SEABIRD: A FLEXIBLE AND INTUITIVE PLANETARY DATAMINING INFRASTRUCTURE. R. Politi[1], F. Capaccioni[1], M. Giardino[2], S. Fonte[1], M.T. Capria[1], D. Turrini[1], M.C. De Sanctis[1] and G. Piccioni[1], [1]INAF-IAPS Rome Italy (romolo.politi@iaps.inaf.it, Via Fosso del Cavaliere 100, I00133 Rome Italy), [2]ASI.

Introduction: Research activities on planetary science, from data analysis to theoretical modelling, need to deal with a large and extremely diversified amount of data. The heterogeneity of the data arises from the following factors:

- Data sources (Space missions, ground observations, laboratory experiments, numerical simulations);
- Data formats (PDS 3 and 4, Fits, ASCII tables, VO-compliant exchange protocols, etc.);
- Data types (spectra, images, iperspectral images, temperatures, etc.).

When searching for a specific target (e.g. specific planetary body or region) the difficulty of the task increases, because a deep knowledge is required to deal with different formats. The search for specific information in this sea of data is therefore becoming increasing complex and time expensive.

SeaBIRD: To simplify the data search process and to remove from the user the burden of handling the different formats and technical specifications of the data, we have developed the SeaBIRD (Searchable and Browsable Infrastructure for Repository of Data)[1] software and hardware infrastructure.

SeaBIRD Structure and implementation stategy: The SeaBIRD approach is the creation of a data space, called "atomic space", in which each single data unit is described by a metadata created by a semantic interpreter, starting from the original metadata, using a synonyms dictionary and a set of derivation rules. In this space, SeaBIRD clusters data by dimensions and defines a new "molecular space" where these dimensions, through suitable surjective mappings, are used to remap the clustered data. In the current release of SeaBIRD we translated this data description approach in a more complex and hierarchical data model generating a ORDBMS (Object Relational Database Management System) schema. Our implementation uses PostgreSQL with the GIS (Geographic Information System) extension.

SeaBIRD's Capabilities: SeaBIRD has been supplied with a web-based GUI (Graphical User Interface) to allow the users to easily search and access both the data and the associated metadata. SeaBIRD allows for two data mining approaches. On one hand, it allows the user to retrieve the original archived files containing the sought data ("*archive mining*" mode). In other hand, SeaBIRD can also provide the specific slice of data of interest for the user ("*information mining*" mode) repacked in a file with the desidered format (PDS, ASCII table/CSV, etc.).

The current version of SeaBIRD also allows to directly perform simple data manipulation tasks on the retrieved informations using a set of preimplemented primitive functions. These primitive functions can be combined by the user to create more complex workflows. The results of these manipulations can then be downloaded with the same approach used for the *information mining* mode.

All SeaBIRD's GUI features are developed using the newest Google™ visualization API to provide a powerful but inuitive interface.

SeaBIRD's API: We developed a SeaBIRD API (Application Programming Interface), whose interoperability software layer is based on the Django framework, to allow programs written in other languages to search and retrieve data from SeaBIRD using, either directly or indirectly, the HTTP (HyperText Transfer Protocol) protocol. The data are delived as JSON (JavaScript Object Notation) objects.

SeaBIRD's Data Readers: as a sub-product of its development process, SeaBIRD's data-reading modules are also being repacked into independent Python softwares to be released to the community. We currently released the VIRTIS-Venus Express reader, VIRTISpy[2].

SeaBIRD In Short: SeaBIRD main characteristics are:

1) easy and fast access to data;
2) personalized data slicing capabilities;
3) optimized data transfer volume;
4) solid and personalizable online data manipulation capabilities;
5) science-oriented interface that allows for abstracting data from archiving formats.

The current, fully-functional release of SeaBIRD is online and available to the community to explore the data provided by the intruments VIRTIS-Rosetta[3], VIRTIS-Venus Express[4] and VIR-Dawn[5], onboard the ESA missions Rosetta and Venus Express and the NASA mission Dawn respectively.

References:

[1] Politi R., Piccioni G. SeaBIRD - A VIRTIS-VEX Data repository, European Planetary Science Congress 2010,held 20-24 September in Rome, Italy, p.401. [2] https://github.com/VIRTIS-VEX/VIRTISpy [3] Coradini et al. VIRTIS: An Imaging Spectrometer for the Rosetta Mission (2007). Space Science Reviews, Volume 128, Issue 1-4, pp. 529-559. [4] Piccioni, G., et al., 2007. VIRTIS: The Visible and Infrared Thermal Imaging Spectrometer. ESA Special Publication, SP-1295, 1-27. [5] De Sanctis M.C., Coradini A., Ammannito E., Filacchione G., Capria M.T., Fonte S., Magni G., Barbis A., Bini A. and Dami M. The VIR Spectrometer, Space Science Reviews, Volume 163, Issue 1-4, p. 329-369