Astronomy and Computing 11 (2015) 161–168

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

Observatory/data centre partnerships and the VO-centric archive: The JCMT Science Archive experience

Frossie Economou[a,*], S. Gaudet[b], T. Jenness[c,d], R.O. Redman[d], S. Goliath[b], P. Dowler[b], M.J. Currie[d], G.S. Bell[d], S.F. Graves[d], J. Ouellette[b], D. Johnstone[d,e,f], D. Schade[b], A. Chrysostomou[g,d]

[a] LSST Project Office, 933 N. Cherry Ave, Tucson, AZ 85721, USA
[b] Canadian Astronomy Data Centre, National Research Council of Canada, 5071 West Saanich Road., Victoria, BC V9E 2E7, Canada
[c] Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[d] Joint Astronomy Centre, 660 N. A'ohōkū Place, Hilo, HI 96720, USA
[e] NRC-Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada
[f] Department of Physics and Astronomy, University of Victoria, PO Box 3055 STN CSC, Victoria, BC V8W 3P6, Canada
[g] Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK

ARTICLE INFO

Article history:
Received 10 June 2014
Received in revised form
4 December 2014
Accepted 14 December 2014
Available online 23 December 2014

Keywords:
Facilities: JCMT
Virtual Observatory tools
Astronomical databases: misc
ISM: individual objects (G34.3)

ABSTRACT

We present, as a case study, a description of the partnership between an observatory (JCMT) and a data centre (CADC) that led to the development of the JCMT Science Archive (JSA). The JSA is a successful example of a service designed to use Virtual Observatory (VO) technologies from the start. We describe the motivation, process and lessons learned from this approach.

1. Origins

The James Clerk Maxwell Telescope (JCMT) has collaborated with the Canadian Astronomy Data Centre (CADC) to create the JCMT Science Archive (JSA) which provides raw and reduced JCMT data to the astronomical community (Gaudet et al., 2008b; Economou et al., 2008; Gaudet et al., 2008a; Economou et al., 2011). As a new generation of instruments was being developed for the JCMT in the early 2000s (HARP/ACSIS & SCUBA-2; Dent et al., 2000; Holland et al., 2003), it became clear that the data rates from these instruments, of order 10 MB/s, were going to be significantly higher than earlier submillimetre instrumentation. In particular SCUBA-2 was the first generation of submillimetre camera that could be considered to be suitable for use as a large-scale survey instrument. Exploratory discussions on the JSA between JCMT and CADC began in 2003 and culminated in a decision to approve the collaboration in May 2005 (Davis, 2005). Development effort was obtained in-house and also from the addition of two programmers recruited from the UK Starlink project (Disney and Wallace, 1982), which had recently been closed.

The commitment to a JCMT Science Archive was followed shortly afterwards by the approval of the JCMT Legacy Survey programme in July 2005 (Davis, 2005). To ensure survey participation in the JSA the JCMT Data Users' Group (JDUG) was created in early 2006 to provide stakeholder input into the pipeline operation and advanced data products (Redman, 2006).

2. Motivation: observatory

Submillimetre data has traditionally been rather esoteric, closer to radio than the optical/infrared regime familiar to most astronomers. Raw data is typically in time series format (Fig. 1), and requires in-house algorithms for transformation to science-ready formats such as spectra or images. Calibration is difficult due to the dominant and highly variable effect of the water vapour in Earth's atmosphere (e.g., Archibald et al., 2002; Dempsey et al., 2013a).

* Corresponding author.
E-mail address: frossie@lsst.org (F. Economou).

http://dx.doi.org/10.1016/j.ascom.2014.12.005
2213-1337/

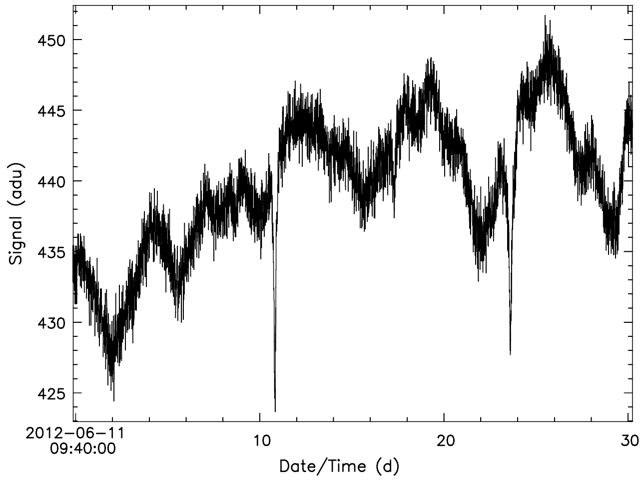

Fig. 1. Single bolometer time-series from a subset of a SCUBA-2 observation of G34.3 from 2012 June 11th. The final image is shown in Fig. 2. The negative spikes are the detections of the bright central source.

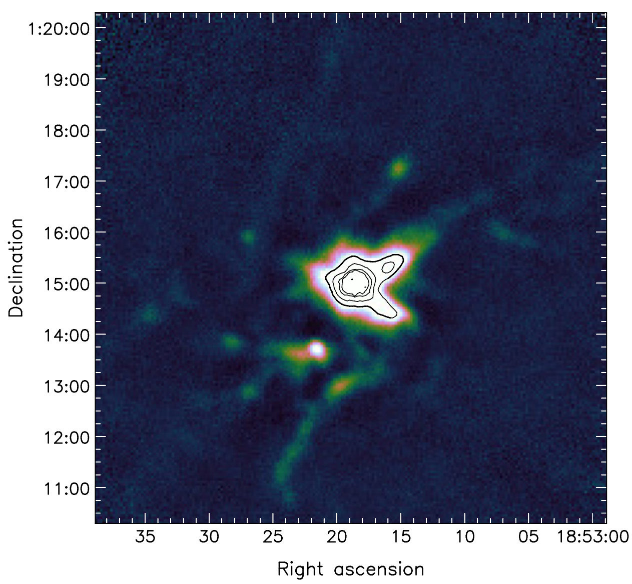

Fig. 2. SCUBA-2 observation of G34.3.

JCMT invested significant effort in automated data reduction based on the ORAC-DR pipeline framework (Economou et al., 1999; Jenness and Economou, 1999; Gibb et al., 2005; Jenness et al., 2008; Jenness and Economou, 2015, ascl:1310.001). In many cases these automatically generated products were publication quality, and thanks to a constantly updated calibration model, better than what an inexperienced astronomer could be expected to achieve on their own. Moreover with the advent of large bolometer arrays such as SCUBA-2 (Holland et al., 2013), this data could be processed in maps that resulted in image data that could be readily understood by non-submm specialists, an example of which can be seen in Fig. 2.

The JCMT had in-house experience with setting up a data archive in the "filing cabinet" sense of allowing users to search and retrieve raw data, but apart from a prototype involving the on-demand generation of SCUBA data products (Jenness et al., 2002), had not tackled the integration of data processing with data product distribution in a full science archive environment. Indeed, distribution of publication-quality data became an issue of the highest priority with the advent of the JCMT Legacy Survey Programme (Chrysostomou, 2010; Economou et al., 2008) using the SCUBA-2 and HARP/ACSIS (Buckle et al., 2009) instruments. Aside from the normal desire to provide a uniformly reduced product to the survey teams, the processing demands for this data required a non-trivial IT infrastructure. The complex iterative map-maker algorithm used to reduce SCUBA-2 data (SMURF; Chapin et al., 2013, ascl:1310.007) was expected to generate higher fidelity maps when more of an observation could be fitted into memory at one time. It was estimated that at minimum a machine with 64 GB of RAM would be required (and 128 GB is the current recommendation) but circa 2008 machines of this size were not readily available to the typical JCMT observer. So there were intrinsic reasons to have a JCMT Science Archive allowing the survey consortia to download the processed products. Ultimately, usage of such a standalone archive would be dominated by JCMT users retrieving their own data, or after the proprietary period elapsed, other JCMT users working in the same scientific areas who were explicitly searching for JCMT data.

JCMT formed a strong interest in going further, and exposing its high-value data product to data-mining astronomers who would not have a priori knowledge either of JCMT in particular or sub-mm astronomy in general. To that end, the Virtual Observatory (VO) data discovery and publication protocols seemed like a natural choice for reaching the large parts of the astronomical community that were oblivious to its existence. VO publication would also have the advantage of exposing the JCMT data sets to workhorse tools that VO-savvy astronomers already used, such as TOPCAT (Taylor, 2005, ascl:1101.010) and Aladin (Ochsenbein et al., 2005, ascl:1112.019).

However, despite being convinced of the desirability of leveraging the VO tools and services for JCMT data, the observatory had the usual constraints of time and effort. The small Scientific Computing Group was busy with supporting the entire non-hardware-controlling software suite at both JCMT and UKIRT (see e.g., Economou et al., 2002; Jenness and Economou, 2011, with both telescopes operated by the same organization), as well as developing data reduction for new instruments, helping with their commissioning, and supporting the JCMT Legacy Surveys. The ability to develop a VO-aware data centre and support the demands of the hoped-for increased usage base was just not there.

What JCMT had, however, was a pre-existing collaboration with CADC, which hosted the older JCMT data archive (Tilanus et al., 1997) for the benefit of the Canadian astronomical community, Canada being one of the three international partners funding the JCMT (the other two being the United Kingdom and the Netherlands). CADC had early involvement in VO protocols (Schade et al., 2002; Dowler et al., in preparation), was a productive developer and enthusiastic supporter of VO standards, and was known to "eat its own dog food"[1] by using many of these interfaces and services internally.

3. Motivation: data centre

CADC already had a varied collection of data from several telescopes and space missions (Crabtree et al., 1994; Gaudet et al., 2008b). Keen to be able to extend its holdings to new observatories and data sets while requiring only a small and well-defined effort, CADC developed the Common Archive Observation Model (CAOM: Dowler et al., 2007, 2008). CAOM defines an extensive and versatile data model that classifies every data file using a common

[1] See http://en.wikipedia.org/wiki/Eating_your_own_dog_food and Economou et al. (2014) for more information.

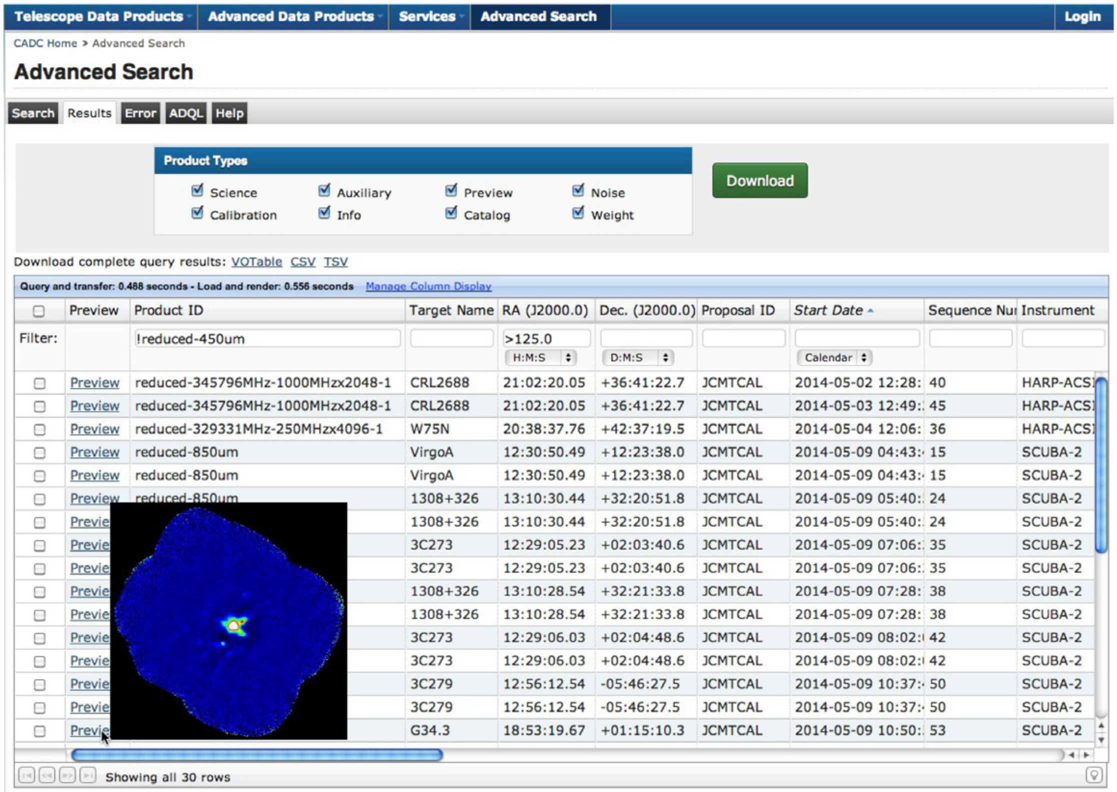

Fig. 3. AdvancedSearch results.

set of physical, observational, organizational, and processing metadata. This allows a generic VO search tool, such as AdvancedSearch, to search the entire set of CADC archives for data relevant to a chosen target in the sky.

One of the main attractions of the JCMT data set was its significant departure from many of the common forms of other astronomical data, that predominantly came from optical and IR instrumentation. Examples include:

- The "photon energy" axis for optical observations is normally described in wavelength units like Ångströms or microns, whereas most radio observations are defined in frequency units like MHz and GHz. To ingest and search for JCMT observations it was necessary to enhance the tools to handle both wavelength and frequency units, with the consequence that CADC interfaces now handle transparently most standard conversions amongst frequency, energy and wavelength units.
- At the start of the collaboration, most optical data consisted of two dimensional RA/Dec images and sets of spectra. Even at that time, JCMT data came in RA/Dec, Galactic and offset coordinates, with up to 4 dimensions (2 spatial, wavelength and polarization). The JCMT standard pipeline generates a diverse set of products, including spectra, data cubes, maps, previews showing both spectral and spatial images, and catalogues for point sources, emission peaks and clumps (extended regions of non-uniform emission).
- Since most detector technologies only allow a photon to be detected once, it can be safely assumed for optical instruments with multiple detectors that the data products from different detectors will not overlap in WCS space. The ability at radio wavelengths to amplify the detected signal and feed it into multiple spectrometers allows the output of the JCMT multi-subsystem spectrometer ACSIS to include spectra and data cubes that overlap in a variety of ways, sometimes with different frequency resolution, sometimes overlapping just at the ends of the spectra to allow a much wider frequency coverage for a given frequency resolution than could be managed by any single subsystem.

The JCMT therefore provided an excellent stretch to the model, and continues to do so; if JCMT data could be described in CAOM, CADC would be in the unprecedented position of being able to accept almost any data set from future observatories with minimal changes to their system.

Another advantage in working with JCMT on its data sets, was the high level of completion and accuracy that JCMT provided in its metadata. Even modern instruments on some older telescopes follow metadata conventions established by the observatory long before the FITS World Coordinate System (WCS) conventions were agreed upon. At the start of the collaboration, the CADC would assign an "archive scientist" to each archive, whose job description included learning all the idiosyncrasies of the observatory. A major part of that effort involved working around poor or incomplete metadata that made astronomical data archiving problematic, especially if the observatory tended to change their data products and headers without warning. Maintaining a proper "Science Archive" requires that both power users and astronomers unfamiliar with an observatory's internal conventions must be able to find and download science-ready data products without mastering an arcane interface or guessing how to interpret the metadata that it presents. JCMT's dedication to high-fidelity metadata and quick response in the rare case of problems made this an attractive test data set.

The success of this approach can be seen from the screen shot in Fig. 3, which shows the reduced (Calibration Level 2) data from May 2014, sorted by observation date, filtered to remove reduced-450 μm data (since the atmosphere at 450 μm is often very opaque) and to include observations with RA $>$ 125.0°. A pop-up preview of G34.3 is shown; clicking would bring up a larger version of the preview in a new tab. The `productID` column shows the kind of data that can be downloaded for each selection, giving the product type (reduced data files in this example) and basic wavelength information (filter for continuum observations, rest frequency and spectrometer configuration for heterodyne observations).

4. VO standards used in the JSA

CAOM: Common Archive Observation Model—This is the data model used in all archives at the CADC. It was designed to be a superset of VO data models so that VO data models and services could be easily implemented on top of CAOM. While CAOM is not a VO data model per se, it was designed and is used as the metadata interface between archives and standard VO data models (Dowler et al., 2007; Redman and Dowler, 2013).

ObsCore: Observation Data Model Core Components—This VO data model is designed to support data discovery specifically by supporting the exact same queries to TAP services run by all data centres. In the JSA, this is simply a view of CAOM as it contains a subset of CAOM metadata (Louys et al., 2011).

SIA: Simple Image Access—Version 1.0 is an early VO service interface that supports positional searching and retrieval of 2D images (Tody et al., 2009). Version 2.0 (Dowler et al., 2014b) is a new VO service interface that supports data discovery of multi-dimensional data sets (images and data cubes) using the ObsCore data model. Both of these are implemented using CAOM and TAP (below).

TAP: Table Access Protocol—This VO service interface supports ad-hoc querying of the CAOM metadata and standard views like ObsCore. All JSA science data is discoverable through this interface (Dowler et al., 2011; Nandrekar-Heinis et al., 2014).

ADQL: Astronomical Data Query Language—Queries to the TAP service are formatted in ADQL, which is designed to closely resemble the popular SQL syntax used by many relational database systems (Ortiz et al., 2011).

DataLink: DataLink Service—This VO service interface allows users and client software to drill-down from discovered data sets to the list of files to download and to services that can operate on the data. The SIA-2.0 and TAP services use this interface to provide access to JSA data files and services (Dowler et al., 2014a).

AccessData: Access Data Prototype—This prototype VO service interface allows users to perform cutouts on data files in a standard set of world coordinates.

CDP: Credential Delegation Protocol—This VO service interface enables CADC services to call other services on behalf of the user so that the correct identity and access rights are enforced. In the JSA, this allows the user interface (AdvancedSearch) to pass the authenticated user identity to the TAP service so that query results will include metadata and access information for proprietary observations the user can access (Graham et al., 2010).

VOTable: Virtual Observatory Table Format—This is a common tabular format used to exchange metadata between clients and services. It is the standard output format in SIA, TAP, and DataLink (Ochsenbein et al., 2013).

5. Evolution of the data flow

The system that moves data from the JCMT to the CADC and on to our users has been under continuous development since the start of the collaboration. Fig. 4 shows the current development goal, which should have been attained by the time this paper is published. Data files sent to be stored in the "Archive Directory" (AD) system at the CADC enter through the "Data Web Service" interface. File metadata in the databases comprising the "JSA CAOM Metadata" system are managed using the "CAOM Repository" interface and can be read through the "TAP" service. Similarly, users access data and metadata through the "Data Web Service" and TAP interfaces. The use of a small number of well tested interfaces improves the reliability of the service and makes it easier to maintain on a limited budget. Using the same interfaces that our users rely on ensures that problems are discovered and addressed quickly.

The system was initially quite different. Before the advent of CAOM, every archive maintained a custom database. Each file was stored in AD and ingested into the database as it arrived through e-transfer.[2] The JCMT supplied by replication a set of observatory databases that contained file metadata for raw data, and published an interface control document (ICD) describing the file headers in reduced data products. The JCMT committed itself to follow strict FITS standards for file headers and WCS, and for raw data reproduced a set of columns in the "File Metadata" database that was nearly identical to the set of headers in the reduced data for single observations. The CADC archive scientist was responsible for the design of software that read the metadata from the replicated database or from the reduced data headers. Writing and maintaining the software to ingest the metadata into the "JSA CAOM Metadata" database required a team of software developers at the CADC. The successful operation of this system required close collaboration of the JCMT with the JSA team at the CADC, with weekly progress videocons and regular (often annual) face-to-face meetings to discuss larger issues. Although the system worked, it was cumbersome and expensive. A leaner and more versatile system was clearly desirable.

The container labelled "Portable Processes" in Fig. 4 illustrates how the leaner system was implemented. The custom software for each archive was refactored into a set of simpler processes. Data processing ran at the CADC for easy access to the stored data, but was developed and maintained by the JAC. This encouraged a clean separation between the "Data Processing Queue" and "Data Processing" itself. The JSA was an early adopter of CAOM, which allowed raw and processed data ingestion to be factored out as separate processes. Since raw data ingestion applies to whole observations, the "Raw Data Discovery Agent" verifies that all of the raw data for an observation is stored in AD before starting the "Raw Data Ingestion" process. Originally, "Processed Data Ingestion" had its own discovery agent, but it is now controlled by the "Data Processing Queue".

The refactored system is quite modular and deployment is extremely flexible. These processes were deployed at the CADC for most of the last decade, but over the last year have migrated to the JAC. Data processing is currently run at the JAC using a queue system with database tables similar to those used by CADC's original interface to Sun Grid Engine. This has allowed the associated software to run with minimal changes. The new system has a web interface which is tailored to the JCMT, including a facility for in-house quality assurance. It is anticipated that data processing might move onto a CANFAR Virtual Machine in the near future and be orchestrated by the current queuing system. Ingestion can now run on any node that can access the "CAOM Repository", read

2 For an introduction to the e-transfer system see Melnychuk et al. (2005).

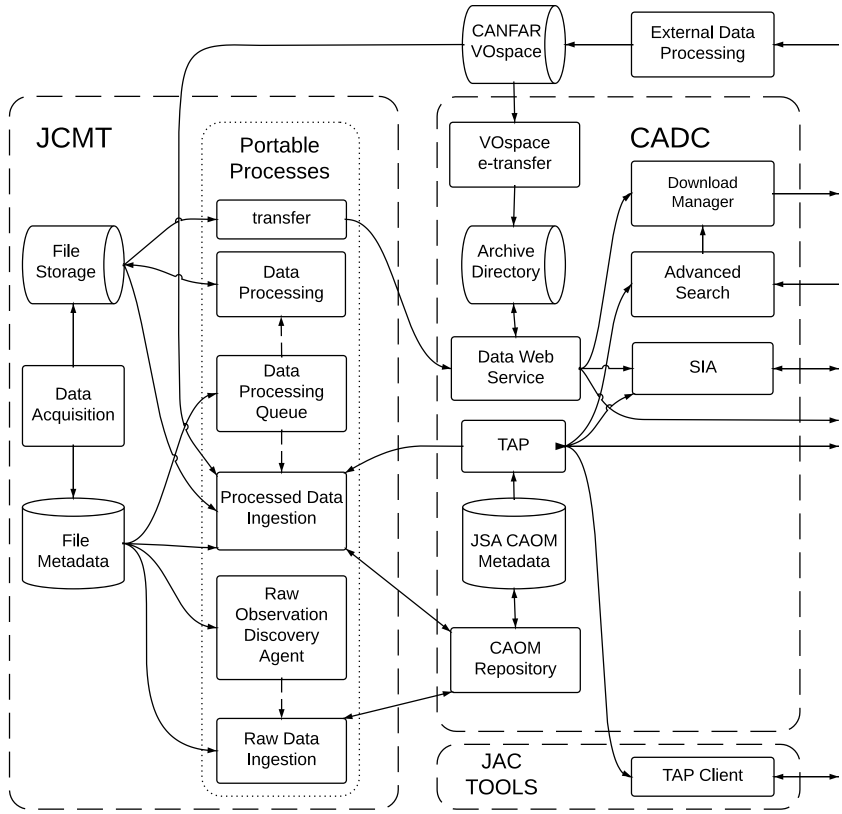

Fig. 4. Data and metadata flow through the JSA as it is intended to be in early 2015. The JCMT and CADC processes are arranged in four columns, with the JCMT-specific processes on the left, "portable processes" (data processing, file transfer and metadata ingestion) in the dotted box in the centre-left, CADC processes in the centre-right, and client processes running on archive users machines on the right. The "vertical drums" in the figure represent relational databases. The "horizontal drums" represent file storage, but do not specify the technology used to implement the storage (disk drives at the JCMT, databases at the CADC and for the CANFAR VOSpace). Where the software is developed, maintained and run by the JCMT/JAC or CADC, this is indicated by dashed container boxes. Manual operations by JAC and CADC staff have been elided; user interactions are shown as arrows on the right side of the figure.

existing metadata through the TAP service and optionally access the "File Metadata" service at the JCMT. This extraordinary flexibility allows JCMT staff who best understand the data to handle all data reduction and CADC staff who best understand the archive to maintain those services.

6. A continuous data release model

Using CADC's data processing infrastructure and the capabilities of JCMT/UKIRT's ORAC-DR automated data reduction, the JCMT Science Archive adopted a model of continuous release (Economou et al., 2011). As data was taken it was pushed for reduction and was ingested at CADC in the same 24-h period it was observed. Thus, high-quality science products were published in the VO as soon as the PI had access to them. Moreover, with every major improvement in the data reduction software, data could be reprocessed and again immediately released.

Proprietary data goes into CAOM and becomes available via VO interfaces almost immediately. Proprietary metadata and data restrictions are enforced on all TAP queries and authentication will permit authorized users to discover and download such data. Either AdvancedSearch or direct TAP queries can be used by PIs and JCMT legacy survey teams to find and download new data using this authenticated access. For example, the Cosmology Legacy Survey team (Geach et al., 2013) runs a script using the TAP interface to keep track of new observations as they arrive in the archive. For year 2014, approximately 40% of all queries to the JSA came through the TAP interface.

Continuous release made the VO publication mechanisms even more useful than they are in the normal data discovery process, as product availability is, from the point of view of the astronomer, unpredictable rather than coming in fixed, scheduled, announced "data releases". An interested user can therefore run regular automated TAP queries with the expectation that newly-reduced data can appear from their field of interest at any time.

7. Post-observatory

Meanwhile, CADC was working on the Canadian Advanced Network for Astronomical Research (CANFAR; Gaudet et al., 2010; Dowler et al., in preparation) project aiming to support a cloud-like model for astronomical data reduction. The system is based on giving the user a Virtual Machine (VM) that is then customized to provide the appropriate software, environment and data access. The user then defines a number of jobs that are serviced on a Condor compute platform composed of customized VM copies.

This service has been of great utility to the Canadian astronomical community dealing with large data volumes, with downloads of raw data from the JSA to CANFAR processing nodes accounting for more than 40% of all JSA raw data downloads in 2014. The Gould Belt Legacy Survey (GBS; Ward-Thompson et al., 2007) make use of CANFAR, and the GBS data processing lifecycle is supported at every step by VO-compliant services. Raw data is retrieved from VO-compliant discovery and delivery services, that data is processed on the customized VMs provisioned on CANFAR, and the resulting products are shared among survey members in VO-compliant storage services using VOSpace (Graham et al., 2013). The total VOSpace usage by the survey teams is currently approaching 1 TB and this has proven to be a critical part of the collaboration infrastructure when dealing with teams spread over Canada, Hawaii and Europe.

The existence of the VOSpace system at CADC has also led to them taking on the role of data publisher for JCMT science papers. JSA data products and externally reduced products can be copied to a VOSpace directory and associated with a Digital Object Identifier. The first two data sets making use of this functionality were Wilson et al. (2012) and Dempsey et al. (2013b).

8. Extending VO for radio astronomy

In the early days of the Virtual Observatory, the focus was specifically on simple protocols (Tody et al., 2009; Williams et al., 2008) to replace pre-existing web services for image retrieval and cone search; with retrieval of individual spectra coming somewhat later in VO developments (Tody et al., 2012; Škoda et al., 2014). These were the pressing issues of the optical community and this discussion dominated early protocol development.

Data cubes were seen as a task for the future as it was felt that they were products that were not yet in the mainstream and optical/IR instruments generating such cubes (such as the UIST IFU or TAURUS imaging Fabry–Perot spectrometer; Ramsay Howat et al., 2004; Atherton et al., 1982) were seen as something of niche interest to be tackled later. This was frustrating given that JCMT heterodyne observations regularly generated cubes and with the arrival of ACSIS in 2006, gigabyte data cubes were commonplace. There was no standard available for making all these cubes available to the VO and it is only recently (e.g., Tody et al., 2014) that a cube access protocol has been approached with any seriousness, driven mainly, in the USA, by ALMA and JWST developments (e.g., MIRI; Wright et al., 2010). The proposed recommendation for SIA-2.0 (Dowler et al., 2014b) will be able to handle the many data cubes generated by the JCMT over the last two decades.

In Table 1, the line labelled "TAP querying for Spectra" and "TAP querying for Cubes" indicate the number of 1-D spectra and data cubes in the JCMT collection. These can easily be found using the CADC AdvancedSearch interface, or directly using a TAP query. The full positional and photon energy WCS are provided for these, even when the positional axes are degenerate. SIA-2.0 should be able to find all of these data, once it has been implemented.

Another peculiarity of submillimetre data is the lack of point sources. Most Galactic objects are extended and dust and gas from large clouds, outflows and filamentary structures are missed by standard source extraction algorithms such as SExtractor (Bertin and Arnouts, 1996, ascl:1010.064). Instead, algorithms such as FellWalker (Berry, 2015; Berry et al., 2007, ascl:1311.007) and Clumpfind (Williams et al., 1994, ascl:1107.014), which detect source emission in irregularly shaped clumps, were used when doing source finding. VO ConeSearch was not set up for this eventuality and the best we could hope for was to provide a catalogue that indicated the peak of the emission. To work around this problem clump catalogues are generated with the clump outline approximated by a polygon specified in STC-S format (Berry and Draper, 2010). These outlines can then be retrieved using TAP for analysis or plotting. This is certainly less convenient for the end user than a clump equivalent of ConeSearch so we are extending the facilities in GAIA (Draper et al., 2009, ascl:1403.024) to hide the TAP interface. We hope a variant of ConeSearch will be developed that works for extended irregular sources. It should be sufficient for an enhanced ConeSearch query to return the results as catalogues with STC-S columns representing the shape of the object that matches, and for a match to be defined as an overlap between the region specified by the caller and the region defining the object. In this manner all existing ConeSearch services could simply return objects with circular regions with size corresponding to the point spread function.

9. Current status

Fig. 5 demonstrates that between 2010 and 2013 more than half of the refereed papers published containing JCMT data, obtained data from the JSA. Table 1 provides the current size of the data holdings accessible via a variety of VO protocols, and Table 2 provides details of how the downloads from the JSA are split between raw and reduced data.

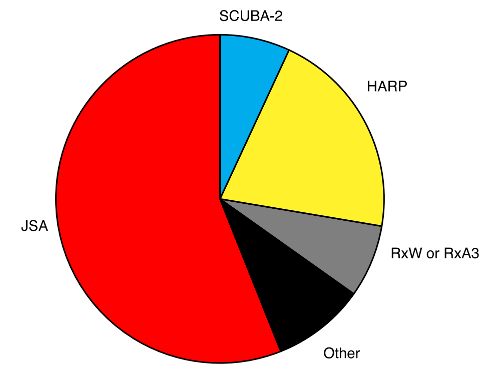

Fig. 5. Breakdown of the 405 JCMT refereed publications between 2010 and 2013 indicating the fraction using data from the JCMT Science Archive. The remaining segments are from papers only using instrument data directly.
Source: Figure derived from Bell et al. (2014b).

Table 1
Data holdings in the JCMT science archive available over VO protocols as of 2014 November.

Data model	Data sets available
TAP with CAOM (AdvancedSearch)	1 279 617
TAP with ObsCore	1 103 787
TAP querying for cubes	102 392
TAP querying for spectra	227 839
SIA	335 185

Table 2
Downloads of raw and reduced data from the JSA for the first 11 months of 2014. 40% of the raw downloads are to CANFAR processing nodes. When interpreting the relative count of raw and processed files, note that SCUBA-2 generates 480 discrete data files every half hour, which may result in only two output maps (one for each wavelength, depending on tiling scheme).

	Number of files	Data volume (GB)
Processed	72 730	1 764
Raw	4 427 478	63 611

The collaboration has proven so successful that the opportunity was taken to transfer the UKIRT raw data from the Cassegrain instruments to CADC (Bell et al., 2004a). It has been possible to re-use the JSA processing infrastructure for UKIRT data processing as the pipeline environment is identical (Jenness and Economou, 2015). Similarly, the ingestion software initially developed for the JSA was easily adapted to ingest data from several other CADC collections, including BLAST (Balloon-borne Large Aperture Submillimeter Telescope), CGPS (Canadian Galactic Plane Survey), IRIS (Improved Reprocessing of the IRAS Survey), and VGPS (VLA Galactic Plane Survey).

The JSA data processing continues to be improved (Johnstone, 2014) and the current plan is to reduce all the public HARP/ACSIS and SCUBA-2 data using an "all-sky" HEALPix projection (Górski et al., 2005; Bell et al., 2014b; Bell, 2014). This processing will also result in catalogue products that are specifically designed to answer the question of whether the JCMT saw any emission in a particular part of the sky. This is achieved by doing a two-pass approach to clump finding where first the emission outline is determined, and ultimately represented by an STC-S polygon, and then the individual peaks are located (Graves, 2014).

There is also an intent to expand the holdings of the JSA to include heterodyne data taken in an older format by the DAS (Bos, 1986) and AOS-C backends. Data from those instruments is being converted from the GSD format (Jenness et al., 1999) to the newer ACSIS format and this allows all the standard processing infrastructure to be used to create reduced data products and make them available to the VO for the first time.

The JSA pioneered the use of CAOM at the CADC being implemented in both CAOM-0.9 and CAOM-1. The latest version, CAOM-2 (Redman and Dowler, 2013; Dowler, 2012), was released for general use on 2014 May 1 and includes clarifications and improvements due to lessons learned from the earlier models. The metadata that is available for searching is richer, more complete, and easier to understand than anything that has been available previously. A full description of CAOM is in the early stages of preparation, but the earlier references cited above still describe the core philosophy of the design, and the current database schema can be found online.[3]

10. Lessons learned

The JCMT Science Archive collaboration was a high successful foray into VO publication via an observatory-data centre collaboration.

Elements that we believe led to this success:

- VO publication was a common goal with significant organizational buy-in for both parties from the start, and was a primary technical goal of the collaboration rather than an afterthought.
- Within that shared vision, there was a clear division of expertise and responsibilities for each side, allowing each organization to focus on its proximate technical goals. Both organizations had "skin in the game" that was served by the technical work undertaken, which allowed this work to be carried out without any kind of external agency funding (each institution supported its own share of the work out of its normal budgetary process).
- Each organization worked from a position of strength based on an advanced, robust and mature software architecture, allowing development to focus on new functionality and interfaces between the two systems. This minimized the communication overheads commonly associated with distributed projects.
- The role of "data engineer" responsible for developing software to ingest new data into a CAOM archive no longer requires special privileges at the CADC. It does require an expert knowledgeable about both the CAOM model and the products generated by the data reduction system, but the tools developed for CAOM allow this role to be assigned to the best available expert regardless of their location or institutional association. Thus, for UKIRT, the Joint Astronomy Centre has been able to assign one of their own staff to this role, and for the JSA a retired CADC staff member currently fills the role.
- There was a high level of pre-existing trust between the two groups from their previous relationship leading to minimal need for contractual language or management oversight. Indeed the entire collaboration's only official governance document was a two paragraph memorandum of understanding.

11. Recommendations

In the general case, for observatories that do not understand the mechanisms or benefits of VO publication, collaboration with a motivated VO-involved data centre that has the appropriate infrastructure and keeps up to date with the IVOA standards process is a far more effective choice than trying to develop those capabilities in-house, especially since there seems to be confusion in the observatory community as to what "VO publication" involves and what are the merits of doing it.

However, in order to be able to properly leverage the capability of a modern multi-mission data centre, a fanatical devotion to correct and complete metadata should be considered a pre-requisite.

Good communications within the team of collaborators is essential. Regular weekly or bi-weekly teleconferences and occasional face-to-face meetings have been important to keeping everyone aware of issues and working to common purposes.

[3] http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/caom2/.

Acknowledgements

The James Clerk Maxwell Telescope has historically been 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 the Netherlands Organisation for Scientific Research. Additional funds for the construction of SCUBA-2 were provided by the Canada Foundation for Innovation. The Canadian Astronomy Data Centre is operated by the National Research Council of Canada with the support of the Canadian Space Agency. This research has made use of NASA's Astrophysics Data System.

References

Archibald, E.N., et al., 2002. On the atmospheric limitations of ground-based submillimetre astronomy using array receivers. Mon. Not. R. Astron. Soc. 336, 1. doi:10.1046/j.1365-8711.2002.05582.x, arXiv:astro-ph/0204444.
Atherton, P.D., Taylor, K., Pike, C.D., Harmer, C.F.W., Parker, N.M., Hook, R.N., 1982. TAURUS: a wide-field imaging Fabry–Perot spectrometer for astronomy. Mon. Not. R. Astron. Soc. 201, 661–696.
Bell, G., 2014. Tiling scheme for the JSA. JCMT Newslett. 35, 20.
Bell, G.S., Currie, M.J., Redman, R.O., Purves, M., Jenness, T., 2014a. A new archive of UKIRT legacy data at CADC. In: Manset, N., Forshay, P. (Eds.), Astronomical Data Analysis Software and Systems XXIII. In: ASP Conf. Ser., vol. 485. p. 143.
Bell, G.S., et al., 2014b. Advanced data products for the JCMT science archive. In: Chiozzi, G., Radziwill, N.M. (Eds.), Software and Cyberinfrastructure for Astronomy III. In: Proc. SPIE, vol. 9152. p. 91522J. doi:10.1117/12.2054983.
Berry, D.S., 2015. FellWalker—a clump identification algorithm. Astron. Comput. in press, doi:10.1016/j.ascom.2014.11.004, arXiv:1411.6267.
Berry, D., Draper, P., 2010. Using the AST library to create and use STC-S region descriptions. In: Mizumoto, Y., Morita, K.I., Ohishi, M. (Eds.), Astronomical Data Analysis Software and Systems XIX. In: ASP Conf. Ser., vol. 434. p. 213.
Berry, D.S., Reinhold, K., Jenness, T., Economou, F., 2007. CUPID: a clump identification and analysis package. In: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), Astronomical Data Analysis Software and Systems XVI. In: ASP Conf. Ser., vol. 376. p. 425.
Bertin, E., Arnouts, S., 1996. SExtractor: software for source extraction. Astron. Astrophys. Suppl. 117, 393.
Bos, A., 1986. Backend spectrometers—a review. In: Kollberg, E. (Ed.), Instrumentation for submillimeter spectroscopy. In: Proc. SPIE, vol. 598. p. 134. doi:10.1117/12.952332.
Buckle, J.V., et al., 2009. HARP/ACSIS: a submillimetre spectral imaging system on the James Clerk Maxwell Telescope. Mon. Not. R. Astron. Soc. 399, 1026. doi:10.1111/j.1365-2966.2009.15347.x, arXiv:0907.3610.
Chapin, E.L., Berry, D.S., Gibb, A.G., Jenness, T., Scott, D., Tilanus, R.P.J., Economou, F., Holland, W.S., 2013. SCUBA-2: iterative map-making with the Sub-Millimetre User Reduction Facility. Mon. Not. R. Astron. Soc. 430, 2545. doi:10.1093/mnras/stt052, arXiv:1301.3652.
Chrysostomou, A., 2010. The JCMT legacy survey: mapping the milky way in the submillimetre. Highlights Astron. 15, 797. doi:10.1017/S1743921310011750.
Crabtree, D., Durand, D., Fisher, W., Gaudet, S., Hill, N., Justice, G., Morris, S., Woodsworth, A., 1994. The Canadian astronomy data centre. In: Crabtree, D.R., Hanisch, R.J., Barnes, J. (Eds.), Astronomical Data Analysis Software and Systems III. In: ASP Conf. Ser., vol. 61. p. 123.
Davis, G.R., 2005. JCMT Newslett. 23. URL: http://www.jach.hawaii.edu/JCMT/publications/newsletter/n23/jcmt-n23.pdf.
Dempsey, J.T., et al., 2013a. SCUBA-2: on-sky calibration using submillimetre standard sources. Mon. Not. R. Astron. Soc. 430, 2534. doi:10.1093/mnras/stt090, arXiv:1301.3773.
Dempsey, J.T., Thomas, H.S., Currie, M.J., 2013b. CO (3-2) High-resolution survey of the galactic plane: R1. ApJS 209, 8. doi:10.1088/0067-0049/209/1/8.
Dent, W., et al., 2000. HARP and ACSIS on the JCMT. In: Mangum, J.G., Radford, S.J.E. (Eds.), Imaging at Radio through Submillimeter Wavelengths. In: ASP Conf. Ser., vol. 217. p. 33.
Disney, M.J., Wallace, P.T., 1982. Starlink. QJRAS 23, 485.
Dowler, P., 2012. CAOM-2.0: the inevitable evolution of a data model. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. In: ASP Conf. Ser., vol. 461. p. 339.
Dowler, P., et al., 2015. Use of virtual observatory standards and technology at the Canadian astronomy data center and CANFAR. Astron. Comput. in preparation.

Dowler, P., Bonnarel, F., Michel, L., Demleitner, M., 2014a. IVOA Proposed Recommendation: IVOA DataLink Version 1.0. URL: http://www.ivoa.net/documents/DataLink/.

Dowler, P.D., Gaudet, S., Durand, D., Redman, R., Hill, N., Goliath, S., 2007. Common archive observation model. In: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), Astronomical Data Analysis Software and Systems XVI. In: ASP Conf. Ser., vol. 376. p. 347.

Dowler, P., Gaudet, S., Durand, D., Redman, R., Hill, N., Goliath, S., 2008. Common archive observation model. In: Argyle, R.W., Bunclark, P.S., Lewis, J.R. (Eds.), Astronomical Data Analysis Software and Systems XVII. In: ASP Conf. Ser., vol. 394. p. 426.

Dowler, P., Rixon, G., Tody, D., 2011. IVOA Recommendation: Table Access Protocol Version 1.0. URL: http://www.ivoa.net/documents/TAP/, arXiv:1110.0497.

Dowler, P., Tody, D., Bonnarel, F., Plante, R., 2014b. IVOA: Proposed Recommendataion: IVOA Simple Image Access Version 2.0. URL: http://www.ivoa.net/documents/SIA/20141024/.

Draper, P.W., Berry, D.S., Jenness, T., Economou, F., 2009. GAIA—version 4. In: Bohlender, D.A., Durand, D., Dowler, P. (Eds.), Astronomical Data Analysis Software and Systems XVIII. In: ASP Conf. Ser., vol. 411. p. 575.

Economou, F., Bridger, A., Wright, G.S., Jenness, T., Currie, M.J., Adamson, A., 1999. ORAC-DR: pipelining with other people's code. In: Mehringer, D.M., Plante, R.L., Roberts, D.A. (Eds.), Astronomical Data Analysis Software and Systems VIII. In: ASP Conf. Ser., vol. 172. p. 11.

Economou, F., Hoblitt, J.C., Norris, P., 2014. Your data is your dogfood: DevOps in the astronomical observatory. ArXiv e-prints arXiv:1407.6463.

Economou, F., et al., 2011. The JSA and the grid: how infinite computing power enables a new archive model for PI-led observatories. In: Evans, I.N., Accomazzi, A., Mink, D.J., Rots, A.H. (Eds.), Astronomical Data Analysis Software and Systems XX. In: ASP Conf. Ser., vol. 442. p. 203.

Economou, F., Jenness, T., Chrysostomou, A., Cavanagh, B., Redman, R., Berry, D.S., 2008. The JCMT legacy survey: the challenges of the JCMT science archive. In: Argyle, R.W., Bunclark, P.S., Lewis, J.R. (Eds.), Astronomical Data Analysis Software and Systems XVII. In: ASP Conf. Ser., vol. 394. p. 450.

Economou, F., Jenness, T., Rees, N.P., 2002. Sharing code and support between heterogeneous telescopes: the UKIRT and JCMT joint software projects. In: Quinn, P.J. (Ed.), Observatory Operations to Optimize Scientific Return III. In: Proc. SPIE, vol. 4844. p. 321. doi:10.1117/12.460642.

Gaudet, S., Dowler, P., Goliath, S., Redman, R., 2008a. The JCMT science archive and the virtual observatory. In: Argyle, R.W., Bunclark, P.S., Lewis, J.R. (Eds.), Astronomical Data Analysis Software and Systems XVII. In: ASP Conf. Ser., vol. 394. p. 135.

Gaudet, S., Durand, D., Schade, D., 2008b. The evolving role of a data centre in support of observatory operations and community needs. In: Brissenden, R.J., Silva, D.R. (Eds.), Observatory Operations: Strategies, Processes, and Systems II. In: Proc. SPIE, vol. 7016. p. 70160J. doi:10.1117/12.790295.

Gaudet, S., et al., 2010. CANFAR: the Canadian advanced network for astronomical research. In: Radziwill, N.M., Bridger, A. (Eds.), Software and Cyberinfrastructure for Astronomy. In: Proc. SPIE, vol. 7740. p. 77401I. doi:10.1117/12.858026.

Geach, J.E., et al., 2013. The SCUBA-2 cosmology legacy survey: blank-field number counts of 450-μm-selected galaxies and their contribution to the cosmic infrared background. Mon. Not. R. Astron. Soc. 432, 53. doi:10.1093/mnras/stt352, arXiv:1211.6668.

Gibb, A.G., Scott, D., Jenness, T., Economou, F., Kelly, B.D., Holland, W.S., 2005. Design of the SCUBA-2 quick look display and data reduction pipeline. In: Shopbell, P., Britton, M., Ebert, R. (Eds.), Astronomical Data Analysis Software and Systems XIV. In: ASP Conf. Ser., vol. 347. p. 585.

Górski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M., Bartelmann, M., 2005. HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. ApJ 622, 759–771. doi:10.1086/427976, arXiv:astro-ph/0409513.

Graham, M., Morris, D., Rixon, G., Dowler, P., Schaaff, A., Tody, D., 2013. IVOA Recommendation: VOSpace specification Version 2.0. URL: http://www.ivoa.net/documents/VOSpace/20130329/REC-VOSpace-2.0-20130329.pdf, arXiv:1110.0508.

Graham, M., Plante, R., Rixon, G., Taffoni, G., 2010. IVOA Recommendation: IVOA Credential Delegation Protocol Version 1.0. URL: http://www.ivoa.net/Documents/REC/ADQL/ADQL-20081030.pdf, arXiv:1110.0509.

Graves, S., 2014. Source finding: creating a legacy catalogue. JCMT Newslett. 35, 21.

Holland, W.S., et al., 2013. SCUBA-2: the 10 000 pixel bolometer camera on the James Clerk Maxwell Telescope. Mon. Not. R. Astron. Soc. 430, 2513. doi:10.1093/mnras/sts612, arXiv:1301.3650.

Holland, W.S., Duncan, W., Kelly, B.D., Irwin, K.D., Walton, A.J., Ade, P.A.R., Robson, E.I., 2003. SCUBA-2: a new generation submillimeter imager for the James Clerk Maxwell Telescope. In: Phillips, T.G., Zmuidzinas, J. (Eds.), Millimeter and Submillimeter Detectors for Astronomy. In: Proc. SPIE, vol. 4855. p. 1. doi:10.1117/12.459152.

Jenness, T., Bohlender, D.A., Gaudet, S.J., Economou, F., Tilanus, R.P.J., Durand, D., Hill, N.R., 2002. Reducing and calibrating SCUBA data on demand. In: Bohlender, D.A., Durand, D., Handley, T.H. (Eds.), Astronomical Data Analysis Software and Systems XI. In: ASP Conf. Ser., vol. 281. p. 243.

Jenness, T., Cavanagh, B., Economou, F., Berry, D.S., 2008. JCMT science archive: advanced heterodyne data products pipeline. In: Argyle, R.W., Bunclark, P.S., Lewis, J.R. (Eds.), Astronomical Data Analysis Software and Systems XVII. In: ASP Conf. Ser., vol. 394. p. 565.

Jenness, T., Economou, F., 1999. The SCUBA data reduction pipeline: ORAC-DR at the JCMT. In: Mehringer, D.M., Plante, R.L., Roberts, D.A. (Eds.), Astronomical Data Analysis Software and Systems VIII. In: ASP Conf. Ser., vol. 172. p. 171.

Jenness, T., Economou, F., 2011. Data management at the UKIRT and JCMT. In: Gajadhar, S., et al. (Eds.), Telescopes from Afar. p. 42. arXiv:1111.5855.

Jenness, T., Economou, F., 2015. ORAC-DR: a generic data reduction pipeline infrastructure. Astron. Comput. 9, 40. doi:10.1016/j.ascom.2014.10.005, arXiv:1410.7509.

Jenness, T., Tilanus, R.P.J., Meyerdierks, H., Fairclough, J., 1999. The Global Section Datafile (GSD) access library. Starlink User Note 229. Joint Astronomy Centre.

Johnstone, D., 2014. The JCMT science archive: ambitious plans for the JSA. JCMT Newslett. 35, 19.

Louys, M., et al. 2011. IVOA Recommendation: Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.0. URL: http://www.ivoa.net/documents/ObsCore/20111028/REC-ObsCore-v1.0-20111028.pdf, arXiv:1111.1758.

Melnychuk, G., Damian, A., Goliath, S., Shao, L., Hill, N., Gaudet, S., Bohlender, D., Aspin, C., 2005. Gemini science archive automatic data transfer. In: Shopbell, P., Britton, M., Ebert, R. (Eds.), Astronomical Data Analysis Software and Systems XIV. In: ASP Conf. Ser., vol. 347. p. 647.

Nandrekar-Heinis, D., Michel, L., Louys, M., Bonnarel, F., 2014. The table access protocol: providing standard access to astronomical data. Astron. Comput. 7–8, 37–44. (Special Issue on The Virtual Observatory: I) doi:10.1016/j.ascom.2014.08.002.

Ochsenbein, F., Fernique, P., Bonnarel, F., Allen, M., Boch, T., Genova, F., Schaaff, A., 2005. Interoperability in action: the Aladin experience. In: Shopbell, P., Britton, M., Ebert, R. (Eds.), Astronomical Data Analysis Software and Systems XIV. In: ASP Conf. Ser., vol. 347. p. 193.

Ochsenbein, F., et al. 2013. IVOA Recommendation: VOTable Format Definition Version 1.3. URL: http://www.ivoa.net/documents/VOTable/20130920/REC-VOTable-1.3-20130920.html.

Ortiz, I., et al. 2011. IVOA Recommendation: IVOA Astronomical Data Query Language Version 2.00. URL: http://www.ivoa.net/Documents/CredentialDelegation/20100218/REC-CredentialDelegation-1.0.html, arXiv:1110.0503.

Ramsay Howat, S.K., et al., 2004. The commissioning of and first results from the UIST imager spectrometer. In: Moorwood, A.F.M., Iye, M. (Eds.), Ground-based Instrumentation for Astronomy. In: Proc. SPIE, vol. 5492. p. 1160. doi:10.1117/12.551673.

Redman, R., 2006. An introduction to the JCMT data management system. JCMT Newslett. 24, 22. URL: http://www.jach.hawaii.edu/JCMT/publications/newsletter/n24/jcmt-n24.pdf.

Redman, R.O., Dowler, P., 2013. Implementing a common database architecture at the CADC using CAOM-2. In: Friedel, D. (Ed.), Astronomical Data Analysis Software and Systems XXII. In: ASP Conf. Ser., vol. 475. p. 159.

Schade, D., Dowler, P., Durand, D., Hill, N., Gaudet, S., Bohlender, D., 2002. Data mining and the virtual observatory: a data centre viewpoint. In: Bohlender, D.A., Durand, D., Handley, T.H. (Eds.), Astronomical Data Analysis Software and Systems XI. In: ASP Conf. Ser., vol. 281. p. 36.

Škoda, P., Draper, P.W., Neves, M.C., Andrešič, D., Jenness, T., 2014. Spectroscopic analysis in the virtual observatory environment with SPLAT-VO. Astron. Comput. 7, 108. doi:10.1016/j.ascom.2014.06.001, arXiv:1407.1765.

Taylor, M.B., 2005. TOPCAT & STIL: starlink table/VOTable processing software. In: Shopbell, P., Britton, M., Ebert, R. (Eds.), Astronomical Data Analysis Software and Systems XIV. In: ASP Conf. Ser., vol. 347. p. 29.

Tilanus, R.P.J., Jenness, T., Economou, F., Cockayne, S., 1997. The JCMT telescope management system. In: Hunt, G., Payne, H. (Eds.), Astronomical Data Analysis Software and Systems VI. In: ASP Conf. Ser., vol. 125. p. 397.

Tody, D., et al. 2012. Simple Spectral Access Protocol Version 1.1 IVOA Recommendation. URL: http://www.ivoa.net/documents/SSA/20120210/REC-SSA-1.1-20120210.pdf, arXiv:1203.5725.

Tody, D., Plante, R., Harrison, P., 2009. IVOA Recommendation: Simple Image Access Specification Version 1.0. URL: http://www.ivoa.net/documents/SIA/20091116/REC-SIA-1.0.html, arXiv:1110.0499.

Tody, D., et al., 2014. Accessing multi-dimensional images and data cubes in the virtual observatory. In: American Astronomical Society Meeting #223. In: American Astronomical Society Meeting Abstracts, vol. 223. p. #255.05.

Ward-Thompson, D., et al., 2007. The James Clerk Maxwell Telescope legacy survey of nearby star-forming regions in the gould belt. PASP 119, 855. doi:10.1086/521277, arXiv:0707.0169.

Williams, J.P., de Geus, E.J., Blitz, L., 1994. Determining structure in molecular clouds. ApJ 428, 693. doi:10.1086/174279.

Williams, R., Hanisch, R., Szalay, A., Plante, R., 2008. IVOA Recommendation: Simple Cone Search Version 1.03. URL: http://www.ivoa.net/documents/REC/DAL/ConeSearch-20080222.html, arXiv:1110.0498.

Wilson, C.D., et al., 2012. The JCMT nearby galaxies legacy survey—VIII. CO data and the $L_{CO(3-2)}$–L_{FIR} correlation in the SINGS sample. Mon. Not. R. Astron. Soc. 424, 3050. doi:10.1111/j.1365-2966.2012.21453.x, arXiv:1206.1629.

Wright, G.S., et al., 2010. Progress with the design and development of MIRI, the mid-IR instrument for JWST. In: Oschmann Jr., Jacobus M., Clampin, Mark C., MacEwen, Howard A. (Eds.), Space Telescopes and Instrumentation 2010: Optical, Infrared, and Millimeter Wave. In: Proc. SPIE, vol. 7731. p. 77310E. doi:10.1117/12.857262.