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 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 northen 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, equiping 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 catagorising 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 at 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 unserpervised ML algorithms to quickly remove obious 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 astronomica 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 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 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 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 should the project be funded (see Letters of Support). At this stage of the project we prefer to limit the participants (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 intention 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 greater number of local businesses and organisations.

1.3   Description of work to be undertaken

1.4   Maximising the impact of the project

1.5   Secondary benefits of the project

2   Management plan

3   Track record of applicants

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