

## Evolutionary Map of the Universe: large-n astronomy

---

**Ray P. Norris\***

*CSIRO Astronomy & Space Science*

*& Western Sydney University*

*E-mail: [raypnorris@gmail.com](mailto:raypnorris@gmail.com)*

EMU is an all-sky radio survey, and one of the two key projects primarily driving the design and construction of the new Australian SKA Pathfinder (ASKAP). EMU represents an enormous leap in the global radio-astronomy capability, increasing the total number of known radio sources to about 70 million, from the 2.5 million detected over the entire previous history of radio-astronomy. Next-generation radio continuum surveys such as EMU break through a sensitivity threshold such that they will be equally dominated by star formation and Active Galactic Nuclei (AGN), whereas earlier radio continuum surveys were completely dominated by AGN. The impact is even higher because the millions of sources enable fine-grained statistical studies: the large numbers enable the sample to be chopped up along many different orthogonal axes, so even subtle correlations can be detected. EMU will also almost certainly discover unexpected phenomena. While the first ASKAP observations, using prototype receivers, have already been published, the first deep observations (ASKAP early science) are expected to start in 2016, and we hope the full survey will start about a year after that. A key feature of EMU is that, as well as its key science goals, it is being designed to search for unexpected discoveries.

*EXTRA-RADSUR2015 (\*)*

*20–23 October 2015*

*Bologna, Italy*

---

(\*) This conference has been organized with the support of the Ministry of Foreign Affairs and International Cooperation, Directorate General for the Country Promotion (Bilateral Grant Agreement ZA14GR02 - Mapping the Universe on the Pathway to SKA)

---

\*Speaker.

## 1. Introduction

The new Australian SKA Pathfinder (ASKAP) telescope is approaching completion in Western Australia. It is primarily a survey radio-telescope, and its performance is being optimised for the ten Science Survey Projects that were selected to run on it initially, and particularly for the two key Science Survey projects: EMU and WALLABY. ASKAP will be the world's fastest survey telescope at least until the completion of the Square Kilometre Array (SKA) almost a decade from now, and EMU will be unsurpassed by any other radio continuum survey until at least that time.

EMU represents an enormous leap in the global radio-astronomy capability, increasing the total number of known radio sources to about 70 million, from the 2.5 million detected over the entire previous history of radio-astronomy. Next-generation radio continuum surveys such as EMU also break through a sensitivity threshold such that they will be equally dominated by star formation and Active Galactic Nuclei (AGN), whereas earlier radio continuum surveys were completely dominated by AGN.

The impact is even higher because the millions of sources enable fine-grained statistical studies: the large numbers enable the sample to be chopped up along many different orthogonal axes, so even subtle correlations can be detected. Such “large-n” astronomy also enables new approaches to traditional questions. For example, to study the properties of a population of galaxies discovered in the radio, it may not be necessary to know the redshift of every galaxy. Instead, in some circumstances, the properties of the population may be determined if the redshift distribution of the sample is accurately known.

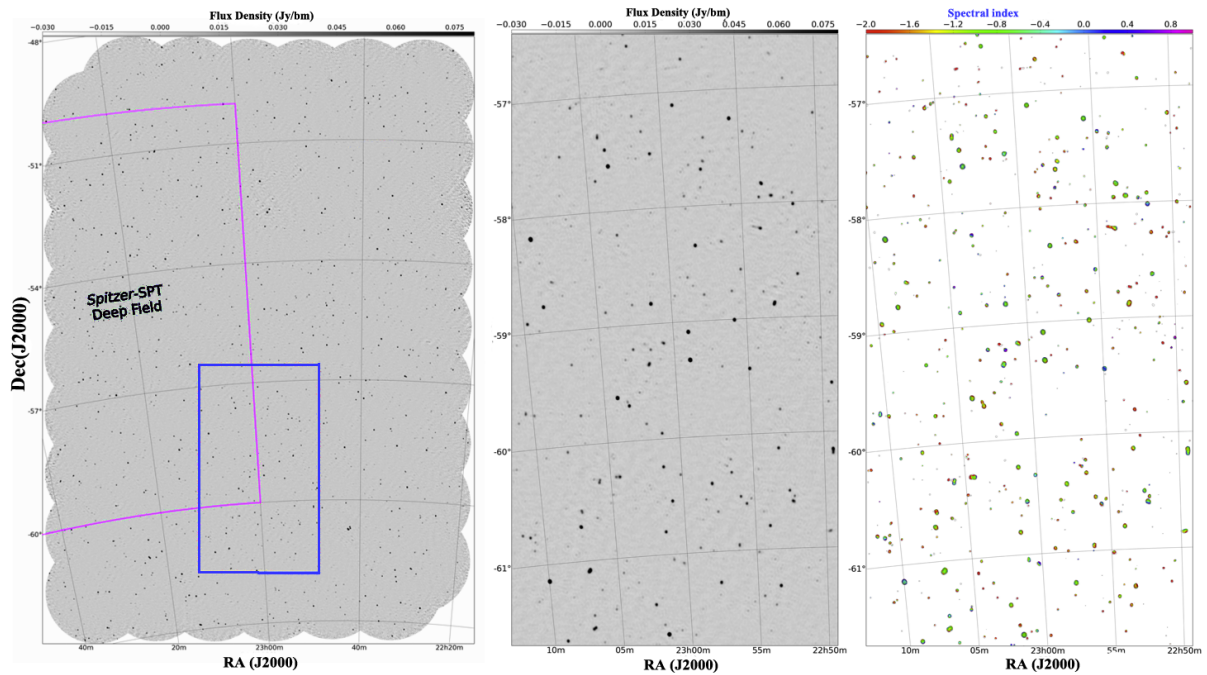
Like most surveys, the EMU survey is being optimised to answer specific science goals, such as measuring the global star formation rate as a function of cosmic time. Such well-defined questions can be termed the “known-unknowns”. However, experience shows that the greatest impact from major new telescopes or astronomical surveys is often the discovery of unexpected phenomena, such as pulsars, or quasars, or dark energy. These may be termed the “unknown-unknowns”.

Here I describe the ASKAP telescope, the EMU project, some of its science goals and the WTF project which will mine data for the “unknown-unknowns”.

## 2. ASKAP: the Australian SKA Pathfinder

The new \$165m Australian SKA Pathfinder (ASKAP) telescope is currently approaching completion in Western Australia. ASKAP is revolutionary in that it uses 94-pixel dual-polarisation phased array feeds (PAFs) at the centre of each antenna, rather than conventional single-pixel receivers. This enables ASKAP to observe 30 sq. deg. of sky in a single pointing, making it the fastest survey telescope in the world [1]. Another revolutionary feature is the use of three-axis antennas, enabling the dish to rotate to keep a constant parallactic angle on the sky, to minimise calibration artefacts and maximise dynamic range. One consequence of the phased array feeds is a large data rate, with a total bandwidth of 2.1 THz on each antenna, and a total processed data volume of 70 PB/year. After correlation at the ASKAP site, the data are sent by optical fibre to Perth for processing.

The data are processed in a real-time pipeline running on the *galaxy* system (consisting of 472 Cray XC30 compute nodes) at the Pawsey centre in Perth. The pipeline processor, using



**Figure 1:** (Left) An early mosaiced image, adapted from [13], taken with the six-antenna “BETA” ASKAP engineering prototype array, which is already the second-fastest survey telescope in the world, beaten only by the JVLA. The image is taken from a 36-hour observation, and covers 150 square degrees, to a mean rms of  $375 \mu\text{Jy}/\text{beam}$ , and detects 3722 sources at  $10 \sigma$ . The blue rectangle marks the area shown in the centre panel. (Centre) An enlargement of a region  $3^\circ \times 5^\circ$  from the left-hand panel. (Right) The spectral-index map corresponding to the centre panel.

the ASKAPsoft software, includes calibration, deconvolution, and source extraction. After data validation, these data are placed in the public domain. Further post-processing by the survey teams result, in most cases, in “value-added data” which will have a proprietary period.

ASKAP construction started in 2009, and the first images from the commissioning (BETA) array were obtained in 2014, an example of which is shown in Figure 1. All 36 of the ASKAP antennas, and the entire infrastructure, are complete, and the antennas are now progressively being equipped with the final (MkII) PAFs. ASKAP early science (using just 12 antennas) is expected to start in mid-late 2016, and the main surveys, using 36 antennas, are expected to start in 2017, depending on progress with commissioning.

When the ASKAP project started, proposals for survey science projects were invited from the community. Out of 38 submitted survey science proposals, ten were selected, of which two (EMU and WALLABY) were selected to be the two key projects which are primarily driving ASKAP design, and which will be given the highest priority for observing when ASKAP is completed.

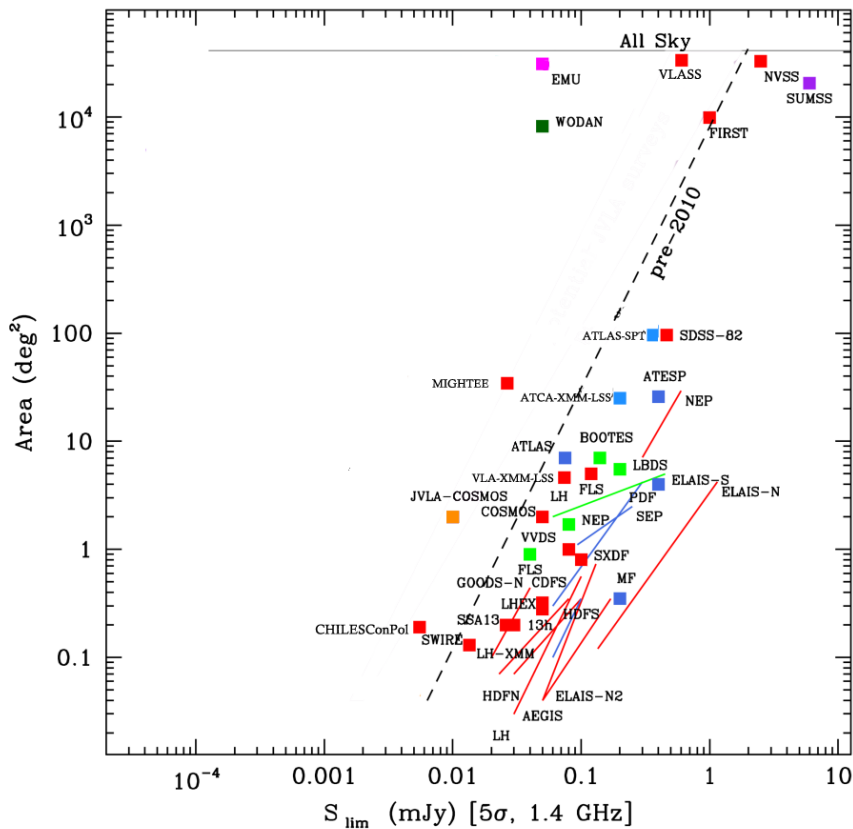
### 3. EMU: Evolutionary Map of the Universe

Evolutionary Map of the Universe (EMU) [18] primarily addresses questions about the origin and evolution of galaxies. EMU is an enormous leap in the global radio-astronomy capability,

increasing the total number of radio sources from the 2.5 million detected over the entire history of radio-astronomy, to about 70 million. It will also detect diffuse emission from hundreds or thousands of clusters of galaxies, compared to the  $\sim 50$  currently known. EMU is an international project, and the EMU team consists of about 300 astronomers in 21 countries.

EMU will make a deep (rms  $\sim 10 \mu\text{Jy}/\text{beam}$ ) radio continuum survey of the entire Southern sky at 1.3 GHz, extending as far North as  $+30^\circ$  declination, with a resolution of 10 arcsec. About half the radio sources detected by EMU will be star-forming galaxies, and the other half will be “Active Galactic Nuclei”(AGN). A small fraction (hundreds of thousands) will be Galactic objects and the diffuse emission from clusters of galaxies. Importantly, because of the short baselines in the ASKAP array, EMU will be very sensitive to extended structures as well as compact objects.

Figure 2 shows a comparison of EMU to other radio continuum survey projects.



**Figure 2:** A diagram showing EMU compared to other major 18cm continuum surveys, based on an earlier diagram by Isabella Prandoni. The position of MIGHTEE is uncertain, as its parameters are under discussion. Surveys at other frequencies have been converted to 1.4 GHz assuming  $S \propto \nu^\alpha$

EMU differs from earlier radio surveys in four important respects:

- Its scale is transformational - it will multiply the number of known radio sources by a factor of  $\sim 30$ .

- It will not be dominated by AGN, which are rare at other wavelengths, so will have an unusually high impact on science done at other wavelengths.
- It is more than just a survey: the EMU project includes cross-identification with multiwavelength catalogues, obtaining redshifts, and includes the key science projects as an integral part of the project.
- It explicitly includes “discovering the unexpected” as described below.

EMU’s main overall science driver is to understand the formation and evolution of galaxies. The large number of galaxies detected by EMU means that even rare or short-lived phases of galaxy evolution can be studied. In addition to planning and conducting the radio survey itself, the EMU project also includes:

- Development Projects, which develop and optimise the tools needed to generate the science from EMU data. These include source extraction and classification [14], cross-identification with multi-wavelength catalogues[9], and redshift determination (Salvato et al., in preparation).
- Key Science Projects, which will deliver the key science goals from EMU. These include Galaxy Evolution, Cosmology, Galaxy Clusters, the Galactic Plane, and Radio Stars. EMU is primarily observational, but has strong links to theory within these projects.
- Collaboration Projects, which develop and maintain collaborations with other large survey projects such as Meerkat- MIGHTEE, MWA-GLEAM, LOFAR, SkyMapper, Taipan, WISE, VHS, and eROSITA.
- EMU Pathfinder projects, on which EMU techniques and science are tested prior to the EMU survey. These include the Australia Telescope Large Area Survey [17, 11] which is a radio survey with a similar sensitivity and resolution to EMU, but over a much smaller area, and the ATLAS-SPT project (O’Brien et al, this volume) to study the clusters in the South Pole Telescope field.

The EMU sensitivity is not expected to change, as the approved EMU proposal requests a final sensitivity, not a fixed number of hours. For example, should receiver temperatures be slightly worse than expected, then it is expected that the observing time will be increased rather than permitting the sensitivity to degrade.

Table 1 shows the Key Science projects of EMU, most of which have the underlying theme of understanding the evolution of galaxies over cosmic time, and are discussed in some detail in [18]. The rest of this paper discusses a small selection of the Key Science Projects.

#### 4. Clusters of Galaxies

Galaxy clusters, the largest bound objects in the universe, consist of galaxies in a cloud of intergalactic gas. Once viewed predominantly as relaxed spherically symmetric systems with a cooling gas flow towards a central potential well, they are now more often viewed as the complex

**Table 1:** EMU Key Science Projects

<b>Key Science Project</b>	<b>Project Leader</b>
EVACAT: Value-Added Catalogue	Nick Seymour
Characterising the Radio Sky	Ian Heywood
Clusters of Galaxies	Melanie Johnston-Hollitt & Chiara Ferrari
Cosmic star formation history	Andrew Hopkins
Cosmic web	Shea Brown
Cosmology	David Parkinson
Galactic Plane	Roland Kothes
Local Universe	Josh Marvil & Michael Brown
Radio AGN in the EoR	Jose Afonso
Radio-loud AGN	Anna Kapinska
Radio-quiet AGN	Isabella Prandoni
SCORPIO: Radio Stars	Grazia Umata
WTF: Mining Data for the Unexpected	Ray Norris

intersections of filaments and sheets of the cosmic web (the large-scale structure of the Universe), with a continuing infall of gas and galaxies that have been streaming along the filaments of the web. Clusters are now known to be rich in dynamics, shocks, interactions, and even molecular gas [8]. The galaxies in each cluster account for only a small fraction of the total baryonic mass, which is dominated by the intra-cluster medium (ICM).

The radio emission from clusters is not well understood [4], and consists of three components: (a) the radio emission from constituent galaxies, including bent-tail galaxies that are interacting with the ICM, (b) relics, caused by shock-excited electrons, and (c) the diffuse cluster-wide halo emission [5]. However, the radio study of clusters has been hampered by inadequate radio data and inadequate algorithms to detect and extract diffuse radio emission. For example, most of the  $\sim 60$  known radio haloes [3] have been found in clusters detected first at X-ray wavelengths, and our knowledge may be biased by selection effects.

EMU will detect tens to hundreds of thousands of clusters via bent-tail galaxies [6] and will detect diffuse haloes from at least a few thousand of them. EMU will increase the radio data on clusters by at least an order of magnitude, potentially revolutionising cluster science.

Bent-tail galaxies are particularly effective tracers of clusters, as they are potentially easily identifiable in large radio surveys, and modeling [16] show they are mainly confined to high-mass ( $M > 10^{14} M_{\odot}$ ) clusters. However, there is a discord between theory and experiment, because the ATLAS radio survey shows about 10 per square degree [6], implying that EMU will detect tens to hundreds of thousands, whilst modeling predicts that the number of high-mass clusters should be far less. Potential explanations may reside in the models, observations, interpretation, or the production of bent-tail galaxies by other mechanisms, but we do not yet have the data to distinguish between these explanations.

It is known that not all clusters have radio haloes, but it is unclear what determines whether or not a cluster has halo emission. There is good evidence [10] that cluster radio haloes are distributed

bimodally as a function of X-ray luminosity, with radio haloes found only in clusters that have recently undergone a merger. However, this may be partly a selection effect, as there are no large sensitive unbiased radio surveys of clusters, and nearly all radio observations have been of clusters chosen from X-ray selected samples (e.g. [21]). There is also limited knowledge as to how this changes with redshift, as no wide-field survey exists with the necessary resolution and sensitivity to detect radio haloes out to redshifts of  $z \sim 0.5$ , where major cluster mergers are commonplace. Furthermore, at least one halo has been detected in a cluster with low X-ray emission [2]. In summary, existing models are based on a small, potentially biased, subset of clusters, and the EMU observations may reveal a very different picture.

## 5. SCORPIO: Radio Stars

Radio emission has been discovered in a wide variety of stellar objects, often revealing astrophysical phenomena not detectable at other wavelengths. Curiously, there have not been any deep wide surveys for radio stars, so our knowledge of radio stars is based on targeted observations of stars thought to be good candidates for radio emission. Our knowledge of radio stars may therefore be very biased, and EMU may turn up an unexpected number of radio stars.

In the same way that the ATLAS survey was used to survey a small patch of extragalactic sky to help plan EMU observations, the SCORPIO survey [20] is surveying about  $4 \text{ deg}^2$  of the Galactic plane to a sensitivity of  $\sim 25 \text{ } \mu\text{Jy}/\text{beam}$  with a resolution of 10 arcsec. Preliminary results show a significant number of unresolved galactic objects in addition to the expected extragalactic background objects, with a much wider range of spectral indices than are seen in the extragalactic objects (Cavallero et al., in preparation). The vast majority are unidentified, but spectral indices suggest many will be planetary nebulae of ultra-compact HII regions.

Preliminary statistics suggest that EMU will provide an unbiased sample of several thousand stellar radio sources, enabling a detailed investigation into the physics of radio stars. This will include stellar physical parameters, magnetic activity, the fraction of active single stars and binary objects that show coherent emission, the time-scales of its variability and their relationship to stellar parameters.

## 6. WTF: Mining EMU data for the unexpected

At least half the major discoveries in astronomy are unplanned, and are typically made by surveying the Universe in a new way, rather than by testing a hypothesis or conducting an investigation with planned outcomes [12, 7, 15, 22]. For example, of the 10 greatest discoveries by HST, only one was listed in its key science goals [19]. So if a major new telescope merely achieves its stated science goals, it is probably performing well below its potential scientific productivity.

The case study of the Nobel-prize-winning discovery of pulsars by Jocelyn Bell is instructive. A talented and persistent PhD student studying interstellar scintillation (and thus expanding the observational phase space), and who knew her instrument intimately, recognised that “bits of scruff” on the chart recorder could not be terrestrial interference, but represented a new type of astronomical object [1]. As a result, she discovered pulsars.



Next generation radio continuum surveys such as EMU will significantly expand the volume of observational phase space (as shown in Figure 2), so are likely in principle to discover unexpected new phenomena and new types of object. Although it is impossible to predict their nature, we might reasonably expect new classes of galaxy, or perhaps even stumble across new Galactic populations. However, the complexity of ASKAP and the large data volumes mean that it may be non-trivial to identify them. For example, a present-day Jocelyn Bell is unlikely to understand the instrument well enough to distinguish astrophysical phenomena from instrumental effects, and would not be able to sift through the petabytes by hand, searching for something unusual. On the other hand, failure to identify unexpected effects may mean missing out on the most important science to emerge from ASKAP. It is therefore necessary to plan explicitly to build techniques that may make unexpected discoveries, rather than hoping to stumble across them.

We have therefore started a pilot project within EMU, Widefield outlier Finder (WTF), to develop techniques for mining large volumes of astronomical data for unexpected results, using machine-learning techniques and algorithms. For example, one approach looks for groups of properties in an n-dimensional plot with axes such as flux density, spectral index, and IR-to-radio ratio. Some groups will correspond to known types of object (e.g. stars, galaxies, quasars) but unexpected groups may correspond to unknown classes of object. More sophisticated algorithms are described below. Although targeted specifically at EMU, such approaches are expected to have broad applicability to astronomical survey data.

WTF is currently operating on an Amazon Web Services (AWS) cloud environment. On this platform, simulated discoveries (“EMU eggs”) are hidden in images and catalogs of real and simulated data, and a challenge is being issued to machine-learning and other algorithms to discover them. A number of algorithms are being tried in-house, but we extend an invitation to external groups who would like to try their algorithms on the challenge. The data challenge will gradually grow to encompass ASKAP early science data and EMU data, and we expect the discovery of simulated eggs to be overtaken by real discoveries from real data.

## 7. Acknowledgement

RPN acknowledges traveling support of the Ministry of Foreign Affairs and International Cooperation, Directorate General for the Country Promotion (Bilateral Grant Agreement ZA14GR02 - Mapping the Universe on the Pathway to SKA)

## References

- [1] Bell-Burnell, J., 2009, in *Accelerating the Rate of Astronomical Discovery*, PoS (sps5) , 014
- [2] Bonafede, A., Intema, H., Brügger, M., et al. 2015, MNRAS, 454, 3391
- [3] Brunetti, G., & Jones, T. W. 2015, Ap & Space Sci. Library, 407, 557
- [4] Cassano, R., Brunetti, G., Norris, R. P., et al. 2012, A&A, 548, A100
- [5] Cassano, R., Bernardi, G., Brunetti, G., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array*, PoS (AASKA14) , 73
- [6] Dehghan, S., et al., 2014, AJ, 148, 75



- [7] Ekers, R. D. , 2009, in *Accelerating the Rate of Astronomical Discovery*, [PoS \(sps5\) , 007](#)
- [8] Emonts, B., et al., 2015, *MNRAS*, 451, 1025
- [9] Fan, D., Budavári, T., Norris, R. P., & Hopkins, A. M. 2015, *MNRAS*, 451, 1299
- [10] Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, *A&A Rev.*, 20, 54
- [11] Franzen, T. M. O., Banfield, J. K., Hales, C. A., et al. 2015, *MNRAS*, 453, 4020
- [12] Harwit, M. 1981, *Cosmic Discovery*, Brighton: Harvester Press
- [13] Heywood, I., Bannister, K. W., Marvil, J., et al. 2016, [arXiv:1601.05857](#)
- [14] Hopkins, A. M., Whiting, M. T., Seymour, N., et al. 2015, *PASA*, 32, e037
- [15] Kellermann, K.I., et al., , 2009, in *Accelerating the Rate of Astronomical Discovery*, [PoS \(sps5\) , 005](#)
- [16] Mguda, Z., et al., 2015, *MNRAS*, 446, 3310
- [17] Norris, R. P., Afonso, J., Appleton, P. N., et al. 2006, *AJ*, 132, 2409
- [18] Norris, R. P., Hopkins, A. M., Afonso, J., et al. 2011, *PASA*, 28, 215
- [19] Norris, R., Basu, K., Brown, M., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array*, [PoS \(AASKA14\) , 86](#)
- [20] Umana, G., Trigilio, C., Franzen, T. M. O., et al. 2015, *MNRAS*, 454, 902
- [21] Venturi, T., Giacintucci, S., Dallacasa, D., et al. 2008, *A&A*, 484, 327
- [22] Wilkinson, P. N., Kellermann, K. I., Ekers, R. D., Cordes, J. M., & W. Lazio, T. J. 2004, *New Astr. Rev.*, 48, 1551