

User requirements for an Optical interferometry database

October 2013

JMMC-SPE-3000-0001

Xavier Haubois

Abstract

We present a database project that aims to archive interferometric data using Virtual Observatory techniques. Here we introduce the key-concepts, the tools, the functioning and the implementation plan for this database. This document acts as a proposal to the community to design the most useful database before starting the technical implementation. We therefore invite all interested readers to send their ideas and comments on the project, especially about the use terms and the access policy.

Introduction

The Virtual Observatory (VO) is opening up new ways of exploiting the huge amount of data provided by the ever-growing number of ground-based and space facilities. Specially created to develop software for the exploitation of interferometric facilities, the Jean Marie Mariotti Center (JMMC) aims at building VO tools to increase the diffusion and operability of interferometric data.

As a first part of a long-term database project, the JMMC is developing an OIFITS archive service and portal that allow users to query interferometric data (first-version OIFITS) obtained with a wide range of interferometric instruments. We want this database to be as complete as possible and easy to use for non-specialists in order to promote the use of interferometric data and develop collaborations. A proof of concept of the OIFITS archive is already available here: <http://apps.jmmc.fr/oidb>

The purpose of this document is to summarise the user requirements and some design aspects of the database. **Before starting the technical implementation, we require feedback from the future users and data providers of the database. This is the main goal of this document.** Once a first version is operational, we will invite PIs of interferometric instruments and interferometric data owners to link/upload the data they want to share in the database.

1. Description of the database

The database will store mainly OIFITS files (or urls for remote ones) for a wide distribution to the astronomical community. The data providers, hereafter named dataPIs (see below), fill a simple online form that links (or uploads) OIFITS data to the database and add contact and publication details. The database offers an archive service for uploaded files and these files will remain exactly as they have been deposited. Any user, interferometrist or not, will then be able to query OIFITS files in the database with different criteria (e.g. target, target type, observation date, instrument/instrument mode, wavelength range, etc). Several use cases are presented in the appendix. Information and graphics (e.g. UV coverage plots) will be available to allow her/him to identify the OIFITS files that are suitable for her/his research, to download them or to contact the team that owns these OIFITS files.

Our first priority will be to archive published interferometric data, followed at a later stage by other levels of data reduction. On Fig. 1 and Fig. 2., we depict the intended role of the archive service.

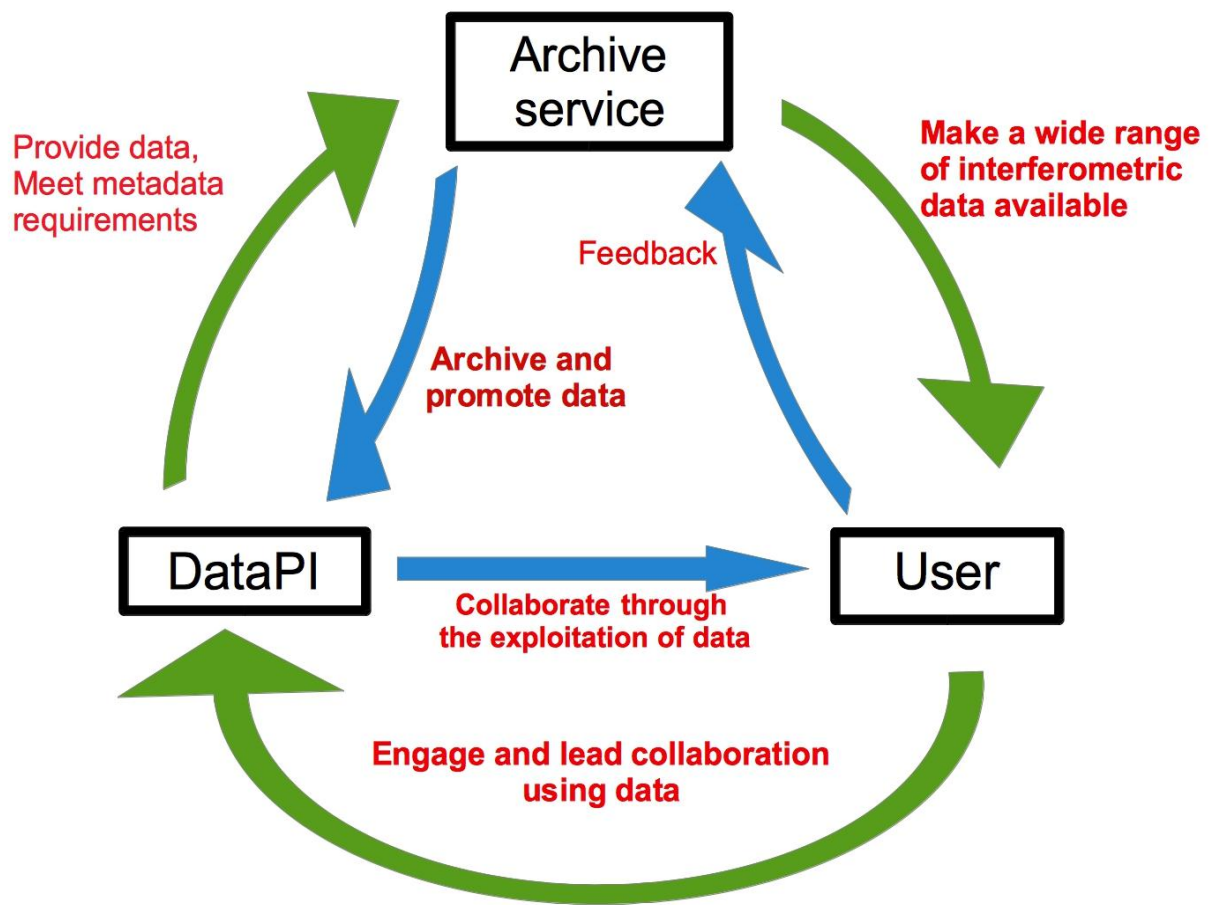


Figure 1: General diagram of the archive service

2. Key concepts

2.1 The dataPI

The dataPI is the person in charge of the scientific program that led to the observations and to the subsequent production of the calibrated data that will be shared via the database.

The dataPI is responsible for the data and metadata* she/he shares on the database.

* additional information describing the data included in the OIFITS files.

2.2 What's in the database ?

The database stores OIFITS files (data) and metadata. We define different levels of reduced data and store them in the database in the following way:

Table 1: Definition of the data reduction levels and their storage in the database

Data reduction level	Content	What is stored in the database	Access Policy
Level-0 - L0	Raw data + metadata	Metadata only	Public
Level-1 - L1	Reduced uncalibrated data + metadata	None	–
Level-2 - L2	Reduced, non-published, calibrated data + metadata	All	To be defined: restricted access during the proprietary period OR full public access ?
Level-3 - L3	Published data + metadata	All	Public
Level-3+ - L3+	Byproducts (spectra, fitting, reconstructed images)	All	Public

* The L0 metadata are useful to have an updated status of the existing data on a given object. The purpose of including them in the database is to allow the creation of collaborations as soon as the raw data are obtained. In this case, the interested user can directly contact the dataPI to enquire about the possibility of using these data.

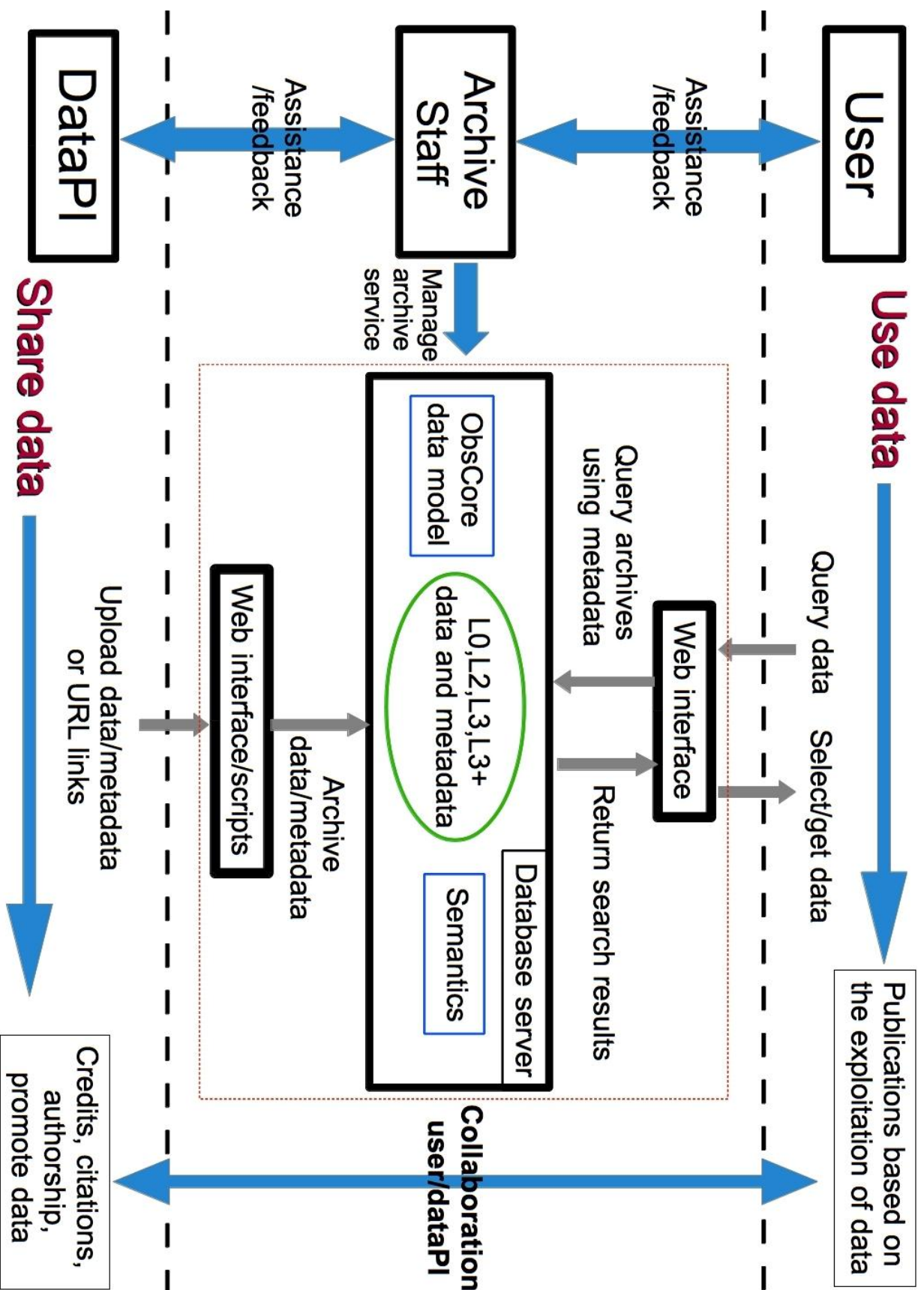


Figure 2: Functional diagram of the archive service

2.3 How to feed the database ?

For L0 metadata, under the responsibility of the **instrument PI**:

The database will provide a basic web interface to upload the L0 metadata by scripts from the client side. This protocol will be set up in collaboration with the archive staff and the person in charge of data maintenance for each instrument. This will allow L0 metadata to be automatically uploaded on a weekly basis and be visible to the database users a short time after the observation. As a first step, we will include the L0 metadata from the ESO database.

For L2/L3/L3+ data and metadata, under the responsibility of the **dataPI**:

The dataPIs link/upload the data via a secured web interface. According to the level of data reduction, they fill a different form. This procedure will be as easy/fast as possible. Most of the metadata are therefore automatically retrieved with routines that search for keywords in the OIFITS files and the number of fields to manually fill on the online form is limited (Table 2).

Once the first version will be operational, the next update will allow the database to be automatically fed by L2 data produced from robust data reduction pipelines. Thus, a few months after starting the technical implementation, we will automatically reduce all public PIONIER (and most probably AMBER as well) raw data and make their L2 (reduced) data outputs available on the database.

Table 2: Definition of metadata per level of data reduction

Data reduction level	Metadata input by the dataPI on the online form	Automatically retrieved metadata in the OIFITS data: oi_array + oi_target + oi_wave + oi_vis/oi_vis2/oi_t3
L0	Object, telescopes, number of telescopes + dataPI email address + seeing, t ₀ (optional)	—
L2	Quality flag, run or program ID, dataPI email address + seeing, t ₀ (optional)	Target, Dateobs, MJD, insName, effWave, telName, staName, staXYZ
L3/L3+	Paper reference (DOI), run or program ID number, dataPI email address	Target, Dateobs, MJD, insName, effWave, telName, staName, staXYZ

2.4 The life cycle of an archive

Once it is created, a data file follows an evolution cycle that includes processing, analysing, using, archiving, merging with other data, transformation, re-using, etc.

On the long-term, we want our database to be more than a simple repository space. We aim at enabling data discovery and retrieval on top of maintaining data and metadata quality along this evolution cycle. For example, we want to be able to show the L0 (raw) metadata and L2 data that were used to produce L3 (published) data.

In order to meet that goal, we want to adopt the concept of curation¹, which is a set of processes to assemble the information needed to access, manipulate, and manage data. This concept is an active topic within the International Virtual Observatory Alliance (IVOA) from which we can use

¹ <http://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaCP>

tools and processes.

2.5 Metadata quality

Metadata are critical to assess the relevance of archives. When possible, their content will therefore be checked and a quality criterion will be provided to the user (from 0 to 3).

Metadata quality criterion (for each archived OIFITS):

0- Some basic metadata are provided but are not necessarily consistent (i.e. compliant to the datamodel used in the database, see 3.1).

1- Essential metadata are provided and are consistent.

2- Metadata are provided in a complete and consistent way.

3- In addition to meeting the level 2 definition, metadata are cross-checked with independent sources of information.

When a file doesn't properly include the basic OIFITS metadata (fields of the oi_array, oi_target and oi_wave arrays), the searchability of this file on the database will not be optimal and consequently, the probability for this file to be used will be low.

2.6 Terms of use: agreement between users, dataPIs and the archive staff

All users, dataPIs and archive staff must agree to the following terms before using the database:

DataPIs:

- To the dataPI's knowledge, all data that are provided to the database are authentic and accurate in the sense that they really are the products of interferometric observations as described by their metadata.

Users:

- The user has a commitment to use the data for public astronomical research only.
- The user has a commitment to get in touch with the dataPI before presenting (in article, conferences) any work including the exploitation of his/her data.
- All publication based on the data taken from the database must give credits and citations in agreement with the dataPI and the database acknowledgement rules.

Archive staff:

- In the case where archives are uploaded to the database (not url linked) the archive service commits in making sure the data and metadata remain as they were uploaded.
- The archive staff commits in providing help to users and dataPIs to answer questions and fixing issues related to the database project.

2.7 Archive statistics

Statistical quantities will be provided to users and data providers as an additional information on the archives. For example, the server will keep track of the download counts for each archive and tell which archives are preferred by users. The download history will also allow to suggest alternative archives that were downloaded by previous users that might be relevant for the user's research. Finally, this service will help the archive staff to diagnose the way the database is being used and improve its functioning.

3. Technical Implementation

3.1 VO tools/Datamodel

Virtual Observatory (VO) concepts may help us to list and choose metadata format. Using common semantics will make the process of exposing our specific data through the VO much easier (no translation codes, standardised concepts). Such basic levels have been defined to describe common observations. ObsCore (<http://www.ivoa.net/documents/ObsCore/>) is one of the best examples. Table 3 presents the basic ObsCore mapping for our metadata on top of generic concepts.

Table 3: Parameters of the ObsCore model

<i>Column Name</i>	<i>Unit</i>	<i>Type</i>	<i>Description</i>	<i>_Optical_Interferometry_</i>
dataprodukt_type	unitless	string	Logical data product type (image etc.)	OpticalInterferometryData
o_calib_status	unitless	string	Calibrator or Science {absolute, relative, normalized, any}	OpticalInterferometryData
calib_level	unitless	enum integer	<u>ObsCore</u> Calibration level {0, 1, 2, 3}	Database provided, between 1 and 2
obs_collection	unitless	string	Name of the data collection	Database provided
obs_id	unitless	string	Observation ID	Database provided
obs_publisher_did	unitless	string	Dataset identifier given by the publisher	Database provided
access_url	unitless	string	URL used to access (download) dataset	OIFits url provided by database
access_format	unitless	string	File content format	Fits
access_estsize	kbyte	integer	Estimated size of dataset in kilo bytes	OIFits filesize
target_name	unitless	string	Astronomical object observed, if any	OI_TARGET.TARGET
s_ra	deg	double	Central right ascension, ICRS	OI_TARGET.Ra
s_dec	deg	double	Central declination, ICRS	OI_TARGET.Dec
s_fov	deg	double	Diameter (bounds) of the covered region	TBD
s_region	unitless	Astro Coord Area	Region covered as specified in STC or ADQL	TBD
s_resolution	arcsec	float	Spatial resolution of data as FWHM	TBD
t_min	d	double	Start time in MJD	Min OI_DATAs.MJD
t_max	d	double	Stop time in MJD	Max OI_DATAs.MJD
t_exptime	s	float	Total exposure time	OI_VIS2s.INT_TIME
t_resolution	s	float	Temporal resolution	TBD

			FWHM	
em_min	m	double	Start in spectral coordinates	Min OI_WAVELENGTHs .EFF_WAVE
em_max	m	double	Stop in spectral coordinates	Max OI_WAVELENGTHs .EFF_WAVE
em_res_power	unitless	double	Spectral resolving power	$\text{abs}((\text{EFF_WAVE}[1]-\text{EFF_WAVE}[0])/\text{EFF_WAVE}[0])$
o_ucd	unitless	string	UCD of observable (e.g. phot.flux.density)	Request one for optical interferometry data
pol_states	unitless	string	List of polarization states or NULL if not applicable	NULL
facility_name	unitless	string	Name of the facility used for this observation	Facility of the instrument
instrument_name	unitless	string	Name of the instrument used for this observation	One record per distinct OI_WAVELENGTH.INSNAME

This table and extended parameters will be detailed in the specification document of the database.

[3.2 Implementation plan](#)

The first version of the database will provide the following functions for the L3 data and L0 metadata:

- Query database and display its content through web pages
- Ingest L3 OIFITS through a web form (limited to one target/one instrument per file)
- Ingest L0 lists
- Handle L3 OIFITS with multiple target/multiple instrument

When completed,

- Accept L2 level and handle
 - by script to accept huge dataset provided by public programs (e.g. PIONIER and AMBER public data)
 - by web forms to enable collaborations with private OIFITS
- Improve curation and preservation of archives
- Improve metadata quality
- Accept L3+ level

3.3 Implementation examples: what the database could look like

A proof of concept has already been conceived and can be found here:

<http://apps.jmmc.fr/oidb/>

The VEGA team also developed their own database that you can see here:

<http://vegaobs-ws.oca.eu/>

Inspired from these examples, a first layout of the database web interface will be designed soon. However, here would be the main components of the query and search result panels:

Query panels :

- Target
- Target type (as defined in Simbad, e.g. Alf Cen: Double or multiple star)
- Date of Observation
- Instrument + Instrument mode
- Wavelength range
- Data reduction level (L0,L2,L3,L3+)
- DataPI name

Example of search results panels :

L3+ data:

ObjectName Identifiers	Observation Date / MJD	Instrument + instrument mode	Wavelength range (microns)	Run/Program ID	Number of telescopes	Telescopes configuration	Description of the archive	Special remarks	DOI	PI contact details
Betelgeuse** HD 39801	14/03/11 55634.123	PIONIER-POLA1	1.4-1.8	89.A-2954(B)	4	J0-H0-G1-K0*	Reconstructed image (with MIRA)	Quadratic regularisation, etc...	XXXX**	dataPI@institute.country

L3 data:

ObjectName Identifiers	Observation Date / MJD	Instrument + instrument mode	Wavelength range (microns)	Run/Program ID	Number of telescopes	Telescopes configuration	Special remarks	DOI	PI contact details
Betelgeuse** HD 39801	14/03/11 55634.123	PIONIER-POLA1	1.4-1.8	89.A-2954(B)	4	J0-H0-G1-K0*	----	XXXX***	dataPI@institute.country

L2 data:

ObjectName Identifiers	Observation Date / MJD	Instrument + instrument mode	Wavelength range (microns)	Run/Program ID	Number of telescopes	Telescopes configuration	Observation notes	Special remarks	Quality flag	PI contact details
Betelgeuse** HD 39801	14/03/11 55634.123	PIONIER-POLA1	1.4-1.8	89.A-2954(B)	4	J0-H0-G1-K0*	ro=1' to = 2ms, wind	----	5	dataPI@institute.country

L0 metadata:

ObjectName Identifiers	Observation Date / MJD	Instrument + instrument mode	Wavelength range (microns)	Number of telescopes	Telescopes configuration	Observation notes	DataPI contact details
Betelgeuse** HD 39801	14/03/11 55634.123	PIONIER-POLA1	1.4-1.8	4	J0-H0-G1-K0*	ro=1' to = 2ms	dataPI@institute.country

* Click for a UV-plan plot, ** Link to simbad, *** Hyperlink to ADS

Conclusion

We presented various technical aspects of the database as well as an implementation plan. Yet, we lack some feedback from the community to finalise the design. We would therefore like to get comments on the whole concept but also on these particular points:

- What do you think this database should be able to do on top of what we present in this document ?
- Do you have any requirement about the use terms: dataPI definition, ownership, licensing, etc ? About the data access policy ?
- For L2 data: do you think this database should enable private access to the data so that you can share data with the people you choose during a proprietary period ?

What's next:

- 1) After some feedback and iterations, we will finalize the design of the database and write a technical specification document.
- 2) In the coming weeks, an IT engineer will be hired at JMMC in Grenoble and will start the technical implementation following these specifications.
- 3) The first version, news and updates of the database will be released via the olbin.vo mailing list (see below).

For more details about this project :

- The twiki site of the project:

<http://www.jmmc.fr/twiki/bin/view/Jmmc/VirtualObservatoryForOpticalInterferometry>

- The IT engineer job offer (in French):

http://www.jmmc.fr/job_offers.htm

- The mailing list olbin.vo:

<https://listes.ujf-grenoble.fr/sympa/info/olbin.vo>

Appendix

A.1 Why such a database is useful: use cases

The first goal of this project is to increase the use of interferometric data and create collaborations with non-interferometry specialists. A number of different scenarios can be thought of. The database could be used in the following examples:

- 1) Query all published interferometric measurements obtained on Betelgeuse in H-band
→ The user fills in the following fields on the query form:
Target: Betelgeuse
Target type: Semi-regular pulsating star
Date of Observation: Any
Instrument: Any
Instrument mode: Any
Wavelength range: H-band
DataPI name: Any
Data reduction level: L3

- 2) Query reduced AMBER LR measurements of a list of targets observed between 03/02/2012 and 13/02/2012
→ The user fills in the following fields on the query form:
Target: browse a list of targets (with identifiers valid for SIMBAD)
Target type: Any
Date of Observation: 03/02/2012 - 13/02/2012
Instrument: AMBER
Instrument mode: LR
Wavelength range: J-H-K
DataPI name: Any
Data reduction level: L2

- 3) Query MIDI measurements in N-band for objects located 1' around (RA,DEC)=(02 42 40.771, -00 00 47.84) whose dataPI is Gilles Duvert
→ The user fills in the following fields on the query form:
Target: **02 42 40.771 -00 00 47.84 1'**
Target type: Any
Date of Observation: Any
Instrument: MIDI
Instrument mode: Any
Wavelength range: N-band
DataPI name: Gilles Duvert
Data reduction level: Any

- 4) Query recent non-reduced interferometric VEGA measurements made on Delta Sco
→ The user fills in the following fields on the query form:
Target: Delta Sco
Target type: Be star
Date of Observation : 1 month ago-Today
Instrument : VEGA
Instrument mode : Any
Wavelength range: Any
DataPI name: Any
Data reduction level: L0