Careers of Data Scientists: Report from 13 Academic Institutions

October 2020

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 U.S. universities and by data scientists themselves. Of the 72 candidates suggested, 46 met our inclusion criteria; of those, 33 (72%) agreed to a telephone interview. The interviews were conducted in the Summer and Fall of 2019. The data presented in this report are based on 28 interviews.²

The study 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)
- Harvard University (n=1)
- 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)
- Fermilab (n=1)³

Within the universities, respondents were based at:

- Research center (n=15)
- Department (n=2)
- Research institute (n=2)
- University administrative unit such as dean’s office (n=4)
- Dual affiliation (n=4)⁴

---

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

² 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 postdoctoral students.

³ 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.

⁴ Typically a research institute and a department.
Almost 75% (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)
- 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 (60% of the sample) and consultants (40%, Exhibit 1). The majority of researchers had a doctorate (82%) and a Principal Investigator (PI) status (59%) and expressed some interest in a tenure-track position (54%). Approximately 40% spent most of their time on research activities and all worked either mostly on their own projects or on a mix of their projects and those initiated by others. We believe that the researcher group was similar in duties and professional qualifications to a fairly common academic research scientist position, but with more involvement in software development, on which 24% spent most of their time. We also found that 67% worked in multiple disciplines.

**Exhibit 1: Participant characteristics by career track**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Researchers (n=17)</th>
<th>Consultants (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have a PhD</td>
<td>82</td>
<td>55</td>
</tr>
<tr>
<td>Have a faculty title</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Have a PI status</td>
<td>59</td>
<td>0</td>
</tr>
<tr>
<td>Interested in a faculty position</td>
<td>54</td>
<td>22</td>
</tr>
<tr>
<td>Prefer to remain in academia over going to industry</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>Work mostly on projects for others</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Work mostly on own projects</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Work mostly on a mix of projects</td>
<td>76</td>
<td>0</td>
</tr>
<tr>
<td>Spend most time on research</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>Spend most time on software development</td>
<td>24</td>
<td>45</td>
</tr>
<tr>
<td>Have full flexibility in how time is spent</td>
<td>59</td>
<td>45</td>
</tr>
<tr>
<td>Work in multiple disciplines</td>
<td>67</td>
<td>44</td>
</tr>
<tr>
<td>Have a clear path to promotion</td>
<td>33</td>
<td>11</td>
</tr>
<tr>
<td>Have a permanent position</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Fully funded through grants</td>
<td>41</td>
<td>36</td>
</tr>
</tbody>
</table>

The remaining 40% of the sample appeared to be on a different track, which we call “consultants,” as they spent most of the time supporting others. Consultants were less likely than researchers to
have a PhD (55% vs 82%), PI status (0% vs 59%), a 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 most of their time on software development (45% vs 24%).

**Career paths and current positions**

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 few 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 (e.g., Research Scientist, Senior Data Scientist, Research Computing Lead, Data Engineer, Data Science Architect, Senior Software Engineer, 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. Thirty-three percent of the participants had two titles; in most of these cases this was because their official titles did not reflect their duties and so they added their own.

**Exhibit 2: Study participant titles**

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

Most participants were hired after responding to a job announcement and 75% said that their duties were consistent with their expectations. A few participants who were recruited internally helped craft their job descriptions. 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 interviewees, mostly in the researcher group, said that their positions only partially met their expectations. In some cases, this was either because the job was not well-defined when they were hired and they had few expectations. In other cases, respondents would have preferred a different balance of duties – 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 her work turned out to be largely managerial. This respondent was trying to reshape the position to better fit her interests and aspirations.

Of the 28 respondents, 5 (18%) had a permanent position, 16 (57%) an open-ended position, and 7 (25%) a term-limited position. However, no concerns were voiced about short-term job security. On the contrary, several respondents commented that the demand for their services exceeded supply and that money would probably be found to cover their salaries if their current funding ended. We found that 11 (39%) of the respondents were supported entirely through grants; 4 (14%) entirely by their universities; and the rest with a mix of sources, typically including grants. The funding streams were similar for researchers and consultants, and respondents in both groups felt pressure to win grants to cover their own salary and/or their staff salaries.

In our previous work, we found that many data scientists were interested in having PI status, but some encountered obstacles to obtaining it. Of the 28 participants in this study, 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 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 somewhat interested but uncertain how to obtain it. The remaining four (three researchers and one consultant) wished they had the status, but considered themselves ineligible because they did not have a PhD or faculty position. Two of these four respondents said that without PI status they did not always get credit for winning grants and had no path to independence. We gathered from the interviews that policies regarding PI status varied widely between universities and even between schools within the same university.

We asked participants to estimate how much time they spent on software development, research,5 consultations, training, teaching, management, and other tasks. We note that some respondents were reluctant to distinguish between research and software development work, and therefore these data are meant to offer a qualitative description of respondents’ duties. For roughly one-quarter of the participants, it was a mix of all of the activities listed above, with no one task reaching 50%. Of the remaining respondents, 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 consultants to say that they were primarily involved in research (41% vs 9%) and less likely in software development (24% vs 45%).

Fifteen of the 28 interviewees (59% of researchers and 45% of consultants) reported having complete flexibility professionally. This flexibility was defined very differently by respondents – from having “the luxury” to turn down projects that did not interest them to the ability to pursue a fully independent research program. A few also acknowledged that they occasionally were involved in projects that were of limited interest but well-funded, were a priority for their supervisor, or could strengthen a particular collaboration. Twelve participants (41% of researchers and 45% of consultants) had partial flexibility, and one consultant had limited flexibility.

5 Defined for them as running analyses, attending talks, reading papers, and writing grants.
We also explored whether the work was self-directed or initiated by others. We found that one-quarter of researchers worked mostly on their own projects and for the remaining, it was a mix, although they made substantive contribution to work initiated by others. Many researchers felt strongly about maintaining at least some level of intellectual independence (e.g., one said that he was “militant” about devoting 85% of the time to his own projects). In contrast, consultants made a deliberate choice to provide services or perform work for others over having their own research program.

Virtually all interviewees were encouraged to develop networks and several saw this as a job requirement. Most told us that they began 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 switched to maintaining their existing relationships. Interviewees were in agreement that it took initiative and energy to become known at their institution; some found this aspect of their job challenging, while others enjoyed it.

All but two respondents had a supervisor, mentor, or both; and half were getting what they needed from these relationships. Some study participants made a distinction between technical and career mentorship, with the former being much easier to access. Those whose needs were not fully met were looking for support with grant writing, and for greater clarity about job expectations and their performance. Finally, a few did not have a mentor because they were too senior or their professional situation was unique.

Most respondents were required to participate in annual or semi-annual performance evaluations, during which they discussed progress toward the previous year’s goals with their supervisor. Example goals included providing consultations, establishing new collaborations, publishing papers, giving presentations, winning grants, and creating resources for the community. However, many study participants characterized performance reviews as a formality and of limited value to their career development. A few had additional, informal check-ins with their supervisor/mentor, which they found more helpful.

**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 work in data science, and being in a position to create knowledge and help others. Collaborations with industry were appreciated by the few researchers who had them.

Several weaknesses of the positions also emerged, however. One of these was funding uncertainty, which was identified as a problem by 50% of researchers and 20% of consultants. 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 few interviewees also mentioned antiquated and/or inefficient

---

6 Two recent-hire researchers had not yet gone through the review process.
institutional policies or practices. Problems in this area included slow hiring processes, outdated technology, and a lack of intellectual property policies applicable to data science (such as how to handle open-source tools).

Challenges related to pay and promotion were raised by one-half of the researchers and consultants. Respondents said that their salaries were too low given their qualifications and corresponding compensation in industry, and questioned how long they could justify remaining in academia. To boost salaries, one interviewee suggested that staff who have grant funding be allowed to draw a summer salary, and be eligible for bonuses or supplemental pay by industry partners.

Opportunities for career advancement was another weakness of the positions: only 11% 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 they were unsure how to reach the next career level, or there was no funding to support a higher salary even if they could be promoted. Consequently, many interviewees did not plan to remain in their jobs in the long-term.

Those for whom promotion was feasible, anticipated that it would involve more management responsibilities. For researchers, this meant the expansion of their program, more independence, and the ability to have students – all welcome developments. In contrast, consultants were less enthusiastic, as they were concerned that taking on more management duties might mean giving up software 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. We heard 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, four of the researchers, but 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 the inability to serve as a PI and mentor students. Some of these interviewees argued that they are very valuable to their universities and deserve greater support and recognition.

---

7 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.
Conclusion

We concluded that the differences in the experiences and career aspirations captured in our study reflect two emerging career tracks for data scientists in academia: researchers, a track similar to research scientist; and consultants, a track more focused on service and community support. Despite being satisfied with their present professional situation, many respondents on both tracks described their positions as temporary due to low salaries, limited opportunities for advancement, and being underappreciated for their contributions. Therefore, to retain data scientists, universities and funders need to develop programs and policies to make their career paths in academia more attractive.