

Careers of Data Scientists: Report from 13 Academic Institutions

October 2019

Prepared for:

Joshua Greenberg, PhD
Alfred P. Sloan Foundation

Prepared by:

Luba Katz, PhD
Abt Associates



Introduction and Methods

The goal of this study was to document professional duties, challenges, and aspirations of staff data scientists in academia. The study subjects were recommended by leaders of data science entities at dozens of US universities and by data scientists.¹ Of the 72 candidates suggested, 46 met our inclusion criteria² and 33 (72%) agreed to a telephone interview. The data presented in this report are based on 28 interviews.³

The 28 participants were affiliated with 13 institutions: Boston University (n=2), California Institute of Technology (n=2), Columbia University (n=2), Georgia Institute of Technology (n=3), University of Illinois (n=2), Northwestern University (n=1), University of North Carolina (n=2), Stanford University (n=1), University of Michigan (n=5), University of Virginia (n=3), University of Washington (n=3), Harvard University (n=1), and Fermilab (n=1).⁴ Within the universities, respondents were based at a research center (n=15); a department (n=2); a research institute (n=2); an administrative unit such as an information technology department, library, or dean's office (n=4); or had a dual affiliation, typically at a research center and department (n=4). Almost three-quarters (n=20) held a doctorate degree in the following fields: physics/astrophysics (n=5), computer science (n=6), mathematics (n=1), statistics (n=2), biology (n=2), engineering (n=1), environmental science (n=1), and political science (n=2). The majority of participants (86%) were male.

Findings

Career groups

Based on the data collected in our interviews, we divided participants into two career groups: researchers and consultants (Exhibit 1). Almost two-thirds of the study participants were in the first category. The majority of researchers had a PhD and Principal Investigator (PI) status, and expressed at least some interest in a tenure-track position; 24% had a faculty title. Approximately 40% spent most of their time on research activities and all worked primarily on their own projects or on a mix of their own projects and those initiated by others. We believe that researchers were similar in duties and professional qualifications to a fairly common position of research scientist, although they were probably more involved in software development (88% spent some time and

¹ We did not provide a definition of data scientist in our requests.

² Inclusion criteria were: (1) identify as a staff data scientist or similar; (2) not a tenure-track faculty; (3) affiliated with a US university. Postdocs and senior administrators were excluded. For universities with multiple nominations, we selected individuals with titles and public profiles that best fit our criteria.

³ Five individuals interviewed were excluded for the following reasons: one agreed to an interview but was unwilling to answer our questions, and another four were determined to be ineligible during the interview because they were either administrators or postdocs.

⁴ While our goal was to focus only on universities, we included one data scientist from a national laboratory because he identified himself as an academic and his organization was very similar to a university-affiliated research institute.

24% spent most time) and more likely to work in multiple disciplines (67%) than is typical. These differences were not surprising for individuals who identified as data scientists.

Exhibit 1: Participant characteristics by career track

	Percent respondents	
	Researchers (n=17)	Consultants (n=11)
PhD	82	55
Work mostly on projects for others	0	100
Work mostly on own projects	24	0
Work mostly on mix of projects	76	0
Faculty title	24	0
Majority of time on research	41	9
Majority of time on software development	24	45
Some time on software development	88	64
Full flexibility to choose tasks/projects	59	45
Work in multiple disciplines	67	44
Clear path to promotion	33	11
Permanent position	6	27
Funded through grants only	41	36
PI status	59	0
Preference to remain in academia	53	50
Some or clear interest in tenure-track position	54	22

The remaining one-third of the respondents in our sample appeared to be on a somewhat different track. For lack of a better term, we called them consultants because they spent most of their time helping others with projects. Consultants were less likely than researchers to have a PhD (55% vs 82%), PI status (0% vs 59%), faculty title (0% vs 24%), interest in a tenure-track position (22% vs 54%), and a clear path to promotion (11% vs 33%); but were more likely to be in a permanent position (27% vs 6%) and spend the majority of their time on software development (45% vs 24%).

While acknowledging both within-group variation and between-group overlap, we believe that the differences in the experiences of these two groups captured in our study reflect two emerging career tracks for data scientists in academia. In the remaining sections, we describe these experiences in more detail.

Path to position and its description

Many study participants were PhDs who wished to remain in academia, but had either never been interested in or became disillusioned with a tenure-track career path. Some transitioned to their current position after earning a degree or completing postdoctoral training at the same university. A small number came from industry in search of more meaningful or interesting projects, or from a similar position at another university. The majority were external hires.

Participant titles varied broadly: some examples include Research Scientist, Senior Data Scientist, Research Computing Lead, Data Engineer, Data Science Architect, Senior Software Engineer, and Software Developer (Exhibit 2). Consultants tended to have more diverse titles that contained terms such as *specialist*, *services*, or *developer*; and researchers were more likely to be called *scientists*.

Exhibit 2: Titles of study participants



Source: <https://www.jasondavies.com/wordcloud/>.

One-third of the participants had two titles. In four cases, all in the researcher group, this second title was a professorship. In many of the remaining cases, this second title was chosen by the participants themselves because they felt that their “official” title was too general (e.g., Statistician II) and/or did not adequately describe their role. Virtually all respondents said that at least one of their titles was accurate.

The majority of participants told us that they were hired after responding to an announcement and that their daily duties were consistent with their expectations. Many of those who were recruited internally either directly participated in the development of the job announcement or it was designed to fit their qualifications. Some respondents were the first employees of their unit and actively participated in defining the overall direction of their organization and their own role – which they enjoyed.

Approximately one-quarter of the respondents, most in the researcher group, said that the position only partially met their expectations. In some cases, this was either because the job was not well-defined when they were hired and, thus, they did not know what to expect. In other cases, respondents would have liked to have a different balance of activities (generally, to spend more time on research or software development) or to have more projects in a particular content area. One interviewee, who had been recently hired, appeared disappointed in the position because the announcement called for advanced data science skills, but the work turned out to be largely managerial. This respondent was trying to reshape the position to involve more time on research and/or data science.

Terms of employment and sources of support

Of the 28 respondents, 5 (18%) had a permanent position, 16 (57%) an open-ended position, and 7 (25%) a term-limited position. None of the interviewees in temporary positions expressed concerns about short-term job security. On the contrary, several commented that demand for their services exceeded supply and that money would probably be found to cover their salaries if their current funding ended.

Eleven of the 28 respondents (39%) were supported fully through grants; 4 (14%) by their university; and the rest with a mix of sources, typically including grants. The sources of support were similar for researchers and consultants. One participant in a “hard money” position was able to fully cover his salary through grants. Regardless of the type of position, many respondents felt under pressure to bring in grants to cover themselves and/or their staff.

PI status

In our previous work, we found that many data scientists were interested in having PI status, but some struggled to obtain it. We explored whether individuals in this sample also experienced this challenge. Of the 28 participants, 10 (36%) had PI status, of whom 3 said that it was very important to their independence and career satisfaction (1 ensured he was hired with PI status because of his previous experience). All respondents with PI status were in the researcher group.

Of the 18 data scientists without PI status, 8 said that it was unnecessary for their position and another 6 were weakly interested, but uncertain how to obtain it. The remaining four (three researchers and one consultant) wished they had PI status, but thought that they were ineligible for it because they did not have a PhD or were not a faculty member. Two of these four said that without PI status they did not reliably get credit for bringing in funding and had no path to independence. It appeared from the interviews that policies regarding PI status eligibility varied widely between universities and even between departments at the same university.

Professional duties

We asked participants to estimate how much time they spent on software development, research (which we defined for them as running analyses, attending talks, reading papers, and writing grants), consultations, training, teaching, management, and other tasks. We note that some respondents were reluctant to allocate time to research versus software development, which they saw as connected. Therefore, these data are meant to give a general sense of the respondents’ duties rather than specific numbers. For roughly one-quarter of the participants, it was a mix of the activities listed above, with none reaching 50%. Of the remaining, eight spent the majority of their time on research, nine on software development, two on consultations, and one on teaching. Individuals in the researcher group were more likely than the consultant group to say that they spent the majority of their time on research (41% vs 9%) and the reverse was true for software development (24% vs 45%).

Fifteen of the 28 interviewees (59% of researchers and 45% of consultants) said that they had full professional flexibility. For some, this meant having the option to turn down requests for help with projects that did not interest them (which was possible for those who had developed a client base), while for others it meant pursuing an independent research program. Some respondents who

characterized themselves as independent qualified this term by saying that their work had to be consistent with the mission of their unit and/or the aims of the grants that support them. A few also acknowledged that they occasionally get involved in projects that are not of particular interest if they are well-funded, are a priority for their supervisor, or can strengthen a collaboration. Twelve participants (41% of researchers and 45% of consultants) said that they had partial flexibility, and one consultant had limited flexibility.

We also explored whether the work of our study subjects was self-directed or initiated by others. All consultants and none of the researchers worked primarily under the direction of other scientists – we used this variable as the primary determinant for group assignment. Interestingly, only one-quarter of researchers worked mostly on their own projects; for the remaining it was a mix. We note that a few respondents said that they substantively contributed to the projects initiated by others and therefore considered these their own (in these cases we characterized their work as a mix). Several researchers felt strongly about this aspect of their position (e.g., one said that he was “militant” about devoting 85% of the time to his own projects). Conversely, some consultants told us that they had made a deliberate choice to provide service/do work for others rather than being an intellectual lead for their own program.

Virtually all interviewees were encouraged to develop networks and several saw this as part of their responsibilities. In a follow-up question, we inquired what venues for establishing collaborations they found most effective. Respondents consistently told us that at first they had to engage in active outreach by presenting at faculty meetings and student orientations, drawing on personal connections, hosting workshops, advertising through listservs, and even going door-to-door to introduce themselves and their skills. Within a year or two, the volume of requests had exceeded their capacity to respond, and they began to passively maintain networks through existing relationships. Interviewees were in agreement that it took much initiative and energy to become known at their institution. Some found this aspect of their job challenging, while others enjoyed it; one respondent told us that he had taken the position in part for this role.

Mentorship and performance evaluation

All but two respondents had a supervisor, mentor, or both. Approximately one-half of the participants said that they were getting what they needed from these relationships and a few described themselves as highly satisfied. Some participants made a distinction between technical and career help, with the former much easier to access. Respondents were also looking for support with grant writing as well as for greater clarity about job expectations and performance. Finally, a few said that no one was in a position to mentor them because they were too senior or their professional situation was unique.

Twenty-five of the 28 interviewees were required to participate in annual or bi-annual performance evaluations, during which they discussed progress toward the previous year’s goals with their supervisor.⁵ Example goals offered by respondents included providing consultations, establishing new collaborations, publishing papers, giving presentations, winning grants, and creating resources for the community. However, many study participants characterized

⁵ Two researchers, recent hires, had not yet gone through the review process.

performance reviews as a formality of limited value to their career development. A small number had additional, off-the-record check-ins with their supervisor/mentor, which they found more useful.

Strengths and weaknesses of positions

Many interviewees described themselves as “very happy” and none had immediate plans to move on. When asked to reflect on what they liked most about their positions, respondents spoke about the intellectual freedom, great colleagues, exposure to many fields, opportunities to do data science, and being in a position to create knowledge and help others. Collaborations with industry were appreciated by the small number of researchers who had them.

Several weaknesses of the positions also emerged. One of these was funding uncertainty, which was identified as a problem by 7 of the 15 researchers and 2 of the 10 consultants who responded to this question. We observed that even the staff whose jobs were not threatened by the loss of grants were concerned that their freedom to choose projects could be compromised or that they would be unable to support the salary of someone in their group.

A small number of interviewees also mentioned antiquated and/or inefficient institutional policies or practices. Problems in this area included being able to quickly hire staff to meet the demand for services and the slow decision-making by the administration. In addition, participants brought up outdated technology and the lack of intellectual property polices applicable to modern data science, such as how to handle open-source software.

Challenges related to pay and promotion were brought up by one-half of the researchers and consultants. Respondents said that their salaries were too low given their qualifications and compensation in industry, and questioned how long they could remain in academia. To boost salaries, one interviewee suggested that universities allow staff with grants to draw summer salaries (as do faculty), introduce bonuses, and let industry partners pay higher rates.

Opportunities for career advancement was another weakness of the positions: only 11% percent of consultants and 33% of researchers indicated that they had a clear path to promotion. For the majority, there was either no next step, or there was a next level but they were unsure how to reach it, or they knew how to reach it but there was no funding to support a higher salary. Consequently, many interviewees did not plan to remain in their jobs in the long-term, even though they were mostly happy with their present situation.

Those for whom promotion was feasible anticipated that it would involve more management responsibilities. Interestingly, for researchers this meant the expansion of their program, more independence, and ability to have students – all welcome developments. In contrast, consultants were less enthusiastic about taking on additional management duties, as they were concerned that this could mean giving up software development. (In acknowledgment of these types of preferences, one center was creating two career tracks: one for data scientists who wish to specialize in a technical area without a management role; and another more focused on leadership, but less on tool development.)

We also explored whether interviewees were interested in and eligible for tenure-track faculty positions. Over 50% of the researchers and 20% of the consultants had some or significant interest

in this option, and one researcher was actively pursuing it. Those who were not interested said that they did not have a PhD, were not competitive, actively disliked the “publish or perish” ethos of faculty life, were geographically bound, or preferred a consulting/tool development role to independent research. Some viewed their positions as the “best of both worlds” because they could stay at a university and do what they liked without having to deal with the pressures of tenure track. A few respondents mentioned the position of principal research scientist in the context of this discussion. They said that achieving this level was comparable to getting tenure, but without gaining the security of a lifelong appointment. Some of those who had considered this option decided that they might as well move to tenure track, while others were working toward this position despite its limitations.

Finally, 4 of the 15 researchers who answered the question (and none of the consultants) felt that they were treated as a “second class citizen.” When asked to give specific examples, they spoke about the lack of spousal benefits and visa sponsorship, exclusion from important committees, and inability to serve as a PI and mentor students. Some of these interviewees argued that they are very valuable to their universities and deserve more support and recognition.

Conclusion

We found that while data scientists in academia are generally satisfied with their present professional situation, many see their positions as temporary due to relatively low pay, limited opportunities for advancement, being undervalued by faculty, and other reasons. If data scientists are an asset to the academic community – as was suggested in interviews – universities and funders should make their career paths more attractive. Additional studies to further understand the career tracks of data scientists as they evolve, as well as the nature and extent of their contribution to discovery, are merited.