine learning, bioinformatics, advanced data visualization, computer vision/image processing, or text analysis.
- Experience using a range of techniques and technologies to clean, analyze, and model complex or messy data in a variety of formats.
- Ability to scope and manage projects by gathering requirements, defining the problem, choosing appropriate approaches and resources, estimating necessary resources, effectively communicating with collaborators, and prioritizing work on multiple projects.
- Demonstrated customer service excellence.
- Demonstrated interpersonal and communication skills, with the ability to build relationships and effectively communicate with a wide range of people in the [UNIVERSITY] community.
- *[Same equal opportunity statement.]*

Lead Data Scientist

As a Lead Data Scientist with [team], you will support world-class research from across the university by implementing practical solutions to researchers' data science challenges. In this position, you will:

- Lead the research data science consultation service helping researchers with a wide range of data challenges,
- Coordinate the workshop and training program for students, faculty, and staff on data science and visualization skills,
- Collaborate with researchers on projects requiring data science and visualization expertise, and
- Supervise staff and student data science and visualization consultants.

By actively developing relationships with the research community, you will identify and lead the provision of the data science services, resources, and support researchers need. By partnering with other university groups, you will foster the development of the data science community at [UNIVERSITY].

[Team] consists of specialists who collaborate to support the full scope of computational and data needs of researchers. We facilitate data-enabled research and training across disciplines and fields. We value continual learning, problem-solving, and sharing our diverse backgrounds and skills with each other and researchers.

Responsibilities:

- **Lead service operations and development**: Coordinate data science and visualization service operations and collaborate with other service leads to solve researchers' technical challenges across service areas including computing, data storage, and data workflows. Lead projects developing new services, improving existing services, and identifying and implementing process improvements. Support efforts to measure and report on service impact and usage.
- **Provide data science support**: Independently assess, and provide guidance to staff, on researchers' data challenges or problems. Develop and communicate strategies to address those problems. Lead cross-team and large researcher support projects involving collecting and organizing data, writing scripts and software programs, analyzing and visualizing data, or other data science services. Delegate work to team members as appropriate.
- **Develop and deliver data science training**: Develop and teach workshops on data science and data visualization topics to build faculty and students' data science skills.
- **Supervise staff**: Supervise other data science and visualization staff, including assigning and supervising work and providing technical guidance and subject matter expertise for consultations and projects. Provide input for performance management, hiring, evaluation, promotion, and termination of employees.
- **Lead student program**: Select, supervise, and mentor undergraduate and graduate student data science consultants working with the team.
- **Engage with the research community**: Foster relationships with researchers and schools/colleges to understand new research trends, identify skill gaps and data science needs, and raise awareness of services.
- **Team collaboration**: Work together with Research Computing Services colleagues, school/college partners, other IT groups, the library, and other university data science groups to provide integrated support for the

research computing and data needs of Northwestern researchers.
- Professional development: Continue professional development in areas necessary to perform duties related to the job. Engage with data science-relevant communities within and outside of Northwestern.

Required knowledge, skills, and abilities:

- A Bachelor's degree in a related field; a master's degree or Ph.D. is preferred.
- 7 or more years combined work experience and/or post-baccalaureate education in a related field, including experience leading data science projects in a research environment.
- Demonstrated experience with Python and/or R, as well as two or more additional computer programs, languages, or skills such as, but not limited to Stata, SAS, MATLAB, JavaScript, SQL, git, bash, high-performance computing (HPC), machine learning, bioinformatics, advanced data visualization, computer vision/image processing, or text analysis.
- Experience supervising or mentoring students or staff.
- Ability to act on own initiative to further organizational and University goals.
- Experience scoping and managing projects by gathering requirements, defining the problem, choosing appropriate approaches and resources, estimating necessary resources, effectively communicating with collaborators, and prioritizing work on multiple projects.
- Demonstrated customer service excellence.
- Demonstrated interpersonal and communication skills, with the ability to build relationships and effectively communicate with a wide range of people in the Northwestern community.
- *[Same equal opportunity statement.]*

Example 2

Context: Public R1 university with a long-standing data science institute.

Data Scientist

The [INSTITUTE] has an outstanding opportunity for a researcher with strengths in data-intensive methods to join our team as a Data Scientist. As a member of our core research team, you will participate in expanding campus-wide opportunities at the intersection of data science and disciplinary domains through collaborative research, consultation, community building, and educational activities. You will also have the opportunity to explore your own independent research agenda to address questions in your chosen field through the use of data, data-intensive methods, and computation. The [INSTITUTE] empowers researchers and students in all fields across the [UNIVERSITY] to answer fundamental and transformative questions. As the hub of data-intensive discovery on campus, we lead a community of innovators in the techniques, technologies, and best practices of data science.

Data Scientist roles in the [INSTITUTE] are full-time, permanent (not term-limited), appointments where you can build a career with advancement opportunities and a rich set of professional development and mentorship opportunities.

The values of diversity, equity, and inclusion are integral to the success of our research enterprise and are embedded in the culture of who we are as an institution and employer. The [INSTITUTE] shares [University's] commitment to combat inequities and racism in higher education and research.

Women, people of color, individuals with disabilities, and veterans are highly encouraged to apply.

Position Purpose: Data Scientists in the [INSTITUTE] provide guidance and expertise to the [UNIVERSITY] community, including students, post-docs, staff, and faculty on the best practices for, and appropriate application of, data science methods. In addition, they have the opportunity to build their own independent research agenda. These interactions are often structured around a set of programmatic and thematic activities that are core to the [INSTITUTE] mission, which include: [list of programs the Data Scientist may be involved in]

Position Complexities: This position requires an individual who is interested in continuous learning opportunities in a higher education research environment. The ideal candidate will have substantive training or experience in one or more areas related to data science, demonstrated experience in using and building open-source software, and, ideally, demonstrated ability to use these skills to solve problems in an applied domain discipline. The [INSTITUTE] is a highly interdisciplinary environment and our work involves communicating and collaborating across varied disciplinary, theoretical, and methodological traditions. You will be joining a team that is respected across campus for their abilities to work with other researchers in both support and leadership roles. The [INSTITUTE] highly values open science. This includes open-source software development, engaging in open data practices, and using best practices for research reproducibility.

Position Dimensions and Impact on the University: As a member of the team of 15+ data scientists and researchers in the [INSTITUTE], you will have the opportunity to have a substantial impact on the development of data science research at the [UNIVERSITY] and beyond. The duties and responsibilities that impact the [UNIVERSITY] include working with students, post-docs, staff, and faculty to help them understand, adopt and engage in data science in a variety of informal and formal education, research, and community-building settings and activities. In addition, data scientists in the [INSTITUTE] build and manage their own research programs.

CORE RESPONSIBILITIES:

Programmatic Support (40%)

- Participate in core programs of the Institute, such as [listing of core programs]
- Hold open office hours for students, faculty, staff, and post-docs

Organizational Service (10%)

- Participate in core team meetings and activities
- Serve on standing and/or ad hoc teams such as hiring committees, equity teams, etc.
- Foster and build research collaborations with other researchers at [UNIVERSITY]

Independent research (50%)

- Develop a research program rooted in applications of data-intensive and data science methods to an applied domain discipline

JOB REQUIREMENTS:

We are looking for applicants who demonstrate the skills listed below through a combination of their work, education, and life experience as demonstrated in their application materials.

Thinking you don't qualify? Studies have shown that women and people of color are less likely to apply for jobs unless they meet every qualification. If you are excited about this job but your experience doesn't perfectly check every box in the job description, we encourage you to apply anyway.

MINIMUM REQUIREMENTS:

- Ph.D. or MS with strong computational background, regardless of specific discipline
- Two years of experience planning, implementing, and assessing the efficacy of data science techniques for studying research questions, these years of experience may derive from graduate research
- Commitment to working with open science methodologies and reproducibility

Equivalent education/experience will substitute for all minimum qualifications.

DESIRED QUALIFICATIONS:

- Demonstrated commitment to valuing diversity and contributing to an inclusive working and learning environment
- Previous experience using R or Python
- Previous experience working with open-source software workflows
- Previous experience working with researchers from a range of disciplinary backgrounds
- Demonstrated experience in software engineering and/or advanced statistical methods

Equivalent education/experience can substitute for all minimum qualifications except when there are legal requirements, such as a license/certification/registration.

This position is open to anyone authorized to work in the US. The [UNIVERSITY] is not able to sponsor visas for staff positions.

Research Software Engineer Position Descriptions

Example Set 1

Context: Public R1 university with a large RSE group supported through grant funding. The group maintains permanently open positions listed at multiple levels in the same listing.

Shared Introduction

[Team] is currently seeking one or more Associate Research Software Engineer/ Research Software Engineer/ and/or Senior Research Software Engineer (RSEs), to provide scientific software development on a long-term basis. In this role, you will work on multiple research projects based on your skill set and the projects' needs. You will interact with stakeholders from a wide range of scholarly disciplines at the [UNIVERSITY] and other universities, as well as government and industry. Domains may include geoscience, civil and environmental engineering, biology, materials science, agriculture, medicine, astronomy, and the humanities. More information on software at [team], including specific projects, is available on our website.

Successful candidates can come from industry or academia and can come from computer science or a domain science with a background in development. The key to a successful candidate will be a willingness to embrace both the professional software development aspects as well as the academic aims, priorities, and research aspects of the position. Our team strives to mentor new RSEs and provide opportunities for training and learning on the job, and successful candidates will identify and support emerging RSEs through their career paths.

In terms of technical experience, we seek talented professionals that have designed, implemented, and tested software systems and architectures across the stack. While ideally, your experience includes research software development, [team] provides support and expertise to incoming staff thanks to its longstanding trajectory of success in scientific computing.

Some examples of the work you might do include:

- Translation of research into specialized scientific software
- Architecture design and implementation of new software systems serving multiple scientific and community stakeholders
- Devising methods for data ingestion and integration from heterogeneous sources
- Data analysis
- General software development
- Web front end
- Visualization
- Back end services
- Infrastructure and DevOps

Creativity, independence, a strong work ethic, as well as coordination with the project principal investigator(s) will be critical for this role. More senior roles may represent the group at meetings, give presentations at conferences or other venues, and contribute to publications and grant proposals.

Appointment Information: This is a 100% full-time Academic Professional position, appointed on a 12-month basis. The expected start date is as soon as possible after position closing. Salary is commensurate with experience.

Associate Research Software Engineer

Major Duties and Responsibilities:

- **Research Software and Application Development:**
  - Develop software that addresses real-world complex research problems with interfaces to powerful backend tools and interdisciplinary analyses such as science, engineering, the humanities, and social sciences.
  - Architect and design software programs to help researchers define and solve heterogeneous problems; provide consulting expertise to researchers across a variety of domain sciences to leverage synergies across projects; and design and support shared cyberinfrastructure to support a broad range of communities.
  - Integrate collaborative input and design new approaches and techniques to address project-specific needs and critical technological challenges.
  - Track research activity in relevant fields both programmatic and domain specific.
  - Evaluate the strengths and weaknesses between varieties of novel approaches to research and scientific software development problems.
- **Research Communication:**
  - Contribute to publications, technical reports, and documentation in collaboration with domain scientists.
  - Communicate findings from research development activities above with project colleagues, including primary investigators, to be used as needed in current programming and design efforts.

- **System Architecture and Design:**
  - Develop research-oriented code which supports an environment of secure systems, and design and develop security-oriented programs using established software development methodologies (i.e. agile, lean, rapid application, etc.) which by nature defend against disruption or misdirection of services.

Required Education and Experience:

- Bachelor's degree required, preferably in computer science or related field. Alternative degree fields will be considered if accompanied by equivalent experience (depending on nature and depth of experience as it relates to current project needs).
- Software development and/or programming experience, which can include a combination of course-level and professional experience

Preferred Experience:

- Programming in three or more of the following: C/C++, Java, PHP, Python, Scala, Ruby, Javascript, or other similar languages, with an ability to explain/decide why one would be utilized over another in different development scenarios.
- Subject matter expert on one or more technologies.
- Experience in one or (preferably) more of the following:
  - Web front-end development (e.g. server-side scripting, client-side frameworks, HTML5, CSS, REST, OAuth).
  - Machine learning or data mining.
  - Natural language processing.
  - Geospatial data management and programming.
  - High-speed, real-time data analysis.
  - Computer vision or graphics.
  - HPC environments, cloud computing, and/or systems administration.
  - Databases (e.g. MySQL, MongoDB, PostgreSQL).
  - Linux DevOps experience.
  - Ability to establish a software development effort from the ground up (create software from scratch).
- Experience working with domain sciences.
- Contributions towards research publications.
- Ability to clearly communicate results and their importance (verbally and in writing).
- Ability to provide input for reports, presentations, and grant proposals.

Research Software Engineer

Major Duties and Responsibilities:

- **Research Software and Application Development:**
  - Engineer and develop software that addresses real-world complex research problems with interfaces to powerful backend tools and interdisciplinary analyses such as science, engineering, the humanities, and social sciences.
  - Architect and design software programs to help researchers define and solve heterogeneous problems; provide consulting expertise to researchers across a variety of domain sciences to leverage synergies across projects; and design and support shared cyberinfrastructure to support a broad range of communities.
  - Evaluate and integrate collaborative input and design new approaches and techniques to address project-specific needs and critical technological challenges.

    - Track research activity in relevant fields both programmatic and domain-specific.
    - Evaluate the strengths and weaknesses between varieties of novel approaches to research and scientific software development problems.
- **Research Communication:**
    - Publish and present results of scientific software research and development at national and international conferences.
    - Contribute to publications, technical reports, and documentation in collaboration with domain scientists.
    - Communicate findings from research development activities above with project colleagues, including primary investigators, to be used as needed in current programming and design efforts.
- **System Architecture and Design:**
    - Develop research-oriented code which supports an environment of secure systems, and design and develop security-oriented programs using established software development methodologies (i.e. agile, lean, rapid application, etc.) which by nature defend against disruption or misdirection of services.

Required Education and Experience:

- Bachelor's degree required, preferably in computer science or related field. Alternative degree fields will be considered if accompanied by equivalent experience (depending on nature and depth of experience as it relates to current project needs).
- At least 3 years of experience in relevant software development.

Preferred Experience:

- Programming in three or more of the following: C/C++, Java, PHP, Python, Scala, Ruby, Javascript, or other similar languages, with an ability to explain/decide why one would be utilized over another in different development scenarios.
- Subject matter expert on one or more technologies.
- Experience in one or (preferably) more of the following:
    - Web front-end development (e.g. server-side scripting, client-side frameworks, HTML5, CSS, REST, OAuth).
    - Machine learning or data mining.
    - Natural language processing.
    - Geospatial data management and programming.
    - High-speed, real-time data analysis.
    - Computer vision or graphics.
    - HPC environments, cloud computing, and/or systems administration.
    - Databases (e.g. MySQL, MongoDB, PostgreSQL).
    - Linux DevOps experience.
    - Ability to establish a software development effort from the ground up (create software from scratch).
- Experience working with domain sciences.
- Contributions towards research publications.
- Ability to clearly communicate results and their importance (verbally and in writing).
- Ability to provide input for reports, presentations, and grant proposals.

Senior Research Software Engineer

Major Duties and Responsibilities:

- **System Architecture and Design:**
  - Architect and design software programs to help researchers define and solve heterogeneous problems; provide consulting expertise to researchers across a variety of domain sciences to leverage synergies across projects; and design and support shared cyberinfrastructure to support a broad range of communities.
  - Design and build the infrastructure necessary to support cutting-edge research projects. Engineer the underlying architecture of the project.
  - Develop research-oriented code which supports an environment of secure systems, and design and develop security-oriented programs using established software development methodologies (i.e. agile, lean, rapid application, etc.) which by nature defend against disruption or misdirection of services.
- **Research Software and Application Development:**
  - Engineer and develop software that addresses real-world complex research problems with interfaces to powerful backend tools and interdisciplinary analyses such as science, engineering, the humanities, and social sciences.
  - Drive research software development projects. This involves guaranteeing milestones are achieved, coordinating efforts among project members, recognizing emergent project needs, and taking the initiative to ensure these are addressed and resolved.
  - Evaluate and integrate collaborative input and design new approaches and techniques to address project specific needs and critical technological challenges.
  - Track research activity in relevant fields both programmatic and domain specific.
  - Evaluate the strengths and weaknesses between varieties of novel approaches to research and scientific software development problems.
  - May direct and train hourly employees and interns.
- **Research Communication:**
  - Publish and present results of scientific software research and development at national and international conferences.
  - Contribute to publications, technical reports, and documentation in collaboration with domain scientists.
  - Communicate findings from research development activities above with project colleagues, including primary investigators, to be used as needed in current programming and design efforts.

Required Education and Experience:

- Bachelor's degree required, preferably in computer science or related field. Alternative degree fields will be considered if accompanied by equivalent experience (depending on nature and depth of experience as it relates to current project needs).
- At least 5 years software development and/or programming experience, which can include a combination of course-level and professional experience.

Preferred Experience:

- Programming in three or more of the following: C/C++, Java, PHP, Python, Scala, Ruby, Javascript, or other similar languages, with an ability to explain/decide why one would be utilized over another in different development scenarios.
- Subject matter expert on one or more technologies.
- Experience in one or (preferably) more of the following:
  - Web front-end development (e.g. server-side scripting, client-side frameworks, HTML5, CSS, REST, OAuth).
  - Machine learning or data mining.
  - Natural language processing.
  - Geospatial data management and programming.
  - High-speed, real-time data analysis.
  - Computer vision or graphics.
  - HPC environments, cloud computing, and/or systems administration.
  - Databases (e.g. MySQL, MongoDB, PostgreSQL).
  - Linux DevOps experience.
- Ability to establish a software development effort from the ground up (create software from scratch).
- Experience working with domain sciences.
- Contributions towards research publications.
- Ability to clearly communicate results and their importance (verbally and in writing).
- Ability to provide input for reports, presentations, and grant proposals.

Example Set 2

Context: Private R1 university with a large RSE group that places RSEs with specific research teams or centers. Job postings, which are customized to each opening, are not the same as the underlying job description, which is general.

Research Software Engineer (Social Science)

Do you have a background in computational research and love to write code? Do you want to help enable and advance truly pioneering computational social science research? If so, the [INSTITUTE] at [UNIVERSITY] is recruiting a Research Software Engineer (RSE) to join our growing team.

The [INSTITUTE] is a new initiative that supports research at the technical forefront of quantitative inquiry in the social sciences. Because [INSTITUTE] funds projects that span the social science disciplines, this position will require forward thinking and versatility in computational research expertise. Project requirements can vary from advanced applications of machine learning, the development of scalable and efficient research code, the creation of tailored software or APIs, the construction of queryable databases within a secure environment, and the versatile usage of algorithms to meet additional computational needs. Candidates must demonstrate the ability to help build a repository of advanced programming techniques that supports groundbreaking research at [UNIVERSITY].

Professional development opportunities include biweekly meetings with the [university RSE Group]. The RSE Group is committed to building collaborative environments in which the best software engineering practices are valued, and sharing and applying cross-disciplinary computational techniques in new and emerging areas.

If you have a strong background in scientific programming or high-performance computing and academic research, along with a penchant for forward-thinking and expanding the boundaries of computational social science, you have the right skill set to make an immediate impact on multiple high-profile research projects. You will be poised to grow and expand your programming and data analytics expertise through an evolving set of research problems.

The Research Software Engineer will be an integral team member, reporting to the [INSTITUTE] Executive Director and working closely with the Senior Research Specialist. They will provide technical support for a variety of computational social science research projects for faculty, student/ postdoctoral researchers, and technical staff to enable and accelerate their research computing efforts.

Responsibilities:

- Efficiently translate research objectives into software solutions that contribute to ongoing research project(s)
- Develop an understanding of complex computational research questions at a level sufficient to converse with [UNIVERSITY]'s world-class researchers.
- Enable long-term maintainability of solutions by documenting projects in a descriptive and detailed manner
- Apply modern software development best practices (e.g., version control, continuous integration, and continuous delivery)
- Provide technical expertise and guidance for improving the performance and quality of new and existing code bases

- Regularly meet with, listen to, and ask questions of researchers to ensure that engineered solutions fit research needs
- Parallelize, debug, port, and tune existing research code to meet criteria-determined research objectives
- Develop software tools that allow researchers to analyze large, complex data sets

Qualifications:

- 5+ years of strong Python or R programming skills in a work environment or academic/research setting
- Knowledge of a high-performance language (e.g., C++, C, Julia)
- Experience working with large, complex data sets
- Basic working knowledge of version control (i.e., Git)
- Ability to create professional software with clear documentation
- Ability to learn new programming languages and frameworks
- Ability to translate research needs into technical requirements
- Ability to communicate clearly with technical and non-technical colleagues
- Ability to work independently
- A graduate degree in computational social science, computer science, mathematics, engineering, or related field. Preferred Qualifications
- Experience with distributed computing frameworks (e.g., Spark, Dask)
- Experience with high-performance computing (HPC) or public clusters (e.g., AWS)
- Experience developing open-source research software
- Experience with machine learning frameworks (e.g., TensorFlow, PyTorch)
- Proficiency in database language and tools (e.g., Postgres, MongoDB)
- Experience writing shell scripts for process automation
- Background in quantitative social science

[UNIVERSITY] is an Equal Opportunity/ Affirmative Action Employer and all qualified applicants will receive consideration for employment without regard to age, race, color, religion, sex, sexual orientation, gender identity or expression, national origin, disability status, protected veteran status, or any other characteristic protected by law. EEO IS THE LAW.

Research Software Engineer II (Computational Biology/Genomics)

Do you have a background in computational research and love to write code? Do you want to help enable and advance groundbreaking computational research? If so, [UNIVERSITY]'s Research Computing department is recruiting a Research Software Engineer to join the fast-growing Research Software Engineering (RSE) Group.

In the RSE Group, we collectively provide computational research expertise to multiple divisions within the University. As a central team of software experts, we are focused on improving the quality, performance, and sustainability of [UNIVERSITY]'s computational research software. Our group is committed to building collaborative environments in which the best software engineering practices are valued, and to sharing and applying cross-disciplinary computational techniques in new and emerging areas.

In this position, you will be an integral member of multiple research teams focused on cutting-edge computational biology. You will join teams of researchers associated with a new [computational biology center] and [genomics institute], where you will contribute to the development of efficient and scalable research code by providing computational expertise in software development, algorithm selection, and optimization. Research Software Engineers II

work closely with a team of researchers and Research Software Engineers to leverage their communication and problem-solving skills to build complete software solutions crucial to the advancement of research.

If you have a strong background in scientific programming, high-performance computing, academic research, and an interest in computational biology and genomics, you have the right skill set to make an immediate impact on multiple high-profile research projects. You will be poised to grow and expand your programming and data analytics expertise into a dynamic new set of research problems.

This position will require you to work closely with colleagues in [IT] as well as with faculty, student/postdoctoral researchers, and technical staff in [genomics institute] to enable and accelerate their research computing efforts. This role functions within a dynamic, supportive team environ-ment that permits diverse backgrounds to thrive, including those wanting to make a career change and those with non-traditional career tracks, educational paths, or life experiences. If this environment sounds like a strong match or even an exciting challenge, we encourage you to apply and use your cover letter to explain why you would be a good fit for the role.

Responsibilities:

- Lead and/or co-lead the design and development of complex research software for computational biology and genomics.
- Fully understands the role within the research domain and working towards advanced proficiency in the underlying science, math, statistics, data analysis, and algorithms of computational research questions at a level sufficient to converse with Princeton's world-class researchers to support the ongoing work. This will consist of independent research (reading publications etc.) and/or studying existing code bases.
- Working independently or in collaboration with a team, initiate and/or maintain open collaboration with researchers. Regularly meet with, listen to, and ask questions of researchers to ensure that engineered solutions fit the research need. Understand and address software engineering questions that arise in research planning.
- Apply appropriate domain-specific algorithms, techniques, and code to advance software engineering in the research field.
- Working independently with minimal guidance to understand and translate research priorities into flexible software solutions
- Independently or in collaboration with a team, use researcher-provided requirements and desired end state to build complete software solutions. To achieve this, RSEs are expected to figure out the problem through independent or team research, build complete software solutions, and provide full documentation for usage by the research team.
- Identify solutions for each project, establish a set of applicable best practices uniquely appropriate for that project (e.g., version control, continuous integration, continuous delivery, software design, programming model, etc.), and enable long-term maintainability and sustainability by documenting the projects in a descriptive and appropriately detailed manner. Independently or in collaboration with a team, provide technical expertise and guidance for improving the performance and quality of new and existing code bases through hands-on work with ongoing research.

- Responding to evolving research needs, apply research software engineering experience to develop robust software solutions to solve challenging research problems. Port, debug, tune, and potentially parallelize existing research code to meet criteria set by the research needs.
- Develop software tools that allow researchers to interact in flexible ways with extremely large data sets.
- Develop software tools that allow researchers to interact in flexible ways with extremely large data sets.
- Independently or in collaboration with a team, develops scope and project management plans, meets milestone delivery timelines, and communicates with the research team. Communicate software engineering concepts with project teams consisting of domain experts with varying degrees of software engineering knowledge.
- Actively expanding technical skill set and expertise to include software development tools and techniques, software engineering best practices, programming languages, high-performance computing hardware, and computational research solutions.

Qualifications

Essential Qualifications

- 4+ years' experience as a Research Software Engineer or equivalent experience (e.g., graduate school, industry experience, open-source software development, etc.).
- Exhibits programming skills, particularly in Python and C/C++ (and experience with the R programming language is a plus).
- Demonstrated success:
- Consistently using conventional and readable coding style.
- Creating comprehensive and well-written documentation.
- Using version control systems.
- Demonstrated success contributing to a collaborative research team.
    - Ability to work independently.
    - Ability to learn new programming languages and technologies beyond area of core knowledge.
    - Ability to communicate effectively with a diverse user base having varied levels of technical proficiencies.
- Experience working in an academic research environment.
- Education: A Bachelor's degree in computer science, engineering, sciences, or related computational field required. A Master's/Ph.D. in a relevant field with a strong computational focus or equivalent experience in a research setting preferred.

Preferred Qualifications

- Experience tuning and optimizing research software and algorithms.
- Experience developing research software outside of core domain knowledge.
- Academic research experience.
- Background in computational biology, genomics, or a related domain is helpful, but not required.
- Experience using HPC systems and job schedulers (e.g., Slurm).

- Experience with standard bioinformatics tools (e.g., SAMtools, bedtools, BWA, FastQC, Picard tools).
- Experience writing and using workflow management systems (e.g., Snakemake, Cromwell).
- Experience with cloud computing systems (e.g., Terra).
- Familiarity with GATK and best practice workflows.
- Knowledge of modern python tooling (pytest, nox, mypy, Flake8).
- Experience with containers and virtual environments for development and deployment.

[Standard university equal opportunity statement.]

Research Software Engineer II (High Energy Physics)

Do you want to help enable and advance groundbreaking High Energy Physics (HEP) research through software development? Do you have a background in research and love to write code? Are you looking to apply your scientific programming expertise to a novel set of research problems? If so, [UNIVERSITY]'s Research Computing department is recruiting a Research Software Engineer to join the fast-growing Research Software Engineering (RSE) Group.

In the RSE Group, we collectively provide computational research expertise to multiple divisions within the University. As a central team of software experts, we focus on improving the quality, performance, and sustainability of [UNIVERSITY]'s computational research software. Our group is committed to building collaborative and inclusive environments in which the best software engineering practices are valued, and to sharing and applying cross-disciplinary computational techniques in new and emerging areas.

In this position, you will be an integral member of multiple projects associated with the [physics institute], a software institute funded by the National Science Foundation. It aims to develop the state-of-the-art software cyberinfrastructure required for the challenges of data-intensive scientific research at the [physics center], and other planned experiments of the 2020s. You will collaborate with researchers to provide computational expertise in algorithm development and software design to create high-quality and sustainable research code.

If you have a strong background in scientific programming, academic research, and an interest in High Energy Physics you have the right skill set to make an immediate impact on multiple high-profile research projects.

You'll have the opportunity to regularly work closely with colleagues associated with [physics center], [IT], as well as with faculty, student/postdoctoral researchers, and technical staff in [physics department] to enable and accelerate research software development.

For additional information, please contact [faculty PIs].

Responsibilities:

- Lead the design and construction of increasingly complex research software systems needed to acquire, process, and analyze data from HEP experiments.
- Establish a set of applicable best practices uniquely appropriate for each project (e.g. version control, continuous integration, continuous delivery, software design, programming model, etc.), and enable long-term maintainability and sustainability by documenting the projects in a descriptive and appropriately detailed manner.

- Respond to evolving research needs and apply research software engineering experience to develop robust software solutions to solve challenging research problems. Port, debug, tune, and potentially parallelize existing research code to meet criteria set by the research needs.
- Transfer knowledge, expertise, and methodologies by providing technical assistance and mentorship to graduate students and postdoctoral researchers.
- Independently or in collaboration with a team, develop scope and project management plans, meet milestone delivery timelines, and communicate with the research team. Communicate software engineering concepts with project teams consisting of domain experts with varying degrees of software engineering knowledge.
- Apply appropriate domain-specific algorithms, techniques, and code to advance software engineering in HEP.
- Initiate and/or maintain open collaboration with the rest of the IRIS-HEP team and other HEP researchers as needed. Regularly meet with, listen to, and ask questions of the relevant collaborators for active projects. Understand and address software engineering questions that arise in research planning.
- Through independent research, (reading publications, etc.) and/or studying existing code bases understand the necessary elements of High Energy Physics to converse with HEP researchers
- Maintain and expand knowledge of current and future software development tools and techniques, programming languages, and computing hardware.

Qualifications

Essential Qualifications

- A minimum of 4 years as a Research Software Engineer or equivalent experience (e.g., graduate school, industry experience, open-source software development, etc.)
- Strong programming skills, particularly in languages used in High Energy Physics applications (e.g., Python and C++)
- Demonstrated success:
  - Consistently using conventional and readable coding style.
  - Creating comprehensive and well-written documentation.
  - Developing and maintaining reproducible build systems.
  - Using version control systems.
- Demonstrated successes working in a collaborative software development environment ("team science", often geographically distributed) as well as independently.
- Ability to learn new concepts and technologies beyond the area of core knowledge.
- Ability to communicate effectively with a diverse user base with varied technical proficiencies.
- Experience working in an academic research environment.
- Education: A bachelor's degree in computer science, engineering, physics, or a related computational field is required.

Preferred Qualifications

- Experience developing scientific software with significant emphasis on performance
- GPU programming experience (e.g., CUDA, OpenCL) and/or experience with FPGAs
- Parallel programming expertise (e.g., OpenMP, Kokkos, etc.)
- Experience developing research software outside of core domain knowledge

- Academic research experience.
- Background in high energy/nuclear/astroparticle physics, computer science, engineering, or related fields.
- A Master's/Ph.D. in physics, computer science, or other applied domain sciences with a strong computational focus is strongly preferred.

[Standard university equal opportunity statement.]

Example Set 3

Context: Public R1 university with a rapidly growing RSE group joining an established data science group. Only the Associate and Senior level positions were available.

Associate Research Software Engineer

As a [UNIVERSITY] employee, you have a unique opportunity to change lives on our campuses, in our state, and around the world. [UNIVERSITY] employees offer their boundless energy, creative problem-solving skills, and dedication to building stronger minds and a healthier world.

[UNIVERSITY] faculty and staff also enjoy outstanding benefits, professional growth opportunities, and unique resources in an environment noted for diversity, intellectual excitement, artistic pursuits, and natural beauty.

[Background on the specific team and current plans for team growth.]

The [team] is seeking outstanding candidates for the position of Associate Software Engineer (multiple openings).

Software engineering is vital to [region's] emergence as a tech leader driving a local climate of innovation. The new [center] at [team] will leverage local software engineering talent to advance scientific frontiers that will change the world around us and respond to the pressing challenges that face humanity.

We are building a diverse team of software engineers who will bring their unique backgrounds and expertise to the [UNIVERSITY] community. These software engineers will work on impactful research projects, infusing them with the software industry's best practices, and delivering

reusable, open-source software that will accelerate future research in areas like climate change, health, energy, and basic science.

We are looking for qualified software engineers with dual backgrounds in science and technology, who will be central to the [team] mission to support data-driven research by enabling the development of new software tools and user communities serving greater scientific goals.

RESPONSIBILITIES:
This position will be part of a new endeavor to create a collegial, creative team, collaborating with University researchers to improve efficiency and reproducibility of research outcomes. Responsibilities will focus primarily on software application design and implementation, and will also be involved with research design and collaboration, problem resolution, and project work. The position reports to the Head of Engineering.

Design and Implement Applications
- Assist with design and implementation of applications/systems that offer a wide range of functionalities to the user population.
- Support the creation of content and contribute to initiatives led by other software engineers. Contribute to other deliverables as designed by [team] leadership.

Collaborate With Research Design
- Collaborate with researchers in the design, planning, and implementation of software that enriches research productivity and reliability.
- Build understanding of research activities through regular engagements.
- Support more senior software engineers by contributing to internal code design and development guides for future contributors.

Problem Resolution/Project Tasks
- Perform analysis and troubleshooting for application issues and process challenges.
- Responsible for the successful completion of assigned project tasks with supervision.
- Provide routine project status updates to the [team lead].
- Participate in all aspects of improving the team, including education/training of other team members and contributing to process/communication improvement initiatives.
- Work with a manager to set professional goals for career development.

REQUIREMENTS
- Bachelor’s Degree in Computer Science, Electrical Engineering, or a related field.
- Minimum of 3 years of previous experience in software engineering, software QA/QC, or similar.
- Experience cultivating positive and productive professional relationships with individuals from diverse social, cultural, and political contexts and ability to build rapport quickly.
- Experience with design and implementation of applications/systems.
- Experience programming with at least one modern language such as Python, Java, C++ including object-oriented design.
- Experience using Git for source control, continuous integration, and release management

Equivalent education and/or experience may substitute for minimum requirements.

DESIRED

- Experience with software infrastructure, cloud deployment, high-performance computing, and scalable architectures.
- Strong understanding of most phases of software development or lifecycle.
- Experience documenting Operational Systems / User Guides.
- Knowledge of database structures, data definitions, and data relationships.
- Excellent written and oral communication skills and excellent problem-solving skills.

CONDITIONS

- Periodically available for phone calls/meetings in the early morning and late evenings to accommodate nationwide time differences.
- Deadline-driven environment, occasionally requiring extended hours and/or work on weekends.

Committed to attracting and retaining a diverse staff, the [UNIVERSITY] will honor your experiences, perspectives, and unique identity. Together, our community strives to create and maintain working and learning environments that are inclusive, equitable, and welcoming.

[UNIVERSITY] is an affirmative action and equal opportunity employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, gender expression, national origin, age, protected veteran or disabled status, or genetic information.

To request disability accommodation in the application process, contact the Disability Services Office at [contact details].

Senior Research Software Engineer

[Same introduction as above, except updating the name of the position.]

RESPONSIBILITIES:

This position will be part of a new endeavor to create a collegial, creative team, collaborating with University researchers to improve the efficiency and reproducibility of research outcomes. Responsibilities will include application design and implementation, research design and collaboration, problem resolution, and projects. This position will also lead, mentor, and coach [team] teammates. The position reports to the Head of Engineering.

Design and Implement Applications

- Design and implement applications/systems that offer a wide range of functionalities to the user population.
- Exhibit versatility, working with researchers on multiple and shifting platforms.
- Create content and contribute to initiatives designed by the Center Software Engineering Lead. Contribute to other deliverables as designed by [team] leadership.

Collaborate With Research Design

- Collaborate with researchers in the design, planning, and implementation of software that enriches research productivity and reliability.
- Build understanding of research activities through regular engagements.

Lead, Mentor, Coach Teammates

- Support a creative, open environment for vibrant research development.
- Assist junior colleagues with documentation, communication, and time management.

Problem Resolution/Project Tasks

- Maintain issue tracking, internally and with researchers.
- Mentor others on the team toward completion of goals and assignments.
- Provide routine project status updates to the [team lead].

REQUIREMENTS

- Bachelor's Degree in Computer Science, Electrical Engineering, or a related field.
- Minimum of 4 years of previous experience in software engineering, software QA/QC, or similar.
- Experience cultivating positive and productive professional relationships with individuals from diverse social, cultural, and political contexts and ability to build rapport quickly.
- Strong understanding of all phases of software development or lifecycle.
- Experience developing software applications using Python, JavaScript, Go, or similar language in a Linux environment.
- Strong acumen for software testing and documentation, with a commitment to quality.
- Experience using Git for source control, continuous integration, and release management.
- Excellent written and oral communication skills and excellent problem-solving skills.

Equivalent education and/or experience may substitute for minimum requirements.

DESIRED

- Experience architecting and implementing solutions that integrate multiple technologies.
- Experience with Agile development practices, including test-driven development.
- Knowledge of database structures, data definitions, and data relationships.
- Experience working with Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform.
- Experience implementing data access and application security policies.

[Same additional information as above.]

Memorandum of Understanding Example

Context: Public R1 university with a long-standing data science institute. This MOU provides terms and conditions for a joint appointment between the DS/RSE unit and another academic unit.

Memorandum of Understanding between The [INSTITUTE] and [DEPARTMENT NAME]

This Memorandum of Understanding (MOU) is for the purpose of establishing the terms and conditions for the joint appointment of [RESEARCH SCIENTIST NAME]'s appointment between the [INSTITUTE] (administered by [COLLEGE OR OTHER ADMINISTRATIE UNIT]) and [DEPARTMENT NAME] (administered by the [COLLEGE OR OTHER ADMINISTRATIvE UNIT]), effective [DATE].

Goals

The [INSTITUTE]'s mission is to engage researchers across disciplines in developing and applying advanced computational methods and tools to real-world problems in data-driven science and research. The Institute's Researcher Scientists have demonstrated accomplishments in at least one disciplinary application area **AND** at least one area related to data science (machine learning, statistics, database management, large-scale data systems, visualization, etc.). These are interdisciplinary research roles that seek to advance the state of the art in both the techniques and technologies of data science and the fields that depend on them.

The [INSTITUTE] Research Scientists will both plan and execute research programs in collaboration with [INSTITUTE] affiliate faculty, through [INSTITUTE] leadership, and through independent research programs. The Research Scientists will publish their research results and pursue external funding, and will also be called on to participate in Institute programs involving software development, technical consulting, workshops and training, and other activities designed to advance the research and practice of data-driven and data-intensive discovery across fields.

Research scientists will typically hold a joint appointment and an office in a home department on campus.

[Hiring department]'s mission is [include mission statement or description of research focus area].

[RESEARCH SCIENTIST NAME]'s expertise lies in [description of expertise]. As an [INSTITUTE] Research Scientist, [he/she/they] will be working jointly with [UNIT NAME] and the [INSTITUTE] as described above for the duration of this MOU.

Primary Appointment

[RESEARCH SCIENTIST NAME] is currently and will remain a Research Scientist in [UNIT NAME], [his/her/their] home department. This appointment will remain the primary and controlling employment relationship with the [UNIVERSITY]. Salary changes (e.g. merit, in-grade) may be proposed by the [INSTITUTE] management and must have the concurrence of the [INSTITUTE] management. In addition, the [INSTITUTE] will supply feedback for annual reviews.

Funding

[RESEARCH SCIENTIST NAME] will be funded by the [INSTITUTE] at [XX]% FTE (salary and benefits) for the duration of this MOU. Shifts in % FTE funded by [INSTITUTE] may be requested by [RESEARCH SCIENTIST NAME] during this time period but changes in % FTE must be approved by [RESEARCH SCIENTIST NAME], the home department, and the [INSTITUTE]. The home department will be responsible for all payroll actions and coordination of funding sources.

Travel Reimbursement

Reimbursement of up to $5,000 per calendar year will be provided by the [INSTITUTE] for expenses incurred for work-related travel by [RESEARCH SCIENTIST NAME], as long as travel is in accordance with [INSTITUTE] and [UNIVERSITY] Travel Policies; see [travel policy website]. The [INSTITUTE] will be responsible for processing, tracking, and monitoring travel reimbursements using [INSTITUTE] budgets.

Workday/Leave Management

The home department will be responsible for monthly time entry into the Workday system, including documentation and obtaining appropriate approvals for all types of leave (e.g. FMLA).

Facilities

The home department will provide office space. However, it is understood that [RESEARCH SCIENTIST NAME] will typically spend 1-2 days per week working in the [DATA SCIENCE UNIT'S OFFICES].

Grant Management and Indirect Cost Distribution

It is expected that [RESEARCH SCIENTIST NAME] will secure external funding. In general, proposals and grant management for which [RESEARCH SCIENTIST NAME] is the PI will be the responsibility of the home department.

Duration

This MOU is in effect from [START DATE] through [END DATE].

This agreement can be changed only with the concurrence of the staff member and the appointing departments. All parties agree that [RESEARCH SCIENTIST NAME]'s [INSTITUTE] Research Scientist appointment and the associated responsibilities as described above will remain in effect for the duration of this MOU.

MOU Authorizations

[Signatures including the RESEARCH SCIENTIST, DS/RSE UNIT DIRECTOR, COLLABORATING DEPARTMENT DIRECTOR, and other administrative signatories as required - such as deans]

Buyback Program Example

Context: Public R1 university with a long-standing data science institute. This example illustrates one approach to providing incentives and rewards for DS/RSEs who acquire external funding for projects

[UNIVERSITY] [INSTITUTE] Buyback Program

Availability: The [INSTITUTE] Buyback Program is available to Data Scientists, hired and funded by [INSTITUTE], who attract external funding which covers a portion of their own salaries.

Policy: [INSTITUTE] Data Scientists, who have external funding that covers some fraction of their salary, will have the option of using the dollar equivalent of 50% of that salary offset towards their own research-related expenses.

- For example, for an offset of $20k of salary, the Data Scientist will have access to $10k of funding to be used on research-related expenses.
- This does not include costs associated with benefits or overhead.

What types of funding count: Personal salary paid for by grants on which the DS is a named PI, Co-PI, Senior Personnel, or unnamed personnel that are run through [INSTITUTE] and/or other departments at [UNIVERSITY] and/or outside of [UNIVERSITY] (through a sub-award mechanism, etc.).

What is the time frame for utilizing buyback funds: Buyback funds need to be utilized within 6 months of the end date of the associated award, so timely is highly recommended. Buyback funds will be tracked and updated quarterly by the [INSTITUTE] Grants Manager, or more frequently upon request. These funds will not be transferred into a separate sub budget, in order to provide the maximum flexibility in utilizing these funds.

Maximum encumbrance: Each individual's buyback account is capped at $25k. For accurate budget planning and projections, [INSTITUTE] cannot encumber funds in excess of $25k per data scientist.

What can funds be spent on: Funds can be used for research-related expenses including but not limited to equipment, salary funding for collaborators/students, travel, publication costs, subscriptions, consultants, and services.

Allow time for review: Work with [GRANT MANAGER] to plan the timing and use of the available funds, and to route any reviews internally if needed. Additional review would be needed for:

- RA or TA support, including tuition
- a research sub-award, outside the [UNIVERSITY]
- an item or product that you intend to retain beyond your tenure at [INSTITUTE]
- hosting an event, with costs related to site rental, catering, and/or hotel accommodations
- arranging entertainment, gifts, or other costs that will be charged to the [INSTITUTE] discretionary budgets
- purchases that include human subjects, such as participant payments, or products and services covered in a service agreement - these may take 4-6 weeks to set in place

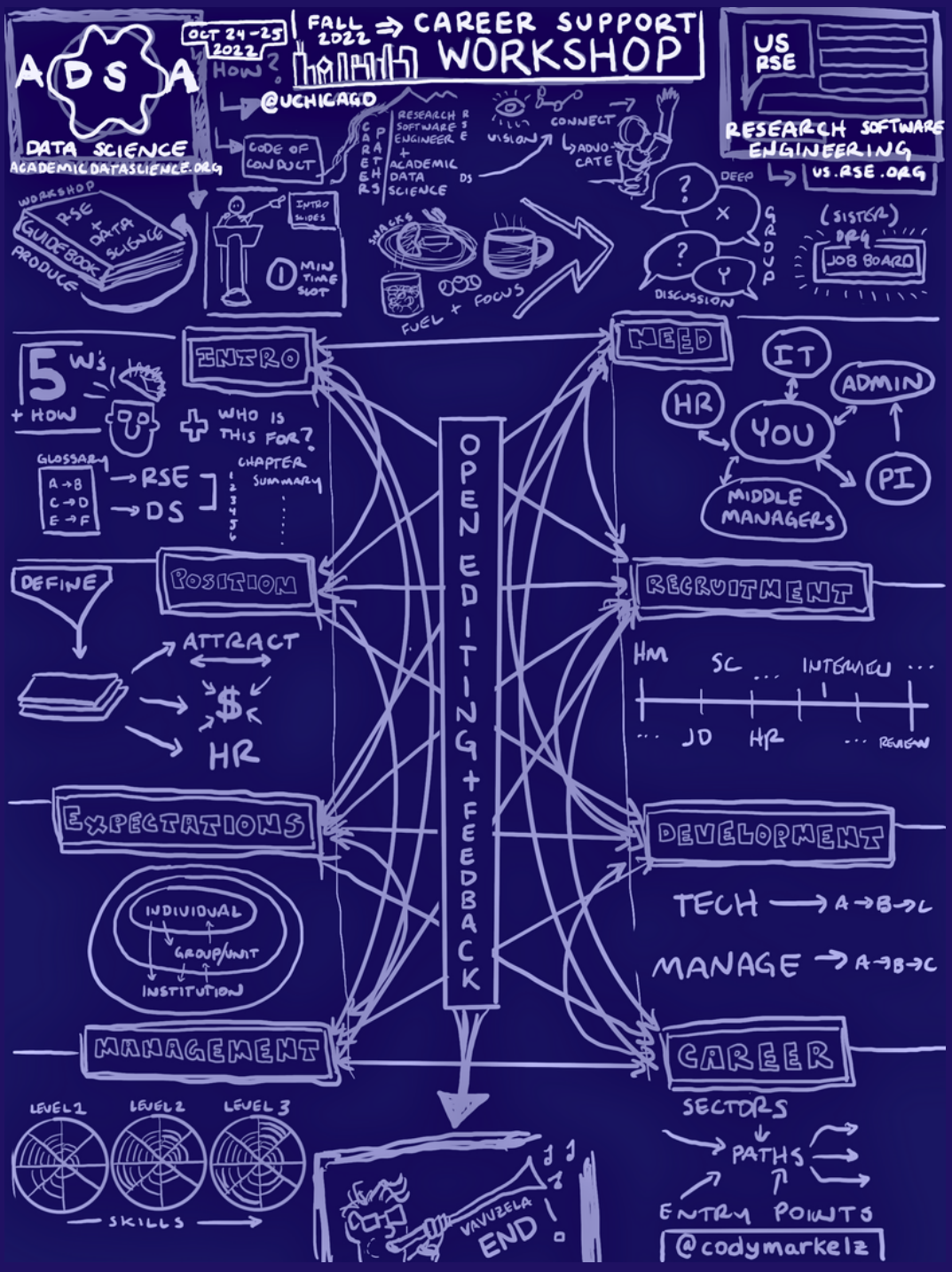
Important notes:

- This buyback is considered to cover a portion of the 50% FTE that Data Scientists have to devote to their own research.
- Having time commitment on a grant does *not* change [INSTITUTE] activity commitments.
- If award time exceeds 50% of your FTE then let's talk.

BACK COVER: Sketchnote illustration by Cody Markelz (www.codymarkelz.com)

Cody Markelz created this sketchnote illustration during the workshop which served as a starting point for this guidebook. The sketchnote shows the major workflow of the meeting. Participants self-selected into small groups to brainstorm outlines and begin writing the chapters of the book. The workshop was similar to a 2-day academic paper writing session with rapid critical feedback.

About the artist: Cody is a scientist and artist living in Northern California who uses data, mathematical models, and illustrations as tools to understand the beauty and complexity of nature.



Sketchnote illustration by Cody Markelz (www.codymarkelz.com)

Academic Data Science Alliance

https://academicdatascience.org/

US RSE

https://us-rse.org/