*Highlights of Astronomy, Volume 16*
*XXVIIIth IAU General Assembly, August 2012*
*T. Montmerle, ed.*
© International Astronomical Union 2015
doi:10.1017/S1743921314012861

# Data Intensive Radio Astronomy en route to the SKA: The Rise of Big Radio Data

**A. R. Taylor**

Department of Physics and Astronomy, University of Calgary, 2500 University Dr. N.W., Calgary, Alberta, Canada T2N 1N4
email: russ@ras.ucalgary.ca

**Abstract.** Advances in both digital processing devices and in technologies to sample the focal and aperture planes of radio antennas is enabling observations of the radio sky with high spectral and spatial resolution combined with large bandwidth and field of view. As a consequence, survey mode radio astronomy generating vast amounts of data and involving globally distributed collaborations is fast becoming a primary tool for scientific advance. The Square Kilometre Array (SKA) will open up a new frontier in data intensive astronomy. Within the next few years SKA precursor telescopes will demonstrate new technologies and take the first major steps toward the SKA. Projects that path find the scientific journey to the SKA with these and other telescopes are currently underway and being planned. The associated exponential growth in data require us to explore new methodologies for collaborative end-to-end execution of data intensive observing programs.

**Keywords.** methods: analytical, instrumentation: miscellaneous, surveys

## 1. Introduction

The first decade of this century has seen a tremendous advance in information and digital technologies which have driven a commensurate advance in the data capacities of radio telescopes. The instantaneous observing bandwidths of both single-dish and array radio telescopes have increased by two orders of magnitude. High dynamic range imaging over wide fields with radio array antennas, and the need to mitigate narrow-band radio frequency interference has driven observing programs to high resolution spectral channelization over broad continuum bandwidths. In combination with the ultra-wide instantaneous fields of view now provided by focal plane horn arrays and phased-array feeds, and by many-element aperture plane arrays, the data rates to the observer sustained by current observing programs are $10^3 - 10^4$ times larger than typical only a few years ago.

Many of our most pressing astrophysical questions require synthesis of information covering large areas of the sky and over a significant range of cosmic history. Survey mode observations, in which large amounts of observing time are devoted to major programs that create vast data sets is becoming an essential approach to observational astronomy at both radio and optical wavelengths. Survey mode observing combined with the new instrumental data capacities is driving an exponential growth in both the rate and the volume of data in radio astronomy. Survey projects dominate the science program for the SKA precursor telescopes in Australia (ASKAP) and South Africa (MeerKAT). Figure 1 illustrates this data trend by showing data output rates and volumes for major observing programs over the course of the past decade and projected into the future en route to the Square Kilometre Array.



https://doi.org/10.1017/S1743921314012861 Published online by Cambridge University Press

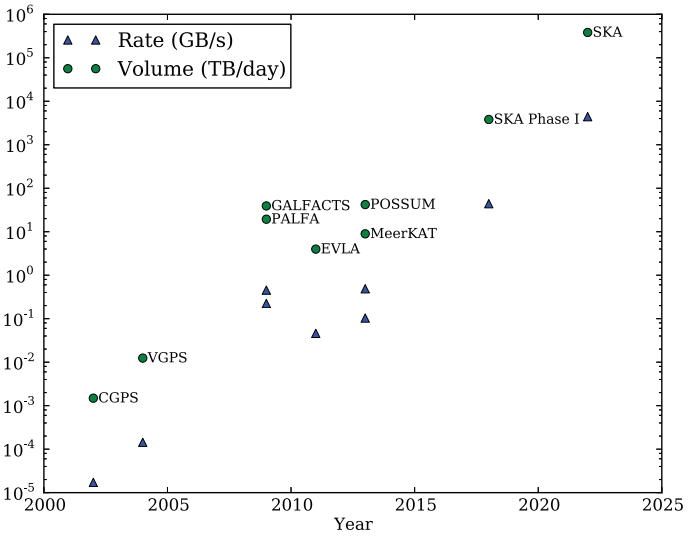


**Figure 1.** The rise of big radio data. The output data rates to the observer are shown for several survey mode observing projects starting with the Canadian Galactic Plane Survey (Taylor *et al.* 2003) to the SKA post 2020. The data rate is increasing on average by 150% per year.

## 2. Cyber Solutions to Data Intensive Radio Astronomy

Survey mode observing is changing the way that science programs are executed and creating new challenges associated with management of the very large data volumes produced, the complexity of processing and analysis of data that may support multiple scientific goals, and collaborative execution of science by large, globally distributed teams.

As we move into the era of the SKA, these challenges may be met with cyber infrastructure platforms that wed emerging web technologies with distributed global resources of cloud-enabled high performance computing and smart data infrastructure. The cyberSKA initiative (Kiddle *et al.* 2011, Grimstrup *et al.* 2012) is a research project to develop a scalable and distributed cyber infrastructure platform to meet the evolving needs of data intensive radio astronomy en route to the SKA. A collaboration between government agencies, universities and industries, cyberSKA is being co-developed with SKA pathfinding survey programs in time-domain and imaging astronomy with Arecibo, the JVLA and GMRT. The platform is accessed via a web portal (www.cyberska.org) and is building e-science infrastructure for collaboration, smart distributed data management, interface to distributed computing for data processing, data visualization and analytics for very large data sets, and an API for community contributed applications that interface to the collaboration metadata and the distributed data system.

## References

Taylor, A. R. *et al.* 2003, *AJ*, 125, 3145.

Grimstrup, A., Mahadevan,V., Eymere, O., Anderson, K., Kiddle, C., Simmonds, R., Rosolowsky, R. & Taylor, A. R. 2012, in *Software and Cyberinfrastructure for Astronomy II*, N. M. Radziwill and G. Chiozzi (eds.), Proc. SPIE 8541, 15.

Kiddle, C., Taylor, A. R., Cordes, J., Eymere, O., Kaspi, V., Pigat, D., Rosolowsky, E., Stairs, S. & Willis, A. G. 2011, in *GCE'11: Proceedings of the 2011 ACM Workshop on Gateway Computing Environments*, (ACM: New York, NY, USA), p 65.

https://doi.org/10.1017/S1743921314012861 Published online by Cambridge University Press