

Chapter 8

Science Archives: Facilitating Survey Science

M. Read, B. Mann, R. Blake, R. Collins, N. Cross, C. Davenhall, M. Holliman, and E. Sutorius

Abstract In this paper we discuss the role of science archives and data centres in supporting survey astronomy. We start by describing the work of the Wide Field Astronomy Unit (WFAU) at the Institute for Astronomy of the University of Edinburgh, which in recent years has built science archives for the current large area and deep infrared surveys (UKIDSS and VISTA). We then go on to briefly discuss, in more general terms, how archives should operate and service the current and ever increasing volume of digital survey data.

8.1 The Wide Field Astronomy Unit

The Wide Field Astronomy Unit (WFAU) is part of the Institute for Astronomy at the University of Edinburgh. Prior to its formation, WFAU had operated under the guise of the UK Schmidt Telescope Unit (UKSTU) and operated the SuperCOSMOS measuring machine. WFAU is currently staffed by a mixture of astronomers, software developers and systems administrators and has been serving data and supporting survey science for over 25 years.

8.1.1 WFAU Science Archives

The first digital science archive created and hosted by WFAU was based on SuperCOSMOS scans of Schmidt photographic survey plates [1]. Using the Sloan Digital Sky Survey (SDSS) [2] as a model, the SuperCOSMOS Science Archive

M. Read (✉) • B. Mann • R. Blake • R. Collins • N. Cross • C. Davenhall • M. Holliman • E. Sutorius

Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

e-mail: mar@roe.ac.uk

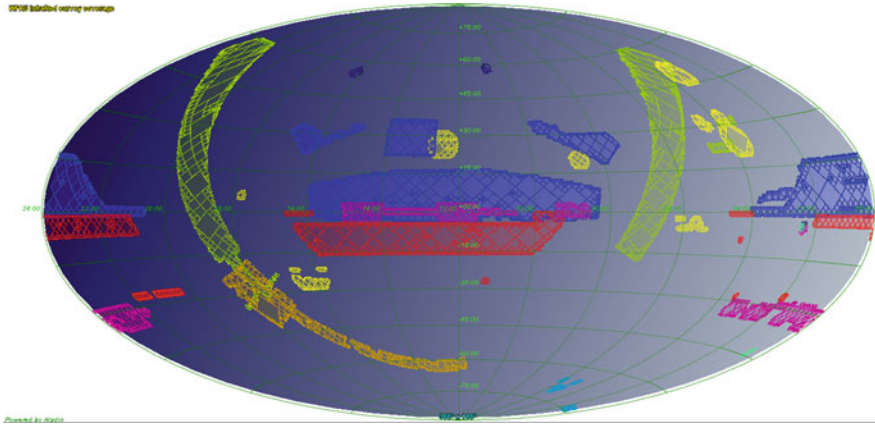


Fig. 8.1 Coverage of the UKIDSS and VISTA surveys archived by WFAU

(SSA¹), is housed in a relational database management system (Microsoft SQL Server). Related to its SSA work, WFAU also designed and hosts the 6dF Galaxy Redshift Survey (6dFGS) archive.²

As part of the VISTA Data Flow System, WFAU then went on to develop archives for the recent and ongoing major infrared surveys carried out by WFCAM on the United Kingdom Infrared Telescope (UKIRT) and by the Visible and Infrared Survey Telescope for Astronomy (VISTA). With the WFCAM Science Archive (WSA³) [3] primarily providing users with access to the UKIRT Infrared Deep Sky Surveys (UKIDSS [4]): Large Area Survey (LAS), Galactic Plane Survey (GPS), Galactic Clusters Survey (GCS), Deep Extragalactic Survey (DXS) and Ultra Deep Survey (UDS) and the VISTA Science Archive (VSA⁴) hosting data from the VISTA public surveys [5]: VISTA Hemisphere Survey (VHS), VISTA Variables in the Via Lactea (VVV), VISTA Kilo-Degree Infrared Galaxy Survey (VIKING), VISTA Kilo-Degree Infrared Galaxy Survey (VMC) and VISTA Deep Extragalactic Observations Survey (VIDEO). In addition to archiving the surveys carried out by UKIRT and VISTA, WFAU supports astronomers by providing data releases for their own programmes and also assists VISTA survey heads in providing catalogue and images for ingest into the ESO Science Archive (Fig. 8.1).

WFAU also operates the OmegaCAM Science Archive (OSA) for the ATLAS survey carried out by the VLT Survey Telescope (VST) and the Gaia-ESO Survey (GES⁵) Science Archive which serves spectroscopic data and analysis. In the near future WFAU will publish data from the UKIRT Hemisphere Survey (UHS).

¹<http://surveys.roe.ac.uk/ssa>

²<http://www-wfau.roe.ac.uk/6dFGS>

³<http://surveys.roe.ac.uk/wsa>

⁴<http://surveys.roe.ac.uk/vsa>

⁵<http://ges.roe.ac.uk>

8.1.2 Data Flow System

For the archives currently been curated by WFAU (WSA, VSA, OSA and GES) the initial, nightly, pipeline processing is carried out by the Cambridge Astronomy Survey Unit (CASU). The photometrically calibrated images and catalogues are then transferred up to Edinburgh where the observational metadata and object catalogues are ingested into the relevant load database. Transfers and ingests are typically run on a monthly basis as this allows for the accumulation of sufficient calibration observations to accurately reduce the data. Following quality control and flagging, additional data products are created, for example band merged catalogues, deep stacks and variability analysis tables and neighbour tables with other surveys.

Periodic data releases are made from the load database which are then copied to a static release database accessed by the user.

8.1.3 Data Volumes and Access

WFAU archives currently total over 1.0 PB in size, with 0.35 Pb of that being held in databases. The tables house over 10^{12} rows and provide calibrated photometry and astrometry for single-band and band-merged sources. Neighbour tables with external surveys are provided to allow fast cross-referencing and variability measurements for multi-epoch programmes are also generated.

The VISTA VVV survey, with its observations of the Galactic Plane, accounts for around 30% of the data and presents the most challenges for curation. The largest table in the archives, with over 19 billion rows, is the catalogue of VVV sources.

Data releases are often initially proprietary and users are required to login to access these databases (there are over 1,000 registered UKIDSS users). Databases are made publicly accessible once any proprietary period has passed.

Users submit over 50,000 queries each month to the archives, with the queries returning over 2 billion rows of results. Users are able to directly submit SQL queries through the browser user-interface thus providing a powerful and flexible way to mine the data. Thousands of images, mainly cut-outs, are also served out daily.

8.1.4 Current Development Work

WFAU is currently focussing its development work on providing users with an even richer environment for exploring and querying its archives. This work is driven by user requirements for developing complex queries iteratively, the need to query across more than one archive and increasing data volumes.

Using Virtual Observatory infrastructure⁶ at its core the system being developed will support distributed queries and allow users to stage and share results in their own user database. An updated, browser based, user-interface will be the access point for the system. A prototype of which is available for the OSA.⁷

8.2 Supporting Survey Science

To allow astronomers to get the best science from current and future surveys, science archives and data centres should exhibit several attributes:

1. High quality science ready data releases e.g. calibrated and quality controlled.
2. Intuitive/powerful user interfaces and access through the Virtual Observatory.
3. Good user support e.g. documentation and helpdesk
4. Add value to the data e.g. neighbour table with other surveys, coverage maps

These requirements are borne out by WFAU's interaction with the astronomical community over two decades.

Larger data volumes and rates produced by current and planned surveys will increase the demands on archives. Flexible queries, the staging of results and querying across archives will go some way in supporting users access these datasets. However more of the analysis will need to be carried out at the data centres and archives will need to develop systems for users to upload and run their own algorithms.

8.3 Conclusion

We have described the archives operated by WFAU and the work in development to further facilitate the science carried out on the survey data held in these archives. Data centres have an increasingly important role to play in getting the best out of the current and future sky surveys. It is important to have involvement throughout the data life-cycle and understand and support users in the way they work.

⁶<http://www.ivoa.net>

⁷<http://osa.roe.ac.uk>

References

1. Hambly, N., MacGillivray, H.T., Read, M.A., et al.: The SuperCOSMOS sky survey – I. Introduction and description. *MNRAS* **326**, 1279–1294 (2001)
2. Szalay, A.S., Kunszt, P.Z., Thakar, A.R., et al.: The Sloan Digital Sky Survey and its archive. *ASPC* **216**, 405 (2000)
3. Hambly, N., Collins, R.S., Cross, N.J., et al.: The WFCAM science archive. *MNRAS* **384**, 637–662 (2008)
4. Dye, S., et al.: The UKIRT infrared deep sky survey early data release. *MNRAS* **372**, 1227 (2006)
5. Emerson, J.P., et al.: The visible & infrared survey telescope for astronomy. *ESO Messenger* **117**, 27–32 (2004)