"I Am Pretty Interested in Coding, Technology, and Infrastructure"

Arfon M. Smith

Arfon M. Smith

Abstract In our interview with Arfon Smith, we discuss how an interest in writing research software can lead to positions with organizations such as Zooniverse and GitHub. Though not yet established at the time, this type of position is characterized by the term 'research software engineer'. Arfon himself has helped pave the way for this career path, through his efforts in founding the *Journal of Open Source Software*. The traditional academic path is not a good fit for everyone, but there are many shared skills involved in PhD research and working as a product manager.

Contents

A. M. Smith (✉)
Space Telescope Science Institute, Baltimore, MD, USA

GitHub Inc., Edinburgh, UK

C. R. Madan (ed.), *Academia and the World Beyond*,
https://doi.org/10.1007/978-3-030-82606-2_10

Chris: Can you introduce yourself and tell me a bit about your current position?

Arfon: My name is Arfon Smith. I work at Space Telescope Science Institute (STScI) in Baltimore, USA. STScI was founded in 1981 to run the science operations of the Hubble Space Telescope. Since 1990, STScI has been the operational interface for Hubble, serving the global astronomical community who make use of this flagship facility. A key part of our work as the *science operations centre* for Hubble and the soon to be launched James Webb Space Telescope (JWST) is to ensure the continued scientific legacy of the missions. Capturing and preserving the data associated with Hubble is the responsibility of the Barbara A. Mikulski Archive for Space Telescopes (MAST), which is the archive for Hubble and more than 20 other mission datasets including Kepler, TESS, IUE and Galex. MAST currently holds data from 21 missions and surveys and, with a data volume of over 2 petabytes, is a major infrastructure support effort in and of itself.

As the head of the Data Science Mission Office at STScI, I am responsible for defining the organization's strategy and roadmap for data management and data science with our multi-mission, multi-petabyte archives, and I am a member of the senior leadership team of the institute (~750 staff). My team manages a portfolio of work for approximately 100 engineers, astronomers and data scientists.

What was the focus of your PhD?

My PhD is from the University of Nottingham (2006) in astrochemistry (strictly speaking just 'chemistry'). I was situated in the chemistry department at Nottingham. The focus of my studies was to probe the small-scale structure of the interstellar medium in our galaxy (the gas and dust between stars) using spectrographs attached to telescopes.

As you were finishing your PhD, what were you thinking about your career plans?

I wasn't too sure to be honest. I had a pretty strong sense that a traditional academic career really wasn't going to work out for me. Clues/signals that led me to this conclusion included me not really being that interested in keeping up with the literature in my field, being more interested in the software I was writing than the results/insights I was producing! I applied for a couple of postdocs because I couldn't really think of anything to do but wasn't successful.

I then embarked upon what became a career building infrastructure/writing software to support academic research (Wellcome Trust Sanger Institute, Zooniverse, GitHub, STScI).

Can you tell us a bit about what day-to-day life is like in your current position?

My current position is pretty mixed. My 'day job' (and roughly 80% of my time) is in a leadership role at STScI. The small team I lead are effectively 'portfolio managers' in the sense that we coordinate a portfolio of work for a small collection of other organizational units (e.g. Hubble Space Telescope (HST), James Webb Space Telescope (JWST), Wide Field Infrared Survey Telescope (WFIRST)). I would say my job roughly breaks down into the following four categories:

1) Spending time with other leaders making and shaping decisions about priorities for the organization, putting together organization-level strategies for data management and data science and developing roadmaps/priorities for individual teams.
2) Working with individual product teams to give feedback on their progress executing on roadmaps.
3) Developing new ideas for projects and exploratory R & D with new technologies and product ideas.
4) Reporting on work: As a government contractor, we have a lot of reporting to do to the government (NASA in our case).

The final category of work is the other 20% of my time. I am fortunate to have 20% research time which I almost exclusively use to support the *Journal of Open Source Software* (JOSS, https://joss.theoj.org) as editor-in-chief.

You've had a few positions between completing your PhD and your current position, can you tell us a bit more about how these positions differed?

I had about 3 years in what I would now call a Research Software Engineer position although that title didn't exist at the time. Firstly, at the Wellcome Trust Sanger Institute in Cambridge building tools to support DNA sequencing experiments. I only held this position for a year before joining the Galaxy Zoo team at the University of Oxford as a postdoc, spending all of my time doing engineering work. Ultimately this led to the development of the Zooniverse platform (https://www.zooniverse.org) which I co-founded and spent the next 3 years leading from a technical standpoint.

I then moved out to the USA to lead a new grant designed to grow the Zooniverse into a true platform for global citizen science. This grant was hosted at the Adler Planetarium, and so I essentially held a dual role here – Director of Citizen Science (at Adler) and Technical Lead of the Zooniverse. This was my first real experience of managing people (I had a team of 12 people at Adler) and leadership (as a member of the management team).

In late 2013, I joined GitHub to lead their engagement with the scientific research community. With a focus on open source software and open data in science, my responsibilities included working with industry partners including journals, data publishers and funding agencies to develop strategies for sharing data and software in academia and improving the GitHub service offerings for academic users. Examples of projects I worked on include Zenodo-GitHub integration (https://guides.github.com/activities/citable-code/), Jupyter Notebook rendering (https://help.github.com/en/articles/working-with-jupyter-notebook-files-on-github) and Software Citation Principles (https://peerj.com/articles/cs-86/).

Later I transitioned to Program Manager for Open Source Data, and in this role, my focus was on ensuring that GitHub remained a responsible host of open source communities and their data and that GitHub maximized opportunities to develop a deeper intellectual understanding of how open source communities work. My core areas of focus were:

1) Working with our platform team to improve open source data products produced by GitHub
2) Facilitating third-party research around GitHub data products and network activity
3) Working with open source communities, enterprises and research groups to develop industry standard metrics for measuring open source health
4) Leading the recruitment of GitHub's first data science team

If someone currently finishing their PhD was considering a position similar to yours, how might they decide if it would be a good fit?

The common thread in my career has been designing and building tools that support the work of academics and researchers. Even though I came late to software (I wasn't into computers growing up, for example), I am pretty interested in coding, technology and infrastructure. I also really like *building things* (I always have a side project on the go, for example).

I think the closest industry term for the sort of thing I've been doing most of my career is *product management* (with my STScI gig being one level more abstracted away as a *portfolio manager*). There's a bunch of articles out there about what it looks like to be a product manager (e.g. https://hbr.org/2017/12/what-it-takes-to-become-a-great-product-manager, http://blogs.nature.com/naturejobs/2018/06/06/how-product-management-could-be-a-route-out-of-academia-for-phds-and-postdocs/). I know a fair number of people ex-academics that have transitioned to product management and are very happy.

If someone was interested in pursuing a similar career path, what would you suggest they do to better prepare themselves?

I think my interest in technology and software has stood me in pretty good stead career-wise. My current role and in part my previous ones too require me to be pretty current with technology trends in both the current research discipline I'm actively supporting (astronomy/astrophysics) and in the wider economy. While this isn't true of my current role, it was at GitHub so I think it's worth sharing here: many companies care much more about your skills and ability to integrate into their workforce than they care about how many papers you've published! So if you do find yourself moving out of academia and into industry, then some skills are much more valuable than others. For example, in an engineering or product company, knowing about Git(Hub), Python and backlogs/Kanban/Scrum/how engineering works is managed – all of these things are much more important than whether you have one, two or three papers in that prestigious journal your field cares about.

While I'm on the topic of transitioning out of academia, I think it's really important that you substantially rewrite your resume if you are looking for positions in industry. Ask someone you already know who has made the transition to help you with this. They should be able to help you reframe your knowledge/skills/expertise for a non-academic audience.

A lot of people like academia because they feel it gives them an opportunity to work on a topic that they deeply care about. Do you think this is also true in your current position?

I do yes, but I think I'm only in this position because of a couple of factors: (1) being willing to move around between jobs semi-regularly and having a family able to support this and (2) some luck (right job, right time opportunities, etc.) There have been a number of times in my career when I had a 'solid' job that there was no good reason to leave (but I did). Each time I've been pleased with that change, but there's been some risk too I think.

Another reason many like academia is that they feel it provides them with more freedom than they think they would get in other positions. How much freedom do you feel you have to work on what you think is interesting?

Hrm, I actually think this is a fallacy for many in academia. You've accidentally hit on a topic I promised myself I'd write a blog post on a few years ago and then failed to (!) so I'll attempt to summarize here….

On the face of it, I agree, academia has lots more freedom. That said, in my experience, many academics (and more importantly their projects) are severely constrained by the funding landscape they exist in. As a result, it's a pretty common game to try and fit your ideas to some kind of funding opportunity, that is, you end up having to *write to the grant*. An extension of this would be actually selecting a potential research direction *because* of the funding opportunities available to them. In the worst-case scenario, your research agenda is effectively being directed or at least heavily influenced by the interests/preferences/whims of your funders. This is obviously an extreme I'm describing here, but I believe I've seen this in action multiple times in a variety of academic settings. So in summary, in the idealized form, academia has lots of freedom, but in reality, I don't think it's quite that 'pure' or straightforward.

In industry, there can be lots of freedom to work on new ideas too, provided the business is doing well and there's a focus on innovation. When a business decides to take on a new challenge, there are often substantially more resources to apply to a problem than in academia. In my experience of industry work and projects, it's often easier to stay true to your ideas and more intellectually honest about the work you want to do. Easier to 'stay true' because you're less resource constrained, and money often doesn't have many restrictions. More intellectually honest because if you've made a pitch for a new product/feature/line of business and customers hate it, then there's a good chance it's not as good an idea as you initially thought….

What do you like most about your work?

I really like working for an organization that has a large impact on the community. For example, because of the nature and scale of the science missions that STScI operates, a significant fraction of the available funds for open source software in astronomy are spent by us. This means that decisions we make as an organization can affect real change on the global astrophysics community. We take this responsibility very seriously, and I'm fortunate to be part of an incredible team of astronomers and engineers responsible for making these decisions.

And what do you like least about your work?

The downside of being at STScI is the overhead of working with the government which is substantial and affects work in interesting and surprising ways. NASA is an amazing organization that has done and continues to do incredibly ambitious things that further our scientific understanding of the universe, but fundamentally, working within government bureaucracy can be tiring and very frustrating at times.

Based on your journey, what advice or suggestions do you want to pass on to someone who's currently finishing their PhD?

If you can, talk regularly to people that know you and whose opinion you trust. One thing I've benefited from immensely is having a small number of people who I've been able to bounce ideas off about possible career moves and professional opportunities. The most important of these people for me is my wonderful partner Laura – her counsel, support and willingness to embark on crazy adventures have been an essential part of my professional journey.

Is there anything else you'd like to tell someone reading this interview?

I consider it part of my job in life to try and help others with their careers, especially people who are thinking that a traditional academic career path might not be right for them. If you want to talk more, feel free to hit me up on Twitter (@arfon).

Thank you so much for telling us about your work, Arfon!

Chris: It's been a few months since your interview and you are in a new position now, can you tell us about your new job?

Arfon: Sure thing! I've actually returned to work at GitHub in the product organization (I previously worked at GitHub from 2013 to 2016). There was absolutely nothing wrong with my job at STScI (I *really* enjoyed it), but ultimately my family and I decided we needed to be back in Europe to be closer to family.

In my new role, I'm a product manager in the Data organization at GitHub. Data includes teams responsible for data engineering, data science/machine learning/AI, analytics and more. The work of these teams is realized in a number of ways including working with product teams not only to build new customer-facing functionality (e.g. data-powered products) but also to support internal teams who want to understand the business through insights derived from data.

As a product manager, my role is to guide the work of these teams, helping us all understand who our customers are, what the most impactful work is that we should be doing right now. Working with leadership to set a vision for the Data organization for the next 12–24 months is also an important part of the role.

Can you tell us more about how your previous GitHub position differs from your new position?

In my previous position, I was in more of an outreach/engagement role with some program management. Ultimately this work was quite mixed – connecting lots of threads across the organization, making sure that GitHub was supporting the needs of particular communities either working on or researching the platform.

You can find lots of articles about the difference between program and product managers (e.g. https://medium.com/pm101/the-difference-between-product-program-and-project-management-64e2f1ee4f01). In truth, there's lots of overlap, and depending upon the organization, I think they can be very similar. In my case though, the biggest difference is that I'm now directly working with a number of engineering and data science teams, which means my input and direction directly influence the ongoing work of these teams. In my previous role, I had very few engineering resources allocated my way so had to spend much more time advocating for a particular approach/direction.

In short, I think my new gig is an upgrade on the last one at GitHub.

How do you think having a PhD helps you succeed in your current position?

Yes, I think it does. There's a fair amount written online, mostly in blog format or presentations, about how academics can make good product managers (e.g. http://blogs.nature.com/naturejobs/2018/06/06/how-product-management-could-be-a-route-out-of-academia-for-phds-and-postdocs/). Some of the skills cited include communicating ideas effectively, working with broad cross-functional teams, having focus in a position of uncertainty and knowing how to ask the right questions and use data to help make decisions. All of these are traits that are generally desirable for product managers but are also core skills for research. While I would never tell anyone to go get a PhD in order to become a product manager, I think there's a surprising amount of overlap in the skills required to be successful across the two roles.

That was very informative. Thank you for the follow-up interview!