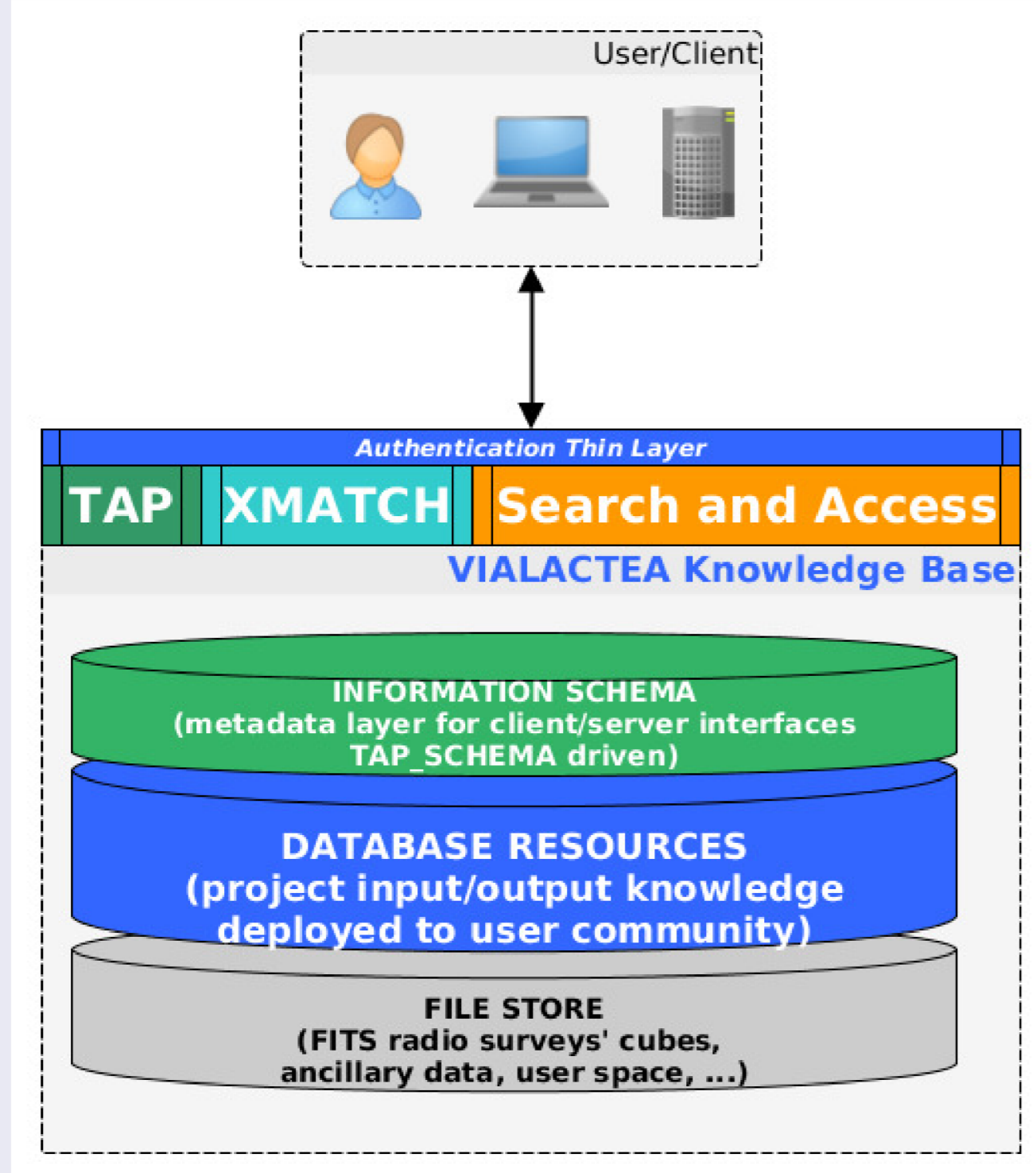


Introduction: the VIALACTEA Knowledge Base

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 on several wavelengths, molecular emission lines, from various surveys are crucial for the project. To facilitate these comparisons, a compact system was built, storing all survey data and catalogues needed by, and produced by the project. It is referred to as "VIALACTEA Knowledge Base". The figure below illustrates this system.



The system contains catalogues stored in SQL databases accessed through "TAP" and "XMATCH" interfaces. And also it contains about 1TB of data in FITS files, containing radiocubes visible through the "Search and Access" interface. This poster presentation describes the "Search and Access" interface, developed and hosted by IA2 (Italian center for Astronomical Archives) at INAF-OATs.

Radiocubes

The VLKB storage contains close to 40000 3-D radio cubes and 2-D images, in form of FITS[1] files. These originate from about 20 different surveys, and so they are rather heterogeneous by their content:

- 2D images in the radio continuum,
- 3D data cubes with radio velocity spectra,
- collection of 3D extinction maps.

And also they differ in encoding, and there are differences in coordinate system and sky frame references; galactic coordinates versus equatorial ones to explicit the most easy one to handle, as well as degenerate axis references and multiple HDUs or non-standard keywords in the FITS header.

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.

Table: 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	70 μ m	166	7.2
Hi-Gal	160 μ m	166	3.7
Hi-Gal	250 μ m	166	2.2
Hi-Gal	350 μ m	166	1.3
Hi-Gal	500 μ m	166	0.6
MIPSGAL	24 μ m	339	13
WISE	3.4 μ m	694	44
WISE	4.6 μ m	694	44
WISE	12 μ m	694	44
WISE	22 μ m	694	44

Table: 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

Table: VIALACTEA VLKB stored 3D extinction maps.

Name	sub-survey	# files	size [MB]
Extinction Maps	5 arcmin resolution	72	76
Extinction Maps	10 arcmin resolution	72	18

Acknowledgements

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 depending 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 section Radiocubes) to be offered to the VIALACTEA members and, at the end of the project, to the astrophysical community.

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.

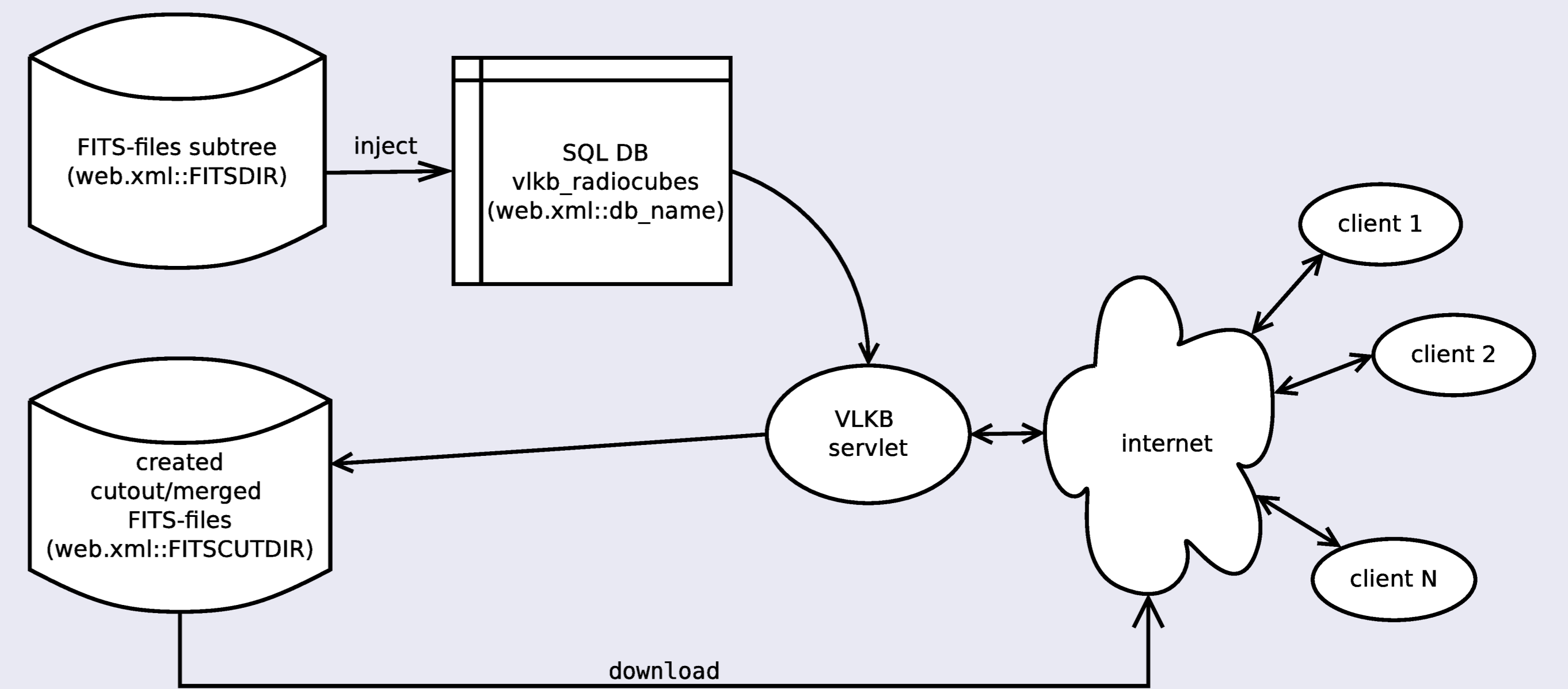
Search and Access services

Remote access to the VLKB's FITS files was built as web-service. The system components are:

- 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).

Figure: Overview of system components



Internally the servlet consists of two parts: interface and the engine. Servlet interface is implemented in Java domain and the engine in C, coupled by JNI (Java Native Interface), so that we cross the JNI boundary only once per client request.

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

The service offers common functionalities:

- search engine to perform data discovery
- cutout engine to create sub-cubes or sub-images
- merge service to unite FITS-files data of adjacent areas on the sky if stored in separate files

The functions are uniformly parametrized by sky- and spectral-coordinates independently of data calibration and data subdivision in the original survey FITS-files.

Search

Search inputs:

- circular or rectangular area in Galactic or ICRS coordinates
- optionally velocity range
- optionally survey name and/or molecular lines and/or transitions to narrow the search

Result is a list of found datacubes, where for each datacube:

- PublisherDID: identifier of the cube (needed for cutout)
- Overlap: whether requested region overlaps partially or fully
- Sub-survey [survey-name, species, transition]: which sub-survey the cube belongs to
- Four vertices: area on the sky covered by the cube
- Bounds: min and max bounds on all axes (2D sky- and 1D-spectral axes)

Cutout

Cutout inputs:

- PublisherDID to identify the chosen datacube
- circular or rectangular area in Galactic or ICRS coordinates
- optionally velocity range

Results:

- original FITS file HDU with coordinate indices where the cut was done
- URL to the generated FITS file ready for download

Merge

Merge inputs:

- sub-survey (e.g. survey name, molecular line, transition)
- circular or rectangular area in Galactic or ICRS coordinates
- optionally velocity range

Merge internally performs search with subsequent cutouts for all hits and then merges the adjacent cubes. So response is the same as for search and cutout above, plus the URL to the merged file ready for download.

Summary

- Achieved unified access to approx 50 different sub-surveys
 - Heterogeneous dataset deployment
 - Metadata database of various datasets
- In-house datahosting
 - chosen because proxying to existing survey data-sources was too heterogeneous, behind custom interfaces, usually web-forms hard to access programmatically
- Integration of already available libraries, standards, tools
 - FITS[1, 2, 3], CFITSIO[4], ATS[5], Montage (<http://montage.ipac.caltech.edu>)
- Providing a solution for machine tools, scripting, which the users can use to access the data

References

- [1] W. D. Pence, L. Chiappetti, C. G. Page, R. A. Shaw, and E. Stobie. Definition of the Flexible Image Transport System (FITS), version 3.0. *Astronomy and Astrophysics*, 524:A42, December 2010.
- [2] M. R. Calabretta and E. W. Greisen. Representations of celestial coordinates in FITS. *Astronomy and Astrophysics*, 395:1077–1122, December 2002.
- [3] E. W. Greisen, M. R. Calabretta, F. G. Valdes, and S. L. Allen. Representations of spectral coordinates in FITS. *Astronomy and Astrophysics*, 446:747–771, February 2006.
- [4] W. D. Pence. CFITSIO: A FITS File Subroutine Library. *Astrophysics Source Code Library*, October 2010.
- [5] D. S. Berry, R. F. Warren-Smith, and T. Jenness. AST: A library for modeling and manipulating coordinate systems. *Astronomy and Computing*, 15:33–49, April 2016.