Astronomical Data Analysis Software and Systems XVII **P4.15**
ASP Conference Series, Vol. 394, © 2008
R. W. Argyle, P. S. Bunclark, and J. R. Lewis, eds.

The JCMT Legacy Survey: The Challenges of the JCMT Science Archive

Frossie Economou, Tim Jenness, Antonio Chrysostomou, Brad Cavanagh

Joint Astronomy Centre, 660 N. A'ohōkū Place, Hilo, HI, 96720, USA

Russell Redman

Canadian Astronomy Data Centre, 5071 West Saanich Road, Victoria, B.C., Canada

David S. Berry

Centre for Astrophysics, University of Central Lancashire, Preston, UK

Abstract. The JCMT Legacy Survey (JLS) is an ambitious programme that will provide the first large-scale survey of the submillimetre sky. From planetary science through to cosmology, the first phase of the campaign will be using the continuum instrument SCUBA-2 and the imaging spectrometer HARP to obtain over 100 Terabytes of raw data. We present our ongoing work to provide a Science Archive aimed at both survey and individual users for the distribution of raw data, processed data, advanced data products and VO integration.

1. Background

The James Clerk Maxwell Telescope is a 15-m submillimetre telescope on the summit of Mauna Kea, Hawaii funded by the United Kingdom, Canada and the Netherlands. It is currently in the process of upgrading its full instrument suite and now has ACSIS, a new heterodyne correlator (Hovey et al. 2000); HARP, a 16 element B-band (320 to 370 GHz) heterodyne array receiver (Smith et al. 2003); and is about to receive SCUBA-2, a new bolometer array (Holland et al. 2006). Due to the extraordinary mapping speed of these instruments (SCUBA-2 maps the sky three orders of magnitude faster than its predecessor), a whole new window on the submillimetre sky has been opened.

The high demand in our community for this instrumentation raised the possibility that there would be a stampede to observe the most popular fields and samples. In order to ensure a coherent use of telescope time, as well as allow for the numerous nights that would be necessary to attain the varied scientific goals, a JCMT Legacy Survey programme (JLS) was proposed that would both produce data of immediate interest and lasting legacy value.

In order to deal with the significant amounts of data expected to arise from this programme (estimated at 100TB), facilitate the collaboration within members of the surveys, and ensure publication in the Virtual Observatory to enhance its legacy value, the JCMT entered into a collaboration with the Canadian Astronomical Center to provide a JCMT Science Archive (Gaudet et al. 2008).

The nature of the data, the structure of the surveys and operational model of JCMT pose a number of interesting challenges to the design and implementation of the JCMT Science Archive (JSA).

2. The Surveys

The surveys were defined by considering individual proposals for telescope time; there is no centralised umbrella group as the UKIDSS consortium for the UKIRT WFCAM surveys (Lawrence et al. 2007).

Seven surveys were allocated time in the following research areas:

- Spectral Legacy survey (SLS; Plume et al. 2007): A spectral imaging survey of the content and distribution of all the molecules detected in the 345 GHz atmospheric window toward a sample of 5 Galactic sources.
- Debris Disk survey (DDS; Matthews et al. 2007): A survey of 500 stars (the nearest 100 of each of 5 spectral types - A to M) to search for debris disks.
- Gould Belt Survey (GBS; Ward-Thompson et al. 2007): An investigation of all accessible low-mass star forming regions within 500pc.
- JCMT survey of the Galactic Plane (JPS; Moore et al. 2005): A survey of the Galactic Plane ($|b| < 1\,\mathrm{deg}$).
- Nearby Galaxy Survey (NGS): A survey of 155 galaxies within 25 Mpc in both continuum and CO 3-2 emission.
- Cosmology Survey (CLS): A number of deep fields at 450μm and 850μm totaling 15 square degrees.
- SCUBA-2 All-Sky Survey (SASSy; Thompson et al. 2005, 2007): A shallow scan across $|b| < 5\,\mathrm{deg}$ and a 10 deg-wide strip through the north celestial pole with the aim to eventually be extended to all the sky visible from JCMT.

These surveys use the majority of JCMT instrumentation including the polarimeter (Bastien, Jenness & Molnar 2005) and the Fourier Transform Spectrometer (FTS) (Naylor, Gom & Zhang 2006), so they will result in a highly heterogeneous data set. Which surveys use which instruments is listed in Table 1.

Table 1. The instrumentation usage by survey.

	HARP	SCUBA-2	FTS-2	POL-2
SLS	×			
DDS		×		
GBS	×	×		×
JPS		×	×	
NGS	×	×		
CLS		×		
SASSy		×		

3. The Challenges of the JCMT Science Archive

The JSA is envisaged as a dynamic collaborative tool instead of a passive repository for data. It contains:

- Raw data: This is the rawest form of data available to us on disk, in the native format provided by our Observatory Control System. In the case of ACSIS, these are time series data. The audience for these data are users who want to use our tool suite to customise the data reduction process to a fundamental level, for example to chose a different gridding algorithm.
- Minimally processed data: These are data that have been processed just far enough to allow export in a standard format. In the case of ACSIS, these are FITS cubes that have been gridded from the time series data. The audience for these data are users who wish to use third-party tools to further process their data.
- Basic Data Products (BDPs): These are reduced products consisting of maps and cubes created by processing, coadding and calibrating raw data. They are created by our data reduction pipeline, ORAC-DR. The audience for these data are PIs and survey users that understand and trust the automated reduction process but wish to further analyse the data by hand to identify sources, lines, etc.
- Advanced Data Products (ADPs): These are advanced products derived from the basic data products. They include large maps from more than one project, catalogues, line lists, noise maps and more (Jenness et al. 2008). The audience for these data are primarily users who are interested in the final results of the observing campaigns, and VO users who lack interest in intermediate forms of the data.

Some challenges that arise from operating such an archive with our data and in our operational environment are:

- The data come from instruments with not only a very high data rate (theoretically up to 0.5TB/night at full speed) but also requiring expensive processing (SCUBA-2 data are reduced using an iterative technique as the analytical solution is computationally prohibitive). Our approach: Create a range of ADPs that can spare individual users from expending resources to reduce the data themselves.
- The highly diverse nature of the JLS science goals of the survey groups and the instrumentation make it unlikely that a single uniform product will be scientifically useful to all concerned. Our approach: Implement a single advanced pipeline infrastructure that runs recipes specific to each survey thus generating products for all fields but using different algorithms.
- The lack of a centralised survey authority makes communication and feedback difficult. Our approach: Create a JCMT Data User Group with a representative from each survey who seek input from their survey membership and provide feedback to the technical working groups.
- The large sizes of the products make casual use of the archive prohibitive to those who won't or can't pay the bandwidth penalty of download. Our approach: a number of different advanced previews of the data that enable the user to judge correctly whether they wish to download the actual product or not.

- The complexity of the data means that automatic and/or iterative analysis of the data, while useful, may not be as optimal as more manual/analytical techniques. Our approach: Make it possible for both surveys and individual users to upload their versions of ADPs to the archive so that they can enhance its legacy value.
- VO protocols, in particular the catalogue models (see e.g. Dowler et al. 2008), are not able to describe complex features such as clumps and filaments that are abundant in our data (see e.g. Berry et al. 2007). Our approach: Calculate masks for the sources. That allows us to specify an elliptical envelope and a velocity extent for VO publication but on retrieval use a mask derived from fitting the data to cut out and return the true outline of the source.

4. Where are we now?

- Raw data distribution for ACSIS data is in place.
- Reduced data distribution for ACSIS data is coming soon.
- The automatic processing infrastructure is under development.
- The HARP components of the surveys are starting this winter.
- We are awaiting delivery of SCUBA-2

References

Bastien, P., Jenness, T., & Molnar, J. 2005, in ASP Conf. Ser. 343, Astronomical Polarimetry: Current Status and Future Directions, ed. A. Adamson, C. Aspin, C. J. Davis, & T. Fujiyoshi (San Francisco: ASP), 69

Berry, D. S., Reinhold, K., Jenness, T., & Economou, F. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco: ASP), 425

Dowler, P., Gaudet, S., Hill, N., Redman, R. 2008, in ASP Conf. Ser. 394, ADASS XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco: ASP), [P4.7]

Gaudet, S., Dowler, D., Goliath, S., & Redman, R. 2008, in ASP Conf. Ser. 394, ADASS XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco: ASP), [O6b.2]

Holland, W. S., et al. 2006 Proc. SPIE, 6275, 45

Hovey, G. J., et al. 2000, Proc. SPIE, 4015, 114

Jenness, T., Cavanagh, B., Berry, D. S., & Economou, F. 2008, in ASP Conf. Ser. 394, ADASS XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco: ASP), [P7.1]

Lawrence, A., et al. 2007, MNRAS, 379, 1599

Matthews, B. C., et al. 2007, PASP, 119, 858

Moore, T. J. T., et al. 2005, in Protostars and Planets V, 8370

Naylor, D. A., Gom, B. G., & Zhang, B. 2006, Proc. SPIE, 6275, 62751Z

Plume, R., et al. 2007, PASP, 119, 102

Smith, H., et al. 2003, Proc. SPIE, 4855, 338

Thompson, M. A., et al. 2005, in Protostars and Planets V, 8145

Thompson, M. A., et al. 2007, PASP, in prep, (astro-ph/0704.3202)

Ward-Thompson, D., et al. 2007, PASP, 119, 855