

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

Pathways to Impact

The goal of the proposed project is to maximise the economic impact of our Newton-funded research. The primary focus of the Newton project has been to use astronomical data to train Thai data science research students in 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. We now request GCRF funds to enable the students and staff to capitalise on these skills and adapt our techniques to increase the productivity of Thai businesses and organisations, thereby promoting Thailand's economic development.

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 terms of getting exposure to working with clients – 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. With the skills, techniques and experience our team will acquire from this project, we will apply for follow-on GCRF funding (from 2019) to support an open call to work with any eligible external partner. Beyond that, our team will have sufficient experience to become self-sustaining, thereby ensuring maximum impact over the long term.

As well as our overall strategy, we have also placed maximising impact at the forefront of each individual element of the workplan. By visiting the UK to present the outcome of their Newton-funded research to the GOTO collaboration, the Thai students involved in the project will gain valuable experience in communicating with their first "client". This will be good training for working with the external partners during the GCRF project.¹ The networking event at the start of the project will enable 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. This will ensure that the whole team is aware of the needs of the client 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 the Newton project, 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 (in that case, the client was the GOTO collaboration). 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 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 member's exposure to working with outside clients – valuable experience for building collaborations with larger numbers of external partners after this initial GCRF grant. Finally, we want to make sure that the data solutions we provide to our clients are effective, which is why we will deliver versions of

¹This visit was not budgetted-for in our Newton project as that earlier project's specific aim was to provide training in handling and analysing large datasets, not training in collaborating with external partners.

our systems at the start of month ten for beta testing. After testing the systems for a month, we will ask our clients to provide feedback and make modifications during month twelve to ensure we meet our client's 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 during the Newton project 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 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. 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 retention. 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.