Starting a Data Science Organisation.
Achieving Sustainability.

Read the full blog post:

*R. Stuart Geiger, Alejandra Gonzalez-Beltran, Robert Haines, James Hetherington, Chris Holdgraf, Heiko Mueller, Martin O'Reilly, Tomas Petricek, Jake VanderPlas, Matthew Archer, Stephen Dowsland, Rosa Filgueira, Sanaz Jabbari Bayandor, David Mawdsley, Tom Redfern, Valentina Staneva, Mark Turner, Kirstie Whitaker**

Data and software have enmeshed themselves in the academic world, many universities wish to improve their ability to create software tools, and spread the use of "data science" methods in the academic community. This has led to a common model: the creation of institutes that act as independent hubs of knowledge that can be deployed across the academic field in a 'science as a service' style consultancy. But setting up such institutes proves challenging as they sit outside the traditional academic style of departmental management. So how do we successfully establish an institute that is sustainable over the long term?

tl;dr:

"Create broad and flexible funding models, build a positive culture and environment where all members feel valued, and define career trajectories that cater to the diverse goals of members within the organisation."

Build flexible funding models.

- **Private funding:** Private philanthropic organisations such as the Moore-Sloan Data Science Initiative.
- **Organisational Grants:** Granting organisations such as NSF and UK Research Councils such as EPSRC who currently fund the Alan Turing Institute.
- **Project Based Grants:** Grants can be made available for the development of software for scientific research such as that by the Sloan Foundation for the NumPy software project.
- **Individual Grants:** Grants can be given to organisations that give their RSEs principal investigator status.
- **Paid Consulting:** Institutes can charge consulting services internally to the university and to external partners.
- **University Funding:** In many cases the university will recognise the benefit of an RSE institute and can provide funding as part of their budget for in-house services for students and researchers.
- **IT Connections:** Institutes can establish themselves within the wider university IT departments as the 'Research Wing' of university IT to benefit from the often large IT budgets.
- **Professors of Practice:** Popular in the US, is the hiring of experienced professionals outside of the traditional remit of an academic, who provides expert knowledge and practical experience in order to support research but who do not have to perform the traditional activities of an academic.
- **Research Librarians:** Parallels can be drawn between how academic libraries have supported long term research library staff who assist in wider research projects and have their own career paths within that area.

Create positive environments.

- **Physical Space:** Spaces that encourage collaboration such as break-out areas and co-location.
- **Get Started Early:** Connect the group to other researchers in the university early on.
- **Celebrate Each Other's Work:** The organisation should recognise many forms of scientific output, including tools and software, coming from data science institutes.
- **Allow Free Headspace:** Individuals should be allowed to work with researchers in other domains and support other research projects while also being able to explore their own research ideas.

Establish career trajectories.

- **Professional Development:** Institutes should prepare their staff with technical and soft skills such as software development best practices and also communication and team work.
- **Prepare for the next step:** Staff must always be preparing for the next step in their careers, be that within academia or into industry.
- **The revolving door:** Foster a 'two way street' movement between academia and industry to allow the movement of people between the two and benefit from the sharing of knowledge and best practice.

Research software engineer or data scientist?

As academia has embraced information technology and data science concepts we have seen the emergence of these two terms. Traditionally an RSE has bridged the gap between scholar and practitioner, creating the tools required for the research, while a data scientist uses the software to perform research and understand data. However, we find that the two roles share many similarities with each role performing interchangeable tasks. A deep understanding of programming, testing, scaling, and data wrangling is shared by both professions along with shared attitudes of multidisciplinary agility. So should we still use these strict titles or should we focus more on the skills required to get the job done?

*This poster was written collaboratively by individuals and is not an official publication of any of the above o