Title: ***Pixel-based optical-to-infrared data fusion in the LSST wide and deep fields***

Proposing team: *Manda Banerji (Cambridge), Matt Jarvis (Oxford), Richard McMahon (Cambridge), Graham Smith (Birmingham), Omar Almaini (Nottingham), Rebecca Bowler (Oxford), Malcolm Bremer (Bristol), Will Hartley (UCL), Paul Hewett (Cambridge), Daniel Mortlock (Imperial), James Mullaney (Sheffield), Kathy Romer (Sussex), Nic Walton (Cambridge)*

Staff Requests: *1 PDRA full-time for 4 years; Dual photometry work package (WP 3.0) will be managed by Banerji (0.2 FTE/yr) and McMahon (0.1 FTE/yr). Banerji is funded by a Royal Society Research Fellowship and does not require FEC contribution. Science-validation (WP 4.0) will be managed by Jarvis (0.1 FTE/yr) and Smith (0.1 FTE/yr).*

LSST Science Collaborations: *Galaxies, AGN, DESC*

Scientific Justification: While LSST will undoubtedly revolutionize ground-based wide-field astronomy, its scientific return will be greatly enhanced by the addition of information at other wavelengths to the LSST optical data. In particular, by adding data at infrared (IR) wavelengths to the photometry in the six LSST optical bands, there is the opportunity to create a unique, legacy dataset that will open up the parameter space of galaxies explored by LSST. IR photometry for LSST stars, galaxies and quasars (and vice versa) will facilitate a very broad range of science. Below are examples of science areas where the addition of IR data is fundamental. See also Task 3.5.1 and Task 3.9 in the LSST Galaxies Science Roadmap [14], which are being directly addressed by this proposal.

- The Epoch of Reionization: The most distant galaxies and quasars at redshifts, $z>7$, corresponding to the Epoch of Reionization and beyond can only be discovered by combining information from LSST (where the galaxies appear as "drop-outs") and IR surveys (where the galaxies are robustly detected). Identification of large numbers of $z\sim7$ galaxies and quasars (e.g. [4],[8],[11],[12]) is an important milestone for both the LSST:AGN and LSST:Galaxies Science Collaborations [14]. It is an area where LSST will have early impact e.g. we expect ~10 new $z\sim7$ quasars and thousands of the most luminous $z\sim7$ galaxies in the LSST Commissioning and Science Verification data. These most distant galaxies and supermassive black-holes can *only* be identified by combining LSST data with IR surveys such as VISTA and Level 3 development (L3 DEV) work in this area is therefore critical.
- The Epoch of Galaxy Formation: At redshifts of ~1-3 corresponding to the main epoch of galaxy formation, IR photometry is essential to obtain robust stellar masses and star formation histories for galaxies (e.g. [10]) and therefore to build up a complete picture of the mass build-up of galaxies over cosmic time. IR data will also aid star-galaxy separation and object classification via spectral energy distribution (SED) modelling – e.g. see Chapter 11 of LSST Observing Strategy Document [7] - thereby enabling cleaner photometric galaxy samples to be produced over the entire redshift range sampled by LSST and benefiting multiple LSST Science Collaborations.
- Photometric Redshifts: Photometric redshifts (photo-z) underpin much of the science to be done with LSST. The addition of IR photometry to optical surveys can significantly improve photo-z performance; both reducing the scatter and catastrophic outlier fraction in photo-z estimates as key spectral indicators for old and evolved stellar populations shift into the IR at redshifts >1 (e.g. [1],[2],[3],[5]). More accurate photo-z's will in turn enhance LSST's ability to use photometric samples to constrain the dark energy equation of state (e.g. [13]).

The above list of science aims is by no means exhaustive and many other science areas will also benefit from IR photometry - e.g. (i) identification of cool brown dwarfs (ii) spatially resolved studies of low-redshift galaxies to connect colour-gradients to different morphological components (iii) studies of Early Type and Brightest Cluster Galaxies – to name a few. In preparation for the Phase B funding call we organized a 3-day workshop on the subject of multi-wavelength data fusion in the LSST fields during September 2017, engaging a large cross-section of the LSST:UK community (40 delegates) as well international leads of the LSST:Galaxies (H. Ferguson) and LSST:AGN (W. N. Brandt) Science Collaborations. As evidenced by the broad range of talks and discussions at that meeting (https://sites.google.com/view/lsstuk-multiwave/home), multi-wavelength data-fusion emerged as among the most important contributions that the UK could make to the LSST project.

The UK astronomy community has a long history in conducting wide-field IR surveys. The VIRCAM camera on the VISTA telescope is among the premier ground-based wide-field IR survey facilities operating in the Southern Hemisphere, and the majority of ESO Public Surveys being conducted with VISTA are led by UK PIs. The VISTA Hemisphere Survey (VHS; PI:McMahon) provides the widest area IR coverage over the entire LSST footprint in the J and Ks-bands. The Ks-band coverage over 20,000 sq-deg will not be superseded even with the launch of *Euclid.* The VISTA VIDEO (PI:Jarvis), VEILS (PI:Banerji) and UltraVISTA surveys provide IR coverage of the four announced LSST deep-drilling fields with deeper sub-regions imaged to unprecedented depths in the IR by surveys such as the UKIDSS-UDS (PI:Almaini). This legacy of IR survey astronomy in the UK will continue into the 2020s with the UK positioned as one of the major partners within *Euclid*. By developing algorithms to effectively combine LSST with the IR we will therefore be leveraging UK expertise in this area.

With this set of work packages (WPs) we propose to build on the UK's leading role in near IR surveys and implement a flexible and effective pipeline for ingesting the IR data in the UK data access centre (DAC) and processing these IR pixels jointly with LSST pixels to produce high-fidelity, science-ready optical+IR photometric catalogues for the community.

Work package description and justification of resources

WP1.0: Project Management (0.5 FTE/yr total; 0.3 FTE/yr funding requested for Jarvis, McMahon & Smith): The PDRA funded by this proposal will be based at Cambridge and the algorithm development and implementation will be managed by Banerji (0.2 FTE/yr; 0% FEC) and McMahon (0.1 FTE/yr). Data validation will be split into the deep-drilling fields (managed by Jarvis at Oxford; 0.1 FTE/yr) and wide-fields (managed by Smith at Birmingham; 0.1 FTE/yr). All science validation will be organised through the LSST Science Working Groups (WGs).

WP2.0: Data Management (0.2 FTE/yr from LUSC-DAC): The proposed work package relies on ingestion and curation of the relevant datasets within a UK data access centre (DAC). This effort is not requested as part of this application as we expect it to be funded from the DAC.

- Multi-wavelength data ingestion (WP 2.1): We will ingest pixel and catalogue data from IR surveys to the UK DAC including re-formatting metadata to comply with LSST stack processing. We will use VISTA, *Spitzer, WISE* and *HST* as the test-bed for ingesting other imaging surveys that will be available towards the end of Phase B, e.g. *Euclid*.
- Optical data ingestion (WP 2.2): Although LSST data will be the priority optical dataset in the DAC, pixel data ingestion of currently available optical surveys e.g. HSC and DES together with their associated catalogues, will be important for benchmarking the LSST outputs as well as facilitating unique science e.g. variability studies in the optical. The first public data releases for both HSC and DES will already be available at the start of Phase B.
- Data Curation & Queryable Database (WP 2.3): The outputs from our dual photometry pipeline (WP 3.0) will need to be curated and ingested into a queryable database thus making these products easily accessible by the community.

Deliverables: (i) Single-epoch and coadd images and associated weight maps and catalogues for all optical and IR surveys accessible to the UK community via the DAC (ii) Postage-stamp server with web-based access infrastructure (iii) Queryable database with dual photometry outputs produced as part of WP3.0 (iv) Documentation

WP3.0: Dual photometry pipeline (0.75 FTE/yr funding requested for new PDRA): A critical aspect of combining data from different surveys is ensuring that the photometry is measured consistently across all wavelengths. This is a challenge when trying to combine LSST data with data at lower angular resolution (e.g. *Spitzer, WISE*) and depth (e.g. VISTA VHS) or indeed higher resolution (e.g. *HST, Euclid*). The science goals outlined above will therefore often require pixel-level (rather than catalogue-level) matching between optical and IR surveys, incorporating a flexible and efficient forced photometry pipeline to deal with non-detections in one or more bands. In keeping with the LSST vision, the software will have a modular structure enabling new algorithms to be easily implemented within the pipeline, and it will be easily adaptable and configurable to be able to process new datasets e.g. *Euclid*.

- <u>LSST Stack Implementation on VISTA (WP3.1)</u>: The main implementation of the pipeline will be based on the forced-photometry module within the LSST stack. This module will be configured to perform photometry on VISTA IR images based on LSST detections and vice versa. The combination with the IR is not a Level 2 deliverable and therefore requires L3 DEV effort.
- <u>Implementation of other photometry pipelines (WP3.2):</u> Existing photometry pipelines (e.g. SExtractor, imcore(list)) used by current surveys such as DES and VISTA, will also be implemented in order to allow benchmarking. The LSST stack is still in developmental stages hence benchmarking is an important step. Direct comparisons between outputs produced using these different approaches will be made.
- <u>Different-Angular Resolution Datasets (WP3.3):</u> An alternative approach may be required to combine LSST pixels with pixels with much higher (e.g. *HST/Euclid)* or lower (e.g. *Spitzer/WISE)* angular resolution. One common method for conducting forced photometry in such cases is to use a surface-brightness model based on the high-resolution imaging, convolve the model with the PSF of the low-resolution data, and then perform consistent photometry between the two datasets. Algorithms such as Tractor [6] and TPHOT [9] both use such an approach. We will test these algorithms on *HST/Spitzer/WISE*, VISTA and LSST-like datasets (e.g. HSC/DES) and ensure the best algorithm (both in terms of accurate photometry and scalability) is implemented.
- <u>HPC Implementation (WP 3.4):</u> The workflow will be ported onto generic High Performance Computing (HPC) systems e.g. DiRAC, CC-IN2P3 (EU DAC).
- <u>Supporting Scientific Validation (WP 3.5):</u> The PDRA will support the scientific validation of data products from the pipeline (WP4.0). We envisage the products from our dual-photometry pipeline having broad applicability to other proposed LSST:UK WPs - e.g. de-blending (Smith et al.) and automated junk-rejection (Mortlock et al.) - and the PDRA can therefore also support interfaces with any funded WPs that will benefit from IR photometry for LSST sources.
- <u>Commissioning (WP 3.6):</u> Pipeline commissioning and testing will overlap with the commissioning and Science Verification phases of the LSST survey, allowing LSST pixels to be directly processed and combined with the VISTA, *WISE* and *Spitzer* data during this phase. The key timelines are (i) Phase 1 Commissioning using ComCam (Jan-Sept 2020) (ii) Phase 2 commissioning using LSSTCam (Oct 2020- Mar 2021) and (iii) Phase 3 Science Verification (Apr-Jun 2021). We will push for fields with good IR coverage to be observed as part of LSST Commisioning and Science Verification.

<u>Deliverables:</u> (i) Multi-wavelength catalogues with source-level metadata, detection and measurement image provenance information and workflow provenance information (e.g. configuration files) (ii) Documentation detailing pipeline implementation (iii) Results of benchmarking tests scoping out computational requirements. All outputs to be ingested into a queryable database within the UK DAC (WP 2.0).

WP4.0: Scientific Validation (0.25 FTE/yr funding requested for new PDRA + 0.6 FTE/yr PDRA effort funded by other sources): Scientific validation WPs will be managed by Oxford (deep-drilling fields) and Birmingham (wide-fields). Initially the data products will be multi-wavelength catalogues

produced by combining current optical photometric surveys such as DES and HSC with VISTA, *Spitzer* and *WISE*, while LSST pixels will be combined with IR pixels towards the end of Phase B. The PDRA funded by this proposal will devote 0.25 FTE/yr to facilitating science validation. Cambridge, Oxford and Birmingham will also contribute 0.2 FTE/yr of postdoctoral effort from other funding sources (e.g. relevant AGP-funded postdocs) to support the science validation effort.

- Data Quality Control (WP4.1): Assessment of noise properties, depths, seeing FWHM, stellar and galaxy colours produced using different approaches to photometry for the LSST, *HST*, HSC, DES, VISTA, *Spitzer* and *WISE* data.
- Star-Galaxy Separation (WP4.2): Assessment of the utility of IR data in improving star-galaxy separation. A key aspect of this work will be to enhance star-galaxy separation by exploiting the new colour information from the IR – e.g. Chapter 11 of LSST Observing Strategy [7].
- Photometric redshifts and SED-based classifications (WP4.3): Photo-z measurements underpin a significant fraction of the goals of the DESC, AGN and Galaxies Science Collaborations. The UK has strong involvement in the photo-z WG within DESC (e.g. Jarvis, Hartley). This proposal will ensure the best possible data is available for improving photo-z accuracy, and the scientific outputs from our photometry software will be validated by testing against the metrics developed by the photo-z SWG within DESC. This aspect will be managed by Jarvis who is also the photo-z link to the LSST:AGN Science Collaboration. SED-fitting will also allow us to derive galaxy types, stellar masses and star formation rates, and AGN reddening and host galaxy contributions.
- Other Datasets (WP4.4): This proposal focuses on the combination of LSST with IR data from VISTA, *HST, Spitzer* and *WISE.* However the UK also has strong interests in other multi-wavelength datasets, which can be combined with LSST during science validation. We expect key X-ray (e.g. *XMM-Newton, Chandra, NuSTAR, e-ROSITA*), FIR (e.g. *Herschel*) and radio (e.g. *MeerKAT, ASKAP*) data to be accessible to the UK community via the DAC and science validation efforts will also involve cross-matches to these other surveys. Ingestion of these multi-wavelength catalogues into the UK DAC will allow the UK community to leverage the optical+IR datasets produced as part of this proposal for maximum scientific impact.

Deliverables: Reports in the form of presentations and publications

Relevant Expertise of Proposing Team: Our team includes PIs of the major wide-area near IR extragalactic surveys that currently overlap the LSST footprint. These are VHS (PI:McMahon), VIDEO (PI:Jarvis) and VEILS (PI:Banerji). The proposers have significant combined expertise in the exploitation of near IR surveys with UKIRT (Almaini, Hartley, Hewett, McMahon, Mortlock) and VISTA (Banerji, Bowler, Jarvis, McMahon) as well as optical surveys such as DES (Banerji, Hartley, McMahon) that will serve as pre-cursors to LSST data. Banerji is the IR Liaison within LSST:UK and the High-Redshift Working Group Coordinator within the LSST:Galaxies Science Collaboration. McMahon is LSST:UK Spectroscopy Liaison, Jarvis is LSST:UK Deep-Drilling Fields Liaison and Smith is the Clusters Point of Contact and Commissioning Coordinator within LSST:UK as well as co-chair of the Clusters & LSS Working Group within LSST:Galaxies. Many of the proposing team are significantly involved in the LSST:Galaxies and LSST:AGN Science Collaborations as well as leading photo-z studies as part of the LSST:DE Science Collaboration. Mortlock will contribute expertise in statistics via links to the LSST Informatics & Statistics Science Collaboration.

References: *[1] Banerji et al. 2008, MNRAS, 386, 1219 [2] Banerji et al. 2015, MNRAS, 446, 2523 [3] Bezanson et al. 2016, ApJ, 822, 30 [4] Bowler et al. 2015, MNRAS, 452, 1817 [5] Jarvis et al. 2013, MNRAS, 428, 1281 [6] Lang et al. 2016, AJ, 151, 36 [7] LSST Science Collaboration 2017, arXiv:1708.04058 [8] Mortlock et al. 2011, Nature, 474, 616 [9] Merlin et al. 2016, A&A, 595, 27 [10] Muzzin et al. 2013, ApJ, 777, 18 [11] Reed et al. 2015, MNRAS, 454, 3952 [12] Reed et al. 2017, MNRAS, 468, 4702 [13] Rhodes et al. 2017, LSST-Euclid Synergy, arXiv:1710.08489[14] Robertson et al. 2017, LSST Galaxies Science Roadmap, arXiv:1708.01617*