Astronomical Data Analysis Software and Systems XXVI
ASP Conference Series, Vol. 521
Marco Molinaro, Keith Shortridge, and Fabio Pasian, eds.

Common Access to 2-D and 3-D Galactic Radio Surveys Within the VIALACTEA Project

Robert Butora, Marco Molinaro, and Riccardo Smareglia

INAF - Osservatorio Astronomico di Trieste,
via G.B. Tiepolo 11, 34143 Trieste, Italy;
butora@oats.inaf.it

Abstract. VIALACTEA is a project to facilitate the study of star formation in our Galaxy based on various galactic surveys from past years. This survey data is available in the form of FITS files which are accessed through a web service developed by IA2 (Italian center for Astronomical Archives) at INAF-OATs. The service handles ≈ 40000 3-D radio cubes and 2-D images, originating from about 20 surveys, each offering data on several molecular emission lines. All the data is accessible from one interface in the form of a web service, offering the common search, cutout and merge functions.

1. Introduction

VIALACTEA is a project to facilitate the study of star formation in our Galaxy. It is based on data collected in many galactic surveys from past years. The comparison of the data at several molecular emission line wavelengths, from various surveys, is crucial for the project. To facilitate these comparisons, a compact system was built, storing all survey data and catalagues called *VIALACTEA Knowledge Base* (Molinaro et al. 2016). The generation and handling of the catalogues is described in Smareglia et al. (2019). This paper describes the *Search and Access* interface, which exposes the surveys' radiocubes.

2. Radiocubes

The VLKB storage contains close to 40000 3-D radio cubes and 2-D images, in form of FITS files, altogether about 1TB of data. These originate from about 20 different surveys (see Acknowledgments), and so they are rather heterogeneous in their content:

- 2D images in the radio continuum,
- 3D data cubes with radio velocity spectra at specific molecular lines,
- collection of 3D extinction maps by distance.

And also heterogeneous in their encoding, because the survey publishers use different software packages to generate the FITS-files. There are differences in coordinate system and sky frame references (like for instance galactic coordinates versus equatorial ones - which are the easiest to handle) as well as degenerate axis references and multiple HDUs or non-standard keywords in the FITS headers.

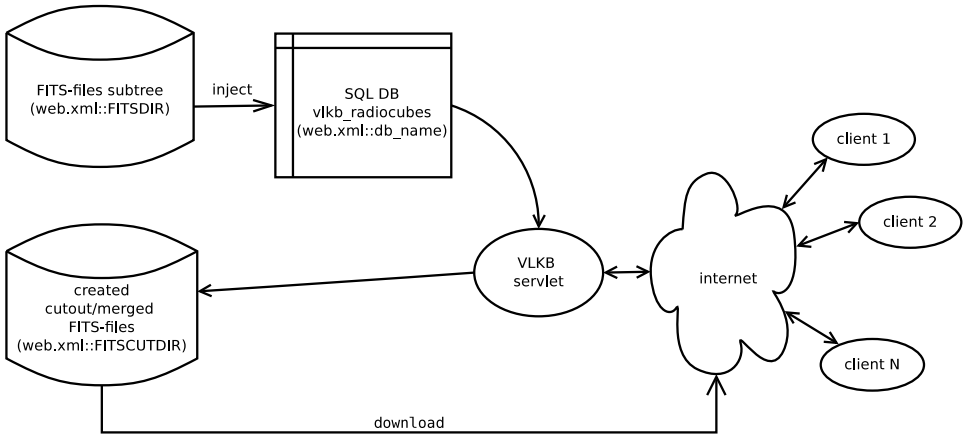

Figure 1. Overview of system components. *VLKB servlet* to interface the requests from clients, *SQL database server* to store metadata of the radiocubes, *storage* to hold the generated datacubes requested by clients and ready for download, *storage* of the original FITS files from surveys, *injection system* which traverses the FITS files' directories and generates the database entries. The components are configurable before deployment (web.xml).

Considering a *sub-survey* (i.e. a collection of data from a specific survey or pointed archive referring to only one single molecular line or band or other specific metadata), about 50 different data collections have been put together to be searched and accessed as a unique resource (see Tables 1 & 2).

Table 1. VIALACTEA VLKB stored data collections consisting of 2D radio continuum images.

Name	**sub-survey**	**# files**	**size [GB]**
CORNISH	5 GHz	1408	84
MAGPIS	1.4GHz	352	1.4
Hi-Gal	70μm	166	7.2
Hi-Gal	160μm	166	3.7
Hi-Gal	250μm	166	2.2
Hi-Gal	350μm	166	1.3
Hi-Gal	500μm	166	0.6
MIPSGAL	24μm	339	13
WISE	3.4μm	694	44
WISE	4.6μm	694	44
WISE	12μm	694	44
WISE	22μm	694	44

Hi-Gal data, the survey tiles listed in Table 1 as well all the data derived from them are private to the project. MOPRA, GRS, NANTEN, OGS have been approved for

Table 2. VIALACTEA VLKB stored surveys consisting of FITS 3D radio cubes.

Name	**sub-survey**	**files**	**size [GB]**	**Name**	**sub-survey**	**files**	**size [GB]**
MOPRA	12CO	52	45	MALT90	HCO+	2012	23
MOPRA	13CO	52	30	MALT90	HCN	2012	23
MOPRA	C17O	51	14	MALT90	N2H+	2012	23
MOPRA	C18O	51	24	MALT90	HNC	2012	23
CHIMPS	13CO	224	18	MALT90	13C34N	2012	23
CHIMPS	C18O	223	20	MALT90	13CS	2012	23
CHaMP	HCO+	16	1.6	MALT90	C2H	2012	23
HOPS	H2O	11	14	MALT90	CH3CN	2012	23
HOPS	NH3 (1-1)	11	5.3	MALT90	H13CO+	2012	23
HOPS	NH3 (2-2)	11	5.3	MALT90	H41alpha	2012	23
FCRAO_GRS	13CO	42	11	MALT90	HC13CCN	2012	23
ThrUMMS	12CO	23	13	MALT90	HC3N	2012	23
ThrUMMS	13CO	22	11	MALT90	HN13C	2012	23
ThrUMMS	C18O	23	11	MALT90	HNCO404	2012	23
ThrUMMS	CN	23	12	MALT90	HNCO413	2012	23
NANTEN	12CO	2	1.1	MALT90	SiO	2012	23
OGS	12CO	4	14	VGPS	HI	13	5.7
OGS	13CO	3	11	CGPS	HI	84	45
JCMT-HARP	12CO	92	24	SGPS	HI	13	4.4

usage inside the VIALACTEA project, but are covered by privacy policy. Additionally, the *Extinction Maps* data are private until the end of the project because they have been produced within the VIALACTEA project itself. All the other data cube surveys and pointed archives listed in Table 1 and 2, are publicly available from locations listed in the appendix to Molinaro et al. (2016).

3. Search and Access services

Remote access to the VLKB's FITS files was built as a web-service. The system components are shown in Figure 1.

The developed service interface now consists of three common services: *search engine* to perform data discovery, *cutout engine* to create sub-cubes, and *merge service* to unite FITS-file data of adjacent areas on the sky if stored in separate files.

These services share the same parametric query solution for positions, bound descriptions, sub-survey selection and are able to be searched and cutout indifferently of the underlying FITS header metadata - at least so long as the WCS descriptors (Calabretta & Greisen 2002; Greisen et al. 2006) are correctly in place and understood by the AST library (Berry et al. 2016). Specifically, the headers of various sub-surveys are passed to the AST library which is used to perform region overlap and boundary computations. The AST library is written in the C-programming language, so JNI (Java

Native Interface) technology is needed to access C-code from the Java domain. Most of the Java-code is performing interfacing. Essential work is done "below" the JNI, in the C-language domain. We process the partial results from AST in the C-domain and then return the final results to the Java domain. So we cross the JNI once per client request.

A similar approach is taken also for the merge service, where the engine used is Montage (http://montage.ipac.caltech.edu). The merging service internally adds 3D-datacubes which are the result of the cutout service.

The system can also operate in distributed mode to scale with the number of clients: in this case the JNI is replaced by the AMQP (http://www.amqp.org) messaging system to distribute the requests between several engines running on separate CPU cores.

Acknowledgments. This work has been developed under the European Community 7th Framework Programme, Grant Agreement 607380, VIALACTEA - The Milky Way as a Star formation engine.

The VLKB resource data and metadata are highly dependent on external public and private sources. We acknowledge the efforts of the staff of the primary repositories and archives from which we retrieved the data (see Tables 1 & 2) to be offered to the VIALACTEA members and, at the end of the project, to the astrophysical community; all the relevant starting points are listed in the appendix to Molinaro et al. (2016).

This research made use of Montage which is funded by the National Science Foundation under Grant Number ACI-1440620, and was previously funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology.

References

Berry, D. S., Warren-Smith, R. F., & Jenness, T. 2016, Astronomy and Computing, 15, 33. 1602.06681

Calabretta, M. R., & Greisen, E. W. 2002, A&A, 395, 1077. astro-ph/0207413

Greisen, E. W., et al. 2006, Astronomy and Astrophysics, 446, 747. astro-ph/0507293

Molinaro, M., et al. 2016, in Software and Cyberinfrastructure for Astronomy IV, vol. 9913 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 99130H. 1608.04526

Smareglia, R., et al. 2019, in ADASS XXVI, edited by M. Molinaro, K. Shortridge, & F. Pasian (San Francisco: ASP), vol. 521 of ASP Conf. Ser., 116