

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 goal, 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. As such, equipping Thai students with high-level data handling and analysis skills will play an important role in the continuation of Thailand's economic development.

Fortunately, there already exists in Thailand a knowledgebase in data handling and analysis that is able to teach the next generation of data scientists. The problem, however, is that Thai students typically only have access to small amounts of data. This places severe limits on their ability to develop the skills and experience necessary to analyse the large amounts of digital data generated by modern industries. To address this problem, we successfully applied for Newton funding (awarded, January, 2017) to train Thai students in databasing and automatically analysing large amounts of astronomical data. The data they are using is being generated by the Gravitational-Wave Optical Transient Observatory (GOTO), of which the PI and some co-Is are members. GOTO surveys the whole night sky every two weeks, delivering data on roughly seven million astronomical sources *every night*. This large, constantly-updated dataset is providing Thai data scientists and their students the experience they need to develop the skills and techniques needed to analyse the kind of "Big Data" generated by high-tech industries.

The primary *astronomy* 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. In meeting these goals we have developed some novel solutions. For example, 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 gaining popularity in 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 project has originated from work associated with STFC's core Science Programme.

1.2 Broadening the impact of our work: A feasibility study

Our Newton-funded project focussed on training Thai postgraduate students in the efficient databasing and ML-based analysis of large digital datasets. Importantly, the experience has also provided the Thai co-Is vital exposure to Big Data analysis, while introducing the UK and Thai astronomers to novel ML and databasing techniques. At present, however, the impact of 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 the wider Thai economy. In this regard our ultimate goal 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 ascertain how best to adapt our research, technology and training to their data handling needs.

Here, we request funds to support partnership-building between our team and Thai businesses and organisations with the express goal of establishing how the technology and skills we have developed during the Newton project can be adapted to meet their data handling needs. Importantly, we have pre-identified a set of **XXX** Thai businesses and organisations – all based in areas local to the Thai co-Is – that have 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 will have a thorough understanding of the goals and expectations of the project from the outset. Our expectation is that, once we have gained experience of working with our pre-identified external partners, we will request follow-on funding to widen the impact of this work to a larger number of local businesses and organisations.

1.3 Description of work to be undertaken

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

- Have the Thai 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.
- Host a networking event in Chiang-Rai, Thailand (where most of the Thai co-Is are based), attended by all UK and Thai team members and representatives from all our pre-identified external partners. During the event, our team will deliver a series of short (5-10 minute) presentations to highlight our skills, using GOTO as a case study to give clear examples of deliverables in terms of storing and analysing large digital datasets. Representatives from our external partners will also be asked give short presentations to describe their business/organisation, what data they currently hold, and what they seek to achieve from the collaboration.
- Each external partner will be assigned at least one primary team member contact (could be a student, in which case a staff researcher will act as a secondary contact). The role of the primary contact is to have on-site fortnightly meetings with the external partner to get to fully understand their data and analysis needs. From our experience with GOTO data, 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.
- As the project progresses (and our understanding of their data and storage/analysis needs improve), our team’s work will increasingly focus 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 client and allow the client to suggest improvements.
- At the start of month eleven of the project, we will deliver to the end user “beta versions” of the systems we have developed. At this time, we will train the end user 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 end user, 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 a number of steps to ensure maximal impact for this stage of the project:

- Independent research has shown that the type of robust data storage and analysis provided by this project is one of the most effective ways of increasing productivity within businesses and organisations.
- Focussing on a limited number of enthusiastic external partners that are local to the Chiang Rai area ensures we can have an active, face-to-face dialogue with the end users throughout the project. From our experience with the GOTO project, this is by far the most effective means of meeting the needs of the end user.
- Working directly with the end user will provide researchers and, in particular, students experience of working with external partners. This will enhance their ability to work with other businesses and organisations in the future (whether as employees or as part of other research projects), thereby ensuring continued impact beyond the lifetime of the proposed project.

1.5 Secondary benefits of the project

Important secondary benefits of the project include:

- Giving UK astronomers involved in the project exposure to a broad range of data storage and analysis techniques. Such skills will become increasingly important with STFC’s involvement in forthcoming data-intensive projects such as the LSST and SKA.
- Providing the UK team members with valuable experience of working with data-intensive businesses. This will help prepare them for working with UK businesses – identified as a key priority of 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 **XXX**. Following the first networking event, and on consultation with the whole team, **XXX** will assign primary contacts by matching team members’ skills and experience to the needs of the external partners. **XXX** will coordinate the fortnightly meetings between the external partners and their primary contact.¹ **S/He** 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, we mitigate the risk that the needs of any external 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. **XXX** and Mullaney, respectively.

3 Track record of applicants

References 1. Fabian 2012, ARA&A,50,455;

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