# Building Communities of Data Scientists: what works and what doesn't

Lesley Wyborn[1],

[1]Geoscience Australia, Canberra, Australia, lesley.wyborn@ga.gov.au

Volumes of scientific data are growing exponentially as substantial improvements in the capacity of instruments have resulted in data being captured at higher resolution and at a greater rate. The resultant data deluge has been impacting for some time putting pressure on the need to build communities of data scientists to work together, preferably at an international level, to collaboratively build communities that will facilitate sharing not only of data and information but also applications to turn these data and information into new knowledge.

Modern communities of data scientists need to comprise people who can work across all boundaries, including organisational, national and discipline, so that science can tackle current key challenges in earth and space sciences, such as climate change, hazard prediction and sustainable development of our natural resources.

Building successful communities of data scientists is difficult. It requires a team of committed and enthusiastic people who have the support of their organisations to take on this somewhat time consuming process of reaching consensus. Successful communities also tend to have an international/national project driver that brings disparate groups of people together to achieve a shared goal (eg, OneGeology).

In an age of highly competitive research funding, collaborative behaviours are not common and communities often tend to compete with each other or try to dictate that their method is best and that everyone else should follow them. Another limitation to the development of communities of data scientists is that many scientists will not willingly share their fundamental data as they feel it undermines their edge in competitive funding regimes (knowledge is power). Not surprisingly, excessive funding of some parts of some communities has proven to be divisive.