

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

Case for support

1 Description of the proposed project

We request funds to research how the data archiving and analysis techniques developed as part of our previous STFC/Newton-funded project can be best adapted to address the needs of Thai businesses and organisations. To achieve this research aim, our multi-disciplinary team of Thai and UK scientists will spend 12 months working closely with a group of pre-identified local businesses and organisations based in northern Thailand that have specific data handling and analysis needs.

1.1 Background: Using astronomy data to train Thai data scientists

As outlined in our ODA statement, there is strong evidence that improving access to advanced data handling and analysis techniques is one of the most effective ways of increasing productivity in businesses and organisations. This is most sustainable when using home-grown talent to provide these services. However, Thai students typically lack access to very large amounts of digital data, restricting their training in “Big Data” science that is so crucial for the development of modern economies. To address this problem, we successfully applied for Newton funding (12-month project starting February, 2017) to train Thai students in databasing and analysing large amounts of astronomical data generated by the Gravitational-Wave Optical Transient Observatory (GOTO). GOTO surveys the full observable night sky every two weeks, delivering data on roughly seven million astronomical sources *every night*. This large, constantly-updated dataset is giving Thai data science students the opportunity to develop the skills and techniques needed to handle and analyse the kinds of Big Data generated by high-tech industries.

The primary *research* goals of our Newton project are to: (a) develop a database that is capable of storing large amounts of data that is updated on a daily basis and (b) develop fully-automated Machine-Learning (ML) algorithms capable of quickly and robustly categorising sources detected by GOTO. Our research into how best to achieve these goals has led to some novel solutions. For the database component we are developing a hybrid system that combines the structure of a relational database (such as that used by SDSS) with the flexibility of a non-relational database (which are popular within tech. industries). For the ML component, we are developing a new two-stage categorization algorithm which uses unsupervised ML algorithms to quickly remove obvious artefacts from our data before passing more ambiguous cases to a supervised ML algorithm. Since both goals are to address challenges associated with the archiving and analysis of large amounts of astronomical data, the science, technology and expertise involved in the proposed GCRF project has originated from work associated with STFC’s core Science Programme.

1.2 Broadening the impact of our research to local businesses and organisations

Our Newton-funded project focussed on using astronomical data to help train Thai postgraduate students in the efficient databasing and ML-based analysis of large digital datasets. At present, the impact of the our Newton project is limited to the people directly involved in the project (i.e., Thai and UK astronomers, Thai data scientists and their students), whereas our long-term ambition has always been to broaden the impact of our research to develop the wider Thai economy. In this regard the goal of this GCRF project is to use the technology and skills we have developed to increase the productivity of Thai businesses and organisations. To do this effectively, however, our team needs to first gain experience in working with such external parties to research how best to adapt our technologies and expertise to their data handling needs.

Here, we request funds to support partnership-building between our team and six, pre-identified Thai businesses and organisations (hereafter, “external partners” or “clients”). With this funding we will conduct research into how the technologies and expertise we have developed during the Newton

project can be adapted to meet our external partners’ data handling and analysis needs. For example, one of our external partners – the Thanapiriya Public Company Limited (<http://thanapiriya.co.th/>) – is a fast-growing retailing business in northern Thailand that wishes to research how ML-based analysis of their sizeable customer database can increase the effectiveness of their marketing campaigns. Importantly, all external partners are based in areas local to the Thai co-Is and have already agreed to work with us during the 12-month project (see Letters of Support). At this stage, we prefer to limit the number of external partners (rather than, for example, having an open call) as it ensures that all parties have a thorough understanding of the goals and expectations of the project from the outset. Our ambition is that, once we have gained experience of working with our pre-identified external partners, we will request follow-on GCRF funding (from 2019) to widen the impact of this work to a larger and more diverse number of external partners.

1.3 Description of work to be undertaken

Over the twelve month period of the grant, we will (in chronological order; see Gantt Chart):

1. Have the Thai staff and students currently working on the project visit the UK to give them further experience of interacting with the GOTO collaboration – the first “client” of the project. The purpose of this visit is provide the students with first-hand experience of presenting to and communicating with – what is to them – an external partner.
2. Host a networking event in Chiang-Rai, Thailand (where most of the Thai co-Is and external partners are based). In attendance will be all UK and Thai team members and representatives from all six external partners. At the event, our team will deliver a series of short (5-10 minute) presentations using GOTO as a case study to highlight our skills and technology. Representatives from our external partners will be asked give short presentations to describe their business/organisation, what data they currently hold, and what they seek to achieve from the partnership.
3. Each external partner will be assigned at least one primary team member contact (this could be a postgraduate student, in which case a staff researcher will act as a secondary contact). The role of the primary contact is to attend on-site fortnightly meetings with the external partner to get to fully understand their data and analysis needs. From our experience, such close interaction is vital in order to discriminate between informative vs. non-informative data and, in some instances, derive more informative features from existing data. After each meeting, the primary contact will report back to the rest of the team to ensure a collaborative effort is maintained.
4. As the project progresses (and team members’ understanding of the external partners’ data and needs improve), our team’s work will focus increasingly on adapting our technology and techniques to the needs of the clients. As this happens, the purpose of the fortnightly meetings will progressively shift toward feedback sessions, during which the primary contact will demonstrate our solutions to the external partner and allow them to suggest improvements.
5. At the start of month eleven of the project, we will deliver to the external partners “beta versions” of the systems we have developed. At this time, we will train the external partner on how to use the systems and collect any immediate feedback they may have. Following this training, the system will go through independent, month-long beta testing by the external partner, who will then report their experience back to the team. Month twelve will be spent addressing any feedback from the beta testing phase to deliver the final product.

1.4 Maximising the impact of the project

We have taken many steps to ensure that the project has maximal impact for the current skills and development level of our technology. These are laid-out in detail in our Pathways to Impact statement. Briefly, our choice of working very closely with a limited the number of external partners is motivated by our desire to focus our efforts to ensure maximal impact both in terms of our partners’ productivity and our exposure to working with outside clients. The latter will prepare our team for working with a larger, more diverse, number of external partners, increasing the potential for greater impact in the future.

In addition to the impact within Thailand, the project promises secondary benefits to the UK. In particular, with STFC’s involvement in forthcoming data-intensive projects such as the LSST and SKA, it is vitally important that UK astronomers gain exposure to advanced data handling and analysis techniques. Further, the proposed work will give UK team members experience of working with data-intensive businesses – valuable preparation for research under the UK Government’s Industrial Strategy.

2 Management plan

The aim of this project is to research how to adapt and build-upon the databasing and machine-learning technologies we have developed to best satisfy the data handling/analysis needs of our six external partners. Our goal is that this research will increase our partners’ productivity, which will first be assessed qualitatively via partner feedback then, after 12 months, quantitatively through data analysis. The project is the next step in our long-term ambition to establish a self-sustaining “center of excellence” that will deliver data solutions to a wide range of businesses and organisations in northern Thailand.

Day-to-day management of the project will be the responsibility of Dr. Boongoen, Lecturer in the School of IT at MFU and Thai PI of the project. Following the first networking event, and on consultation with the whole team, Dr. Boongoen will assign primary contacts by matching team members’ skills and experience to the needs of the external partners. He will coordinate the fortnightly meetings between the external partners and their primary contact.¹ Dr. Boongoen will also ensure that the outcomes of the fortnightly meetings are communicated to the whole team via minuted face-to-face and/or video-conference team meetings. This will ensure that the data needs and challenges of the external partners are shared amongst the whole team, allowing each team member to contribute their own expertise. By pooling our resources and focussing on a limited number of external partners we mitigate the risk that the needs of any one partner will go unmet. This also means that the success of each external partnership is not the sole responsibility of the primary contact. Ultimate responsibility for each external partnership will lie with the Thai and UK PIs: Drs. Boongoen and Mullaney, respectively.

3 Track record of applicants

¹In circumstances where the student is the primary contact, they will be joined by staff researchers for their first two fortnightly meetings.