

# Data Central Science Advisory Committee

30 June 2023

*Present:* Andrew Hopkins (Chair), Simon O'Toole (Head, Data Central), Simon Driver, Bruce Gendre, Cullan Howlett, Natasha Hurley-Walker, Sarah Martell

*Apologies:* Amanda Karakas, Christian Wolf

## Agenda

1. Introduction/Overview
2. Staffing update
3. Current projects update
  - a. Data Central – AAL/NCRIS
  - b. ARC LIEF ANU 2.3m
  - c. ARC LIEF LSST
  - d. ARC Linkage ORICL (Early childhood education)
4. Future funding
  - a. AAL/NCRIS base
  - b. NCRIS extra
  - c. UWA/MQ joint appointment (Simon Driver)
  - d. AusSRC/MQ joint appointment (possible)
5. AOB

## Minutes

Meeting opened at 11:00 AEST with Andrew welcoming the committee members.

### 1. Introduction/Overview

Andrew briefly introduced the meeting, noted that Simon would be providing updates on the agenda items, and that discussion is encouraged. In the event any discussion items might need longer than this meeting can support, they will be taken offline and progressed by email.

### 2. Staffing update

Simon noted that the Data Central team is having staffing limitations. This is in part due to the university structure with staff being on “professional staff” contracts, with no formal promotion pathway. He noted that DC has lost 5 staff this year to roles that have better conditions and salaries elsewhere. Software engineers are in the top 3 most difficult to fill roles, along with nurses and aged care workers. Have filled a casual role recently; the purpose of casual appointments is to assess staff for potential contract positions.

There is an ongoing recruitment drive. The need for C/C++ experience is harder to fill. Python, ML backgrounds are also needed and easier to find.

Simon D: Difficult to find staff, and difficult to retain them. We are never going to be competitive with industry, are finishing PhD students the potential resource?

Simon O'T: Sometimes yes. We have several in our team (James Tocknell, Brent Miszalski, Liz Mannering). We do want to have a variety of backgrounds though, people with industry experience bring a different perspective, and it's about getting a good balance.

Simon D: Do you advertise through the ASA?

Simon O'T: Yes, but also Seek, LinkedIn, and other channels. We are always interested in astronomers with strong software engineering experience. But we do need C/C++ expertise for specific projects, such as the ESO pipelines project. We do have some potential candidates lined up.

### 3. Current projects update

#### a. Data Central – AAL/NCRIS

Primary NCRIS-funded project, funded through AAL. Because of staffing shortages we are delayed in milestones, but we have agreement with AAL that these will roll over into the next financial year. Filling the open two roles will make sure we can meet these milestones in the coming FY. The main thing that we have done is refactoring the way data ingestion is handled to improve speed and scalability. Now needs to be relinked to the front end. This update will be released the week after the ASA. It will include a range of extra datasets, including the original GLEAM data, Gleam4Jy, GALAH, GAMA, SAMI, OzDES, and other legacy datasets, also the KiDS imaging and catalogues, and more. New TAP service will allow users to upload catalogues for server-side cross-matching.

Sarah M: Question about hosting at NCI.

Simon O'T: NCI is hosting the AAT archive, and we have a test-version that runs 2dFDR in an automated way, and in process of deploying that at NCI. Have held off on doing much more because of waiting for SkyMapper DR4, which is delayed.

Sarah M: Other things being moved over there?

Simon O'T: Not so far, but have a new head of cloud services (Andrew Howard), so there may be discussions around others, although that will require new funding. Requires that they give us access to VMs at NCI to install/ingest tools/datasets. Hoping to have another conversation in a month or two. It would allow us to have higher availability, and especially DNS persistence.

#### b. ARC LIEF ANU 2.3m

Project about to start. It is an automation project, proposal management system, to be used for 2.3m as well as the AAT. Also manage the data pipeline and archive (pipeline, called PyWIFES, being developed by others).

#### c. ARC LIEF LSST

Grant awarded at end of 2021, but MIA between 14 universities delayed project start. Three components: (a) IDAC (international data access centre). Will enable data interoperability between LSST and DC. (b) Low surface brightness data reduction pipeline development. (c) Data Central (together with Swinburne) is the team providing the in-kind software contribution (3 FTE) which is the element that gives Australian researchers access to LSST data.

#### d. ARC Linkage ORICL (Early childhood education)

ORICL stands for Observe Reflect Impact in Childhood Learning, and is an international project led by Macquarie with researchers also at Charles Sturt and McMaster (Canada). It's a \$100k contract, to provide infrastructure, data security and privacy systems. They have survey data that has been digitised, and DC is tasked with providing access and security, including MFA, tokens, etc.

Sarah M: How did this come about?

Simon O'T: Were approached by the School of Education at MQ, as DC is one of very few avenues for software effort. DC provided a quote for the proposal, which was not funded by the ARC but supported by MQ's safety-net funding. It integrates with other DC tools (accounts, teams, etc). Planning to roll this functionality into Lens as well.

Sarah M: What other things are in the works along these lines?

Simon O'T: We are on an NHMRC grant with Australian Hearing Hub, industry partnership, about \$85k, project is for de-stigmatising hearing loss.

Sarah M: How do these projects come about? Should it be more formalised?

Simon O'T: Made good connections early on with the eResearch group at MQ. DC is known throughout MQ as a group that can do software/data/web/UX type projects. Some are at the business development level, one is a website for a project on neurodegenerative diseases, but could also turn into data management for them as well.

Andrew H: Very big opportunity to use DC's "scalable and extensible" infrastructure to grow and expand DC into the wider university ecosystem beyond MQ alone, and into the commercial world. We should have a longer conversation among this group around how best to pursue that.

Simon O'T: Anticipating NCRIS funding, hopefully can be announced at the ASA next week. Some of this funding could be directed to non-astronomy dataset funding, as well as potentially support for Euclid data pipelines.

Simon D: Negotiation with Euclid that 2 FTE software effort would provide 5 Euclid membership slots.

#### **4. Future funding**

##### **a. AAL/NCRIS base**

Received a reduced allocation (\$600k), and requires a new umbrella agreement between AAL and MQ. Money from previous FY rolled over will support activity until new agreement finalised. Will invite research teams to nominate new functionality to enhance their data releases, will bring those requests to the Science Advisory Committee to advise on priorities.

##### **b. NCRIS extra**

Potentially an extra ~FTE of funding to cover the shortfall in the nominal NCRIS allocation. A set of itemised activities that this would be allocated to.

##### **c. UWA/MQ joint appointment (Simon Driver)**

Simon D: With 4MOST on the horizon, need to put the input catalogue into the public domain. UWA fund 0.5-FTE, DC fund 0.5-FTE, for 5 years, with the UWA half-funding to support 4MOST/WAVES development within the DC environment. Hoping this might be a useful model, by entangling Data Central with science-focused activities, providing external funding for DC, and ensuring key activities have the priority and resourcing needed.

Simon O'T: Similar to some early DC development. Focus on the software development from both sides is important.

##### **d. AusSRC/MQ joint appointment (possible)**

In discussion with AusSRC as their MIA is being finalised. Potential for a joint appointment with MQ, to support links with AusSRC for data release focus. They are currently doing a lot of data processing but need extra support on data release.

Sarah M: Would be nice to have an overview of the broadening of DC activities.

Simon O'T: Propose to meet in a few months to talk about strategic plans.

**ACTION (AMH):** Circulate poll to identify a meeting time for a focused discussion by this group on DC strategic directions, and funding support, for building beyond "just" astronomy archives.

#### **5. AOB**

Sarah M: Request for a written status update on the current projects.

**ACTION (Simon O'T):** Send status updates to the Committee by email.

Meeting closed at 12:04 AEST.