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

Case for support

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 data scientists and their 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) give Thai data scientists experience of working with 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 has given Thai data scientists the opportunity to develop the skills and techniques needed to handle and analyse the kinds of Big Data generated by many of today's industries.

The primary *research* goals of our Newton project were 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 catagorising sources detected by GOTO. Our research 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 at 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 unsurpervised 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 archiving and analysing 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

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, our team needs to first gain experience of working with such external parties to research how best to adapt our technologies and expertise to their data handling and analysis needs.

Here, we request funds to support partnership-building between our team and five pre-identified Thai businesses and organisations (hereafter, "external partners"; see Table 1). 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. All our external partners are based in areas local to the Thai co-Is and have already agreed to work with us throughout the 12-month project. At this "capacity building" stage, we prefer to limit the number of external partners as it ensures that all parties have a thorough understanding of the goals and expectations of the project from the outset. Our aim is to learn lessons and define commonalities of approach that can be applied when we expand the project to work with larger numbers of external partners. Our ambition is to apply for follow-on GCRF funding to support this expansion.

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. The purpose of this visit is provide the students with first-hand experience of presenting to and communicating with their first external partner – the GOTO collaboration.

Partner	Data handling/analysis needs
Thanapiriya plc	A food retail business that seeks a system that can take multiple factors into account to predict optimum stock volumes. This is a categorisation task in terms of increase/decrease/no-change of product demand.
TAPP Auto	A second-hand car business seeking to develop a system that can predict the depreciation curve of a given vehicle given multiple input factors.
Pibulsongkram Raj. University	The Academic Resoures Office wishes to identify the optimum online resources to meet the needs of different types of users. This can be achieved through ML-based Personalised Service Provision as often used by streaming services.
MFU	The Student Office seeks a system to identify the causes of the high student dropout rates in Thai Universities. This is, in part, a catagorisation problem, since students can be grouped according to different drop-out factors.
M-Store	A retail complex based on MFU's campus seeks to increase footfall by targetting promotions to specific groups. Again, a ML-based clustering analysis will help to identify different groups according to customer information.

Table 1: *Our external partners and their data handling and analysis needs.*

2. Host a networking event in Chiang-Rai, Thailand, where most of the Thai co-Is and external partners are based. All team members and representatives from all external partners will attend. At the event, our team will deliver a series of short presentations to highlight our skills and technology, using GOTO as a case study. Representatives from our external partners will describe their businesses/organisations, the data they 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 primary contact will attend on-site fortnightly meetings with the external partner to get to fully understand their data and analysis needs. Such close interaction is vital in order to discriminate between informative vs. non-informative 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 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 external partners. As this happens, the purpose of the fortnightly meetings will progressively shift toward feedback sessions, during which the primary contact will demonstrate our systems and allow the end user 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.

1.3 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.

2.1 Management plan

The aim of this project is to research how to adapt and buld-upon the databasing and machine-learning technologies we have developed to best satisfy the data handling/analysis needs of our five 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.

2.2 Track record of applicants

Our multi-disciplinary team of reseachers is made up of astronomers, computer and data scientists, a computer hardware specialist, and Lecturer in Business Management and Marketing. The UK PI (Mullaney) is an astronomer with extensive experience of analysing data from large astronomical surveys. He is the UK PI of our Newton-funded project, responsible for ensuring that the research meets the needs of the GOTO collaboration. The Thai PI (Boongoen) is a computer scientist with specialist expertise in developing ML algorithms. He has extensive experience in managing research projects, having PI'd four successful grants in the last four years. Dr. Iam-on is data scientist with expertise in database design, data mining, and developing ML algorithms for automated data analysis. She will oversee the database design elements of the project as well as contributing offering her expertise on the automated analysis aspect. Dr. Eungwanichayapant's background is in high energy astrophysics, with particular expertise in developing unsupervised ML algorithms to analyse data from Gamma Ray telescopes. Drs. Sawangwit and Awiphan are astronomers based in Thailand. Their research expertise lies in analysing large astronomical datasets and time varying data, and thus are very relevant to the project. Their experience of conducting science outreach will be used in our analysis of outreach data callected by NARIT. Mr. Vattayasak will be the team's computer hardware expert: his specialism is in setting up distributed networks of computers to host large, distributed databases. Finally, Ms. Noichankgkid is a Lecturer in Business Management and Marketing, with extensive management and finance experience prior to and during her academic career. This experience will prove invaluable when liasing with and establishing the needs of our business partners.

---

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