From Stars to Baht: Broadening the economic impact of astronomical data handling techniques in Thailand

Pathways to Impact

The primary goals of the proposed project are to expand the economic impact of our previous GCRF-funded research and to build capacity within our research team to work with a broad variety of external businesses and organisations. The primary focus of the Newton project has been to use astronomical data give Thai data scientists and their students experience of handling and analysing significantly larger datasets than they had previously been exposed to. No Newton funds were requested to work with external partners, as we saw training as the first crucial step to our long term goal of developing home-grown talent to promote economic development in Thailand. In the process we have learned new skills and have researched novel databasing and machine learning techniques to overcome some of the challenges inherent to storing and analysing large, constantly-updated digital datasets. Our follow-up GCRF Foundation Award has enabled us establish successful collaborations with a small number of local businesses, during which we have demonstrated our ability to engage students in working with external partners. We now request follow-on GCRF funding to fully capitalise on our skills and experiences to ensure that MFU becomes a self-sustaining centre for research and training in data science within Thailand.

The GCRF project has been carefully designed to maximise impact on both our external partners as well as on team members. For the former, this corresponds to maximising productivity, while for the latter this is in the form of capacity building by getting exposure to working with external partners – valuable experience for setting up future collaborations beyond the lifetime of the GCRF funding. Our overall strategy of working with a limited number of pre-identified external partners is based on maximising the impact of our team's *current* skills and techniques. In doing so, our resources will not be spread too thinly, enabling the team to gain a deep understanding of each partner's data and analysis needs, thereby ensuring the delivery of effective solutions. This strategy also ensures that there is already an understanding among the external partners that part of their role in the collaboration is to provide researchers and students a learning experience of working with outside clients.

As well as our overall strategy, we have also placed maximising impact at the forefront of each individual element of the work plan. The networking event at the start of the project enables our team to showcase our data handling/analysis skills and techniques, using GOTO as an example, while also enabling the external partners to describe the data and analysis needs and their desired outcome of the research. This ensures that the whole team is aware of the needs of the external partners from the outset, not just the primary contact, thereby facilitating greater collaboration within the team leading to greater impact.

Maximising impact is also a major consideration in planning how we will work with the external partners throughout the period of the grant. From the team's own experience from collaborating on our Newton and GCRF Foundation Award projects, regular meetings are by the far the most effective means of making progress when designing data handling and analysis systems to meet a client's needs. This is because the client has a better sense of what data fields are more important and/or informative than others, and this must be effectively communicated to the data scientists in order for them to design an effective database or analysis system. Usually, however, this is an iterative process that evolves over multiple meetings, which is why we have planned for fortnightly meetings throughout most of the period of the project. This also has the benefit of maximising team members' exposure to working with external partners – valuable experience for building collaborations with larger numbers of businesses and organisations after this initial GCRF grant. Finally, we want to make sure that the data solutions we provide to our external partners are effective, which is why we will deliver versions of our systems at the start of month ten for Beta testing. After testing the systems for a month, we will ask our external partners to provide feedback and we will make modifications during months eleven and twelve to ensure we meet our external partners' needs by the end of the grant.

The research component of the project is to adapt and advance the databasing and machine-learning techniques we have developed so far to meet the needs of our external partners. An immediate, qualitative measure of the impact of this work will be in the form of the feedback that we receive from our external partners during the third networking event at the end of the project. Once the systems have been adopted by the external partners, we will be able to quantitatively measure the impact of our research and compare against expectations laid out during the first networking event. In the case of our business partners, this will be in terms of increased productivity (i.e., turnover or profit per hour worked), whereas in the case of the educational organisations this would be in terms student admissions or staff productivity. Clearly, these numbers will only become available some time after the end of the grant, but we are committed to obtaining them since the project may prove to be an eligible REF impact case study. We will, of course, forward this data to STFC for their own impact monitoring.