EPN2020-RI

EUROPLANET2020 Research Infrastructure

H2020-INFRAIA-2014-2015

Grant agreement no: 654208

Deliverable D11.8
SSHADE final infrastructure

Due date of deliverable: 31/01/2018
Actual submission date: 30/01/2018

Start date of project: 01 September 2015 Duration: 48 months

Responsible WP Leader: Observatory of Paris, Stephane Erard

Project funded by the European Union's Horizon 2020 research and innovation programme		
Dissemination level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

Project Number	654208
Project Title	EPN2020 - RI
Project Duration	48 months: 01 September 2015 – 30 August 2019

Deliverable Number	D11.8
Contractual Delivery date	31.01.2018
Actual delivery date	30.01.2018
Title of Deliverable	SSHADE final infrastructure
Contributing Work package (s)	WP 11
Dissemination level	Public
Author (s)	Bernard Schmitt, Philippe Bollard, Alexandre Garenne, Damien Albert

Abstract: SSHADE is a multi-database infrastructure in experimental solid spectroscopy for a VESPA sub-network of 21 European contributor groups (8 countries) + 2 non-European (IN, TW). It puts online a large set of state of the art spectral data of ices, minerals, organic material and extra-terrestrial materials for the astronomical/astrophysical user community allowing them to interpret observations of planets and small body surfaces, aerosols and grains. The SSHADE final interface (https://www.sshade.eu/) with its supporting relational Solid Spectroscopy Data Model (SSDM) has been developed in order to manage a large number of databases and data managers. The SSDM data model inherited from the GhoSST database (http://ghosst.osug.fr) was greatly enhanced in order to expand from the single database structure of GhoSST to the SSHADE multi-database infrastructure with multiple levels of data managements. The development of the infrastructure of SSHADE aimed to design a multi-database architecture and to build dedicated tools for data providers (data import and management, data search, and visualization) and users (data search, visualization and export). The technical choices required a development from scratch in order to provide a reliable and scalable infrastructure. The SSHADE infrastructure has been filled with fundamental data, which are common to all databases, and also with more than 1250 spectral data from 10 active databases. The ability of the SSHADE infrastructure to manage smoothly a large number of databases has thus been demonstrated.

Table of Contents

1 Introduction and aims

1.1 The SSHADE project

SSHADE (lead: IPAG, Consortium of 23 contributor partners: https://wiki.sshade.eu/doku.php?id=sshade:databases) is a multi-database infrastructure in experimental solid spectroscopy for a VESPA sub-network of 21 European contributor groups (+ 2 from India and Taiwan) from 21 different laboratories in 10 different countries. It is an extension from the GhoSST database (http://ghosst.osug.fr) and its data model (SSDM) defined in FP7 programs Europlanet-RI and VAMDC to a much larger service dedicated to a consortium of contributors in the field of experimental solid spectroscopy, including the major ones in Europe.

The **SSHADE databases**:

- can cover laboratory, field, airborne as well as simulated and theoretical spectral data
- is fed with their corresponding spectra and their various types of products (ex: transmission, absorbance, absorption coefficient, optical constants, band list)
- is defined for many different types of solids: ices, snows and molecular solids, minerals, rocks, inorganic solids, natural and synthetics organic matters, meteorites and other cosmomaterials, ...
- is compatible with a wide range of measurement technics: transmission, bidirectional reflection, Raman, fluorescence, microscopy, ...
- cover a wide range of wavelengths: from γ-rays to radio wavelengths

The resulting service not only puts online a large set of spectral databases of ices, minerals, organic material and extra-terrestrial materials, but also gives access to the state of the art laboratory data readily available to the astronomical/astrophysical user community (and well beyond: geology, glaciology, material sciences, optical design, chemistry, …) as references to interpret observations of planets and small bodies surfaces, aerosols and grains, in particular from spacecraft missions such as Mars Express, Cassini, Rosetta, Dawn, New Horizons, BepiColombo, ExoMars, JUICE, …

1.2 The SSHADE infrastructure

The developed SSHADE infrastructure includes tools for data import, search and visualization for the data providers (developed in the first year) in order to allow them to prepare, import and validate data (within VESPA VA WP). It also includes tools for spectral data search, visualization and export for the users (developed in the second year). SSHADE was developed to get the capacity to deal with a large number of databases and data managers (one or two per group).

To do so, a fully reorganized and extended data model, adapted to the specific needs of all partners, and beyond, had first to be developed starting from the one of the GhoSST database (initially designed to accept such an extension). Then on this data model a multi-database architecture complemented by a series of tools was designed

and implemented using the state of the art of available free softwares and technologies. Hardware at the "OSUG-Data Center" (OSUG-DC) of the "Observatoire des Sciences de l'Univers de Grenoble, CNRS-INSU/UGA" has been set in order to provide a flexible development platform and an efficient access for the partners and the users. Then the fundamental (common) data of the database infrastructure have been converted from GhoSST and greatly improved/extended, according to the data model changes, and imported in the SSHADE infrastructure. Finally a wide range of spectral data (over 1250 spectra as for 31st January) have been imported by the partners in 10 different databases in the SSHADE infrastructure. This allowed us to demonstrate the ability of the SSHADE database infrastructure to smoothly manage a large set of data within a large set of databases.

This phase of intensive data documentation, validation (to ensure consistency and data quality) and ingestion of spectral data in the SSHADE databases was performed by the consortium of data providers (SSHADE partners) within the VESPA-VA2 Work Package. Database training of about 60% of the ~35 database managers has already been done over 2016-2017 within VESPA-VA WP and will continue over 2018-2019.

2 SSDM data model

Authors	Development: Bernard Schmitt (manager), Philippe Bollard, Alexandre Garenne, Lydie Bonal

Abstract: This part presents the development of the SSDM data model needed to support the SSHADE infrastructure

The relational Solid Spectroscopy Data Model (SSDM, currently version 0.8.4a, over 880pp. of full description of all keywords) inherited from the GhoSST database (single database of IPAG) was greatly enhanced during these two past years in order to extrapolate it from the single database structure of GhoSST to the SSHADE multi-database infrastructure with multiple levels of data managements. Indeed, the common SSHADE data are managed by IPAG but the spectral and sample data of each database have their own database and scientific managers (see https://wiki.sshade.eu/doku.php?id=sshade:databases)

To adapt SSDM for this new structure many changes were necessary. In particular it was necessary to split the data in several categories and store them in different databases while keeping efficient relational links between all these data categories:

- common data for all databases (fundamental species, fundamental phases, astronomical objects, publications, common 'international' materials, ...)
- specific data of each database (description of database, laboratories, experimentalists and managers, instruments, laboratory specific materials, samples, experiments and spectra)
- a future common "band list" database (band list, bands and states, molecular and band parameters), to be upgraded from GhoSST.

A number of reorganizations and simplifications of SSDM were also implemented to make the data model (i) more adapted to the wide range of data produced by the consortium partners, and (ii) 'user friendly' for the database managers, and the users. Several modules and options were also added to extend SSDM/SSHADE capabilities in terms of types of sample materials (extra-terrestrial materials, micrometeorites, IDPs objects, etc.), measurement techniques (BRDF, microscopy, field measurements, …) and spectral ranges (gamma, X, ... mm and cm waves..).

This new version of SSDM is now fully implemented in the SSHADE databases structure.

Some points specifically developed or strongly improved:

- adaptation to multi-database infrastructure
  - data management rights
  - data validation steps
  - database and data managers identification
  - links between data classes
- improvements / simplifications
  - matters
  - sample processing
  - isotopes of fundamental species
  - …
- extensions
  - fundamental phases
  - extra-terrestrial matters (of all types)
  - micrometeorites and IDPs objects
  - spectral range (now from gamma to VHF)
  - management of data versions
  - ...

3 Database infrastructure

Authors	Development: Philippe Bollard, Damien Albert (manager), Bernard Schmitt

Abstract: This part presents the technical choices of software and hardware implementations and the majors steps of the SSHADE infrastructure development

During the first year, the major actions of the development of the SSHADE infrastructure were to design the multi-database architecture from the new SSDM data model and to build the tools dedicated to data providers (import, search, and visualization of data). The second year was dedicated to develop the tools for users (search, visualization and export of data) and to develop improved data search methods. The technical choices implied a development from scratch in order to provide a reliable and scalable infrastructure.

3.1 Technical choices

The technical base of the SSHADE infrastructure was selected in order to ease the VO integration with other software parts of the Europlanet project, particularly VESPA. Only state of the art and stable free softwares were selected.

PostgreSQL has been chosen as DBMS (database management system) regarding to its reliability and its advanced features.
The programming language Python has been chosen through the framework Pyramid and the ORM (Object-relational mapping) SQLAlchemy in order to develop the whole application around the database. It provides reliability and access to several scientific libraries that can be used for calculation and visualization.
Elasticsearch has been chosen as search engine in order to provide a 'Google-like' search tool based on an efficient data indexation.

3.2 Server infrastructure

The OSUG-DC platform provides two virtual servers for the SSHADE infrastructure.

A first VM hosts the "production" web-services:

- SSHADE interface at https://www.sshade.eu/
- SSHADE Wiki at https://wiki.sshade.eu/ (public, but with a restricted area for partners)
- SSHADE advanced prototype at https://pre.sshade.eu/ (access strictly restricted to SSHADE Consortium partners). Will be probably discontinued after official release of SSHADE.
- SSHADE blog at https://blog.sshade.eu/ (public, for SSHADE users and partners information). Will be probably discontinued when 'News' implemented in SSHADE.

- Cold Surface Spectroscopy Facility at https://cold-spectro.sshade.eu/ (public, data of this facility from TA2 of EPN2020-RI will be stored and made available in SSHADE)

A second VM is dedicated to development and hosts:

- A forge (for project management, tests and continuous integration) at https://forge.sshade.eu/ (access restricted to the development team). Will be probably discontinued and migrated on shared infrastructure of UGA (GriCAD).
- A development version of SSHADE at https://dev.sshade.eu/ (access restricted to the development team)

The OSUG-DC hosting infrastructure is based on three physical servers distributed over two buildings and connected each other by dedicated optical fibers. These servers are operated together through a redundant VMWare cluster providing service continuity and regular backups for all hosted virtual servers (VM). The VMWare cluster uses the "SUMMER" infrastructure provided by UGA for distributed storage and backups.

A command-line tool was developed to export a history of all operations done on SSHADE database (import, correction, UID rename and deletion, user accounts and access rights). These data and all related source files will be regularly copied on another server through a dedicated tool providing a way to reset SSHADE from scratch.

4 Data search, visualization and export for users

Authors	Development: Philippe Bollard, Damien Albert (manager), Bernard Schmitt, Alexandre Garenne

Abstract: This part presents the data search, visualization and export tools for the users.

The second year focused on the development of the GUIs for users: data search, visualization and export.

4.1 Data search for users

Data searches do not query directly the main database but use a dedicated search engine based on 'ElasticSearch' tool. Every kind of search uses a specific index containing a set of documents. A "document" is a custom hierarchical collection of inherent and dependent properties of a data selected from the main database. The 'Google-like' main search field acts as a 'full-text' operator on mostly all properties of a document providing an efficient way to search a complex combination of keywords.

A user can currently search either spectral data or publications through two distinct forms using a simple 'Google-style' search tool that he can complement with a number of specialized filters to refine the search. For the publication he can search according to the following topics: by reference, by content or by published spectrum. For the spectral data he can filter his search according to a series of topics: by experiment, by instrument parameters, by environment, by extra-terrestrial object, by sample, by composition and/or by publication.

Figure 1. User search page for 'Spectra' showing the different search options

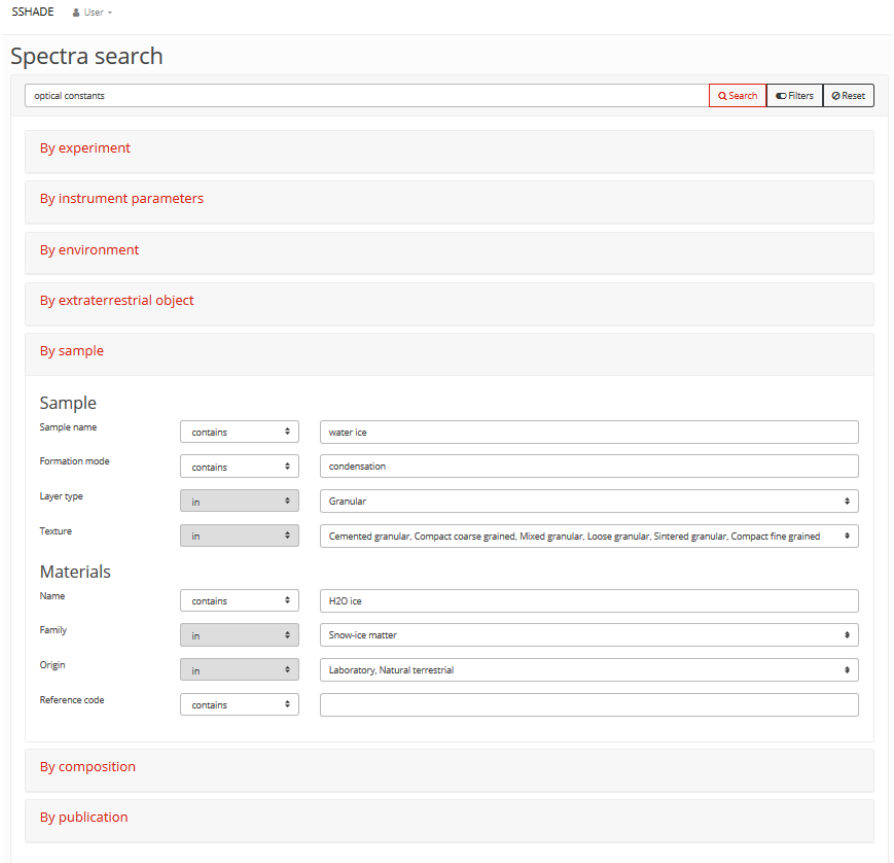

Figure 2. User search page for 'Spectra' showing the different filters for the sample search option

The results of the spectral search are displayed either as spectra or as experiments with an unfold button when several spectra of the experiment fit the search criteria. The user can then have a quick look at each spectrum with a preview popup, but he can also decide to immediately download a spectrum or send it to his shopping basket.

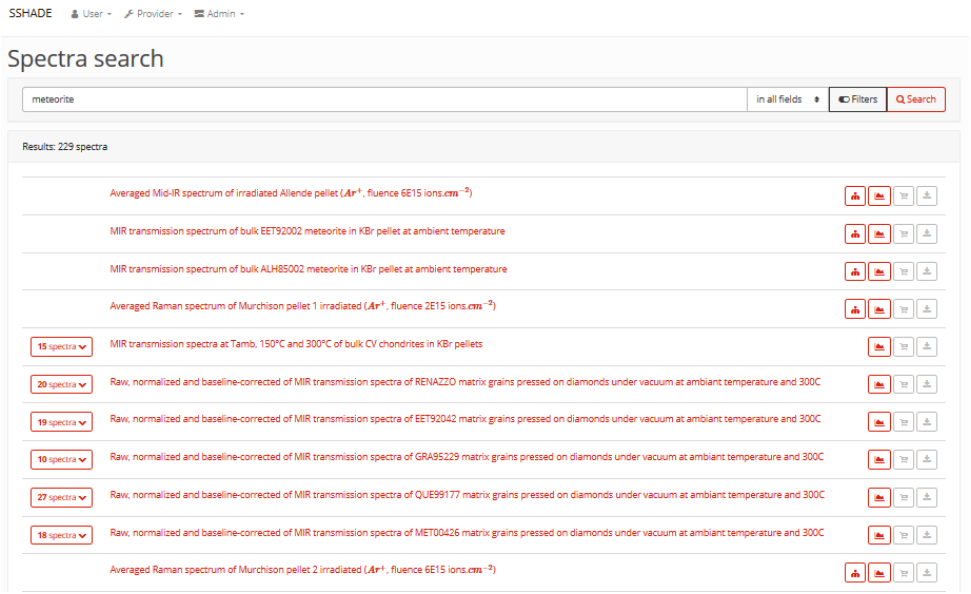

Figure 3. Example of results for a "spectra" search of 'meteorites'. The experiments have a left button with the number of relevant spectra they contain and to unfold them

4.2 Data display for users

The user can select and visualise a spectrum, he will then get a page with the collapsible structure of the experiment/(sub-experiments)/spectra on top of the left part of the page, and the whole collapsible structure of the sample/layer(s)/material(s)/constituent(s) on the bottom of this left part. While on the right part of the page a preview of the spectrum is displayed together with the main information on the spectrum and on the measured sample.

The user can then decide either to visualize the spectrum interactively together with all its associated information, or to look at the detailed information of the experiment or of any part of the sample structure. If the user decides to look at another spectrum of the experiment (possibly on another sample) then all displays of the page adjust accordingly.
The detailed page of each level of the experiment or sample structure contains all the relevant parameters values with different types of links either to another level of the structure, to other information stored in SSHADE (such as publications) or to external pages (such as Wikipedia, …). These links either create a popup window with the information (internal links) or open the information in another tab of the browser (external links).

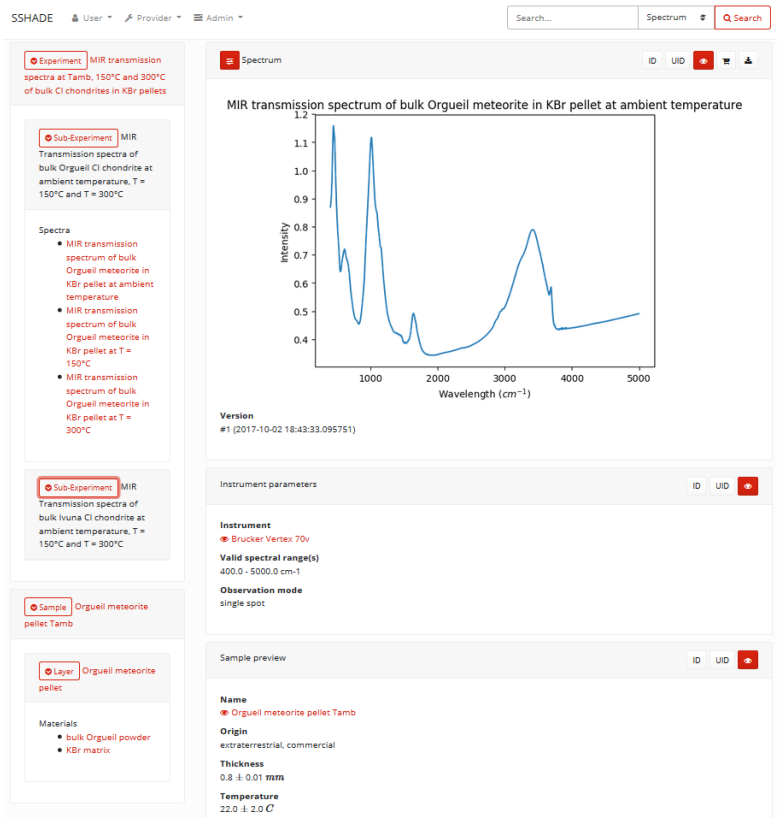

Figure 4. Collapsible structures of the experiment/(sub-experiments)/spectra and of the sample/layer(s)/material(s)/constituent(s) (left) and spectrum preview with its main spectral and sample information below

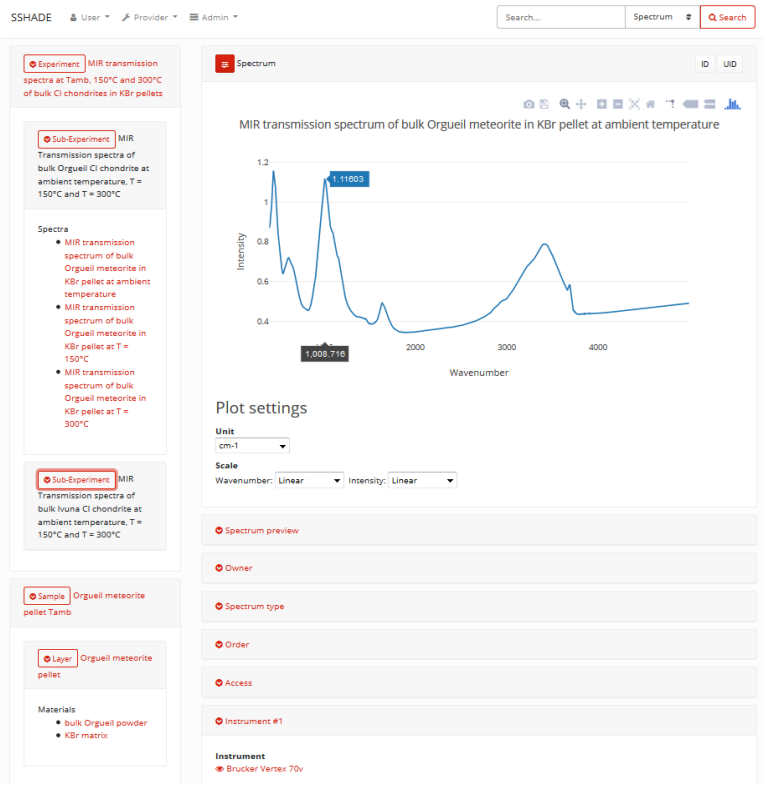

Figure 5. Example of display of a meteorite spectrum (dynamic), with the different categories of spectrum metadata below (left: experiment and sample structures).

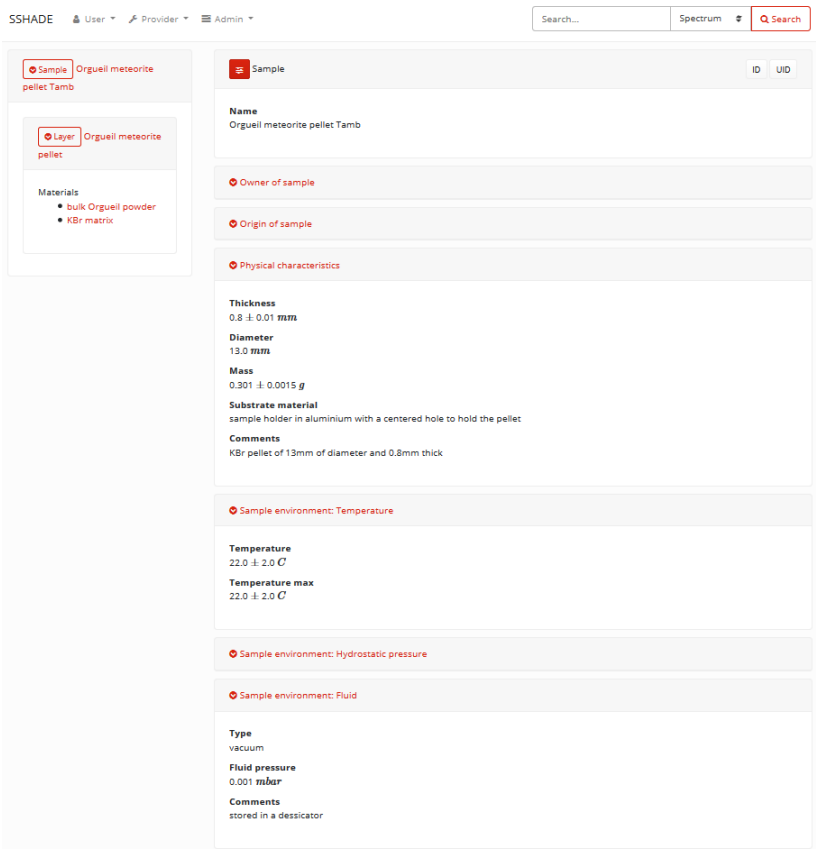

Figure 6. Example of display of part of the detailed information on the Orgueil meteorite sample

4.3 Data export for users

The users can download a spectrum or an experiment from the export page for immediate and individual download. The users may also add a spectrum or an experiment in the 'shopping basket' for future export.
Currently the export page allows to download (i) the spectral data as simple 2 or 3 columns (optical constants) ascii files; and the full metadata of the experiment/spectrum and of the sample in an html file that can be visualized on the user computer.

All data and metadata are delivered in a zip file that also contains a 'citation file' providing the DOI reference of the data, the reference of the paper(s) in which the data are published as well as guidelines on how to use and cite the data.

5 Development of tools for data providers

Authors	Development: Philippe Bollard, Damien Albert (manager), Bernard Schmitt, Alexandre Garenne

Abstract: This part presents the data ingestion tools, the data search and visualization tools and the BibTex conversion tool developed for the data providers.

5.1 Data ingestion tools

The main part of the development focused on data ingestion. The GUI for the ingestion of all types of fundamental (common) data and of spectral and associated data and metadata from data providers was developed and fully tested. This tool is more reliable and more accurate on data error reporting than its GhoSST equivalent. The import tool includes a validator that checks the correctness of the xml structure, the validity of the terms used in keywords with enumerations, the existence in the database of the linked data, the presence of "mandatory" data or optional blocks when specific conditions are satisfied, the format of the spectral files, … . On error, the validator stops the ingestion and provides guidance to solve the problem in the xml file without recording any wrong data into the database. An history is stored for every successful import and provides a way to download the imported files for possible later correction. Data unit conversions are operated on most numerical data (e.g. spectral data => 'cm^{-1}', temperature => 'K', …) during import in order to homogenise the content of the whole set of SSHADE databases.

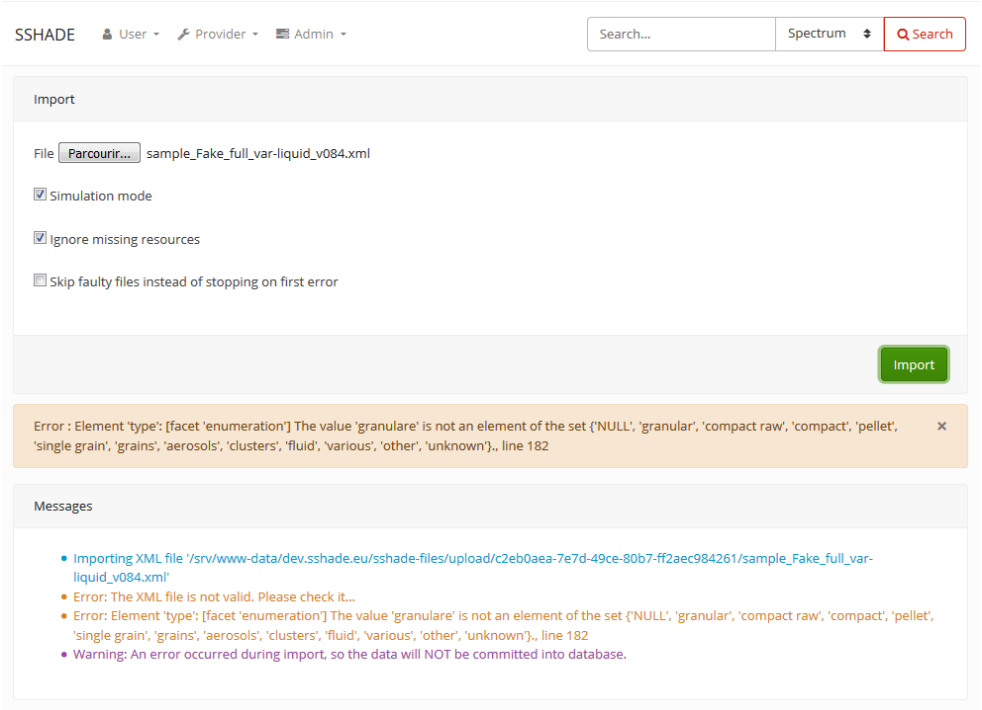

Figure 7. Import interface for providers, displaying a list of errors found in the xml import file after validation check, and guidance to correct them.

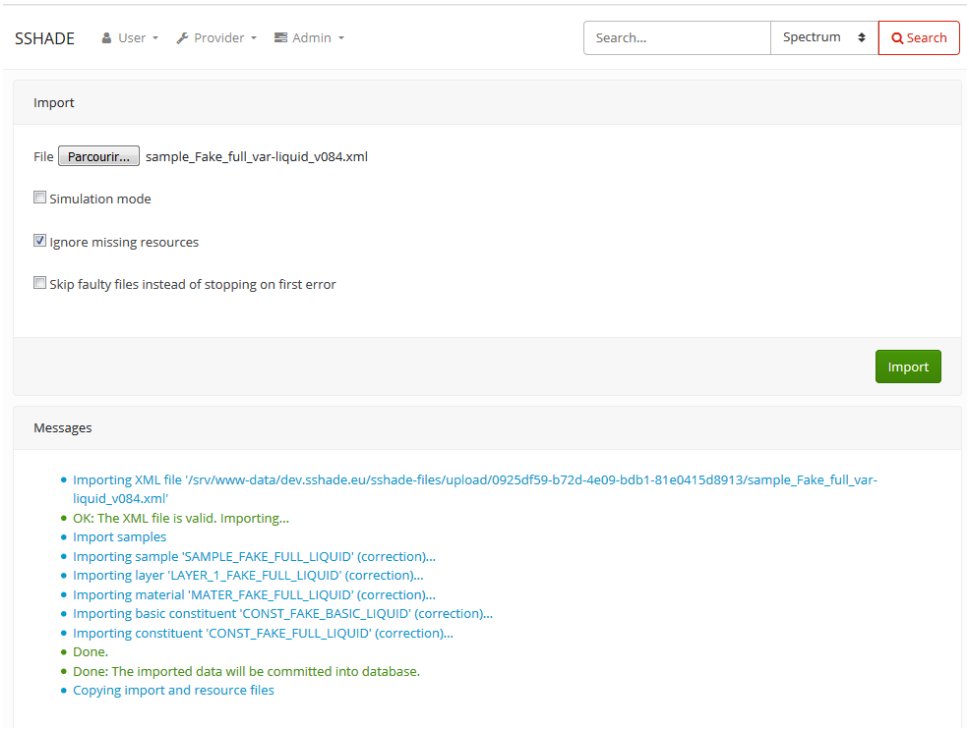

Figure 8. Import interface for providers, displaying the list of data import upon successful validation of the content of the xml import file

The import tool manages the right of data import and correction: a database manager can only import or correct data in the database(s) in which he get the rights to manage, while the SSHADE managers can import/correct all common and specific data.

A set of 11 specific import XML template files (https://wiki.sshade.eu/doku.php?id=sshade:provider:templates) are provided to the SSHADE managers and 10 others to the data providers to prepare and validate the different types of data (common and database-specific). They have been greatly improved (and simplified/homogenised) to provide to the data providers all necessary data information and import rules (complemented by the SSDM reference document and tutorials)

5.2 BibTeX conversion

A tool was developed in order to convert BibTeX files into XML "publication" files for semi-automatic import. It helps data providers to import their bibliography.

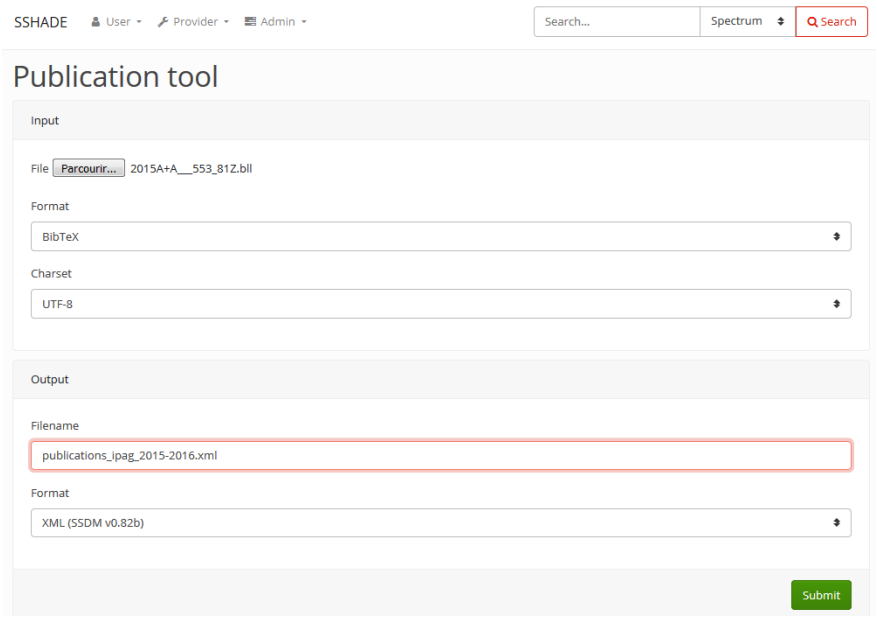

Figure 9. Interface of the conversion tool of publication references from BibTex format to the SSHADE 'publication' xml import file

5.3 Data search and visualization for providers

The GUIs for data providers, data search and visualization, were in fact developed before that for the users due to the need for the partners to be able to import data in SSHADE as early as possible. A data provider need to be able to search all kinds of data stored either in the SSHADE common database or in the individual partner databases in order to prepare his own sample and spectra data (need to link to them): fundamental species (atoms, ions, chemical bonds, chemical functions, and molecules) and phases (minerals, solid phases, liquid phases), database related information (database, laboratories, experimentalists and managers), spectral data (samples and their constituents, specific laboratory materials, instruments, experiments, spectra), related astronomical/geochemical information (meteorites, micrometeorites, IDPs), publications. For this, the data provider has access to several forms using a 'Google-style' search toolbar, using the same 'ElasticSearch' search engine as the data search for users, but it differs mainly on the properties selected to be queried and displayed. This toolbar is complemented with a large set of specialized filters that offer the data provider a powerful tool to explore the whole content of SSHADE. In particular the provider can filter his search either on its own database or over the whole set of databases, including the common data.

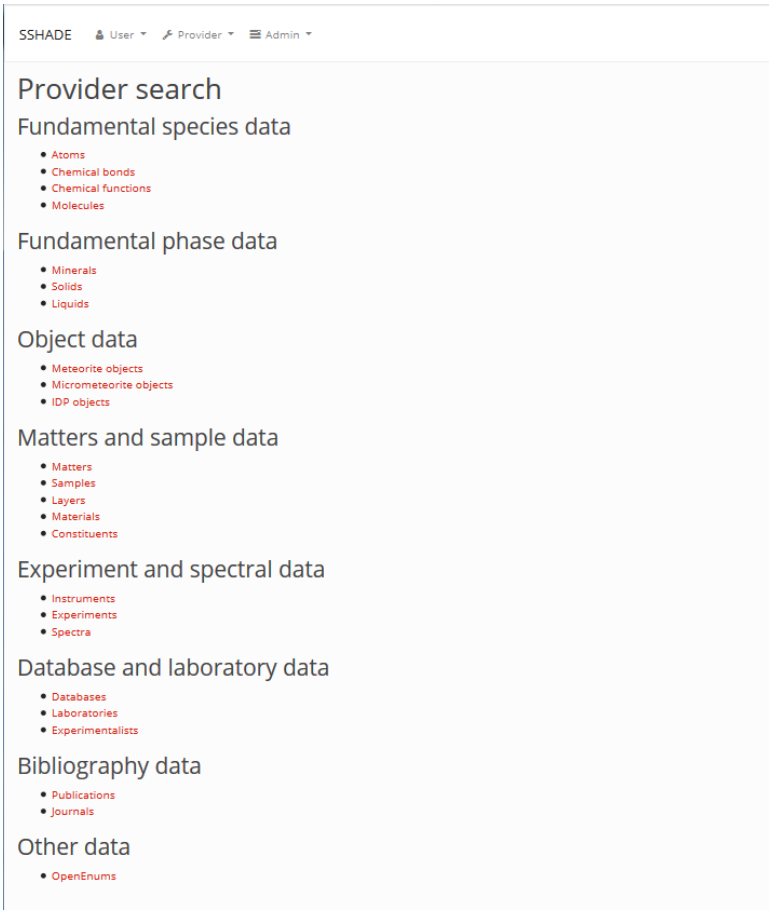

Figure 10. Search interface for providers giving access to all data stored in the SSHADE databases infrastructure

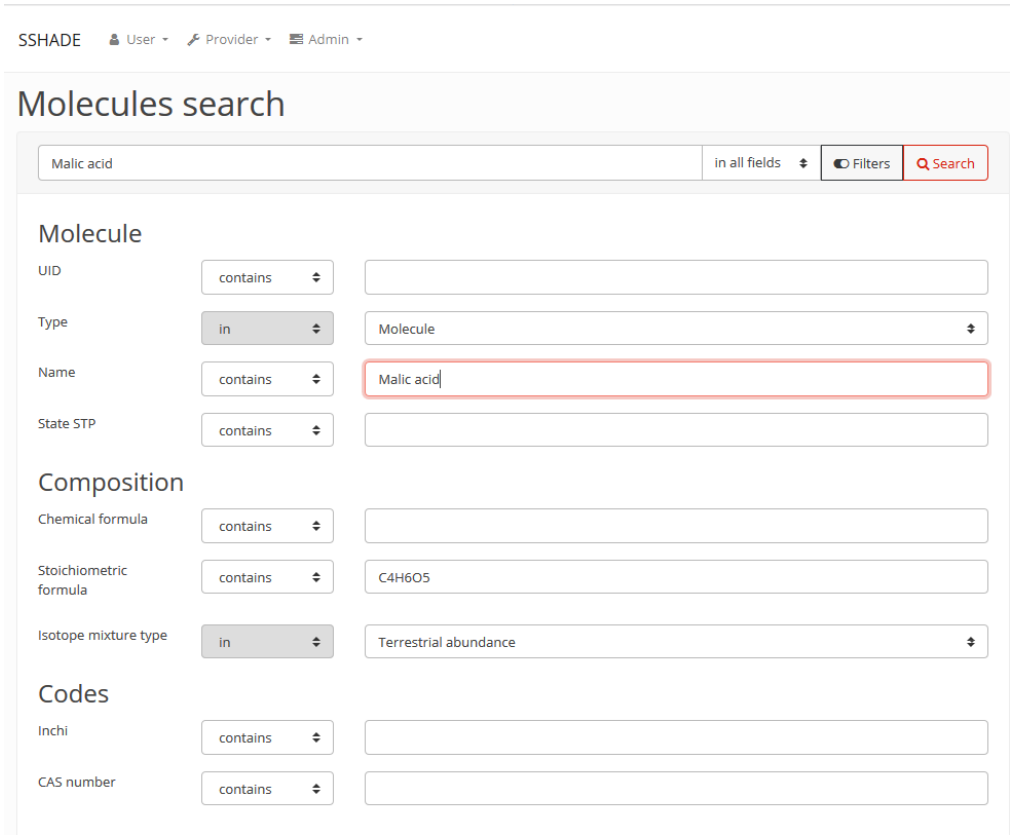

Figure 11. Search page for 'Molecules' showing the different search options

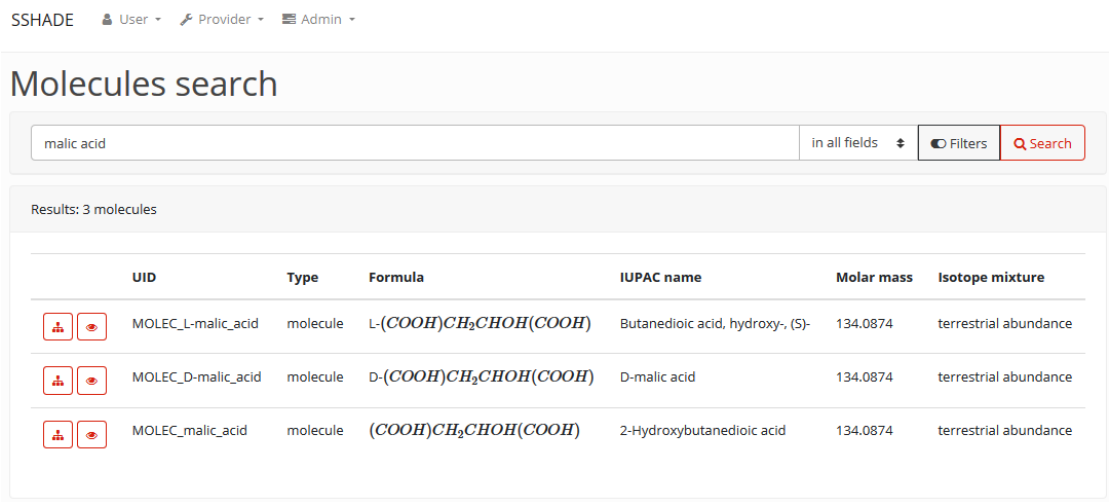

Figure 12. Result page for 'Molecule' search displaying the different results. Here it show the 2 different enantiomers of Malic acid and its racemic mixture.

All metadata can be visualized with all their relational links to other data (e.g. publications, …) and with external sources of information (e.g. Wikipedia, WebMineral, …).

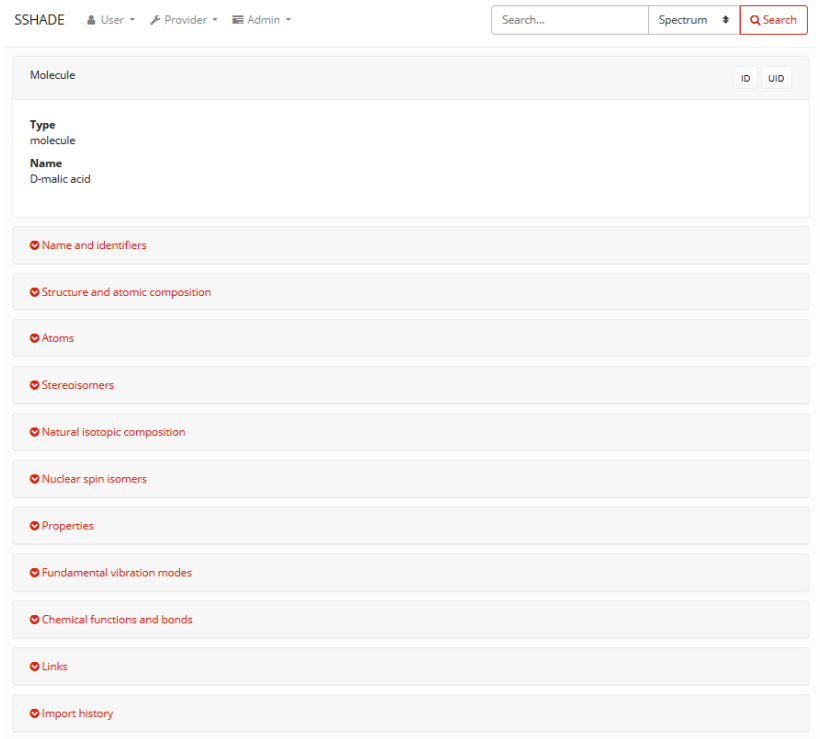

Figure 13. Information page for the selected 'Molecule' showing the different categories of metadata available

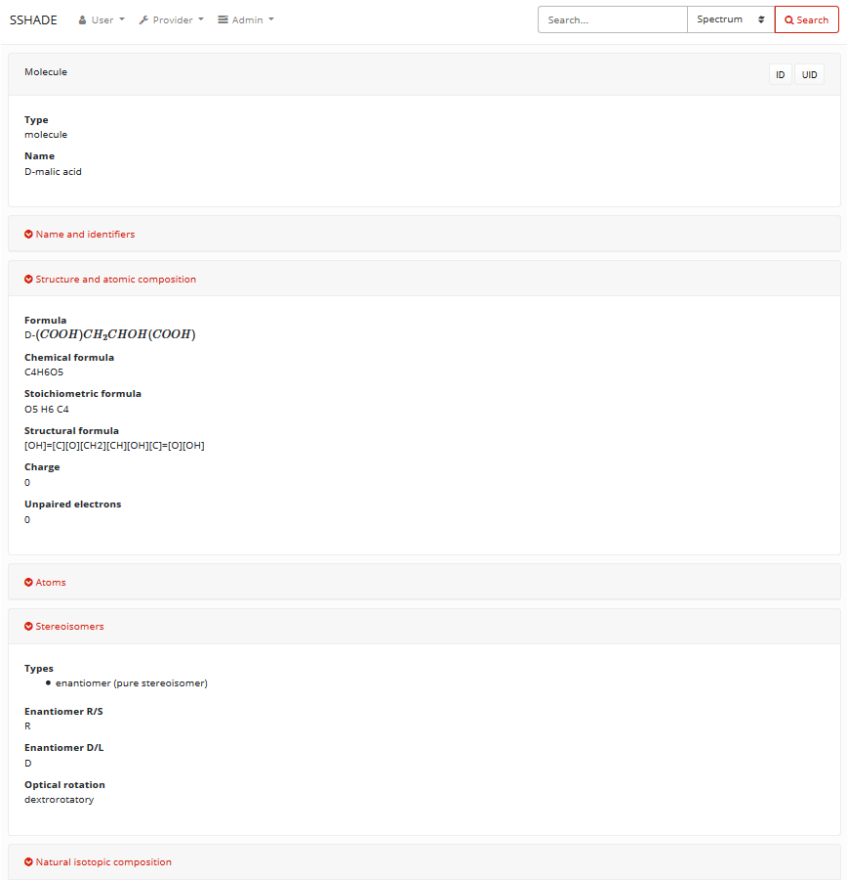

Figure 14. Information page for the selected 'Molecule' displaying information on two categories of metadata

6 Tools for the SSHADE administrators

The SSHADE administrators have access to a number of tools to manage the access rights to the databases. In particular, they can gives various management rights to the 'database manager' and the 'scientific manager' of each database. They can also manage the user accounts.
The SSHADE administrators have also a few special tools to manage the data, such as a 'data backup restoration tool' and a 'rename/delete UID' tool.

Some tools (partially public), will allow to record various standard statistics on the connections to SSHADE web pages, but also on the content of the database, on the new data imports and on the downloaded spectra. These statistics will be useful to monitor the activity of both the SSHADE partners and users.

7 SSHADE wiki

A SSHADE wiki (https://wiki.sshade.eu/) has been developed to provide a number of information and documentations for the users and the partners *(restricted access)* about SSHADE and its use.

It has 4 main sections: 2 dedicated to the users, and 2 to the providers.

For the users:

- 'SSHADE databases' provides the complete list of the active databases plus most of the future databases, with their description as well a number of information on the types of samples and spectra they currently provide in the SSHADE infrastructure and on the instruments and techniques used to record the spectra, and possibly the models used to analyse them.
- 'Interface documentation' provides a reference manual to the interface (which should be already quite intuitive), a data searching guide, and guidelines on how to use and cite SSHADE data. It will also host the future 'user tutorials' and a series of 'user cases', to be done within the VESPA VA activity.

For the partners (data providers):

- 'SSDM & SSHADE documentation' gives access to the last version of the Solid Spectroscopy Data Model (SSDM) as well as to the different presentations and papers on SSDM and SSHADE
- 'Provider documentation' (access restricted to partners) provides to the data providers a set of documentation and files to allow them to import data in SSHADE. In particular it contains the last version of all the templates of XML import files, as well as a number of tutorials on how to fill the XML files.

8 Databases filling

Authors	Bernard Schmitt (manager), Alexandre Garenne, Lydie Bonal and all the

	partners of the following databases and partners: BYPASS (Unibe), DAYSY (IAS), DOCCD (AIU Obs.), FAME (ESRF), GhoSST (IPAG), LSD (CML, IGS-PAS), PaSSTEL (IRAP), SOSYPOL (PGL, IGS-PAS), SSTONE (LPGN) and STOPCODA (IRAP)

Abstract: This part presents the current state of data ingestion in SSHADE both for the common fundamental data and for the database specific samples & spectral data.

The SSHADE infrastructure has been filled with both a large amount of fundamental data (over 1050 data), common for all databases, and with more than 1250 spectral data over 10 active databases.

8.1 Common data

Part of the common data are derived from the GhoSST database but needed a conversion (mostly manual!) to adapt their ingestion xml files to the new version of SSDM and to the new ingestion tool. But many other common data have been specifically prepared and ingested in SSHADE for the need of the partners to describe their samples in their databases.

Currently (as 31/01/18) the following common fundamental data have been ingested:

- Species: atoms and ions (145 data), chemical bonds (197), chemical functions (56), molecules (125).
- Phases: minerals (171), solids (70), liquids (9)
- Objects: meteorite (94), micrometeorites (11), IDPs (0)
- International matters (10)
- Publications (180)

These data continue to be prepared and ingested by the SSHADE manager team on request from the partners when a required fundamental specie, a phase, an object or a matter is not yet present in SSHADE. Nevertheless the partners contribute to the preparation and ingestion of their own publications (but carefully checked by the import parser to avoid duplication).

8.2 Partners data

Ten databases are already active in SSHADE (BYPASS, DAYSY, DOCCD, FAME, GhoSST, LSD, PaSSTEL, SOSYPOL, SSTONE and STOPCODA) and one is just starting (MIA), and **over 1250 spectra are already online** (from 182 experiments on about 830 samples), well over what we expected for the initial delivery of the SSHADE infrastructure.

The table below provides a snapshot of SSHADE data content by the 10 currently active databases (as 31st January 2018), in terms of local matters, samples, experiments and spectra.

Database	Matters	Samples	Experiments	Spectra
BYPASS	12	14	2	18
DAYSY	2	5	4	20
DOCCD	11	42	4	36
FAME	8	55	49	89
GhoSST	86	340	51	472
LSD	63	130	33	145
MIA*	0	1	1	1
PaSSTEL	15	12	13	18
SOSYPOL	2	73	16	274
SSTONE	6	146	6	146
STOPCODA	14	12	3	56
Total	**232**	**829**	**182**	**1275**

Table 1: snapshot of the content of the currently active databases (31/01/18)

Additionally several sets of data describing their database, laboratory(ies), instruments and experimentalists have been prepared and imported by each partner (208 data). The total number of data imported by the 10 partners exceeds 2700.

Note: The development of the SSHADE content will be reported in two future deliverables: 3rd VESPA Annual report (D6.8, August 2018) and 4th VESPA Annual report (D6.14, August 2019). We expect to reach over 3000 spectra from 20 databases at the end of the program and continue to grow in terms of number of databases and number of data in the following years.

8.3 Partners databases

Each partner database has one ‘scientific manager’ and one ‘database manager’ (which can be the same person for small databases). The first is mostly responsible of the scientific content of the database while the second is in charge of preparing, validating and importing the xml files containing the information and spectra.

The 10 already active databases already cover a wide range of samples, spectroscopic techniques and spectral ranges.

- various types of natural and synthetic samples (ices, minerals, rocks, meteorites, carbonaceous material, micrometeorites, prebiotic and bio molecules, …), including some made from 'international matters' such as JSC Mars 1.
- various types of spectral data from transmission spectra, reflectance spectra, to optical constants
- Various spectral ranges: X, Vis, NIR, MIR, FIR, sub-mm and mm.

The current contents of the 10 active databases (+ the one just starting to be fed: *) are described below. Their names link to their wiki page which provide more detailed information.

- BYPASS - **Bern icY Planetary Analogues Solid Spectroscopy**
  - Space Research and Planetary Science Division (WP-Unibe) / Planetary Imaging team - Physikalishes Intitut, University of Bern (Unibe), Bern, Switzerland
    - Scientific Manager: Antoine Pommerol
    - Data Base Manager: Olivier Poch

BYPASS is a database on the reflectance spectroscopy of analogues of planetary or small-bodies surfaces composed of particles of ices, pure or mixed with mineral and organic dusts. Surface samples of typical size and thickness of several centimetres are produced using methods allowing to control the size/shape of the icy grains and the way the refractory components are mixed with the ice. The database contains the reflectance spectra (from 0.4 to 2.5 µm or 4 µm) of these ice-dust surfaces placed under space simulated conditions (low temperature and pressure), including their temporal evolution as they are affected by metamorphism and/or sublimation. Most spectra were measured in the Laboratory for Outflow Studies of Sublimating icy materials (LOSSy) at the University of Bern using the hyperspectral imaging system SCITEAS. Some additional spectra of samples from LOSSy measured in partner laboratories are also included.

- DAYSY - **Database for AstrochemistrY and SpectroscopY**
  - Institut d'Astrophysique Spatiale (IAS) / Astrochimie et Origines team - CNRS / Université Paris-Sud, OSUPS, Orsay, France
    - Scientific Manager: Rosario Brunetto
    - Data Base Managers: Donia Baklouti

DAYSY is a database of spectroscopic characterization of solids of planetary and astrophysical interests. DAYSY covers a wide range of natural and synthetic solid samples: ices, organics and carbonaceous materials, and extra-terrestrial materials (meteorites, micro-meteorites & IDPs). Their spectroscopic properties are essentially measured either at IAS laboratory or at Soleil Synchrotron (SMIS beam line facilities), over a large wavelength range (Visible to Far-IR) with various types of spectroscopic techniques (transmission of thin films and processed thin deposits, bidirectional and diffuse reflectance of grains compressed in pellets and raw surfaces (slabs), microscopy of grains, Raman scattering, ellipsometry…), under controlled conditions and often, in a cell designed to reproduce some astrophysical conditions (vacuum, irradiation, low to high temperature, gas phase composition…). High level products such as optical constants, band lists and vibration mode parameters are also extracted through theoretical analysis and modelling of the data.

- DOCCD - **Database of Optical Constants for Cosmic Dust**
  - Astrophysikalisches Institut und Universitäts-Sternwarte (AIU Observatory) / Laboratory Astrophysics team - Friedrich Schiller University, Jena, Germany
    - Scientific Manager: Harald Mutschke
    - Data Base Manager: Jürgen Weiprecht

The **DOCCD** (Database of Optical Constants for Cosmic Dust) hosted at the AIU, Friedrich Schiller University Jena, Germany, contains complex refractive indices of solid materials relevant for cosmic dust (Silicates, Oxides, Sulphides, Carbonates, Carbides, Carbon materials ...). These data have been determined from specular reflection and transmission spectra of polished samples and thin films, measured in

various wavelength ranges from the UV to the far IR. The data ingested so far into SHHADE comprise optical constants of amorphous silicates and carbonaceous materials over wide wavelength ranges, as well as infrared optical constants of oxides at variable (elevated) temperatures.

- FAME - **French Absorption spectroscopy beamline in Material and Environmental science** (ESRF)
  - European Synchrotron Radiation Facility (ESRF) / FAME line team - Structure d'Exploitation des CRG, OSUG, Grenoble, France
    - Scientific Manager: Denis Testemale
    - Data Base Manager: Isabelle Kieffer

The **FAME** database is composed of transmission and fluorescence XAS spectra acquired at the two French spectroscopy beamlines of the ESRF synchrotron: FAME (BM30B) and FAME-UHD (BM16). It aims at gathering spectra of standards and characteristic samples provided by the beamlines users. In its current state, the database contains spectra collected on the different instruments of the beamlines, and in various experimental conditions, both in terms of sample nature (liquid/crystal/powder) and environment (pressure/temperature). Though still limited (~70 spectra, at Ag, Fe and Zr K edges), this dataset already encompasses most of the experimental conditions encountered at the FAME beamlines. It will be enriched soon with data collected with the crystal analyser spectrometer at the Ce L3 edge and with the spectra of the reference metal foils at all the edges accessible on the beamlines.

- GhoSST - **Grenoble Astrophysics and Planetology Solid Spectroscopy and Thermodynamics**
  - Institut de Planétologie et Astrophysique de Grenoble (IPAG) / Planetologie team - CNRS / Université Grenoble Alpes. OSUG, Grenoble, France
    - Scientific Manager: Bernard Schmitt
    - Data Base Managers: Lydie Bonal, Alexandre Garenne

GhoSST is a database on “spectroscopy of solids” of planetary and astrophysical interests covering a wide range of natural and synthetic solid samples (but also some liquids) with special focuses on low temperature ices & molecular solids, hydrated minerals, organics and carbonaceous materials, and cosmomaterials (meteorites, micro-meteorites & IDPs). Their spectroscopic properties are measured, either at IPAG or in international facilities, over a large wavelength range (Visible to Far-IR, but also XANES) with various types of spectroscopic techniques (transmission of thin films, crystals & grains in pellets, bidirectional reflectance & BRDF of granular or compact surfaces, microscopy of grains, Raman scattering & fluorescence emission of grains, …) mostly in a set of environmental cells designed to reproduce the planetary conditions (low to high temperature, vacuum, gas phase, adsorption, irradiation, …). Several high level products such as optical constants, band lists and vibration mode parameters are also provided through theoretical analysis and modelling of the data.

- LSD - **Layered Silicates Database**
  - Clay Minerals Laboratory (CML) - Institute of Geological Sciences, Polish Academy of Sciences (IGS-PAS), Kraków, Poland
    - Scientific Manager: Artur Kuligiewicz
    - Data Base Manager: Artur Kuligiewicz

The aim of Layered Silicates Database (**LSD**) is to present mid-infrared spectra of layered silicates and other related materials, with emphasis on minerals that are suspected to be present on Mars. As of December 2017, LSD contains 147 spectra of clay minerals, mainly smectites and chlorites. All samples have undergone careful purifying procedures prior to spectra collection in order to ensure the highest possible degree of purity. LSD contains spectra collected under varying temperature or relative humidity, which enables tracking of spectral changes occurring during progressive dehydration and dehydroxylation of clays minerals.

- MIA* - **Marseille Ice Analogues** [starting database]
  - Laboratoire de Physique des Interactions Ioniques et Moléculaires (PIIM) / Spectrométries et Dynamique Moléculaire - Astrochimie team - CNRS / Université Aix-Marseille, Marseille, France
    - Scientific Manager: Patrice Theulé
    - Data Base Manager: Patrice Theulé

MIA is a database specialized on interstellar ice analogues spectra. It mainly encompasses mid-infrared spectra of low-temperature molecular solids. The database will contain spectra of several solid-state complex organic molecules, commercially available or not, in either pure state or diluted into water, at different temperatures. Many of these complex molecules spectra results from reactivity studies, either thermal or photo-induced reactivity, from an initial mixture of reactants. The spectroscopic properties of these molecular solids are measured with Marseille based experimental facilities, using commercial Fourier transform infrared spectrometers (Bruker Vertex 7 and Tensor), under controlled conditions (high-vacuum, UV irradiation).

- PaSSTEL - **PlanetAry Surface Spectroscopy Toulouse Experimental Laboratory**
  - Institut de Recherche en Astrophysique et Planétologie (IRAP) / Planètes, Environnements et Plasmas Spatiaux team (PEPS) - CNRS / Université Paul Sabatier, OMP, Toulouse, France
    - Scientific Manager: Patrick Pinet
    - Data Base Manager: Yves Daydou

PaSSTEL database is aimed at characterizing the optical response of geological targets from 10 x 10 mm to 20 x 20 cm size, ranging from controlled powders to complex slab rock samples. These data are used for testing and improving spectral deconvolution and spectrophotometric modelling to better handle the behaviour of terrestrial and planetary surfaces (soils, regoliths, outcrops …) seen at different scales of analysis, ranging from orbital to in situ spatial scales. This approach supports interpretation of telescopic, spaceborne and/or in situ planetary observations. The database mainly consists of visible-near infrared bidirectional reflectance measurements performed on natural mineral solid solution samples, involving pyroxene and/or olivine minerals, with controlled grain size ranging from 45 µm to 2 mm, with six defined granulometric classes. Samples are derived from natural volcanic products with diverse contents of rock fragments, monocrystals and glass, various shapes and surface aspects. Samples can be homogeneous (e.g., olivine) or heterogeneous, consisting of complex mixtures (e.g., pyroclastics), with matrix contributions.

- SOSYPOL - **SOlar SYstem analogues database POLand**
  - Planetary Geology Laboratory (PGL) - Institute of Geological Sciences, Polish Academy of Sciences (IGS-PAS), Wroclaw, Poland
    - Scientific Manager: Joanna Gurgurewicz
    - Data Base Manager: Joanna Gurgurewicz

SOSYPOL is a database on spectra of analogue materials of Solar System solid body surfaces. The database currently contains: (1) Vis and NIR reflectance spectra of basalts altered in cold (Udokan, Siberia) and hot (Ogaden, Ethiopia) arid environments collected to help interpreting the spectra of the surface of Mars; (2) NIR reflectance spectra of several mixtures of ice and organic matter (tholins) at various temperature and pressure conditions to contribute interpreting the composition of the surface of Pluto and Charon. This dataset can be also used to determine the composition of aerosols in the atmosphere of Titan.

- SSTONE - **Solid Spectroscopy daTabase Of NantEs**
  - Laboratoire de Planétologie et Géodynamique de Nantes (LPG-Nantes) / Surfaces planétaires & Intérieurs planétaires teams - CNRS / Université de Nantes, OSUNA, Nantes, France
    - Scientific Manager: Marion Massé
    - Data Base Manager: Manuel Giraud

SSTONE is a database on "spectroscopy of solids and ices" of planetary interest. It uses various types of spectroscopic techniques: Visible and Near Infrared bidirectionnal reflectance of granular or compact surfaces, microscopy of grains, Raman scattering. Spectra are acquired directly on the field or in the laboratory in various environmental conditions: terrestrial or analogous planetary conditions (low to high temperature, low to high pressure, gas phase, …). These data are mainly used for improving the interpretation of orbital or in-situ planetary spectra on rocks, minerals and ices, but also for terrestrial data. The database currently contains: (1) NIR reflectance spectra of ice mixed with various concentration of sulphate and (2) NIR reflectance spectra sands wet with pure water (various origins and compositions) at different steps of drying. The database will contain soon: Raman spectra of volcanic glasses and icy moons analogous clathrates, NIR reflectance spectra of olivine altered in martian conditions and of pure water ice at different temperatures and grain sizes.

- STOPCODA - **SpecTroscopy and Optical Properties of COsmic Dust Analogues**
  - Institut de Recherche en Astrophysique et Planétologie (IRAP) / Milieu Interstellaire, Cycle de la Matière, Astro-Chimie team (MICMAC) - CNRS / Université Paul Sabatier, OMP, Toulouse, France
    - Scientific Manager: Karine Demyk
    - Data Base Manager: Yves Daydou

STOPCODA is a database gathering infrared spectra of cosmic dust analogues at low temperature. The spectra are recorded on the ESPOIR setup at IRAP and on the experimental setup of the AILES beam line at the synchrotron facility SOLEIL. The spectra are recorded in transmission mode, in the spectral range from 2 µm to 1 mm. The samples may be cool down from 300 K to 10 K. The samples are Mg and Fe rich silicate submicronic grains simulating interstellar and circumstellar dust. They are synthesized with solgel or quenching technics at LPCNO, UMET, ICEHT. The delivered data are mass absorption coefficients (MAC, g/cm^2).

9 Sustainability of the SSHADE infrastructure

The sustainability of the SSHADE infrastructure over the next decades is our main concern for the future and we worked strongly to make it a reality. We passed recently an important milestone towards this objective by obtaining the 'certification' of the 'SSHADE service' by the French CNRS/INSU in the category 'data processing, archiving and dissemination center'. This certification will ensure some support to the SSHADE infrastructure itself and also to all the French databases depending from an observatory. This certification may also lead to an additional CNAP position of 'associate astronomer' dedicating 30% of his working time to SSHADE. It may also help to get an additional engineer position at the OSUG Data Center, specialized in databases. This engineer will help maintain and hopefully continue to develop some tools for the SSHADE infrastructure.

In addition this certification engages OSUG-Data Center to maintain the SSHADE service operational in terms of hosting servers, backups, redundancy of operation and continuity of service. The new "SUMMER" infrastructure provided by UGA is also used for distributed storage and backups.

We are currently elaborating a Memorandum Of Understanding (MOU) of the SSHADE consortium where the respective roles and responsibilities of the managers of SSHADE, the OSUG Data Center and the managers of the databases of the consortium will be formalized. All scientific managers of the SSHADE consortium and their respective institutes will have to sign it before the end of the Europlanet-RI 2020 program.

After the end of the present program, extensions of the SSHADE infrastructure and consortium to other European groups providing laboratory or field spectral data of solids of planetary/astrophysical interest are anticipated, but will depend on funds available on future programs.

10 Deliverable

This deliverable is provided already with more than 1250 spectral data over 10 active databases, together with the associated SSHADE wiki, and the full set of XML templates for data preparation and ingestion.

- The SSHADE infrastructure is released at https://www.sshade.eu/
- Source code of the SSHADE interface is available at: https://dev.sshade.eu/sources/sshade_latest_develop.zip
- The SSHADE wiki is at https://wiki.sshade.eu/
- The XML templates for data ingestion are at https://wiki.sshade.eu/doku.php?id=sshade:provider:templates but are currently in access restricted to the SSHADE partners or upon request for external people.

The ‘provider area’ of the SSHADE infrastructure and of the SSHADE wiki, dedicated to the consortium partners, have restricted access. A temporary login/password with ‘data provider rights’ has been set. It allows to access the SSHADE provider menu with all the specialized search and BibTeX convertor tools for providers, but does not allow to import/correct data. In the wiki, it also provides access to the ‘Provider documentation’ section. It is strictly restricted to the European Commission service and the Europlanet Management Committee.

Note: the final version of the Solid Spectroscopy Data Model (SSDM) will be delivered in D6.10 deliverable “2nd set of standards documentation” (June 2019).

Contacts: Bernard Schmitt (bernard.schmitt@univ-grenoble-alpes.fr)

11 Future developments of the SSHADE infrastructure

As an infrastructure may continuously evolve along the data provider needs and the user expectations, we are committed to keep improving the current infrastructure and its tools. The following main hardware developments will be continued or undertaken after this JRA (on another budget) but they will also depend on the user feedback on the current interface and on the development of the databases content.

11.1 Future SSDM developments

- Some additions to better manage field and remote sensing experiments
- Improvement of import keyword of spectral BRDF data
- New complex data types (spectro-images, ...)
- New extra-terrestrial object categories: asteroids, comets, satellites, … (for Lunar/Hayabusa/Stardust… samples)
- Completion of SSDM upgrade for band list, molecular vibration mode and band parameters

SSDM will be delivered in D6.10: “2nd set of standards documentation”

11.2 Future SSHADE infrastructure developments

- Improvement of interface design and usability for better 'user-friendly" use.
- New search options by keyword for the users will be also added depending on the user needs and on the database content.
- Extension of the user export options in terms of data and file formats
- Development of various tools for logged users (user dashboard): search and download histories, email notification upon new data/new version import …
- Finalization of the pipeline tool for automatic experiment DOI generation and submission (through OSUG-DC)
- Implementation of a user interface (from DACE/PlanetS project) to visualize sets of bidirectional reflectance spectra acquired under varying illumination / observation conditions.
- Development of the user interface for band list search and visualization

- Finalisation of the development of the VO interface with VESPA and development of a new VO interface with VAMDC (an extension of the EPN-TAP protocol, for experimental spectroscopy)

Most of these improvements/extensions for the users will be delivered in D6.8 and D6.14: "3rd & 4th VESPA Annual reports"

11.3 Future SSHADE content

The filling of the databases by the SSHADE consortium of partners is managed under the VESPA-VA activity (including managers & users trainings and online-support, SSHADE documentation, tutorials and user-cases, SSHADE demonstrations at conferences, ...) and will be reported in several future deliverables between May 2018 and August 2019 (D6.8 & D6.14: "3rd & 4th VESPA Annual reports" and D6.13 & D6.15: "3rd & 4th VESPA Training session reports").

The 9 remaining databases that should be created after their managers are trained (training sessions planned in 2018-early 2019) are the following:

- ACID - **AstroChemical Ices Database**
  - Department of Physical Sciences (DPS-OU) / Astronomy team - Faculty of Science, Open University - Milton Keynes, United Kingdom
  - Physical Research Laboratory (PRL) / Indian Institute of Science - Ahmedabad, India
  - National Synchrotron Radiation Research Center (NSRRC) - Hsinchu City, Taiwan
    - Scientific Manager: Nigel Mason
    - Data Base Manager: Bhala Sivaraman

  The **ACID** Database will contains Vacuum UltraViolet (VUV) and InfraRed (IR) spectra collected from experiments that mimic the icy mantles of interstellar dust analogues. Spectra are from molecules (simple inorganic and organic molecules and their mixtures) already identified in the interstellar medium or on solar system bodies, and their isomers. Some molecules are synthesized in-situ starting from ice mixtures containing oxygen, water - carbon dioxide or ammonia-oxygen. Spectra are recorded from 4K or 10K up to sublimation temperatures or 300K for residues.

- CAB - **Centro de Astrobiología database** [Acronym to be defined]
  - Centro de Astrobiología (CAB) – Instituto Nacional de Tecnica Aerospacial (INTA), CSIC, Madrid
    - Scientific Manager: Guillermo Muñoz Caro
    - Data Base Managers: Guillermo Muñoz Caro

  The **CAB** database will focus on VUV-absorption cross section spectra (110-180 nm) of simple molecules and some of their isotopes in the solid phase at very low temperature and pressure. The spectra will be presented together with Mid-IR spectra of the same sample and VUV spectra of the same molecule in the gas phase. Mid-IR spectra of monolayer thin and thin films of pure ices and mixtures will be also part of the database.

- COMEDA - **COmetary, MEteoroids Dust and Analogs**
  - Instituto di Astrofisica e Planetologia Spaziali (IAPS) - INAF, Roma, Italy
  - Laboratory of Cosmic Physics and Planetology, Dipartimento di Scienze e Tecnologie, Università degli studi di Napoli "Parthenope, Napoli, Italy
    - Scientific Manager: Alessandra Rotundi
    - Data Base Manager: Andrea Longobardo

The **COMEDA** database will contain micro-infrared and micro-Raman spectra (with FESEM/EDX analyses) of different types of cosmomaterials: 1) 81P/Wild 2 cometary samples collected by Stardust spacecraft, 2) Meteoric dusts (DUSTER) collected in the upper stratosphere: nm to µm uncontaminated meteoric dust particles, 3) Cometary analogues (used to calibrate GIADA instrument onboard Rosetta): anhydrous silicate minerals and hydrated layer silicates and some selected analogue minerals coated with carbon and/or sodium hexafluorosilicate.

- Handbook - **Handbook of Minerals Raman Spectra**
  - Laboratoire de Géologie de Lyon – Terre, Planètes, Environnement (LGL-TPE) / Terre et Planètes team - ENS-Lyon, CNRS / Université Claude Bernard (UCBL), Obs. Lyon, Lyon, France
    - Scientific Manager: Bruno Reynard
    - Data Base Managers: Gilles Montagnac (exp.), Razvan Caracas (th.)

The **Handbook of Minerals Raman Spectra** database of the Laboratoire de Géologie de Lyon will contain a selection of Raman spectra of minerals representative of some major mineral groups. The spectra provide the main vibrational fingerprints of each main mineral group and related minerals. Most of the spectra were recorded in backscattering geometry with a laser beam focussed through microscope objectives down to a 2 µm spot on sample.

- ISMAD - **database of the Institute for the Structure of Matter MADrid**
  - Instituto de Estructura de la Materia (IEM), Department of Molecular Physics / Molecular Physics of Atmospheres and Plasmas team - CSIC, Madrid, Spain
    - Scientific Manager: Vicente Timón
    - Data Base Manager: Miguel Angel Moreno

The **ISMAD** database will include near- and mid-infrared transmission spectra and optical constants of ices similar to those present in different astrophysical environments, like the surface of interstellar dust grains, the surface of cold transneptunian objects, or in comets nuclei. Those ices contain mainly H2O, CO2, CH4, CO, NH3, N2, and other minor neutral species such as glycine amino acid, ionic species and alcohols, such us methanol and ethanol. It will also contain Raman spectra of clay minerals, phyllosilicates, uranium and iron oxides.

- MTACSFK - **Magyar Tudományos Akadémia CSillagászati és Földtudományi Kutatóközpont database**
  - Research Center for Astronomy and Earth Sciences (MTA CSFK) / Astrophysical and Geochemical Laboratory team (AGC) - Hungarian Academy of Sciences, Budapest, Hungary
    - Scientific Manager: Akos Kereszturi
    - Data Base Manager: Ildiko Gyollai

The **MTACSFK** database will contain mid-IR and Raman spectra of several specimens of three different meteorites analysed using FTIR (600–4000 cm-1) and Raman (532 nm laser) spectrometer, occasionally together with EPMA based elemental compositional analysis. Spectra of the following minerals were recorded in the meteorites: fayalite, forsterite, enstatite, diopside, hedenbergite, melilite, spinel, troilite, kamacite, pentlandite, magnetite, anorthite.

- REFL_SLAB - **REFLectance database of the Spectroscopy LABoratory**
  - Instituto di Astrofisica e Planetologia Spaziali (IAPS) / Spectroscopical Laboratory team (SLAB) - INAF , Roma, Italy
    - Scientific Manager: Fabrizio Capaccioni
    - Data Base Manager: Cristian Carli

Spectra that will populate the **REFL_SLAB** database are bi-directional reflectance spectra acquired with illumination = 30° and emission = 0° in the visible and near infrared (0.35-2.5 µm) using a FieldspecPRO spectrometer. We will also release spectra of rock forming minerals and their mixtures, like plagioclase and mafic minerals (pyroxene and olivine or assemblages of them) as well as igneous rocks and igneous glassy with variable compositions. Spectra will be associated with mineral/bulk chemistry.

- SCOOP - **Spectroscopy of organic COmpounds fOr Planetology**
  - Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA) / Exobiologie et Astrochimie team - CNRS / Université Paris-Est Créteil (UPEC) / Université Paris 7 Diderot, EFLUVE, Créteil, France
    - Scientific Manager: Nicolas Fray
    - Data Base Manager: Nicolas Fray

SCOOP is a database on spectroscopy of pure or mixed ices and of organics of planetary and astrophysical interests. The database will contain transmission spectra from 2.5 to 12 µm measured at temperatures ranging from 20 to 300 K under vacuum conditions. Many of these spectra results from chemical studies, induced by VUV photolysis and/or heating from 20 to 300 K, from an initial mixture of reactants. Some series of spectra show the spectroscopic and chemical changes occurring when the temperature increases.

- SPAN - **Spectroscopy of Planetary Aerosol aNalogues**
  - Laboratoire Atmosphères, Milieux, Observation Spatiales (LATMOS) / Instrumentation, Modélisation en Planétologie, Exobiologie et Comètes team - IPSL, CNRS / Université de Versailles Saint-Quentin-en-Yvelines (UVSQ) / Université Pierre et Marie Curie (UPMC), OVSQ, Guyancourt, France
    - Scientific Manager: Nathalie Carrasco
    - Data Base Manager: Thomas Gautier

The spectra of the **SPAN** database will concern laboratory analogues of planetary atmospheric aerosols. Those are synthesized at LATMOS, either with a cold plasma experiment (Szopa et al. 2006) or with a photochemical reactor (Carrasco et al. 2013). Their structures are sub-micrometric spherical particles or thin organic films. We are interested in characterizing the samples in a large wavelength range, from X-ray up to far IR. The spectra will be in transmission, absorption or complex refractive indices.