

## Proposal for Data Movement Presentation *Stream*

**Brett Rosolen**

AARNet, Sydney, Australia  
Brett.Rosolen@aarnet.edu.au

### GENERAL INFORMATION

Recent investment in eResearch infrastructure in Australia has delivered a world-class toolset into the hands of researchers for data creation, movement, processing and storage. Massive multi-Petabyte data stores have significant data collections of national importance and peak HPC capability has scaled in response to the demand for big data analysis. With the next generation of the research and education network **it is now possible to move 1 Petabyte of data in a single day** from one single resource to another.

However data transfers of this magnitude are not being seen or are rarely attempted – not due to a lack of need to move data, but primarily due to a commonly held belief that achieving this is not possible and/or it is very difficult. Handlers are resorting to trickle feeding data over a long period by ineffective and out-of-date methods and even resorting to “sneakernet” to move large datasets. The result is a reduced potential to take full advantage of the entire eResearch system that the Australian research community has at its fingertips.

### DESCRIPTION

We propose that it is timely to address this in **a stream (possibly a half day)** at eResearch Conference 2015 through a series of presentations that take attendees on a data movement journey by community champions who have “been there, done that,” bringing real world use cases in delivering data to peak research capabilities.

These champions will be sourced from disciplines who have discovered the value in skills and tools that lead to the most efficient and effective way to handle data between collaborators, and who acknowledge that resources most useful for manipulating and storing such data could be anywhere on the planet.

Contributors could include those who have leveraged toolsets (Aspera, GridFTP), those whose need for scale has led to novel new solutions (LHC, SKA), and those who profess to deliver architecture that facilitates data movement (ScienceDMZ). Many of these data champions are known to AARNet. We will approach contributors to bring more structure to this stream subject to the approval of this approach in concept by the programming committee.

We propose to contribute significant network resource such that real world demonstration plays a major role in the stream, including multiple 10GB circuits if desirable. We also consider that we could subsidise speaker attendance to the conference proper.

We acknowledge that the formation of this program needs to happen quickly, so are pursuing a target outline is as follows:

1. Tools – Aspera
2. Tools – GridFTP
3. Architecture – Science DMZ
4. Research specific (SKA or LHC)
5. Panel \*optional\*

## **NEXT STEPS AND TIMING**

Subject to general approval from the program committee we (AARNet) will confirm the contributors and abstracts very quickly. Notably data movement presentations from the community may already be in the process of being submitted! This proposal doesn't seek to alienate them in any way, but rather to encompass and support them to ensure that there is sufficient content to broadly cover the subject matter and equip the community. We recommend to the committee that this proposal be considered in parallel to the submissions being received within the "Tools for collecting, generating and moving data" and request that the committee provide advice to us where the particular topic of data movement is lacking in the coming weeks.

Although we believe that this topic is best addressed in a half day stream in the conference proper, it is also conceivable that this program could belong in a workshop or even BoF format. Although less than ideal we request that the committee take this into consideration if integration into the program proper proves difficult to accommodate.

As we are actively pursuing the makeup of this program, we expect to make good progress and provide more information as required reasonably quickly.

## **WHO SHOULD ATTEND**

Anyone with the potential to handle large datasets

- eResearch infrastructure providers
- Campus eResearch support staff
- Research project facilitators