*Education and Heritage in the era of Big Data in Astronomy*
*Proceedings IAU Symposium No. 367, 2020*
*R. M. Ros, B. Garcia, S. R. Gullberg, J. Moldon & P. Rojo, eds.*
doi:10.1017/S1743921321000879

# Open Astronomy and Big Data Science

**Bärbel S. Koribalski**

CSIRO Astronomy and Space Science, Australia Telescope National Facility,
P.O. Box 76, Epping, NSW 1710, Australia
email: Baerbel.Koribalski@csiro.au

**Abstract.** Open Astronomy is an important and valuable goal, including the availability of refereed science papers and user-friendly public astronomy data archives. The latter allow and encourage interested researchers from around the world to visualise, analyse and possibly download data from many different science and frequency domains. With the enormous growth of data volumes and complexity, open archives are essential to explore ideas and make discoveries. Open source software is equally important for many reasons, including reproducibility and collaboration. I will present examples of open archive and software tools, including the CSIRO ASKAP Science Data Archive (CASDA), the Local Volume HI Survey (LVHIS), the 3D Source Finding Application (SoFiA) and the Busy Function (BF). Astronomy is international and includes or links to an incredibly wide range of sciences, computing, engineering, and education. Its open nature can serve as an example for world-wide interdisciplinary collaborations.

**Keywords.** Astronomical data bases, catalogs, surveys, ISM: galaxies, methods: data analysis

## 1. Introduction

For the purpose of education across the world, trustworthy and up-to-date public webpages, databases, software tools, etc., are essential. Ideally these are provided in many different languages and at a range of entry levels to ensure wide uptake. In our field of astronomy a large range of excellent sites exist and are regularly updated and expanded. Here I list a small selection of Open Astronomy sites that are indispensable in my field:

- ArXiv: *arxiv.org*
- Astrophysics Data System (ADS): *ui.adsabs.harvard.edu*
- NASA/IPAC Extragalactic Database (NED): *ned.ipac.caltech.edu*
- Lyon-Meudon Extragalactic Database (LEDA): *leda.univ-lyon1.fr*
- CDS Portal: *cdsportal.u-strasbg.fr*
- – SIMBAD Astronomical Database: *simbad.u-strasbg.fr*
- – VizieR: *vizier.u-strasbg.fr*
- – Aladin Lite: *aladin.u-strasbg.fr/AladinLite/*
- SkyView: *skyview.gsfc.nasa.gov*
- ESA Sky: *sky.esa.int*
- Astropy: *astropy.org*

Most astronomers would be familiar with many if not all of these sites, but often less well known are **Open Radio Astronomy** observatories, outreach & education resources, etc. In the following I briefly describe data portals from CSIRO's Australia Telescope National Facility (ATNF), which includes the famous 64-m Parkes Radio Telescope, the Australia Telescope Compact Array (ATCA) and the new Australian Square Kilometer Array Pathfinder (ASKAP), each regularly enhanced with innovative instrumentation.

**Figure 1.** ATNF radio telescopes. — From left to right: ASKAP (36 × 12-m dishes equipped with wide-field Phased-Array Feeds (PAFs); 6-km diameter), ATCA (6 × 22-m dishes; 6-km diameter) and the 64-m Parkes Telescope as featured in the ATNF Daily Astronomy Picture (ADAP: *www.atnf.csiro.au/ATNF-DailyImage/* created by B. S. Koribalski).

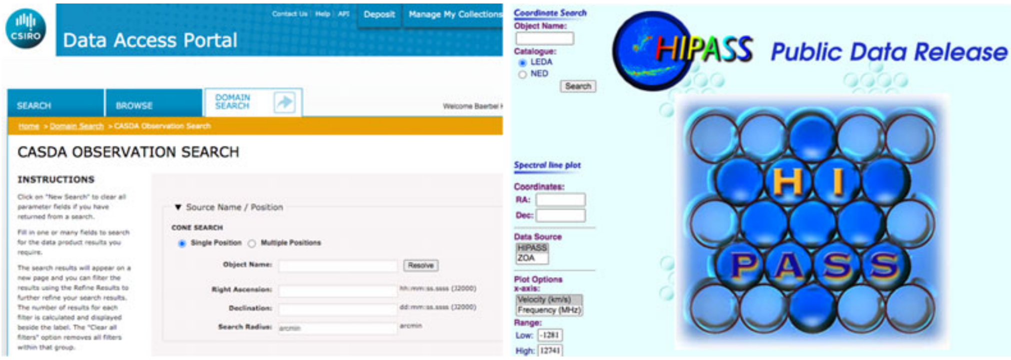


**Figure 2.** Screenshots of the CASDA portal and the HIPASS Spectral Line Server.

## 2. ATNF Data Archives and Software

The Australia Telescope Online Archive (ATOA) contains unprocessed data from the ATCA (visibilities) and the Parkes Telescope (spectral line and continuum surveys) as well as unprocessed data and data products from the 22-m Mopra Telescope. ATCA observations are generally calibrated using the MIRIAD open software package: *www.atnf.csiro.au/computing/software/miriad/*. Observations with the Compact Array Broadband Backend (CABB; Wilson et al. 2011) on average add ∼25 TB $\mathrm{yr}^{-1}$ to the archive compared to ASKAP's expected data rate of 500 PB $\mathrm{yr}^{-1}$.

Our fastest growing database is the CSIRO ASKAP Science Data Archive (CASDA), which is dedicated to holding the pipeline-reduced and validated output products (2D images, 3D image cubes, spectra, catalogs and in some cases calibrated visibilities) from ASKAP (see Figs. 1 & 2). Each ASKAP dish is equipped with Chequerboard PAFs which are typically used to form 36 beams on the sky providing an instantaneous field of view of around 30 square degrees at 1.4 GHz. CASDA contains excellent pilot survey data from WALLABY (Koribalski 2012, Koribalski et al. 2020) and EMU (Norris et al. 2011), including gas-rich galaxies, groups and clusters, distant black holes and giant radio galaxies. A recent addition is the Rapid ASKAP Continuum Survey (RACS; McConnell et al. 2020) conducted at 888 MHz, covering the entire sky south of declination $\delta = +51$ degr at 15 arcsec resolution and a median rms of 250 microJy $\mathrm{beam}^{-1}$.

- CASDA: *data.csiro.au/dap/public/casda/casdaSearch.zul*
- ATOA: *atoa.atnf.csiro.au*
- HIPASS spectral data release: *www.atnf.csiro.au/research/multibeam/release*
- LVHIS Atlas and Database: *www.atnf.csiro.au/research/LVHIS*

https://doi.org/10.1017/S1743921321000879 Published online by Cambridge University Press

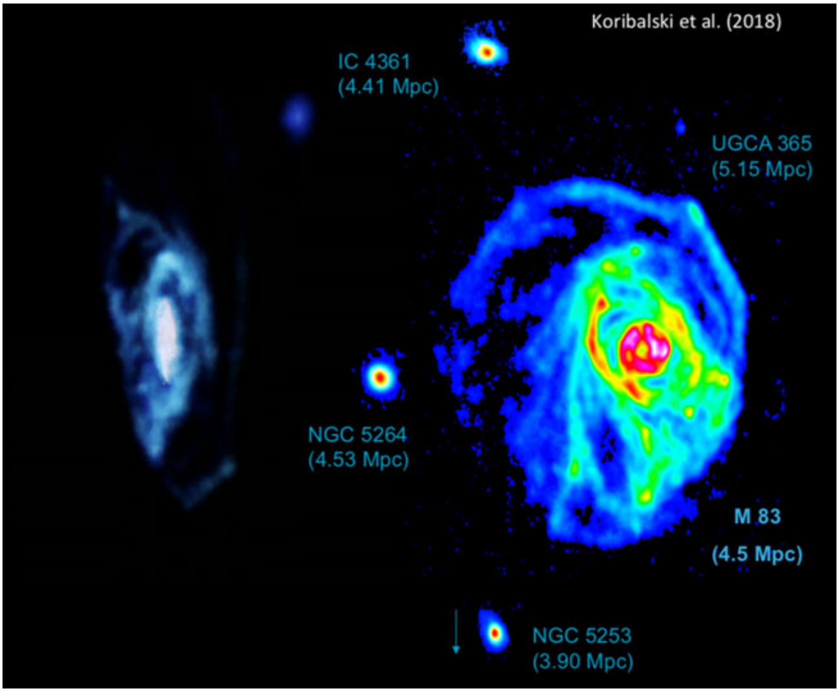


**Figure 3.** The galaxy M 83 and companions. — Left: 3D Visualisation of the stars and gas in the spiral galaxy M 83 using observational data. — Right: ATCA H I intensity distribution of M 83 and its neighboring dwarf galaxies from the LVHIS project (Koribalski et al. 2018).

## 3. The Local Volume HI Survey (LVHIS)

The LVHIS Atlas and Database (Koribalski et al. 2018) contain ATCA H I data cubes and H I moment maps for ∼80 nearby galaxies. These highlight the often very large, fast-rotating gas disks of spiral galaxies, as well as the wide range of dwarf galaxy morphologies. The 21-cm line of atomic hydrogen is an excellent tracer of past and on-going galaxy interactions and transformations (see Figs. 3 & 4). For ATCA-LVHIS we targeted a complete sample of southern gas-rich galaxies within 10 Mpc, selected from HIPASS (Koribalski et al. 2004). The raw (uncalibrated) data can be downloaded from the ATOA. Calibrated data and additional data products can be made available on request to the principal investigator. Figure 3 shows the large spiral galaxy M 83 and its nearest gas-rich neighbours. Figure 4 shows a collage of 20 LVHIS galaxies (not to scale), including M 83, ranging from low-mass dwarf companion galaxies to giant gas-rich spirals. Here we highlight the mean H I velocity fields obtained by combining data from typically three different ATCA configuration (∼30 hours on-source), with emphasis on the diffuse outer galaxy disks. Blue and red colours indicate the galaxies approaching and receding sides, respectively. While some of the displayed galaxies reveal reasonably symmetric velocity fields, most show peculiar features such as warps, bars, and tidal tails.

Similar quality H I images of nearby galaxies will be delivered by the ASKAP H I All Sky Survey (known as WALLABY). Early science and pilot survey data for WALLABY and other ASKAP science projects are already available in CASDA (e.g., For et al. 2019, Kleiner et al. 2019, Lee-Waddell et al. 2019, Koribalski et al. 2020), while the full H I 21-cm sky survey ($\delta < +30$ degr) with a resolution of ∼30 arcsec and 4 km s$^{-1}$) is expected to commence in 2021/2. We expect to detect at least 500 000 H I-rich galaxies out to redshift $z \lesssim 0.2$, including ∼5 000 well-resolved galaxies mostly known from HIPASS.

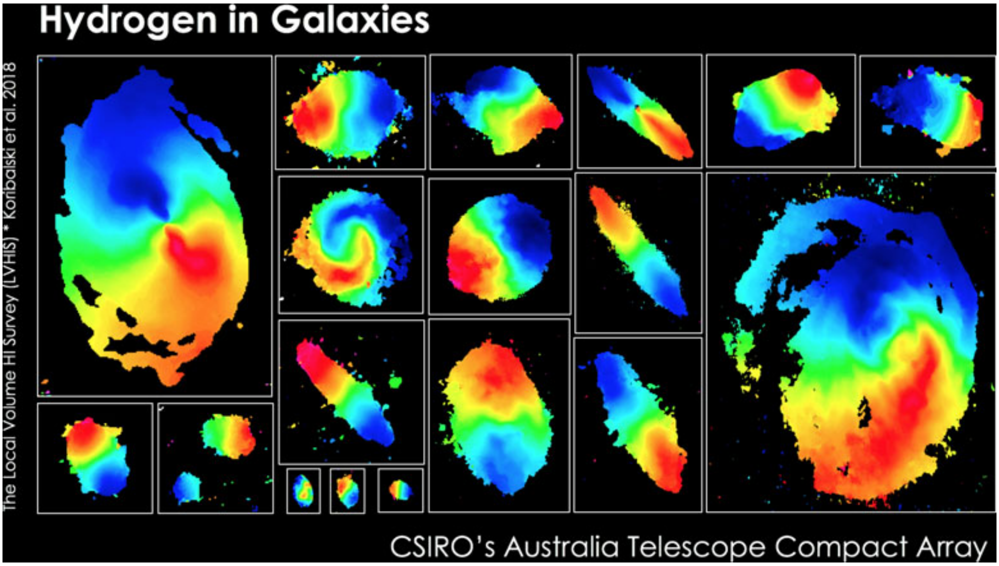


**Figure 4.** Collage of LVHIS galaxy velocity fields (Koribalski et al. 2018).

Our 3D Source Finding Application (SoFiA; Serra et al. 2015), which was developed for WALLABY (Koribalski et al. 2020), is an open-source software package with extensive user support. It is typically used for 3D source finding in spectral line data sets and includes a reliability estimate for every candidate. Among its outputs are source cubes, moment maps, integrated spectra and source properties. To enhance fitting of H I spectra as well as generation of realistic H I spectra for simulations we also developed the Busy Function (BF; Westmeier et al. 2014), which is available stand-alone (BusyFit) or as part of SoFiA (see *www.atnf.csiro.au/people/Tobias.Westmeier/tools.php*). These and other software packages as well as open multi-wavelength data archives are invaluable for research and education, here our ASKAP projects and future SKA projects.

## References

For, B.-Q., et al. 2019, *MNRAS*, 489, 5723
Kleiner, D., et al. 2019, *MNRAS*, 488, 5352
Koribalski, B.S., et al. 2004, *AJ*, 128, 16
Koribalski, B.S. 2012, *PASA*, 29, 359
Koribalski, B.S., et al. 2018, *MNRAS*, 478, 1611
Koribalski, B.S., et al. 2020, *Ap&SS*, 365, 118 *(https://rdcu.be/b5Bfg)*
Lee-Waddell, K., et al. 2019, *MNRAS*, 487, 5248
McConnell, D., et al. 2020, *PASA*, 37, 48
Norris, R.P., et al. 2011, *PASA*, 28, 215
Serra, P., et al. 2015, *MNRAS*, 448, 1922
Westmeier, T., et al. 2014, *MNRAS*, 438, 1176
Wilson, W., et al. 2011, *MNRAS*, 416, 832