eResearch Australasia Conference 2016
Poster Presentations

Title	Authors	Abstract
Data as Social Capital and the Gift Culture in the Scholarly Community	Jens Klump	The value of research data as a resource for present and future research is broadly accepted but more than ten years after the first minting of a Digital Object Identifier (DOI) for data publication and citation sharing data from publications is not yet the norm. Statements by researchers on their willingness to share data come with the best intentions but rarely go beyond lip service and bear little resemblance with the actual release of data. Discussions about sharing data in science often revolve around "finding the right incentives". But are we offering the right incentives to researchers to share their data with others more freely? Maybe there are more fundamental social drivers that influence researchers in their decision to share data, or not.
Working Smarter With Custom Programming Robots	Sara Drieman	The number of tools that assist eResearch is rapidly growing, each one requiring its own skill set and specific syntax. Add to this the multitude of platforms to connect to and authentication required, and basic tasks can become quite time consuming. Many applications try and compensate for this by including API clients which allow quick command line tasks, but not all applications have this functionality, and the commands can be difficult to be made practical. Fortunately developers are constantly on the lookout for automation, and eRSA have implemented the use of bots to handle convoluted and repetitive tasks. Hubot is an opensource GitHub developed automation tool which acts as a broker between the user and the application, making use of API clients and other programming tools. When deployed on a messaging client, Hubot can be given tasks in simple text to invoke its programmed scripts. eRSA have written specific scripts to suit application needs, which will be demonstrated in this session. Together with GitHub's community library containing hundreds of scripts there are a vast number of practical applications for Hubot. Creating an application bot allows for simpler commands as the syntax can be chosen to fit what is logical for the user. By eliminating the user connection to the application, the commands can be run faster and in quick succession. Not only is this an effective use of time, but also a more secure way of managing access; rather than several team members having (potentially vulnerable) accounts, only the bot has the application logons. This session will include a practical demonstration of an eRSA application bot, showcasing the ease of use as well as highlighting potential use cases for Hubot.

Monitoring GPU Requests and Utilisation on the Bragg Accelerator Cluster	Ahmed Shamsul Arefin and Steve McMahon	The CSIRO's GPU accelerator cluster was the first of its kind in Australia when it was launched in 2009 combining traditional central processing units (CPUs) with graphics processing units (GPUs). At that time, it was also one of the fastest supercomputers running a Windows HPC operating system in the world. It not only ranked well on the TOP500 list of the world's fastest and most powerful supercomputers, but also ranked highly on the Green500 list - a ranking determined by energy efficiency, making it one of the most energy efficient supercomputers in Australia. It was further expanded in 2012 and named Bragg, after Australia's first Nobel prize-winners - father-and-son physicists Lawrence and Henry Bragg. According to the latest audit in Nov 2015, the Top500 ranked it 297 among the world's fastest super computers. Current configuration of the Bragg consists of 128 Dual Xeon 8-core E5-2650 Compute Nodes (a total of 2048 compute cores) and 384 Kepler Tesla K20 GPUs (a total of 950,976 CUDA cores). Bragg's workload is managed by an open source tool called Simple Linux Utility for Resource Management (SLURM) [3] and it is provisioned and managed by using a commercial software called Bright Cluster Manager (BCM). Bragg is an expensive system of GPU accelerated computers. Therefore, one of the many questions we receive from its closest stakeholders is, "how much of the Bragg's GPUs are being utilized by its users?". In an attempt to address this, we have retrieved data from BCM, NVIDIA Management Library (NVML) and SLURM and created a simplistic mechanism to measure and view the GPU requests and usage. In this work, we present our solution, its limitations and future works in this area.
monash.figshare - store | manage | collaborate | publish	Patrick Splawa-Neyman	monash.figshare (https://monash.figshare.com) is the new research data repository for Monash University. monash.figshare allows researchers to manage, store, share and publish their research files as citable research outputs with a persistent URL without mediation by the Library. Researchers will be able to control who can see their research files in a variety of ways. All research files will be stored at Monash University.
Leveraging open standards and open data delivery	Pavel Golodoniuc, Robert Woodcock, Lesley Wyborn and Ryan Fraser	The AuScope-developed technology based on open standards has had a profound science impact allowing open access to vast data holdings previously hardly accessible to researchers. The developed data delivery technology erased project boundaries, allowed sharing data with international community initiatives (e.g., INSPIRE, OneGeology), and equipped researchers with tools allowing application of new numerical methods to broader range of available data sets.

Starting the conversation! A graphic approach to talking data management	Lou Connell, Jennifer Murphy, Cameron Barrie, Pam Abalo and Jessica Cork	At Victoria University, the Library and the Office for Research work together to support researchers through the data management planning process. Victoria University has an established data management policy, a data management plan template and a dedicated storage drive for researchers. There are also extensive support materials available on the Research web page. The Research Librarian team runs generic 'Data Management' workshops aimed at researchers at all levels. Beyond this top level tier of support, the Research and College Librarians team identified an opportunity to 'start the conversation' with researchers about their data management at a grassroots level. College Librarians already have established relationships with the researchers in their discipline area so it was a strategic move to start talking about data management in these interactions. This graphic, as represented in the poster, was developed as a visual guide to start the conversation. The graphic was designed for both researchers and librarians, and has steps and options for researchers to progress through, encompassing data management planning, storage and format, and analysis tools. It also provides information to assist researchers to make decisions about collaboration and discoverability options, including open data. As librarians move into the role of providing data management support, they have had to increase their own knowledge and skills in this area. Developing this graphic enabled us to clarify our own understanding of the different facets and stages of data management. The outcome of this project had the dual benefit of developing College Librarians' expertise and providing a resource that they can draw guidance from in their conversations with researchers about data. Additionally, the data management graphic is available through the Victoria University website for researchers to access independently. The 'Plan, Store, Collect, Analyse and Discover' options in the graphic enable researchers to successfully manage their data.

Research Data Management in a Regional University: a Sisyphean battle?	Rebecca Owen and Beth Crawter	The research data experience began at the University of the Sunshine Coast (USC) when the Australian National Data Service (ANDS) approached us to see if we would be part of Seeding the Commons. Before this, USC had no data management in place, with individual researchers and faculties managing data in an ad hoc fashion. The Seeding the Commons project, funded by ANDS in 2011, was an eye-opener. We, the Library, suddenly realised the value of research data and the importance of managing it in a coordinated and professional manner. The challenge then was to convince the rest of the University. Over the ensuing five years, we have had many uphill battles with a few wins. Our biggest hurdles were where we least expected them. We are still trying to communicate the research data message.
Top 10 research data Things for medicine and health: a self directed learning program	Kate Lemay	In 2016 the Australian National Data Service (ANDS) ran the highly valued 23 (research data) Things program. The self-directed program, was designed for anyone who wanted to learn more about research data - what it is, why it is a global hot topic, finding it, re-using it, managing it, and more. ANDS has also been engaging with the medical and health community, through a series of 'Sharing Health-y Data: Challenges and Solutions' workshops across Australia, a 'virtual' health data community group within the 23 Things program, and their resources relevant to medical and health data collated on the 'medical and health data' page of their website. As the active phase of running the 23 (research data) Things program throughout 2016 came to a close, ANDS produced a re-purpose toolkit. The top 10 research data Things for medicine and health is a part of this toolkit. It was developed from the 23 (research data) Things materials; adapting the Things that were most relevant to medical and health data, and also substituting medical and health examples into many of the activities. Each Thing has two or three activities. Some of the activities are an introduction to a topic, and some delve a little deeper, to allow for a range of previous experience and knowledge. The program is available to be reused and adapted as required, and accordingly it has a CC-BY licence. The top 10 research data Things for medicine and health are a flexible, adaptable learning resource for anyone working with medical, clinical or health data.

Wonambi: An Integrated Software Development Sandbox	Adam Brown, Jess Robertson, Jens Klump, Carsten Friedrich, Ryan Fraser, Anusuriya Devaraju and Geoff Squire	INTRODUCTION CSIRO's Geoanalytics Platform (GAP) group have, in concert with colleagues from Data61, designed and developed a web-based system to help lower the barrier to entry for researchers to develop scientific programs and scripts of their own. The system, named Wonambi, harnesses existing cloud infrastructure and aims to integrate with other solutions in GAP's software ecosystem. The proposed poster will communicate the purpose and benefits of Wonambi and describe how and why they interact with other internal and external systems: the Scientific Software Solutions Centre (SSSC), Virtual Laboratories, and git repositories. METHOD Wonambi is to serve as a sandbox environment that allows researchers, who may have only limited exposure to scientific programming, to easily instantiate development environments on existing cloud infrastructure (NeCTAR, AWS etc) and start coding. The development environments will be created with Jupyter Notebooks pre-installed along with any auxiliary libraries that the scientific codes will require. Researchers can develop their codes and test them out in real-time using the Jupyter Notebooks interpreter. Wonambi handles the persistence of code changes into Git repositories (such as GitHub or BitBucket) thus removing this complexity from the user's workflow. Integration with the SSSC will enable Wonambi users to instantiate environments pre-configured for development of software as described by the SSSC. This means that users will be able to learn from and enhance existing code, increasing its quality and reuse. Code developed in Wonambi will also be able to be published into the SSSC for use by other people and systems. For example, the AuScope-funded Virtual Geophysics Laboratory can harvest scientific codes from the SSSC and use them to process datasets that it has access to. This expands the GAP's software capability to cover code creation (Wonambi), software discovery (SSSC) and software execution (VLs). ACKNOWLEDGEMENTS The Virtual Geophysics Laboratory is supported by AuScope through the National Collaborative Research Infrastructure Strategy Program.

The Research Hub: Rethinking the Delivery of Institutional Research Services	Richard Hosking, Cameron McLean, Yvette Wharton, Nick Young and Mark Gahegan	Services available to researchers at our institution are numerous, and are provided from facilities across the university, such as the Library, central IT, and Faculties. Culture and practices within research communities however are often disconnected from those of the services available to them, rather than being seen as interconnected in the research process. We have identified this as a barrier to both sustainable development of novel software and of the research services themselves. Our experience, as a service provider (Data, HPC, Virtual Machines, and Visualisation) and as software developers embedded within research groups, has led us to re-envision how research services can be coordinated through a Research Hub. We offer a novel approach to organise services around the research lifecycle (Figure 1), and touch on the clash of curiosity driven workflows with the more formal processes often required to deliver reliable services. We open-source our code and methodology to aid other institutions and prompt a shared conversation around the sustainable management of projects and institutional research services.
Managing Research Projects with Agility	Majid Alsubaie and Fabiana Santana	Global economic challenges force organisations to constantly seek alternatives to be competitive in today's fast moving landscape. Research organisations are not an exception; they need to drive innovation and have agility in the deployment of new solutions. Adopting Agile methodologies instead of the traditional project management ones is a good start to foster innovation and deliver quality, creative solutions. However, research projects usually have pre-defined deliverables established by contract, which contradicts fundamentals established in the Agile manifest, such as "customer collaboration over contract negotiation" and "responding to change over following a plan". This scenario prevents the adoption of pure Agile methodologies for the management of research projects. Fortunately, some Agile practices are not incompatible with PMBOK (Project Management Body of Knowledge) and PRINCE2 (Projects in a Controlled Environment), and their adoption can considerably improve the agility of such systems. This work introduces a reference guide to incorporate agility to the management of research projects. The reference guide considers the scenario of systems software development and shows how the Scrum methodology can be incorporated into the planning, executing and controlling processes established by the PMBOK. It is expected the new methodology will considerably improve the controls and time-to-deliver of research projects. It is also expected it will reduce resistance in adopting Agile practices for research project management, with significant outcomes for the research community.

The design and implementation of a high-throughput data pipeline for an agricultural proximal sensing platform	James Watson, Andries Potgeiter, Edward Holland, David Jordan and Graeme Hammer	Proximal sensing is becoming a key component of agricultural research. The combination of a changing climate, an increasing global population, and limited arable land means that improving crop productivity is fundamental to future global food security. To achieve necessary productivity gains, crop breeders rely on measurements of a wide variety of traits such as plant height, cover, temperature, photosynthetic capability and yield. These traits are sampled over large spatial and temporal scales, with some experiments consisting of thousands of plots. Measurements of these crop traits in such large experiments are restricted and typically performed manually, which is labour-intensive, expensive, and necessitates sub-sampling. The GECKO tractor-mounted sensing platform has been developed by the Queensland Alliance of Agriculture and Food Innovation (QAAFI) to automate and extend this data collection process, using proximal remote sensing to regularly measure key traits across all plots of an experiment. This platform provides more comprehensive crop coverage, using uniform and auditable data collection protocols. The GECKO collects continuous readings from a hyperspectral camera, 4 point spectrometers, 3 infrared thermometers, 3 sonar rangefinders, a LIDAR, a thermal camera and a real-time corrected GPS (accurate within 2cm). These sensors are operated via a touch-screen user interface located in the tractor cabin. Data from a nearby weather station is also captured. Several hundred gigabytes of data can be generated in a single run, resulting in 4 significant computational challenges: 1. collecting and storing disparate sensor output in the field, 2. transferring and archiving the sensor data, 3. computing crop-relevant information for each plot from sensor data and caching these results, and 4. providing crop experts with the tools to analyze, visualize and share the extracted information. Here we describe the design and implementation of a data pipeline for turning the output of the GECKO sensing platform into useful information.