ICRAR & ICRAR-Pawsey Summer Studentships 2018-2019
Project Proposal

Project Details	
Project Title	**An effective cross-matching framework for catalogues and images.**
Primary Supervisor	John Morgan
Primary Supervisor Availability	**All times except for the first two weeks of December**
Contact Details	john.morgan@icrar.org
Additional Supervisors & Contact Details	Paul Hancock paul.hancock@curtin.edu.au
Additional Resources Required	None
Pawsey Centre Hardware Use	Modest use of galaxy resources (a few hundred cpu hours). Mostly within the capabilities of a desktop machine if Pawsey is not available.
Software Required	List all software requirements here. Student Desktop Requirements: • Aegean • Standard python packages Pawsey Centre software installations required: • GLEAM data processing pipeline
Student Location for project	ICRAR-Curtin
Project Description	The Murchison Widefield Array (MWA) routinely makes observations that are adversely affected by the Earth's ionosphere. This means that the images and catalogues that are produced can be distorted and difficult to compare with data from other telescopes, and even with other MWA observations in different ionospheric conditions. We have developed algorithms to correct source positions for ionospheric effects, however they remain suboptimal, and some human intervention is occasionally required to filter out bad crossmatches. One feature of ionospheric distortions which has not been fully exploited is that the distortions are spatially correlated (i.e. nearby sources are distorted in a similar way). Furthermore, since we now have good MWA catalogues over almost the whole sky, we can also use source brightness or morphology to find the optimum match for each source. The project will therefore involve optimising and adding to our algorithms to allow many thousands of catalogues to be crossmatched automatically. Furthermore, when more than one crossmatch is feasible (or if there is a chance that a new source, not in the master catalogue, has been detected), then the probability of this should be robustly calculated. The project would suit a student who is interested in solving problems on very large datasets. This is a very generic 'big data' problem and the skills learnt by the student will be extremely useful either inside or outside astronomy.

Student Attributes	
Academic Background	At least some programming experience required
Computing Skills	Python programming experience might be useful, but certainly not required.
Training Requirement	The student will be taught to schedule jobs on the super computers, and will be taught techniques for interrogating our datasets and developing new algorithms in python in an interactive manner.
Project Timeline	
Week 1	**Pawsey training (or inductions and project introduction)**
Week 2	Get up to speed with current algorithms
Week 3	Identify bad matches in current data,
Week 4	Measure performance of current data and come up with a metric for how well algorithm performs
Week 5	Explore the use of a hierarchical fitting routine
Week 6	Continue to explore different crossmatching approaches
Week 7	Refine master catalogue
Week 8	Warp images using best-matched catalogues. Generate new GLEAM catalogue
Week 9	Compare with original
Week 10	**Final Presentation and Reporting**