SeaBIRD - A VIRTIS-VEX Data repository

R. Politi and G. Piccioni

INAF-IASF Rome (romolo.politi@iasf-roma.inaf.it / Fax: +39-06-49934188)

Abstract

VIRTIS (Visible and Infrared Thermal Imaging Spectrometer) is an experiment on board the ESA mission Venus Express. It has already performed more than 1600 orbits around Venus, collecting billion of spectra for a data amount of about 500 GB of raw data. In order to select and manipulate this amount of data we have developed SeaBIRD, Search-able and Brows-able Infrastructure for the Repository of Data. SeaBIRD is an hardware and software infrastructure with a web main interface able to select the data of interest at pixel level.

1. Introduction

VIRTIS-VEX is composed by two channels: VIRTIS-M that is a mapping spectrometer with a moderate resolution (R $\sim$ 200) and an high spatial resolution of 0.25 mrad, which uses two detectors a CCD (0.25÷1 μm), known as VIRTIS-M Vis, and an IR FPA (1÷5 μm), known as VIRTIS-M IR; VIRTIS-H that is an echelle high resolution spectrometer (R $\sim$ 1200) using an IR FPA detector (2÷5 μm) [1]. The data collected from VIRTIS are provided in PDS format [2] (attached label), one file for the data and one file for the geometric parameter of the observation. The current VIRTIS data amount of raw data for each channel, updated to May 2010, is shown in Table 1.

Table 1: VIRTIS data amount (May 2010).

Channel	# of Files	Data Volume (GB)
VIRTIS-M Vis	8100	332
VIRTIS-M IR	4735	161
VIRTIS-H	5316	10
Total	18151	503

For the channel VIRTIS-M there are two the usual image resolutions: 250x250 pixels for the high and 64X64 for the low resolution and each pixel is a spectra with 864 bands. It is easy to calculate that the spectra collect by VIRTIS are in the range of billions. For

this reason is really important the develop of an infrastructure able to select the VIRTIS files on which we are interested for the study purpose.

2 SeaBIRD

SeaBIRD (Search-able and Brows-able Infrastructure for the Repository of Data) is an infrastructure for the data distribution. It provides not only the VIRTIS files, performing a selection on the PDS label parameters, but also part of them, performing a selection on the geometry parameters. So we can describe SeaBIRD as a pixel oriented repository. The core of the infrastructure is a MySQL® archive in which we describe each pixel of the VIRTIS data. The pixel represents a single point in our data space, what we call as "atomic space". In this space, the pixels are clustered in some dimensions, so that we defined a new space in which a single dimension, defined by a suitable surjective mapping, takes the place of the clustered dimensions, defining a "molecular space". From this description we can mapping the data using an Entity Relationship (ER) diagram [3] that describes the atomic entity and the relation with its property. In a second step, starting from the ER diagram, we developed the extended entity relationship (EER) diagram, that represent the structure of the relational database [4] (we do not report the diagrams because they will not be readable).

3. SeaBIRD Input and Output

SeaBird is designed in order to be as much as flexible as possible and is optimized to provide the answer to most of query in less then 10 minutes. We chose the limit of 10 minutes because, in our idea, SeaBIRD must be an instrument for the on-line computation, that use solid and well tested procedures for the data access and manipulation. If the latent period would be longer than 10 minutes we could not call it "on fly". Any way the user is able to construct really complex queries that, in some cases, need a long processing time. There are many ways to submit the query to the infrastructure: using the web interface, that we have de-

veloped in DHTML (See Figure 1) in which are reported the most important parameter of the data and provide, also, the query in MySQL® syntax; using an XML file submitted via FTP, witch template is downlodable from the web site; using a user software in any language that have a MySQL® API. The system will

Figure 1: The SeaBIRD web interface.

return the answer printing it on screen as ASCII table, or via file, in CVS or PDS format, directly downlodable, if the elaboration time is shorter then 10 minutes, or via link from the server, if the elaboration time is longer than 10 minutes, which link will be provided via e-mail. As last way is the filling of a variable, but only if the request is done via MySQL® API. In Figure 2 we show a chart where we represent the input output schema of the infrastructure. In Figure 2 is
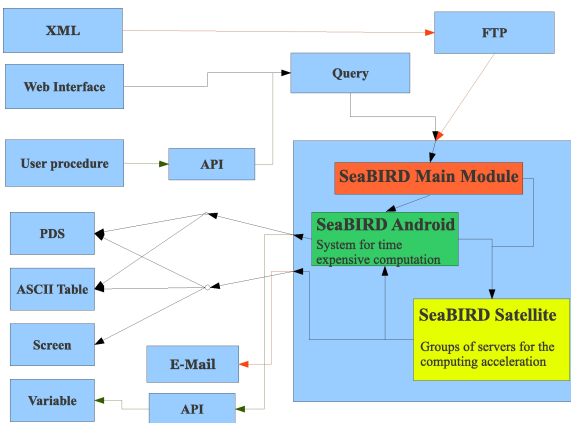

Figure 2: The SeaBIRD input-output schema.

also visible the main modules of SeaBIRD and their interaction. The **SeaBIRD Main Module** performs an analysis of the query for the validation of the syntax and of the the semantics, foresees the computation

time and routes the request to the SeaBIRD Satellite module for the direct computation or to the SeaBIRD Android if special outputs or if the computation time is greater then 10 minutes. The **SeaBIRD Satellite** is the module that perform the real search splitting the query on several servers, called "satellite servers", in order to decrease the computation time. **SeaBIRD Android** operates the queue if the computation time is longer than 10 minutes and sends the e-mail to the user at the end of the off line computation. The last module, not shown in the Figure 2, is the **SeaBIRD Palantir**. It is the module deputy to the data visualization and will manages the connections to the data providers of spectral or cartographic library.

4. Conclusions

At the present time SeaBIRD is in alpha testing in Rome. We have already head satisfactory results. All the test queries at images level have a latent period shorter than 5 minutes, while for queries at pixel level the 80 % have a latent period shorter than 10 minutes. The project schedule foresees the release of the beta version for summer 2010. The plug-in for the connection to external data repositories, for the laboratory data, are under testing, but each repository need of a different plug-in because of the absence of a standard. Also for this reason we are developing some plug-in for the integration in SeaBIRD of other data sources coming from space mission.

References

[1] Piccioni, G. et al VIRTIS: The Visible and Infrared Thermal Imaging Spectrometer, ESA-SP-1295, ESA Publications Division, Noordwijk, The Netherlands, 2007

[2] McMahon, S.K.: Overview of the Planetary Data System, Planetary and Space Science, Volume 44, Issue 1, Planetary data system, pp 3-12, 1996

[3] Chen, P. P.: The entity-relationship model—toward a unified view of data. ACM Trans. Database Syst. Vol. 1, Issue 1, pp 9-36, 1976

[4] Teorey, T.J., Yang, D, JFry, J.P.: A logical design methodology for relational databases using the extended entity-relationship model, ACM Computing Surveys (CSUR), v.18 n.2, pp.197-222,1986