

**UK – Thailand**

**Capacity Building in Software and Hardware Infrastructures and Data Handling through Astronomy 2017**

**Case for Support**

This template should be completed using: Arial (or an equivalent) and a minimum font size of 11. A minimum of single line spacing and standard character spacing must be used. Margins must not be less than 2cm and the document must stay within the page lengths specified for each section.

1. **General Information**

**Project Title** *[up to 150 characters]*

|  |
| --- |
|  |

**Theme**

*Please identify which themes your proposal covers:*

* Mechatronics/telescope control
* VLBI engineering, technology and research
* Data handling
* Outreach to support STEM education programme in schools

All projects should include some element of outreach activity.

|  |
| --- |
| Data handling |

1. **Previous track record of applicants and links between proposed partners**

*Please summarise how the UK and Thai partners will work together and any previous interactions or experience of working internationally and collaboratively. Please take into account track records of the applicants and institutions and provide details of any facilities that are required to undertake the project outside of the host institutions. [Maximum 1 side of A4]*

|  |
| --- |
| The Thai and UK researchers will collaborate to broaden and build upon the success of our previous Newton-funded project: “Using astronomy surveys to train Thai researchers in Big Data analysis”. During that first phase of the project, our multi-disciplinary team has collaborated closely to successfully train Thai graduate students in advanced data handling and analysis techniques, covering machine learning and database management. This has only been possible through the effective pooling the full team’s collective resources and experience. Specifically, the UK and Thai astronomers’ access to and understanding of the subtleties of analysing extremely large datasets combined with the Thai data scientists’ knowledge of database design and machine learning techniques has fostered a highly productive multi-disciplinary training and research environment for the Thai graduate students involved in the project. The skills that the graduate students have acquired through involvement in the project are directly applicable to economic sectors outside astronomy such as information technology, medicine, finance, modern agriculture, security, etc  Here, we request funds to support the next phase of this project in which we will: (a) disseminate the knowledge and skills we have acquired to date through a series of practical workshops aimed at research staff and students based at NARIT and other Thai research institutes, (b) continue our successful graduate training programme in which students gain advanced data-handling techniques through cutting-edge research projects. The knowledge gained from these graduate research projects will be disseminated during practical workshops held in year 2 of funding. A key component of both these project goals will be the establishment of a data centre based in NARIT which will be used as a training aid for staff and students. **Can this actually be achieved? Explore the possibility of cloud-based systems such as AWS?**  Over the past nine months of cross-disciplinary collaboration, the members of our team have learned significant amounts regarding each other’s area of expertise. The UK and Thai astronomers now have a far greater understanding of the strengths and limitations of the tools used by data scientists to organise and analyse data. Similarly, the Thai data scientists have a deep understanding of the type of data used by astronomers, and the challenges associated with extracting meaningful information from that data. As such, we are now in an ideal position to use this knowledge to  **The graduate research component**  The UK and Thai astronomy partners’ access to and understanding of the subtleties of analysing extremely large datasets together with the Thai computer scientists’ knowledge of database design and machine-learning techniques has  During that first phase of the project (lasting twelve months from 1st Feb. 2017) the UK and Thai partners have collaborated closely on two key research and training areas: (a) developing machine learning (ML) algorithms to automatically analyse astronomical survey images and (b) setting-up a data management systems (DMS) to archive the large amounts data produced by astronomical surveys. The primary role of the UK collaborators (both research astronomers) has been to deliver the astronomical datasets, which has predominantly been in the form of data simulated to reproduce that obtained by the Gravitational-wave Optical Transient Observatory (GOTO). When fully operational, GOTO (of which NARIT is a full member) will survey the entire observable night sky every four weeks, delivering measurements of over 7 million sources every night. We chose to work with simulated data as GOTO, while operational, is currently undertaking engineering observations and the data quality and format are subject to significant change. During this time, the simulated data have provided a more stable platform with which the to train Thai MSc and PhD students in data handling and processing.  The primary role of the Thai collaborators has been to train Thai PhD and Masters students in developing (a) ML algorithms to automatically analyse the simulated datasets and (b) a bespoke database management system capable of storing and efficiently retrieving the large amounts of data (in tabular form) measured from the simulated images. Despite the significant learning curve at the start of the project that is inevitable to multi-disciplinary research, this project has been highly productive. For (a) we have researched and developed a ML-based algorithm that is capable of handling extremely the unbalanced data (i.e., where the number of true positives are massively outnumbered by the number of false positives) inherent to astronomical surveys such as that undertaken by GOTO that are searching for extremely rare events. A description of this work – lead-authored by one of the Masters students in the group, Miss Aireen Tabacolde – has been accepted for publication in the proceedings of the "10th International Conference on Machine Learning and Computing”, at which Aireen has also been selected to give an oral presentation. For (b) we have installed and developed two separate DMS – one a traditional relational system based on potsgresql (similar to that used by SDSS), the other a non-relational system based on Hadoop that is more typical to that used in tech industries, but rarely used within astronomy. Our team are currently in the process of testing both approaches to see which is more suited to the frequently updated, large amounts of data delivered by GOTO. |
|  |

1. **Official Development Assistance (ODA) compliance**

*Please provide a statement and* ***evidence explaining******how*** *your proposed research is compliant with Official Development Assistance (ODA) guidelines. Proposals must contribute towards the economic development and welfare of Thailand. For more information on ODA please refer to the* [*http://www.newtonfund.ac.uk/about/what-is-oda/*](http://www.newtonfund.ac.uk/about/what-is-oda/) *and* [*RCUK Newton Fund Guidance*](http://www.rcuk.ac.uk/documents/international/odaguidancercukspecific-pdf)*. Your ODA compliance will be assessed from this statement so please ensure you consider this in detail.* ***If your proposal is not considered ODA compliant it will be rejected.***

*[Maximum 1 A4 page]*

|  |
| --- |
| Pretty much the same as last time |

1. **Detailed Research Information**
2. **Current landscape**

*Describe how the current priorities and challenges in Thailand and the UK will be addressed through this project demonstrating knowledge and understanding of past and current work in the subject area*

*[Maximum 1 side of A4].*

|  |
| --- |
| Similar to last time, I guess. |

1. **Description of proposed work**

*Please describe the research and partnership activities proposed. Please also identify any risks and mitigation strategies.*

*[Maximum 2 sides of A4]*

|  |
| --- |
| Disseminate what we’ve already learned re: ML and databasing via a series of summer workshops.  Install database infrastructure on NARIT’s HPC; providing training on installation and use during the summer workshops (attended by who)?  For research projects:  ML: further research on imbalanced data, deep image-based deep learning  DB: query design and optimization |

1. **Human Participation**

*Please provide the following information* ***if your project involves any kind of human participation*** *(please note that this is a requirement of the call). Any human participation must abide by UK and Thailand standards whether it is taking place in the UK or Thailand. Failure to complete this section may result in your proposal being rejected. If the project does not involve humans, please write ‘Non applicable’:*

* *Please indicate* ***where*** *the recruitment of the human participants/ samples/ tissue will take place and the appropriate agreements.*
* ***Please identify any ethical or health and safety issues arising from any involvement of people, human samples or personal data in the research proposal. Please give details of how these will be addressed and any specific risks mitigated.***
* *Please explain how the proposed research will be carried out to a high ethical standard* ***and how the research will abide by relevant legal requirements in the UK and Thailand.***
* *Please indicate the ethical approvals and* ***research governance arrangements that*** *will be sought/ have been sought and will be in place ahead of starting the grant in both the* ***UK and Thailand****.**(This may include arrangements for supporting and providing expert ethics advice to researchers, should unanticipated ethics issues arise, throughout the lifecycle of the grant.)*
* *If you’re using human samples/tissue please also provide information on the following:*
* *That what is being supplied is suitable for the research being undertaken.*
* *That the quantity of tissue being supplied is suitable, but not excessive for achieving meaningful results.*

*[Maximum 1 side of A4]*

|  |
| --- |
| Not applicable |

1. **Justification of animal use (if applicable)**

*Sufficient information and justification regarding* ***any animal research proposed****, regardless of country, must be provided. Any use of animals must abide by UK and Thailand standards whether it is taking place in the UK or Thailand. If the project does not involve animal use, write ‘Non applicable’*

*Applications including the use of animals should fully justify the animal use, including the following.*

*A statement that:*

* *They will adhere to all relevant national and local regulatory systems in the UK and Thailand.*
* *They will follow the guidelines laid out in the* [*Responsibility in the use of animals in bioscience research document*](https://www.nc3rs.org.uk/sites/default/files/Responsibility%20in%20the%20use%20of%20animals%20in%20bioscience%20research%20-%20July%202015.pdf) *and ensure that work is carried out to UK standards.*
* *Before initiation of the proposed research work, appropriate approvals from Institutional and/or central animal ethics committees will be obtained for experimental protocols to be adopted in their projects from both the UK and* ***Thailand (this is a requirement regardless of where the research is taking place).*** *Successful proposals may be expected to provide copies of these permissions before funding is released.*

*Please also detail the following:*

* *Please indicate* ***where*** *the animal research will take place (UK or overseas) and through which funder the resources are being sought. Applicants should include confirmation that animal welfare standards at the UK and Thailand institutions meet the requirements outlined above.*
* *Justification of the choice of design and numbers of animals and interventions.*
* *Adequate information concerning methodological issues.*
* *Information on the planned procedures to minimise experimental bias (for example, randomisation protocols, blinding) should be outlined or an explanation included as to why such procedures are not appropriate.*
* *Power calculations.*
* *The rationale for the experimental design.*
* *Any additional information which was not included in the proposal document but which is pertinent to the animal research proposed and which the funders should be aware of.*

*[Maximum 1 side of A4]*

|  |
| --- |
| Not applicable |

1. **Work plan**

*Please provide a Gantt chart, or diagrammatic work plan, for the project including milestones [Maximum 1 side of A4]*

**Signed by the UK and Thailand Partner**

*Date UK PI*

*\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Date Thai PI*

*\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Input more as needed*