Advanced data products for the JCMT Science Archive

Graham S. Bell*[a], Sarah F. Graves[a], Malcolm J. Currie[a], David S. Berry[a], Harriet Parsons[a],
Tim Jenness[a,b], Russell O. Redman[a], Jessica T. Dempsey[a], Doug Johnstone[a], Frossie Economou[a,c]

[a]Joint Astronomy Centre, 660 N. A‘ohōkū Place, Hilo, HI 96720, USA;

[b]Department of Astronomy, Cornell University, Ithaca NY, 14853, USA;

[c]National Optical Astronomy Observatory, 950 N Cherry Ave, Tucson, AZ 85719, USA

ABSTRACT

The JCMT Science Archive is a collaboration between the James Clerk Maxwell Telescope and the Canadian Astronomy Data Centre to provide access to raw and reduced data from SCUBA-2 and the telescope's heterodyne instruments. It was designed to include a range of advanced data products, created either by external groups, such as the JCMT Legacy Survey teams, or by the JCMT staff at the Joint Astronomy Centre.

We are currently developing the archive to include a set of advanced data products which combine all of the publicly available data. We have developed a sky tiling scheme based on HEALPix tiles to allow us to construct co-added maps and data cubes on a well-defined grid. There will also be source catalogs both of regions of extended emission and the compact sources detected within these regions.

Keywords: JCMT Science Archive, Catalogs, Surveys, HEALPix, Submillimeter, Virtual Observatory

1. INTRODUCTION

The James Clerk Maxwell Telescope (JCMT) Science Archive[1–3] (JSA) was approved in 2005 to provide a means for researchers quickly to obtain reduced data products from their observations, whilst also integrating the data holdings into the Virtual Observatory as they become public. Early plans for the archive[4,5] introduced a distinction between *basic data products* from the data-reduction pipeline and *advanced data products* generated by applying further processing. These roughly correspond to Calibration Levels 2 and 3 of the IVOA core observation data model.[6]

The archive currently includes basic data products comprising pipeline-reduced maps and data cubes for each observation and for each night, along with corresponding preview images. The next phase of development is the implementation of advanced data products, using the ORAC-DR pipeline[7–9] and the infrastructure of the Canadian Astronomy Data Centre (CADC). In addition to the products discussed here, we expect that further, more specialized, products will be created by external groups, such as the JCMT Legacy Survey (JLS) teams.[10–16]

The primary advanced data projects are co-added "legacy" tiles, containing all available publicly released data, and catalogs generated from them. The catalogs will include both regions of extended emission and also the compact sources detected within these regions. One challenge in producing these tiles and catalogs is the need to use sufficiently generic methods that they are appropriate for observations with very different noise limits and observing strategies taken towards a wide variety of types of target — from dusty circumstellar disks to non-uniform star formation regions to high-redshift point-like cosmological sources. The catalogs can be supplemented by domain-specific catalogs created by the different JLS teams and PI projects, which could classify and derive physical properties for the types of objects that are the focus of each survey.

One way in which the importance of the archive is illustrated is by its impact on the telescope's publication statistics. In Figure 1 we show the proportion of publications which made use of the archive in comparison to those using only new data from the various instruments. The archive is already being used in a large number of publications, and we expect that making advanced data products available will further improve its uptake, especially outside the core user community. This may be aided by the fact that the CADC recently (May, 2014) transitioned to a new unified search interface based on its next-generation data model, CAOM-2.[17,18] This allows astronomers to search for data in all of CADC's collections, including the JCMT Science Archive, simultaneously.

---

[*] E-mail: g.bell@eaobservatory.org

Bell, G. S. et. al., "Advanced data products for the JCMT Science Archive," Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

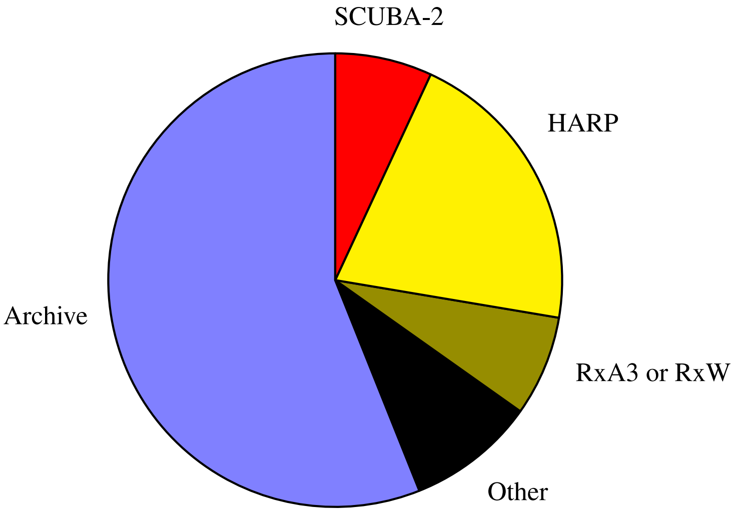


Figure 1. Chart illustrating the proportion of refereed papers using JCMT data published between 2010 and 2013 which made use of the archive. The remaining segments of the chart indicate the proportion using data from each of the current JCMT instruments but not the archive, and those using other, older, instruments.

2. TILES

The advanced data products in the JCMT Science Archive (JSA) will be tiled using the Hierarchical Equal Area isoLatitude Pixelization[19] (HEALPix) scheme. In the base resolution of HEALPix, the sky is divided into twelve diamond-shaped facets arranged in three rings — one around the equator and another around each pole. Each facet is divided into a 2×2 grid of similar smaller diamonds to form the next resolution level, and this process is repeated recursively until the desired cell size is reached. HEALPix therefore offers a well-defined grid over the sky which we will use to specify the boundaries of the tiled data products.

The individual tiles will also use HEALPix as the pixel grid, via the `HPX` projection algorithm code[20] which rotates the HEALPix facets anticlockwise by 45° so that they become squares aligned with the coordinate axes. Using this grid ensures that the pixelization is continuous between adjacent tiles, and the tiles can be joined by abutting them. In this respect, `HPX` offers a distinct advantage over previous tiling schemes using a tangent-plane projection which would introduce distortions towards the edges of large tiles, making it more difficult to combine them. The tiling parameters which we plan to use are given in Table 1. The tile sizes were chosen after investigating how frequently observations would be broken up by tile boundaries. Figure 2 shows the result of this investigation for the selected tile sizes. For HARP we see that the majority of observations will fit into one or two tiles. For SCUBA-2 the graph shows a number of features corresponding to the various sized mapping modes which the instrument employs, the largest of which are labeled. The smallest “daisy” observations will typically occupy three or fewer tiles. We consider this to be an acceptable number of tiles for a user of the archive to download, but of course if they are only interested in a part of the observation then one tile may be sufficient.

Table 1. HEALPix tiling parameters selected for the JCMT Science Archive. The HEALPix resolution, which defines the tile size, is indicated by the N_{side} parameter which gives the number of divisions along the side of each base-resolution facet. It is sometimes expressed as a resolution level, k, where $k = \log_2 N_{\mathrm{side}}$. The linear HEALPix pixel sizes are approximate, because while pixels at a given resolution level are of the same area, they are not necessarily the same shape.

Instrument	N_{side}	**Tile Size**	**Number of Tiles**	**Pixels per Tile**	**Pixel Size**
SCUBA-2 850 µm	64	$\sim 1^\circ$	49,152	1024×1024	$3.22''$
SCUBA-2 450 µm	64	$\sim 1^\circ$	49,152	2048×2048	$1.61''$
HARP	256	$\sim 14'$	786,432	128×128	$6.44''$

Bell, G. S. et. al., “Advanced data products for the JCMT Science Archive,” Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

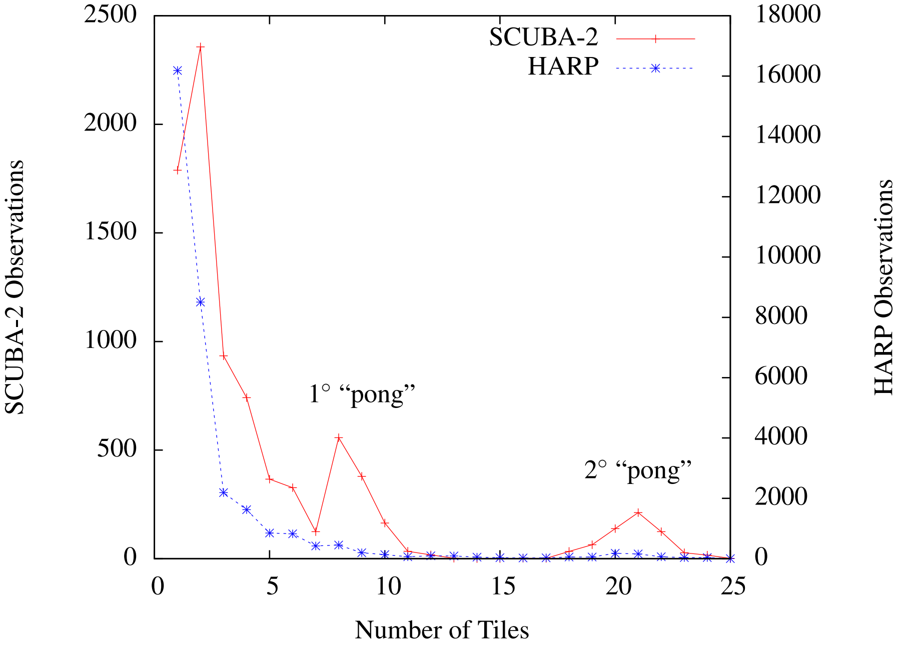


Figure 2. Graph showing the frequency at which observations are expected to be split into different numbers of tiles, using the selected tile sizes. This is based on a conversion of the bounding boxes of SCUBA-2 and HARP observations as stored our database to HEALPix cells. The number of tiles per observation will be slightly over estimated by this technique because the bounding box is not a perfect representation of the area observed — especially in the case of SCUBA-2 whose scan patterns produce roughly circular observations. The empty corners of bounding boxes can cause adjacent tiles to be counted unnecessarily, so this plot represents an upper limit to the number of tiles per observation.

There are two standard numbering systems for HEALPix: the "ring" scheme, where pixels are enumerated along isolatitude rings downwards from the north pole; and the "nested" scheme, which makes use of the hierarchical structure by adding two additional bits at each resolution level to represent the position within the subdivided pixel. We have chosen to use the "nested" scheme as it groups neighboring tiles together and because of its convenience with the grid employed by the `HPX` projection — the bits representing the position along the coordinate axes can be interleaved to determine the index within a base-resolution facet. The "nested" scheme is also being used in Virtual Observatory systems such as Multi-Order Coverage[21] (MOC) maps.

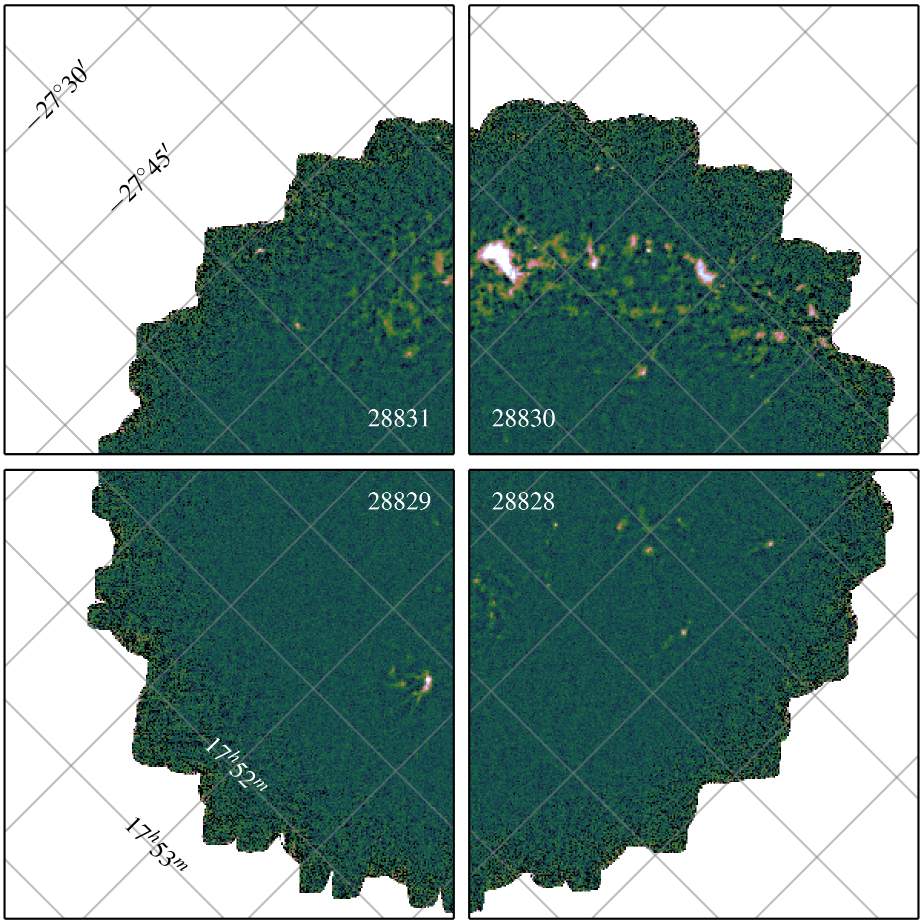
The tiled maps and cubes will be created in two stages. First individual observations will be reduced onto a grid using the `HPX` projection, and stored in the archive as a new type of basic data product. Apart from the different projection, these will differ from the existing reduced observations by the use of a consistent recipe and configuration, rather than settings chosen by each PI for their project. In the second stage, these tiled observation products will be retrieved from the archive and combined to generate the co-added tiles which, with further processing depending on the instrument, will form the new tiled advanced data products. Splitting the process in this way is necessary because many tiles would contain too much data to co-reduce in one processing job. It will also allow us to easily re-reduce or exclude from the co-add single observations in the event of problems, without needing to re-reduce the whole set of data falling within the tile.

2.1 Continuum Maps — SCUBA-2

SCUBA-2[22,23] has two focal-plane arrays, operating at 850 µm and 450 µm, which view the sky simultaneously via a dichroic beam-splitter. During normal SCUBA-2 observations the telescope scans continuously, following either a "daisy" or "pong" pattern.[24] The "daisy" pattern is a circle which offset from the source and orbits around it, and is designed for the observation of compact sources ($< 3'$). In the "pong" pattern, the telescope position bounces around a square which rotates during the observation, with standard map sizes of $15'$, $30'$, $1°$ and $2°$. The data reduction software[25,26] is highly configurable and includes example configurations for various types of maps. In addition most of the JLS teams

Bell, G. S. et. al., "Advanced data products for the JCMT Science Archive," Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

a) SCUBA-2 Pong Observation



b) SCUBA-2 Daisy Observation

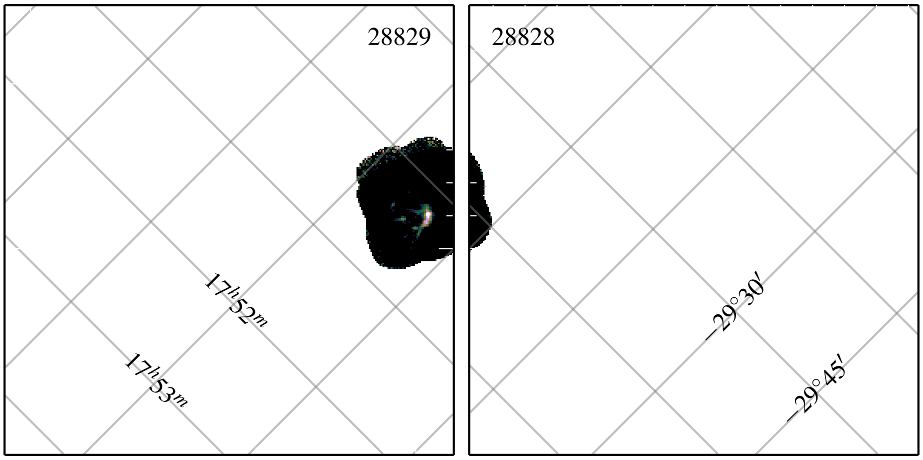


Figure 3. Example SCUBA-2 850 µm observations reduced onto HEALPix tiles. The upper panel (a) shows a single observation taken with a large “pong” scan pattern, and the lower panel (b) shows a separate observation where the “daisy” scan pattern was used. Emission from around the position of the Galactic Center can be seen in tile 28830 and the “Jellyfish” region can be seen in tile 28829.

Bell, G. S. et. al., “Advanced data products for the JCMT Science Archive,” Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

have developed dedicated configuration sets of parameter values for their respective surveys. However, for the legacy tiled data products in the JSA, it is desirable to use a single, consistent map-making configuration for all observations. We have therefore developed a new generic configuration which is intended to be conservative, in the sense that it should avoid the generation of spurious sources at the expense of not being optimal in all circumstances. This is particularly important in light of the plan to use these data reductions for the automated generation of source catalogs as described in Section 3. The generic configuration is designed to be usable with both "daisy" and "pong" observations, allowing us to co-add tiles containing observations of both types. An example of each type of scan pattern reduced in this way, taken from a survey of the Galactic Center, is shown in Figure 3. These observations include a region G 0.55−0.85[27] which resembles a jellyfish in the SCUBA-2 images and is believed to be a star-forming complex at a distance of 2 kpc.[28] Figure 4 illustrates the successful co-addition of these observations with others of the same tile.

Different pixel sizes have been chosen for the two wavelengths, owing to the different beam widths which are approximately $8''$ and $13''$ at 450 µm and 850 µm respectively. The pixel sizes, which are shown in Table 1, have been kept small to avoid distortion by the shape of the HEALPix pixels and because it has been found that the map-maker tends to work best with smaller pixels. The pixel sizes give 3 samples per beam[29] to avoid loss of information from the image. We confirmed that the `HPX` projection was usable with our data reduction system by performing a series of photometry tests, comparing a reduction onto this projection with our standard reduction of the same observation using a tangent plane projection.

SCUBA-2 data are calibrated using a flux conversion factor (FCF) to turn the measured power in picowatts to a flux in jansky per square arcsecond. The FCF should be a constant value[30] but it is recommended that the FCF be determined from calibration observations reduced using the same configuration as the data to be calibrated. Therefore we have decided to defer the application of the FCF to the tile co-adding stage, because there is no need to apply an individual FCF to each observation. The individual observation tiles will therefore be calibrated in picowatts, with the final co-added data products having the full calibration to flux units.

One peculiarity of SCUBA-2 observing is that, to increase efficiency, flux calibrators can be observed without a suitable previous pointing observation. There can thus be pointing errors in these observations, which the pipeline normally corrects by shifting the coordinate system without re-sampling the image, using the calibrator itself as a pointing source. In the case of tiles defined by the HEALPix grid, the coordinate system is effectively fixed and cannot be shifted as this would result in the pixels not being correctly aligned with the tile. Therefore it is necessary to reduce each observation of a calibration source twice — once to derive pointing offsets and a second time to create the properly aligned HEALPix tile. This technique could in general be extended to apply to observations of any sources with known locations, including pointing sources, but must be applied at the observation map-making stage due to the fixed tile grid.

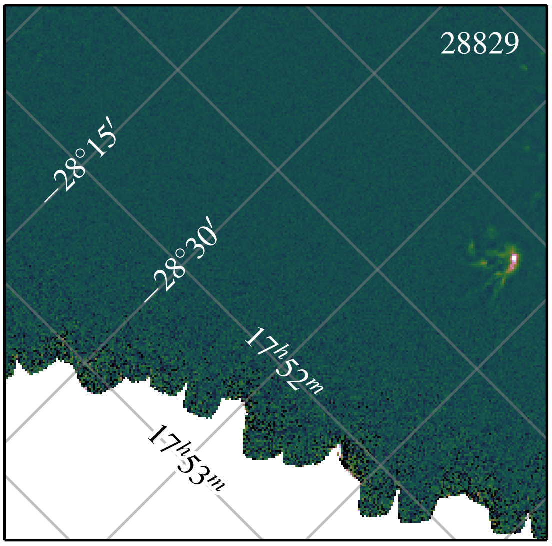


Figure 4. An example co-added SCUBA-2 850 µm HEALPix tile. This map contains data from ten individually-reduced observations which overlap with the area of tile 28829, including the daisy and pong observations shown in Figure 3.

Bell, G. S. et. al., "Advanced data products for the JCMT Science Archive," Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

2.2 Heterodyne Cubes — HARP/ACSIS

HARP[31] is a heterodyne array receiver with 16 receptors operating between 325 GHz and 375 GHz in conjunction with the Auto-Correlation Spectral Imaging System (ACSIS) as the back-end. As with SCUBA-2, the goal for the JSA advanced data products is to be able to co-add data from as wide a range of observation types as possible. We therefore aim to associate data from all science observing modes, excluding frequency-switching, with the same tuned frequency (to 1 MHz precision) and the same basic bandwidth mode. HARP always operates in single side-band mode, and the side-band separation is sufficiently good (measured at commissioning to be better than 19 dB on average[31]) to allow the co-addition of data taken in different side-bands. The observations will be reduced onto HEALPix tiles using a recipe based on our existing narrow-line recipe. Although this may not deal well with broad-line extra-galactic sources, it should be appropriate for nearly all of our large HARP data sets.

ACSIS has two correlators for each HARP receptor, each of which can be used in either a 250 MHz or 1 GHz mode. They can be used separately, "chained" together to increase spectral resolution in the same bandwidth, or overlapped slightly to form a broader "hybrid" band. Table 2 summarizes the bandwidth modes available for use with HARP. We plan to store reductions of individual observations on HEALPix tiles at their full native frequency resolution, but to down-sample them to a common frequency resolution at the co-adding stage. This will allow us to combine all the data taken with each of the basic bandwidth modes, and has the advantage of increasing the signal-to-noise ratio in the legacy tiles. Meanwhile the individual observation tiles will be available for anyone requiring the full spectral resolution.

The observing strategies used with HARP all target specific pixel sizes. There are "jiggle" mapping modes where the telescope's secondary mirror shifts the beam around a small grid of points on the sky to improve sampling within the array footprint giving pixel sizes of $6''$ or $7''$ and a "boustrophedon" (bi-directional raster) scan mode with an effective pixel size of $7.28''$. Since the desire is to be able to combine data from all mapping modes, it was necessary to select a single pixel size for all observations taken with HARP. From the set of available HEALPix sizes, $6.44''$ was chosen as it is the most similar to typical target pixel sizes. The size of tiles was selected as a trade-off between the size of potential data files ($\sim$ 1 GB for the individual observation tiles at the highest spectral resolution) and the desire to avoid having large numbers of excessively small tiles. The tile size is comfortably larger than HARP's field of view ($2'$) so small observations should not be broken into tiles unnecessarily.

3. SOURCE CATALOGS

We intend to produce catalogs of detected emission for every co-added "legacy" tile. The primary aim of these source catalogs is to make it straightforward for an astronomer to answer the question: *Did the JCMT detect any significant emission towards this position?* We also wish to provide some information about the properties of detected emission, without attempting to characterize the physical nature of the underlying sources. Towards this goal, we have chosen to produce two emission catalogs for each of our tiles. First, we identify all the regions of sky where we believe we have detected emission, using a signal-to-noise cutoff. Each spatially connected region above this threshold is identified as an *island* of emission. Second, within each of these islands we identify and characterize the local maxima as *peaks*, based on a representative noise level for the entire island.

Table 2. Table showing the bandwidth configurations of the ACSIS spectrometer which are used with HARP. We will combine data from each of the two basic modes by re-sampling the spectra to a common resolution at the co-adding stage. This resolution is expressed here both as a frequency and an approximately equivalent radial velocity for low-redshift sources.

Basic Mode	**Instrument Configurations**		**Re-sampled Resolution**
250 MHz	Single:	250 MHz × 4096 channels	100 kHz
	"Chained":	250 MHz × 8192 channels	~0.1 km/s
	"Hybrid":	Two 250 MHz × 4096 bands	
1000 MHz	Single:	1000 MHz × 1024 channels	1000 kHz
	"Chained":	1000 MHz × 2048 channels	~1 km/s
	"Hybrid":	Two 1000 MHz × 1024 bands	

Bell, G. S. et. al., "Advanced data products for the JCMT Science Archive," Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

In the context of the JSA, we are constrained to use only reliable techniques which can be applied without supervision and do not require extensive manual checking of each product. Both the *peak* catalog and the *island* catalog use the `FellWalker` algorithm[32] from the Starlink Clump Identification and Analysis Package[33,34] (CUPID) to identify sources. Like the well-known `Clumpfind` algorithm,[35] it segments a data array into clumps without reference to a predefined clump model. `FellWalker` is found to be more robust,[36] aided by the fact that it makes use of the full range of data values rather than imposing a fixed set of contour levels.

For SCUBA-2 continuum data, we will create emission catalogs from the final co-added tiles at 450 µm and 850 µm. Figure 5 shows an example of islands and peaks identified within a section of the 850 µm tile featuring the "Jellyfish" source. In the case of heterodyne observations taken with HARP/ACSIS, we will first collapse the tiled position-position-velocity cubes to create two-dimensional integrated intensity maps, which will then be used to search for emission.

3.1 Island Catalogs

Clump detection algorithms such as `FellWalker` make use of the noise level of the data in order to determine which signal levels constitute a valid detection, and which should be ignored. However, any given JSA tile can be formed from overlapping observations of different sizes, scan patterns and integration times, potentially taken in a variety of atmospheric opacity conditions. As a result, the noise across the image can vary significantly. If we attempted to select a constant representative noise value for the whole map, we would either miss clearly detected but lower-flux emission from low-noise areas, or produce spurious detections of emission in higher-noise areas. To avoid these difficulties, we identify detected emission on the basis of a signal-to-noise ratio (S/N) map for each tile.

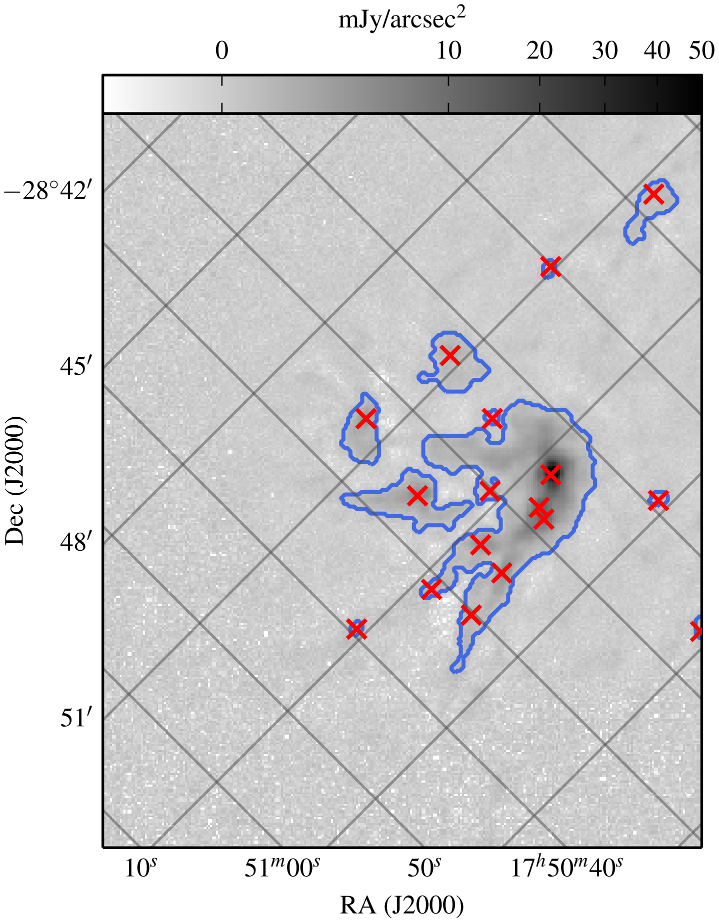


Figure 5. Island and peak catalogs overlaid on a section of the SCUBA-2 850 µm map shown in Figure 4 (tile 28829). The identified islands of emission are outlined in blue and the positions of the local peaks are indicated by red crosses.

The main limitation of an S/N based approach is that it does not allow us easily to give an upper flux limit for regions that are observed, but in which we do not detect emission. Users who require this information will be able to extract it from the co-added tiles as these will include variance information on a pixel-by-pixel basis. In the case of SCUBA-2, the variance of a pixel is estimated from the scatter of input data to that pixel. Therefore before calculating the S/N for SCUBA-2 tiles, we block-average the variance array at the beam size using a median filter. This is not done for heterodyne tiles, where there is a much more accurate measurement of the variance of each pixel.

To be identified as an island, we have decided that a region must have a peak S/N greater than five and the edges of the island will be followed down to an S/N cutoff of three. The island must also be larger than the telescope beam at the appropriate wavelength. We will use the `FellWalker` algorithm to identify the islands using these criteria, as implemented within CUPID. After clump identification, the CUPID software runs two iterations of a cellular automaton designed to produce smoother clump edges.

Each entry in the island catalog will include the following information.

1. Island identifier.
2. Position of the peak pixel within the island (right ascension and declination).
3. Area of the island.
4. Flux at the peak.
5. Total flux contained within the island.
6. Average noise across the island.
7. Cross references if the island continues into an adjoining tile.
8. Polygon approximation to the island outline in Space-Time Coordinate[37, 38] string (STC-S) format.
9. *SCUBA-2 only:* Cross references with islands detected at the other wavelength.

3.2 Peak Catalogs

For each island, a representative (median) noise value is found using the variance associated with the pixels within the island. This representative noise value is used to control the identification of local maxima as peaks, again using the `FellWalker` algorithm as implemented in CUPID. To be identified as a new peak, the peak flux is required to be greater than five times the representative noise for the island, and the minimum required valley between the peak and any neighboring higher peak is also five times the noise. Since we only search for peaks within islands of detected emission, our catalogs will not be contaminated by fake sources, even if the local noise level varies across the island.

We would like to provide some characterization of the local structure of the emission at each peak. Although we do not believe we can provide an accurate automated description of the underlying physical source, due to the diversity of astronomical targets observed by the JCMT, some estimate of the size and shape of the local emission could prove very useful to users of the archive. We are currently investigating different mechanisms for characterizing this structure. Although the first version of the catalog will not directly make use of the velocity information in heterodyne data cubes to detect peaks, we will give the velocity of the brightest spectral channel at the position of each identified peak.

The peak catalog will contain the following information for each entry.

1. Peak identifier.
2. Position of the peak (right ascension and declination).
3. Flux at the peak position.
4. Noise at the peak position.
5. Identifier of the island in which the peak is found.
6. *Heterodyne only:* Radial velocity of the brightest emission at the peak position.

Bell, G. S. et. al., “Advanced data products for the JCMT Science Archive,” Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

3.3 User Interaction with Catalogs

When users search for data in the JCMT Science Archive through the interface provided by the CADC, they will be able to find the catalogs alongside the individual tiles from which they were created. The catalogs can be downloaded together with or separately from the tiled maps and cubes. As well as the catalog files themselves, we will provide a more detailed description of the shape of the islands — we intend to make this information available both as masks in FITS image format and as a MOC representation.

We also hope to enable users to interact with the catalogs as a whole, without having to initially consider which HEALPix tiles they are interested in. Once the catalog files have been created, we, with the CADC, are planning to provide a catalog search service. This will allow people to query the catalog by position and identify which positions fall within an identified island of emission or with a specified radius of an identified peak of emission. The catalog service should support the Table Access Protocol[39] (TAP) so that the catalog can be smoothly accessed from within applications such as the Starlink GAIA[40,41] image viewer. Once GAIA gains the ability to read STC-S shapes from TAP query results (it can already display them[38,42]), this would allow astronomers who are, for example, looking at data sets from other wavelengths to easily overlay the JCMT-detected islands and peaks on their images, without needing to explicitly seek out and download the catalog files.

4. QUALITY ASSURANCE

Data published in the archive undergo a series of quality assurance controls throughout their life cycle. This begins at the telescope, during or shortly after the observation itself, when the telescope operator and visiting observer can assess the data quality. This process is aided by optimized versions of the pipeline recipes, which process the data in real-time.

Observations are typically marked "junk" if they are corrupted such that it is not possible or sensible to attempt to process them. The raw data files are still stored in the archive, in case it subsequently becomes possible to correct the problem, but are not made visible to users. Observations marked "bad" are made available in the archive, both in raw and processed form. However, they are excluded from nightly co-adds and the advanced data products.

4.1 Automatic Controls

The software includes automatic controls of data quality. For SCUBA-2 this is built into the data reduction software.[25] Each observation starts with a flat-field sequence to measure the responsivity of the bolometers to a triangle-wave modulation of the internal heater power. Only bolometers that respond appropriately to this signal are included in the data reduction. Further filtering is applied by the map-maker, such as removing data taken with the telescope scanning too slowly.

For heterodyne data the ORAC-DR pipeline uses a number of quality assurance parameters to control which data to include. Some parameters apply to all of the data from a receptor, for example when the system temperature is too high in absolute terms or in comparison to the other receptors. Others apply to individual spectra, such as a check of the consistency between the noise in the spectra and that expected from the system temperature. Final tests such as a check of the percentage of bad pixels are applied to complete data cubes.

4.2 Human Oversight

The JCMT Science Archive operates in a very dynamic manner[2] where data is constantly flowing into the archive, processed or re-processed as required, and released immediately (to project members only during each observation's proprietary period, and publicly thereafter). The data reduction system is also updated frequently to incorporate the latest improvements to the software and associated configuration parameters. The archive therefore provides the ultimate test for the robustness of the data reduction system as it is thoroughly tested by the large volume of data passing through the archive each day.

Changes in data-reduction techniques, instrument settings, observing modes and conditions create a real need for a visual inspection of the data products. The data processing infrastructure makes available to JCMT and CADC staff an active report page for each processing job. These pages provide information including the status of the job, preview images and log files. We find that jobs which have completed successfully, without software errors and having passed the automatic quality assurance controls, may sometimes not have produced optimal scientific products. A human element can aid in catching rare or consistent problems which may otherwise go undetected.

An example where human monitoring has provided benefits to projects is the case where a better choice of reduction recipe can be suggested to the principle investigator. Ideally a poor choice of recipe would be picked up prior to observing, but it can also be identified by inspecting the nightly reduced data products. Sometimes issues caused by hardware faults — such as microphonic noise on the SCUBA-2 bolometer arrays — can be identified in the preview images and compensated for in a re-reduction.

Human oversight of the archive also provides a final check of the quality flagging of the data products. If not flagged at the telescope when taken, poor data can be retrospectively marked as "bad" allowing exclusion from the nightly co-adds and the legacy advanced data products. In severe cases data can be retrospectively marked "junk" to remove them altogether from the visible part of the archive.

5. CONCLUSIONS

The JCMT Science Archive has been successfully serving raw and reduced data to users of the telescope, and the wider community once the data become public, for a number of years. Publicly available archives are an increasingly important resource for astronomy, and one which we believe is vital in maximizing the scientific rewards to be gained from telescope operations.

To this end, we plan to augment the archive with advanced data products comprising "legacy" maps and spectral cubes of all publicly available data gridded onto a standardized set of tiles, which are defined by the HEALPix scheme and also use the `HPX` projection internally. These tiles will be used to create extended source catalogs which will indicate where significant emission was or was not detected, and catalogs of the peaks of emission.

At the time of writing, the tile-based reduction and co-adding software is complete for the SCUBA-2 bolometer camera and being developed for the HARP heterodyne array. We are testing the generic SCUBA-2 map-making configuration on a wide variety of observed regions and have run a set of test reductions at the CADC to confirm that the infrastructure is in place to manage the tiled data processing. The catalog processing software has been prototyped and is now being integrated into our data processing pipeline.

ACKNOWLEDGMENTS

We thank Linda Gregoire for providing the publication statistics plotted in Figure 1. The example SCUBA-2 data presented in Figures 3–5 were gathered for project M12AJ01 and are used here with the kind permission of Antonio Chrysostomou.

The James Clerk Maxwell Telescope is operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the United Kingdom, the National Research Council of Canada, and (until 31 March 2013) the Netherlands Organisation for Scientific Research.

This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency.

REFERENCES

[1] Gaudet, S., Dowler, P., Goliath, S., and Redman, R., "The JCMT Science Archive and the Virtual Observatory," in [*Astronomical Data Analysis Software and Systems XVII*], Argyle, R. W., Bunclark, P. S., and Lewis, J. R., eds., *ASP Conf. Ser.* **394**, 135 (2008).

[2] Economou, F. et al., "The JSA and the Grid: How "Infinite" Computing Power Enables a New Archive Model for PI-led Observatories," in [*Astronomical Data Analysis Software and Systems XX*], Evans, I. N., Accomazzi, A., Mink, D. J., and Rots, A. H., eds., *ASP Conf. Ser.* **442**, 203 (2011).

[3] Economou, F., Gaudet, S., Jenness, T., Redman, R. O., Goliath, S., Dowler, P., Schade, D., and Chrysostomou, A., "Lessons learned from an observatory/data center partnership: The JCMT Science Archive," *Astronomy & Computing* **in preparation** (2014).

[4] Jenness, T., Cavanagh, B., Economou, F., and Berry, D. S., "JCMT Science Archive: Advanced Heterodyne Data Products Pipeline," in [*Astronomical Data Analysis Software and Systems XVII*], Argyle, R. W., Bunclark, P. S., and Lewis, J. R., eds., *ASP Conf. Ser.* **394**, 565 (2008).

[5] Economou, F., Jenness, T., Chrysostomou, A., Cavanagh, B., Redman, R., and Berry, D. S., "The JCMT Legacy Survey: The challenges of the JCMT Science Archive," in [*Astronomical Data Analysis Software and Systems XVII*], Argyle, R. W., Bunclark, P. S., and Lewis, J. R., eds., *ASP Conf. Ser.* **394**, 450 (2008).

[6] Louys, M., Bonnarel, F., Schade, D., Dowler, P., Micol, A., Durand, D., Tody, D., Michel, L., Salgado, J., Chilingarian, I., Rino, B., Santander-Vela, J. D., and Skoda, P., "IVOA Recommendation: Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.0," *ArXiv e-prints* (2011). arXiv:1111.1758.

[7] Economou, F., Bridger, A., Wright, G. S., Jenness, T., Currie, M. J., and Adamson, A., "ORAC-DR: Pipelining With Other People's Code," in [*Astronomical Data Analysis Software and Systems VIII*], Mehringer, D. M., Plante, R. L., and Roberts, D. A., eds., *ASP Conf. Ser.* **172**, 11 (1999).

[8] Cavanagh, B., Jenness, T., Economou, F., and Currie, M. J., "The ORAC-DR data reduction pipeline," *Astron. Nachr.* **329**, 295–297 (2008).

[9] Jenness, T., Economou, F., Cavanagh, B., Currie, M. J., and Gibb, A., "ORAC-DR: Astronomy data reduction pipeline," (2013). Astrophysics Source Code Library ascl:1310.001.

[10] Moore, T. J. T., Shipman, R. F., Plume, R., Hoare, M. G., and JPS International Collaboration, "Legacy Surveys with the JCMT: The JCMT Plane Survey," in [*Protostars and Planets V*], 8370 (2005).

[11] Plume, R., Fuller, G. A., Helmich, F., van der Tak, F. F. S., Roberts, H., Bowey, J., Buckle, J., Butner, H., Caux, E., Ceccarelli, C., van Dishoeck, E. F., Friberg, P., Gibb, A. G., Hatchell, J., Hogerheijde, M. R., Matthews, H., Millar, T. J., Mitchell, G., Moore, T. J. T., Ossenkopf, V., Rawlings, J. M. C., Richer, J., Roellig, M., Schilke, P., Spaans, M., Tielens, A. G. G. M., Thompson, M. A., Viti, S., Weferling, B., White, G. J., Wouterloot, J., Yates, J., and Zhu, M., "The James Clerk Maxwell Telescope Spectral Legacy Survey," *Publ. Astron. Soc. Pac.* **119**, 102–111 (2007).

[12] Matthews, B. C., Greaves, J. S., Holland, W. S., Wyatt, M. C., Barlow, M. J., Bastien, P., Beichman, C. A., Biggs, A., Butner, H. M., Dent, W. R. F., Di Francesco, J., Dominik, C., Fissel, L., Friberg, P., Gibb, A. G., Halpern, M., Ivison, R. J., Jayawardhana, R., Jenness, T., Johnstone, D., Kavelaars, J. J., Marshall, J. L., Phillips, N., Schieven, G., Snellen, I. A. G., Walker, H. J., Ward-Thompson, D., Weferling, B., White, G. J., Yates, J., Zhu, M., and Craigon, A., "An Unbiased Survey of 500 Nearby Stars for Debris Disks: A JCMT Legacy Program," *Publ. Astron. Soc. Pac.* **119**, 842–854 (2007).

[13] Ward-Thompson, D., Di Francesco, J., Hatchell, J., Hogerheijde, M. R., Nutter, D., Bastien, P., Basu, S., Bonnell, I., Bowey, J., Brunt, C., Buckle, J., Butner, H., Cavanagh, B., Chrysostomou, A., Curtis, E., Davis, C. J., Dent, W. R. F., van Dishoeck, E., Edmunds, M. G., Fich, M., Fiege, J., Fissel, L., Friberg, P., Friesen, R., Frieswijk, W., Fuller, G. A., Gosling, A., Graves, S., Greaves, J. S., Helmich, F., Hills, R. E., Holland, W. S., Houde, M., Jayawardhana, R., Johnstone, D., Joncas, G., Kirk, H., Kirk, J. M., Knee, L. B. G., Matthews, B., Matthews, H., Matzner, C., Moriarty-Schieven, G. H., Naylor, D., Padman, R., Plume, R., Rawlings, J. M. C., Redman, R. O., Reid, M., Richer, J. S., Shipman, R., Simpson, R. J., Spaans, M., Stamatellos, D., Tsamis, Y. G., Viti, S., Weferling, B., White, G. J., Whitworth, A. P., Wouterloot, J., Yates, J., and Zhu, M., "The James Clerk Maxwell Telescope Legacy Survey of Nearby Star-forming Regions in the Gould Belt," *Publ. Astron. Soc. Pac.* **119**, 855–870 (2007).

[14] Thompson, M. A., Serjeant, S., Jenness, T., Scott, D., Ashdown, M., Brunt, C., Butner, H., Chapin, E., Chrysostomou, A. C., Clark, J. S., Clements, D., Collett, J. L., Coppin, K., Coulson, I. M., Dent, W. R. F., Economou, F., Evans, A., Friberg, P., Fuller, G. A., Gibb, A. G., Greaves, J., Hatchell, J., Holland, W. S., Hudson, M., Ivison, R. J., Jaffe, A., Joncas, G., Jones, H. R. A., Knapen, J. H., Leech, J., Mann, R., Matthews, H. E., Moore, T. J. T., Mortier, A., Negrello, M., Nutter, D., Pestalozzi, M. P., Pope, A., Richer, J., Shipman, R., Urquhart, J. S., Vaccari, M., Van Waerbeke, L., Viti, S., Weferling, B., White, G. J., Wouterloot, J., and Zhu, M., "The SCUBA-2 "All-Sky" Survey," *ArXiv e-prints* (2007). arXiv:0704.3202.

[15] Wilson, C. D., Warren, B. E., Israel, F. P., Serjeant, S., Bendo, G., Brinks, E., Clements, D., Courteau, S., Irwin, J., Knapen, J. H., Leech, J., Matthews, H. E., Mühle, S., Mortier, A. M. J., Petitpas, G., Sinukoff, E., Spekkens, K., Tan, B. K., Tilanus, R. P. J., Usero, A., van der Werf, P., Wiegert, T., and Zhu, M., "The James Clerk Maxwell Telescope Nearby Galaxies Legacy Survey. I. Star-Forming Molecular Gas in Virgo Cluster Spiral Galaxies," *Astrophysical Journal* **693**, 1736–1748 (2009).

[16] Geach, J. E., Chapin, E. L., Coppin, K. E. K., Dunlop, J. S., Halpern, M., Smail, I., Werf, P. v. d., Serjeant, S., Farrah, D., Roseboom, I., Targett, T., Arumugam, V., Asboth, V., Blain, A., Chrysostomou, A., Clarke, C., Ivison, R. J., Jones, S. L., Karim, A., Mackenzie, T., Meijerink, R., Michałowski, M. J., Scott, D., Simpson, J. M., Swinbank, A. M., Alexander, D. M., Almaini, O., Aretxaga, I., Best, P., Chapman, S., Clements, D. L., Conselice, C., Danielson, A. L. R., Eales, S., Edge, A. C., Gibb, A. G., Hughes, D., Jenness, T., Knudsen, K. K., Lacey, C. G., Marsden, G.,

McMahon, R., Oliver, S. J., Page, M. J., Peacock, J. A., Rigopoulou, D., Robson, E. I., Spaans, M., Stevens, J., Webb, T. M. A., Willott, C., Wilson, C. D., and Zemcov, M., "The SCUBA-2 Cosmology Legacy Survey: blank-field number counts of 450-μm-selected galaxies and their contribution to the cosmic infrared background," *Monthly Notices of the RAS* **432**, 53–61 (2013).

[17] Dowler, P., "CAOM-2.0: The Inevitable Evolution of a Data Model," in [*Astronomical Data Analysis Software and Systems XXI*], Ballester, P., Egret, D., and Lorente, N. P. F., eds., *ASP Conf. Ser.* **461**, 339 (2012).

[18] Redman, R. O. and Dowler, P., "Implementing a Common Database Architecture at the CADC using CAOM-2," in [*Astronomical Data Analysis Software and Systems XXII*], Friedel, D., ed., *ASP Conf. Ser.* **475**, 159–161 (2013).

[19] Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M., "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere," *Astrophysical Journal* **622**, 759–771 (2005).

[20] Calabretta, M. R. and Roukema, B. F., "Mapping on the HEALPix grid," *Monthly Notices of the RAS* **381**, 865–872 (2007).

[21] Fernique, P., Durand, D., Boch, T., Oberto, A., and Pineau, F., "HEALPix Based Cross-Correlation in Astronomy," in [*Astronomical Data Analysis Software and Systems XXII*], Friedel, D. N., ed., *ASP Conf. Ser.* **475**, 135 (2013).

[22] Holland, W. S., Bintley, D., Chapin, E. L., Chrysostomou, A., Davis, G. R., Dempsey, J. T., Duncan, W. D., Fich, M., Friberg, P., Halpern, M., Irwin, K. D., Jenness, T., Kelly, B. D., MacIntosh, M. J., Robson, E. I., Scott, D., Ade, P. A. R., Atad-Ettedgui, E., Berry, D. S., Craig, S. C., Gao, X., Gibb, A. G., Hilton, G. C., Hollister, M. I., Kycia, J. B., Lunney, D. W., McGregor, H., Montgomery, D., Parkes, W., Tilanus, R. P. J., Ullom, J. N., Walther, C. A., Walton, A. J., Woodcraft, A. L., Amiri, M., Atkinson, D., Burger, B., Chuter, T., Coulson, I. M., Doriese, W. B., Dunare, C., Economou, F., Niemack, M. D., Parsons, H. A. L., Reintsema, C. D., Sibthorpe, B., Smail, I., Sudiwala, R., and Thomas, H. S., "SCUBA-2: the 10 000 pixel bolometer camera on the James Clerk Maxwell Telescope," *Monthly Notices of the RAS* **430**, 2513–2533 (2013).

[23] Bintley, D., Holland, W. S., MacIntosh, M. J., Friberg, P., Bell, G. S., Berke, D. A., Berry, D. S., Berthold, R. M., Coulson, I. M., Currie, M. J., Dempsey, J. T., Gibb, A. G., Gorges, B. H., Graves, S. F., Jenness, T., Johnstone, D. I., Parsons, H. A. L., Thomas, H. S., Walther, C. A., and Wouterloot, J. G., "SCUBA-2: an update on the performance of the 10,000 pixel bolometer camera after 2 years of science operation at the JCMT," in [*Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*], Holland, W. S. and Zmuidzinas, J., eds., *Proc. SPIE* **9153**, 91532 (2014).

[24] Kackley, R., Scott, D., Chapin, E., and Friberg, P., "JCMT Telescope Control System upgrades for SCUBA-2," in [*Software and Cyberinfrastructure for Astronomy*], Radziwill, N. M. and Bridger, A., eds., *Proc. SPIE* **7740**, 77401Z (2010).

[25] Chapin, E. L., Berry, D. S., Gibb, A. G., Jenness, T., Scott, D., Tilanus, R. P. J., Economou, F., and Holland, W. S., "SCUBA-2: iterative map-making with the Sub-Millimetre User Reduction Facility," *Monthly Notices of the RAS* **430**, 2545–2573 (2013).

[26] Jenness, T., Chapin, E. L., Berry, D. S., Gibb, A. G., Tilanus, R. P. J., Balfour, J., Tilanus, V., and Currie, M. J., "SMURF: SubMillimeter User Reduction Facility," (2013). Astrophysics Source Code Library ascl:1310.007.

[27] Walsh, A. J., Macdonald, G. H., Alvey, N. D. S., Burton, M. G., and Lee, J.-K., "Observations of warm dust near methanol masers," *Astronomy and Astrophysics* **410**, 597–610 (2003).

[28] Forster, J. R. and Caswell, J. L., "Radio Continuum Emission at OH and H_2O Maser Sites," *Astrophysical Journal* **530**, 371–386 (2000).

[29] Shepherd, M., "Correct Sampling of Diffraction Limited Images," Tech. Rep. CCAT-TM-109, California Institute of Technology (2012). `http://wiki.astro.cornell.edu/twiki/pub/CCAT/CCAT_Memos/DiffractionLimitedSampling111212.pdf`.

[30] Dempsey, J. T., Friberg, P., Jenness, T., Tilanus, R. P. J., Thomas, H. S., Holland, W. S., Bintley, D., Berry, D. S., Chapin, E. L., Chrysostomou, A., Davis, G. R., Gibb, A. G., Parsons, H., and Robson, E. I., "SCUBA-2: on-sky calibration using submillimetre standard sources," *Monthly Notices of the RAS* **430**, 2534–2544 (2013).

[31] Buckle, J. V., Hills, R. E., Smith, H., Dent, W. R. F., Bell, G., Curtis, E. I., Dace, R., Gibson, H., Graves, S. F., Leech, J., Richer, J. S., Williamson, R., Withington, S., Yassin, G., Bennett, R., Hastings, P., Laidlaw, I., Lightfoot, J. F., Burgess, T., Dewdney, P. E., Hovey, G., Willis, A. G., Redman, R., Wooff, B., Berry, D. S., Cavanagh, B., Davis, G. R., Dempsey, J., Friberg, P., Jenness, T., Kackley, R., Rees, N. P., Tilanus, R., Walther, C., Zwart, W.,

Bell, G. S. et. al., "Advanced data products for the JCMT Science Archive," Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983

Klapwijk, T. M., Kroug, M., and Zijlstra, T., “HARP/ACSIS: a submillimetre spectral imaging system on the James Clerk Maxwell Telescope,” *Monthly Notices of the RAS* **399**, 1026–1043 (2009).

[32] Berry, D. S., “FellWalker – a Clump Identification Algorithm,” *Astronomy & Computing* **in preparation** (2015).

[33] Berry, D. S., Reinhold, K., Jenness, T., and Economou, F., “CUPID: A Clump Identification and Analysis Package,” in [*Astronomical Data Analysis Software and Systems XVI*], Shaw, R. A., Hill, F., and Bell, D. J., eds., *ASP Conf. Ser.* **376**, 425 (2007).

[34] Berry, D. S., Reinhold, K., Jenness, T., and Economou, F., “CUPID: Clump Identification and Analysis Package,” (2013). Astrophysics Source Code Library ascl:1311.007.

[35] Williams, J. P., de Geus, E. J., and Blitz, L., “Determining structure in molecular clouds,” *Astrophysical Journal* **428**, 693–712 (1994).

[36] Watson, M., *Assessing the Performance of Sub-Millimetre Compact Object Detection Algorithms*, Master’s thesis, University of Hertfordshire (2010).

[37] Rots, A. H., “IVOA Recommendation: Space-Time Coordinate Metadata for the Virtual Observatory Version 1.33,” *ArXiv e-prints* (2011). arXiv:1110.0504.

[38] Berry, D. and Draper, P., “Using the AST Library to Create and Use STC-S Region Descriptions,” in [*Astronomical Data Analysis Software and Systems XIX*], Mizumoto, Y., Morita, K.-I., and Ohishi, M., eds., *ASP Conf. Ser.* **434**, 213 (2010).

[39] Dowler, P., Rixon, G., and Tody, D., “IVOA Recommendation: Table Access Protocol Version 1.0,” *ArXiv e-prints* (2011). arXiv:1110.0497.

[40] Draper, P. W., Berry, D. S., Jenness, T., and Economou, F., “GAIA – Version 4,” in [*Astronomical Data Analysis Software and Systems XVIII*], Bohlender, D. A., Durand, D., and Dowler, P., eds., *ASP Conf. Ser.* **411**, 575 (2009).

[41] Draper, P. W., Gray, N., Berry, D. S., and Taylor, M., “GAIA: Graphical Astronomy and Image Analysis Tool,” (2014). Astrophysics Source Code Library ascl:1403.024.

[42] Currie, M. J., Berry, D. S., Jenness, T., Gibb, A. G., Bell, G. S., and Draper, P. W., “Starlink Software in 2013,” in [*Astronomical Data Analysis Software and Systems XXIII*], Manset, N. and Forshay, P., eds., *ASP Conf. Ser.* **485**, 391–394 (2014).

Bell, G. S. et. al., “Advanced data products for the JCMT Science Archive,” Software and Cyberinfrastructure for Astronomy III, Editors Chiozzi, G. and Radziwill, N. M., Proc. SPIE 9152, 91522J (2014). http://dx.doi.org/10.1117/12.2054983