Big data, big science, the data deluge, these are topics we are hearing about more and more in our research pursuits. Then, through media hype, comes cloud computing, the saviour that is going to resolve our big data issues. However, it is difficult to pinpoint exactly what researchers can actually do with data and with clouds, and how they get to exactly solve their big data problems.

Since the beginning of 2012, the NeCTAR Research Cloud has been running at the University of Melbourne, attracting over 1,400 users from around the country. This has not only provided an unprecedented opportunity for researchers to employ clouds in their research, but it has also given us an opportunity to clearly understand how researchers can more easily solve their big data problems. The cloud is now used daily, from running web servers and blog sites, through to hosting virtual laboratories that can automatically create hundreds of servers depending on research demand.

So what's the recipe for solving big data problems on cloud infrastructure? Two of the key ingredients are community and research platforms: Research is built on collaboration, connection and community, and researchers employ platforms daily, whether as bio-imaging platforms, computational platforms or cloud platforms (like DropBox). There are some other important features, too.

Firstly, the borders to collaboration are eased, allowing communities to access infrastructure that can be instantly built to be completely open, through to completely closed, all managed securely through (nationally) standardised interfaces. Secondly, it is free and easy to build servers and infrastructure, but it is also cheap to fail, allowing for experimentation not only at a code-level, but at a server or infrastructure level as well. Thirdly, this (virtual) infrastructure can be shared with collaborators, moving the practice of collaboration from sharing papers and code to sharing servers, pre-configured and ready to go. And finally, the underlying infrastructure is built with big data in mind, co-located with major data storage infrastructure and high-performance computers, and interconnected with high-speed networks nationally to research instruments.

The research cloud is fundamentally new in that it easily allows communities of researchers (often connected by common geography (research precincts), discipline or long-term established collaborations, to build open, collaborative platforms. These open, sharable, and repeatable platforms encourage co-ordinated use and development,

evolving to common community-oriented methods for big data access and data manipulation.

In this paper we will discuss in detail the specifics of engineering these communities, as well as some outcomes as a result of these communities and collaboration enabling platforms. These will draw on real world examples that cover the disciplines of radio astronomy, breast cancer screening, genomics, particle physics and history and culture.