PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

VIALACTEA knowledge base homogenizing access to Milky Way data

Molinaro, Marco, Butora, Robert, Bandieramonte, Marilena, Becciani, Ugo, Brescia, Massimo, et al.

Marco Molinaro, Robert Butora, Marilena Bandieramonte, Ugo Becciani, Massimo Brescia, Stefano Cavuoti, Alessandro Costa, Anna M. Di Giorgio, Davide Elia, Akos Hajnal, Hermann Gabor, Peter Kacsuk, Scige J. Liu, Sergio Molinari, Giuseppe Riccio, Eugenio Schisano, Eva Sciacca, Riccardo Smareglia, Fabio Vitello, "VIALACTEA knowledge base homogenizing access to Milky Way data," Proc. SPIE 9913, Software and Cyberinfrastructure for Astronomy IV, 99130H (8 August 2016); doi: 10.1117/12.2231674

Event: SPIE Astronomical Telescopes + Instrumentation, 2016, Edinburgh, United Kingdom

SPIE.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

VIALACTEA knowledge base homogenizing access to Milky Way data

Marco Molinaro[a], Robert Butora[a], Marilena Bandieramonte[b], Ugo Becciani[b], Massimo Brescia[c], Stefano Cavuoti[c], Alessandro Costa[b], Anna M. Di Giorgio[d], Davide Elia[d], Akos Hajnal[e], Hermann Gabor[e], Peter Kacsuk[e], Scige J. Liu[d], Sergio Molinari[d], Giuseppe Riccio[c], Eugenio Schisano[d], Eva Sciacca[b], Riccardo Smareglia[a], and Fabio Vitello[b]

[a]INAF - Osservatorio Astronomico di Trieste, via G.B. Tiepolo 11 - 34143 Trieste, Italy
[b]INAF - Osservatorio Astrofisico di Catania, Via S. Sofia 78 - 95123 Catania, Italy
[c]INAF - Osservatorio Astronomico di Capodimonte, Salita Moiariello 16 - 80131 Napoli, Italy
[d]INAF - Istituto di Astrofisica e Planetologia Spaziali, Area di Ricerca di Tor Vergata, via Fosso del Cavaliere 100 - 00133 Roma, Italy
[e]MTA-SZTAKI, 1111 Budapest, Kende u. 13-17, Hungary

ABSTRACT

The VIALACTEA project has a work package dedicated to "Tools and Infrastructure" and, inside it, a task for the "Database and Virtual Observatory Infrastructure". This task aims at providing an infrastructure to store all the resources needed by the, more purposely, scientific work packages of the project itself. This infrastructure includes a combination of: storage facilities, relational databases and web services on top of them, and has taken, as a whole, the name of VIALACTEA Knowledge Base (VLKB). This contribution illustrates the current status of this VLKB. It details the set of data resources put together; describes the database that allows data discovery through VO inspired metadata maintenance; illustrates the discovery, cutout and access services built on top of the former two for the users to exploit the data content.

Keywords: Databases, VO, Data Access, Milky Way

1. INTRODUCTION

The VIALACTEA project goal is to bring to a common forum the major new-generation surveys of the Galactic Plane in the radio band, both in thermal continuum and in atomic and molecular lines, from Europe-funded space missions and ground-based facilities, to engage one of the fundamental challenges in Galactic astronomy: to quantify in our galaxy the relationship between the physical agents responsible for the onset and the regulation of star formation in a spiral galaxy and the resulting rate and efficiency of star formation, and obtain a "star formation recipe" that will be a cornerstone to trace the star formation history of galaxies back to their formation.

Among the work packages the project is composed of, one incorporates the activities needed to ensure that a computing infrastructure and a set of data-mining, machine-learning, visual analytics and 3D visualization tools are in place according to astronomers specifications.

The input data that the computational tasks should start from, as well as the output results the various tasks will provide form a knowledge base that is, in turn, the goal of one task of this latter work package.

This contribution focuses on this task, the creation and maintenance, as well as the access interfaces, of the VIALACTEA Knowledge Base (VLKB).

Section 2 will briefly describe the project work packages and tasks and give a more detailed view of the "tools and infrastructure" work package (Sec. 2.1), with specific attention to the VLKB (Sec. 2.1.1). Section 3 will then focus on the heterogeneous content of the storage (Sec. 3.1) available to the VIALACTEA community and

Further author information: (Send correspondence to Marco Molinaro)
Marco Molinaro: E-mail: molinaro@oats.inaf.it, Telephone: +39 040 3199 152

Software and Cyberinfrastructure for Astronomy IV, edited by Gianluca Chiozzi, Juan C. Guzman, Proc. of SPIE Vol. 9913, 99130H · © 2016 SPIE
CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2231674

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

the database (Sec. 3.2) that contains the storage metadata descriptions as well as catalogued data consumed or provided by the project's scientific tasks. Next section (Sec. 4) will then describe the search and access interfaces that live on top of the above data collections; Sec. 4.1 focuses on the generic TAP interface (Table Access Protocol[1] , an IVOA standard) that exposes the database content in an interoperable way while Sec. 4.2 describes in details the dedicated discovery and access interfaces that allow filtering of data collections and retrieval of the actual data files.

2. PROJECT OVERVIEW

The VIALACTEA project consists of four scientific work packages (WPs), a technical one (WP 5, that we will touch in more details in Sec. 2.1) and other packages dedicated to dissemination, management and coordination activities. Scientific packages' tasks span from diffuse and compact structure analysis (WP 1 & 2) to distance estimation (WP 3) and their outputs are then combined into a global scenario of the galaxy as a star formation engine within WP 4 tasks. All the scientific tasks need some data to start from and produce some output to be preserved:

- diffuse structure analysis produce catalogues of filamentary and bubble structure out of continuum radio surveys using morphological analysis and radiative transfer modeling;
- compact source analysis provides a catalogue of compact sources out of continuum surveys and combines it with catalogues at different wavelengths to have a multi band catalogue to fit against synthetic protocluster models (another project output);
- distance estimation uses kinematical analysis of existing radio spectroscopic surveys and project generated 3D extinction maps to produce distance estimation of compact sources or other diffuse regions of the galaxy plane.

All of these tasks output are then combined into WP4 tasks for the final scientific goals of the project.

The goal of WP 5, "Tools and Infrastructures" is to provide technical support in running and developing the tools and recipes needed by the scientific counterpart.

2.1 VIALACTEA Tools and Infrastructures

There are four main activities in this work package:

Database and Virtual Observatory Infrastructure: to have a common set of resources to be used for scientific input and collect the project outputs. This is what this paper is for, the main product of this activity being the setup and maintenance of the VLKB and its interfaces;

Data Mining Systems: to develop intelligent integrated systems directly supporting scientific decision making and situation awareness by dynamically integrating, correlating, fusing and analysing extremely large volumes of disparate data resources and streams;

3D visual analytics systems: allowing the astronomer to easily conduct research activities using advanced visual methods for multidimensional data and information visualization, real-time data interaction to carry out complex tasks for multi-criteria data/metadata queries for subsample selection and further analysis, or real-time control of data fitting to theoretical models;

Science Gateway: to enable workflows based on the analysis tools and recipes and run them through the WS-PGRADE/gUSE gateway system.

The tools and infrastructure developed through these activities should bring the researcher from the input data to the result in an easy reproducible way, moving, at least partially, the steps of data retrieval, decision making, visual inspection of the information and/or results and job runs from manual tasks to automated ones.

The VLKB acts in this scenario as the data resource repository with interoperable capabilities and custom access interfaces. As already mentioned, the resources collection will be reported in Sec. 3 and the interfaces on top of them in the subsequent Sec. 4.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

2.1.1 Database and Virtual Observatory Infrastructure

The task activity within which the VLKB has been developed was defined as the one to carry out the archival goals of the project as well as the data repository for all of the initial data resources needed by the scientific work packages.

The idea was to have a common resource to follow the VO guidelines for interoperability while paying specific attention to the project needs. One of the main issues in keeping the pace with the IVOA standards has been to develop a project on the most recent specifications the VO was discussing upon. This would have been of great help in the VIALACTEA scenario if some specific internal requirements were not in conflict. The result is that the VLKB and its interfaces are currently not completely developed using IVOA protocols, but are ready to be mapped onto them given some time resource (which may be tricky since VIALACTEA is at the end of its life cycle).

The actual result of this task activity (which will be better summarized in Sec. 5) is a resource composed of a set of heterogeneous data collections of observational data, in form of images or multi-dimensional datasets and catalogues, alongside database schemata and relations dedicated both to metadescription of the mentioned data collections and catalogues produced by the project work package activities.

The discovery, access and retrieval solutions on top of these common resource (the VLKB) is based on an IVOA TAP service for all of the database content that needs to be exposed to the community plus some dedicated search, cutout and merge solutions for the 2D and 3D datasets available through the project. This latter secures the datasets (which are a mix of public and private policy ones) and allows search also for project specific datasets, like the 3D extinction maps of the galaxy produced inside the project itself.

3. VLKB DATA COLLECTIONS

The data content of the VLKB comes in two main archival formats:

- storage FITS files;
- relational database schemata.

In this section we describe them both as are currently available (or planned by the end of the project) following the above distinction, instead of following an input vs. output categorization or a task-based one. Descriptions will take into account:

- dataset origin: from within the project or mirror/redistribution of an existing resource;
- resource weight (in terms of datasets, file sizes, catalogue records);

The interfaces to discovery, access and retrieval of the resources is left to the following section (Sec. 4).

3.1 Storage Content

The storage contains files in FITS[2] format only, but they're rather heterogeneous. Those FITS files, indeed, span from 2D images in the radio continuum to 3D FITS cubes containing radio velocity spectra at specific molecular lines and also a collection of 3D extinction maps.

But the number of dimensions and type of axis is not the only difference between the various data collections deployed. 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.

Most prominent, however, were the actual observational differences, that is, the various surveys, and observational bands within them, that have been used in the scientific analysis of the VIALACTEA project. Taking into accout what has been labeled as a *sub-survey* (i.e. a collection of data from a specific survey or pointed archive

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

referring to only one single molecular or band or other specific metadata), about 50 different data collections have been put together to be searched and accessed as a unique resource.

Apart from the Hi-Gal sub-surveys, the core resource from which most of the VIALACTEA primary products are derived, all of the other resources were retrieved from already public repositories or released to the VIALACTEA community but retaining their private data policy. That's the reason why access to the VLKB infrastructure has been secured and allowed only to the project's members.

Tables 1, 2 and 3 summarize the sub-surveys data collections reporting minimal figures for the various FITS file sets.

Table 1. VIALACTEA VLKB stored surveys consisting of FITS data 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

Hi-Gal data, the survey tiles listed in Table 2 as well all the data derived from them (e.g. filaments, bubbles and single band catalogue listed in Table 4) are private to the project. MOPRA, GRS, NANTEN, OGS have been granted for usage inside the VIALACTEA project, but are covered by privacy policy. Extinction Map 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 archive listed in Table 1 and 2, are publicly available starting from locations listed in Appendix A.

All of the above files, that sum up to about 1 TB and nearly 40 000 files, are of course only the archival resource that needs to be made discoverable, accessible and retrievable to the community to enable the workflows and analysis that are the goal of the project.

The first, obvious, step in this direction is the ingestion of the needed metadata in a database back end to allow faster discovery and endpoint retrieval for accessibility. This is part of the content of the next subsection.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Table 2. 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

Table 3. VIALACTEA VLKB stored data collections.

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

3.2 Database Content

Alongside the FITS data collections, a relational database (RDB) completes the VLKB resource content in terms of data information. This RDB (using a MySQL RDBMS) contains various types and blocks of information needed by the project. To mimic the logic underlying the project work packages subdivision, a set of schemata has been prepared to host the information.

Filaments & Bubbles. This database schema holds all the information related to the diffuse objects identified from Hi-Gal continuum tiles (for filaments) and Hi-Gal and CORNISH tiles (for bubbles).

Filamentary structures are described using a set of 3 tables, identifying *filaments* as the primary object, *branches* as the "linear" components within them and their *spines*, and *nodes* as the connection points of the various branch segments that compose a filament.

Bubbles are described as a unique catalogue of diffuse objectS, because their roundish shape doesn't require further relationships among their components.

Both filaments and bubbles tables are completed with positional and global details plus contour information. This latter is represented in two ways: an ordered sequence of sky positions (i.e. a polygon outlining the diffuse structure) and a reference to a FITS file in MOC format (Multi-Order Coverage Map,[3] an IVOA Recommendation), that is an HEALPix[4] tessellation of the diffuse object's area to be used for easier cross-match with other positional features in the VLKB or other databases.

Compact Sources. This schema contains all the single band catalogues used by the project to build up a *bandmerged* catalogue of compact sources. The primary input of the band merge task are the 5 single band catalogues derived from the Hi-Gal images through the CuTEx[5] tool (refined as a deliverable of the VIALACTEA project).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Alongside this catalogues, other have been used to build up a multi band catalogue of the galactic compact sources. The full single band catalogue listing includes: Hi-Gal (PACS and SPIR bands), ATLASGAL, BGPS, MIPSGAL, MSX and WISE (bands W3 and W4) catalogues for the galactic plane. The *bandmerged* catalogue is produced from these ones using data mining techniques included in the QFullTree tool delivered by the VIALACTEA project itself.

This schema includes also other content, not used in the band-merging effort of the project, but used in the SED scenario investigated by the project. One table is devoted to the grid of synthetic protocluster evolutionary models (based on Robitaille[6] models and delivered through the project); this sums up to 20 million SED records to synthesize energy distributions in the bands used by the band-merging effort. Another one is meant to keep track of the velocity information processed using the radio cubes FITS and database through the Peak Finder tool developed by VIALACTEA WP3 members to be later used in the distance estimation process of the various sources on the galactic plane.

Radio Cubes. All the metadata information about the stored FITS files and needed for the search and access interfaces is contained in this schema. This means, currently, all the sky frame boundaries needed for the discovery phase plus the actual location in the storage for data processing (cutout and merge) and retrieval.

Alongside the datasets minimal descriptions this schema includes also a table with a the description of all the sub-surveys to let the consumer applications (currently the Visual Analytics one developed within the project) identify the relevant sub-survey for the user.

This is probably the schema that may see more changes in the last phases of the project, if resources allow to move to a better VO compliant scenario. In that case probably an ObsCore[7] table may take the role of part of the current FITS datasets meta-descriptions.

TAP_SCHEMA. A last schema is kept in the VLKB to describe all of the schemata, tables, columns and their keys and indexes to be deployed through the TAP interface. This is the schema that is part of the IVOA TAP specification and acts as a counterpart to the information schema that is built in the RDBMS systems. Every change in the exposed database architecture that the VLKB wants to make visible to the user as a description counterpart in this schema.

Table 4 summarizes the various catalogues ingested in the VLKB database part including the number of records. Not all of the database table details have been included, because this is highly dependent on the usage and, given the project is still ongoing, they will vary a lot.

As for the datasets, some of these are publicly available, other are covered by a privacy policy. Filaments and bubbles catalogues and SED models, being products of the project, are private until the end of the project itself. Hi-Gal derived data is part of the core of the project, and thus it is also covered by privacy policy. The other catalogues are public and available as listed in Appendix A.

Please note that, since the VIALACTEA project is not yet finished and delivered, the above figures and contents may not be exact, but anyway they give a good idea of the contents and goals of the VLKB.

4. VLKB INTERFACES

All the data and metadata contents described in Section 3 would be, of course, useless if not somehow discoverable and accessible. This is the goal of the interfaces that have been put on top of the database and storage system. The following subsections describe the two consuming interface systems that have been set up for the VIALACTEA community: a TAP interface (Sec. 4.1) and a set of dedicated interfaces (Sec. 4.2) to consume to stored FITS datasets.

4.1 TAP interface

The IVOA Table Access Protocol is a specification to allow generic table-sets to be deployed as VO resources. It aims at homogenizing database access in the astrophysical research field, adding specific metadata content for discoverability and interoperability among different resources.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Table 4. VIALACTEA VLKB catalogue figures (rounded). Please note that figures may be partial or may differ from the final VLKB version because work is still in progress on the project at the time of this writing.

Catalogue	**# records**
Filaments	30 K
Filament Branches	150 K
Bubbles	5 K
SED Models	20 M
WISE	28 M
ATLASGAL	10 K
BGPS	8.6 K
Hi-Gal (70 μm)	160 K
Hi-Gal (160 μm)	600 K
Hi-Gal (250 μm)	470 K
Hi-Gal (350 μm)	250 K
Hi-Gal (500 μm)	130 K
MIPSGAL	2.6 M
MSX	400 K

The choice to have a TAP interface on top of the VLKB database part was made to allow both positional searches (that could have been done using a Simple Cone Search[8] protocol) and generic searches through the fields of the ingested data, also considering cross correlations between multiple tables available through the service.

The TAP service has been deployed using the IA2 TAP implementation*. It runs inside a GlassFish JAVA EE web container and it has been developed on top of the openCADC† libraries by the Canadian Astronomy Data Center‡.

Since the data ingested in the database is partly public and partly private data, the access to the interface has been secured. The authentication mechanism is currently being re-shaped to allow single user access (as opposite to the current project membership credentials), also in view of the user space related to the output products of the cutout and merge services described hereafter.

The TAP service, in short, serves everything included in the filaments, bubbles, compact sources, sed_models and velocity peak store resource span and, in the future, it is intended to support also the radio cubes and images metadata. The visualization tool developed within the VIALACTEA project consumes this TAP service to provide the user with the relevant information to display alongside images and cubes.

4.2 Discovery, Cutout & Merge interfaces

While the TAP interface is useful for generic tabular access, it is less powerful (at least alone) to perform positional discovery and is not designed to provide direct access to data files (unless one is willing only to retrieve a full file).

However, the requirements of the project were also to be able to:

* IA2TAP is part of the IA2 Data Access Layer service generation tools: http://www.ia2.inaf.it/vo-services/vo-dance

†https://github.com/opencadc

‡http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

- identify what files' content overlap a region along a specific line of sight on the celestial sphere and described by a circle or a rectangular range around it;
- cutout the files on positional constraints and also on the velocity axis (if 3D cubes were investigated) to allow for efficient data transport on the net and work only on the relevant part of data;
- merge adjacent files (from the same sub-survey, i.e. no information post-processing during the merge phase) into a single image/cube based on positional and velocity bounds.

IVOA specifications are currently covering the second point (see SODA[§]), were in the phase of describing the first one when the VIALACTEA project was kicked-off (the SIAv2[9] protocol) and foresee the third one in the near future (probably a minor revision of SODA).

Thus, even if keeping an eye on what IVOA was providing, a custom solution was undertook to have the discovery and access interfaces for VIALACTEA in place. This custom solution allowed also to work directly in galactic coordinates, while in the IVOA interfaces equatorial coordinates are prioritized.

The developed interface architecture is now composed of three services that share the same parametric query solution for positions, bound descriptions, sub-survey selection and are able to search and cutout indifferently of the underlying FITS header metadata (at least as far the WCS[10,11] descriptors are correctly in place and understood by the AST[¶] libraries). Specifically, the headers of various sub-surveys are passed to AST[12] library which is used to perform region overlap, and boundary computations. The AST library is a C-library so the JNI technology is used to access C-code from Java domain. Most of the Java-code is performing interfacing. Essential work is done "below" JNI in C. We process the partial results from AST in C-domain and then return the final results to Java domain. So we cross the JNI once per request.

Similar approach is taken also for the merge service, where the used engine is Montage[‖]. The merging service internally adds 3D-datacubes which are the result of the cutout service.

Issues on header metadata content as been solved, as far as possible, through corrected information ingestion at database level, trying not to change the original retrieved FITS, especially in the case of mirroring pre- existing resources. Data content of the FITS files also has been kept to a minimum, untouched in the cutout service, at most regridded along the third axis in the merge service if the adjacent FITS had different binning on that axis.

The set of three services, secured like the TAP interface to prevent unauthorized access to private data, deploys to the project's members all the images, cubes and extinction maps summarized in Tables 1, 2 and 3. The VIALACTEA visualization tool as well as other task tools make use of these services endpoints to prepare the input data for their tasks.

5. PLANNED VLKB SETUP

The data resources described in Section 3 and the interfaces reported in the previous Section 4 build up what was the goal of the VIALACTEA Work Package 5 Task 1. Currently the development is in finalization, but all what as been reported is planned to be ready and consumable by the end of September 2016 (i.e. the end of the funded project). Figure 1 gives a quick view of the architecture of the full system.

Depending mainly on time resource availability, more efficient discovery and access solutions, based on a modular and distributed architecture, may be set in place and a better VO compliance can be achieved.

One major point, not detailed in this paper, but planned to be ready by the end of the project, is the improvement of positional searches with the usage of HEALPix tessellation, in particular a dedicated service should be set up to allow direct, quick, cross match on positions between diffuse objects (filaments and bubbles) and compact sources. Given the large amount of records in the DB, this would be a welcome addition to the available interfaces.

[§]SODA Proposed Recommendation http://www.ivoa.net/documents/SODA/index.html
[¶]http://starlink.eao.hawaii.edu/starlink/AST
[‖]http://montage.ipac.caltech.edu/

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

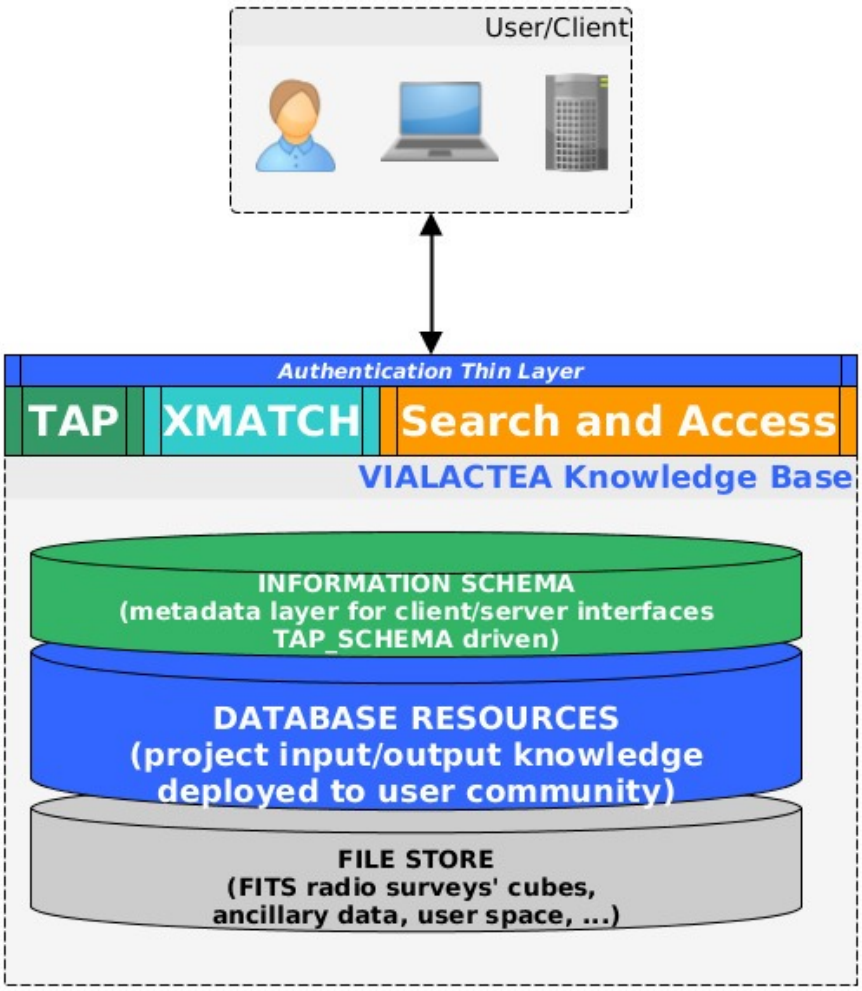

Figure 1. VLKB content and interfaces overview. Shown, on the interfaces layer, the positional cross match service based on HEALPix tessellation, not yet fully developed.

Also planned is the change from a membership-wide user to a single-based user authentication. This as the goal also to help maintain the staging area for the cutout and merged FITS outputs of the access interfaces, that otherwise may grow indefinitely and with no choice, for the user, to retrieve a product already calculated.

6. CONCLUSIONS

Building a data infrastructure to cover large part of the radio data for the galactic plane has been quite a big challenge, especially if undertaken with minimal resources and trying to cope with the short schedule of a three-year project with frequently updated requirements.

In this view the VLKB development has profited from the knowledge and experience available at the INAF IA2 data center, not last the VO experience that was highly useful in defining the solutions for discovery and access to the data, even if not directly available in terms of IVOA Recommendations.

One main critic that may be addressed to the solutions put in place is that large part of the data has been mirrored in a single location to be again deployed to the community. Indeed, initially the idea was to build as much as possible over existing resources and services; however most of the data re-used fell behind custom

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

interfaces, usually graphical web forms, that were difficult to access programmatically and attach to a common interface solution.

Considering that the VIALACTEA project is now running to its final deadline, it is difficult to foresee what improvements may be made to the VLKB, besides minimal maintenance; however, if further development and progress were to be foreseen, for sure developing a real VO compliant architecture will be definitely of a benefit to the community not to have future projects develop again mirrored interfaces to already existing resources.

APPENDIX A. PUBLIC DATA

Here follows the listing of all the public data used in developing the VLKB content.

A.1 FITS cubes

Radio cubes surveys and pointed archives.

CHIMP: http://dx.doi.org/10.11570/16.0001

CHaMP: http://www.astro.ufl.edu/~peterb/research/champ/rbank/

HOPS: http://awalsh.ivec.org/hops/public/data_cubes.php

ThrUMMS: http://www.astro.ufl.edu/~peterb/research/thrumms/rbank/

JCMT-HARPS: http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/jcmt/

MALT90: http://atoa.atnf.csiro.au/MALT90

VGPS: http://www.ras.ucalgary.ca/VGPS/

CGPS: http://www.ras.ucalgary.ca/CGPS/products/

SPGS: http://www.atnf.csiro.au/research/HI/sgps/queryForm.html

A.2 FITS images

Continuum surveys.

CORNISH: http://cornish.leeds.ac.uk/public/index.php

MAGPIS: http://third.ucllnl.org/gps/

MIPSGAL: http://irsa.ipac.caltech.edu/data/SPITZER/MIPSGAL/

WISE: http://irsa.ipac.caltech.edu/Missions/wise.html

A.3 Catalogues

Single band catalogues used in band-merging.

WISE: http://irsa.ipac.caltech.edu/cgi-bin/Gator/nph-dd

ATLASGAL: http://atlasgal.mpifr-bonn.mpg.de/cgi-bin/ATLASGAL_DATABASE.cgi

BGPS: http://irsa.ipac.caltech.edu/cgi-bin/Gator/nph-scan?submit=Select&projshort=BOLOCAM

MIPSGAL: http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=J/AJ/149/64

MSX: http://irsa.ipac.caltech.edu/applications/MSX/MSX/mission.htm

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

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 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 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 Appendix A).

This research made use of Montage. It 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

[1] Dowler, P., Rixon, G., and Tody, D., "Table Access Protocol Version 1.0." IVOA Recommendation 27 March 2010 (Mar. 2010).

[2] Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., and Stobie, E., "Definition of the Flexible Image Transport System (FITS), version 3.0," *Astronomy and Astrophysics* **524**, A42 (Dec. 2010).

[3] Fernique, P., Boch, T., Donaldson, T., Durand, D., O'Mullane, W., Reinecke, M., and Taylor, M., "MOC - HEALPix Multi-Order Coverage map Version 1.0." IVOA Recommendation 02 June 2014 (June 2014).

[4] 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 (Apr. 2005).

[5] Molinari, S., Schisano, E., Faustini, F., Pestalozzi, M., di Giorgio, A. M., and Liu, S., "Source extraction and photometry for the far-infrared and sub-millimeter continuum in the presence of complex backgrounds," *Astronomy and Astrophysics* **530**, A133 (June 2011).

[6] Robitaille, T. P., Whitney, B. A., Indebetouw, R., Wood, K., and Denzmore, P., "Interpreting Spectral Energy Distributions from Young Stellar Objects. I. A Grid of 200,000 YSO Model SEDs," *Astrophysical Journal Suppl.* **167**, 256–285 (Dec. 2006).

[7] Tody, D., Micol, A., Durand, D., Louys, M., Bonnarel, F., Schade, D., Dowler, P., Michel, L., Salgado, J., Chilingarian, I., Rino, B., de Dios Santander, J., and Skoda, P., "Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.0." IVOA Recommendation 28 October 2011 (Oct. 2011).

[8] Plante, R., Williams, R., Hanisch, R., and Szalay, A., "Simple Cone Search Version 1.03." IVOA Recommendation 22 February 2008 (Feb. 2008).

[9] Dowler, P., Bonnarel, F., and Tody, D., "IVOA Simple Image Access Version 2.0." IVOA Recommendation 23 December 2015 (Dec. 2015).

[10] Calabretta, M. R. and Greisen, E. W., "Representations of celestial coordinates in FITS," *Astronomy and Astrophysics* **395**, 1077–1122 (Dec. 2002).

[11] Greisen, E. W., Calabretta, M. R., Valdes, F. G., and Allen, S. L., "Representations of spectral coordinates in FITS," *Astronomy and Astrophysics* **446**, 747–771 (Feb. 2006).

[12] Berry, D. S., Warren-Smith, R. F., and Jenness, T., "AST: A library for modelling and manipulating coordinate systems," *Astronomy and Computing* **15**, 33–49 (Apr. 2016).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 20 Jul 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use