*Journées ASOV*
*Paris – 10 March 2020*

# L'infrastructure de données SSHADE de spectroscopie des solides: état des lieux de son développement et évolutions futures

https://www.sshade.eu

Bernard Schmitt, Philippe Bollard, Damien Albert, Alexandre Garenne,
Lydie Bonal and the SSHADE Consortium Partners

# The SSHADE database infrastructure

## for Astrophysics, Planetary sciences and Geosciences

**Promote** the creation of databases and **develop** the tools to provide on-line experimental data on **spectra of solids** in the **electromagnetic spectrum**

➔ Set of databases from a Consortium of laboratories

GhoSST

FAME

LayeredSilicatesDatabase

BYPASS
Bern icY Planetary Analogues

...

hosted by ***OSUG Data Center/UGA*** in Grenoble, France

# *SSHADE* European Consortium of Data Providers

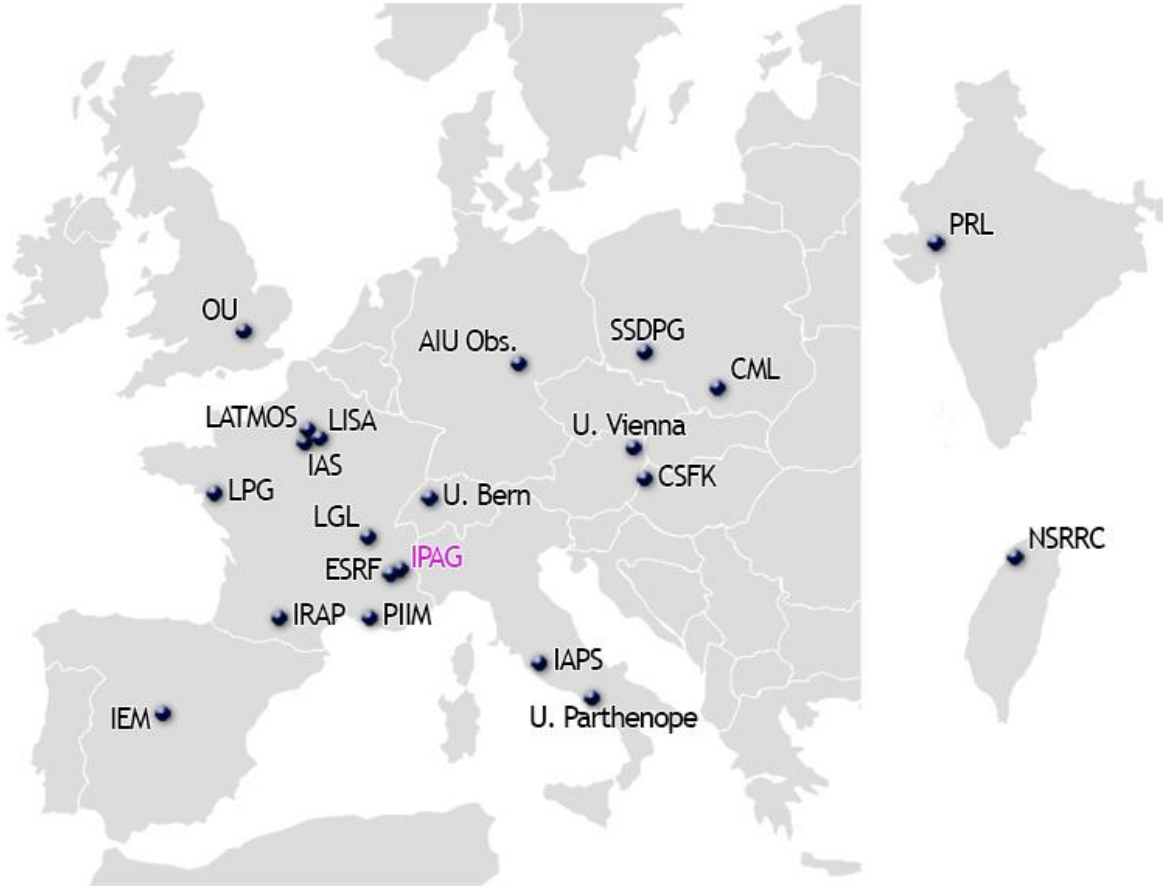
Data from **23** solid spectroscopy experimental groups
in **8** European countries (F, PL, D, GB, CH, E, I, HU) + India + Taiwan
**~75** researchers

**Each with particular expertise on:**

- some wavelength ranges
- type of materials
- physico-chemical conditions
- specific techniques
- type of data and products, …

**17 active databases + 2 starting**

**SSHADE Wiki** : https://wiki.sshade.eu



# A little bit of history: from past to future

- 2002-2006: Idea ... Concept ... Content demonstrator: STSP
- 2007-2008: First “solid spectroscopy” datamodel, Dev. technical demonstrator (OSUG, ...)
- ✓ **2009-2012: Full developments (Europlanet + VAMDC – FP7) of:**
  **SSDM (Solid Spectroscopy Data Model) and GhoSST database infrastructure**
- July 2011 GhoSST functional prototype
- ✓ **25 Sept. 2012: GhoSST opened to the public** (v0.5 beta-version )
- 2013-2015: Continuing SSDM and GhoSST developments, GhoSST data feeding
- 2014 Preparation and opening of a pre-SSHADE database

- ✓ **2015-2019: Development of SSHADE infrastructure under EPN@2020-RI (VESPA JRA)**
  **Opening of SSHADE to participating European (+Indian) partners (VESPA VA)**
- ✓ **1 Feb. 2018: SSHADE online with 10 databases (1250 spectra)**
- Feb. 2020: SSHADE with 17 active databases (> 2700 spectra)

- ✓ **2020-2023 Europlanet-2024 RI : development of ‘band list’ database**
  **Addition of 10-12 databases from around the world**

# **S**olid **S**pectroscopy **D**ata **M**odel

## Recent SSDM Changes / Improvements

# Major upgrade of SSDM (v 0.9.0)

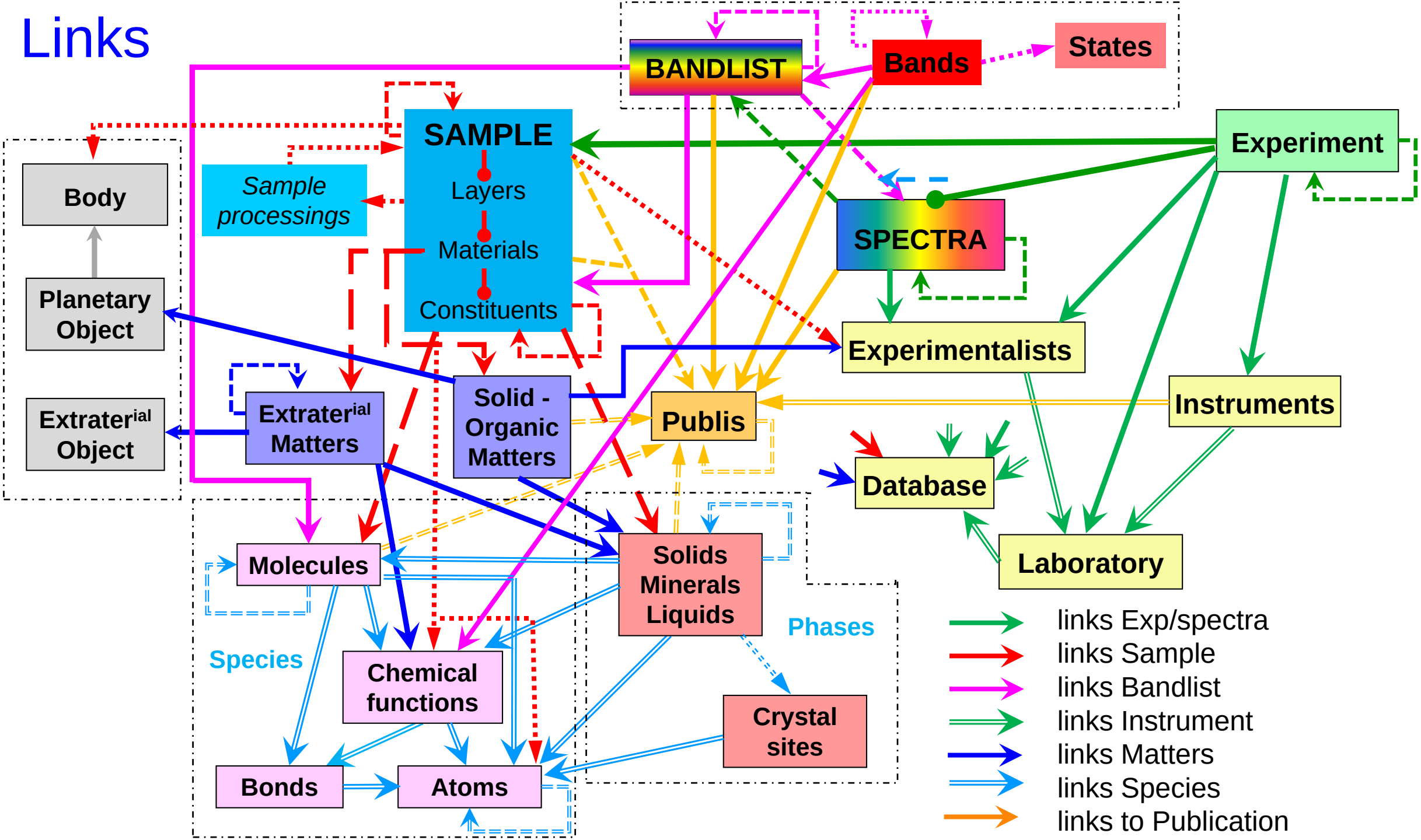
- **Databases:**
  - added KW to fully manage your database entry page @ SSHADE
  - added options to complete DOI information
- **Molecules**:
  - better and simpler description of ‘stereo-isomers’ and ‘nuclear spin’ isomers
- **Phases**:
  - better description of crystal sites (atoms, molecules), and polymers
- **Bodies**:
  - new table describing **planetary bodies** w. some physical parameters (linked from Objects)
- **Objects**:
  - new ‘planetary objects’ (**samples collected on planetary bodies**) [to be completed]
  - possibility of geolocation of objects (planetary or extraterrestrial*)
- **Matters**:
  - some extensions for planetary matters (sample return)
  - possibility of **geolocation** of matters (planetary, including Earth)
- **Sample**:
  - added matter/material grain size median + width. Added crystals (sizes, …) in Constituent, …
- **Publications**:
  - added other publication identifiers type and code (ex: ArXiv, …) + free URL

# Major upgrade of SSDM (v 0.9.0)

- **Experiment & matters**
  - added **DOI support** (completed)
  - Inclusion of field and airborne measurements
  - Addition of geolocation of natural samples (collected or field measurement)
- **Spectra:**
  - Better description and import of **multi-angle data** (BRDF, ...)
  - Improvement of experiment/spectra **version management** (to be finalized)
  - Addition of several preview options for experiment and spectra
  - **Extension** of the spectral range and spectrum types **to radio wavelength**
  - Extension to Polarized spectra  (specific import format to be finalized)
  - Extension to Scattering measurements
  - Extension to reflectance model parameters: (but not yet specific import format for set of n parameters)

- **Everywhere:**
  - removed some unnecessary mandatory or improved conditions
  - extended several Enum/OpenEnum with your suggestions
  - extended size of some text KW
  - improved description & comments in xml and SSDM, … … …

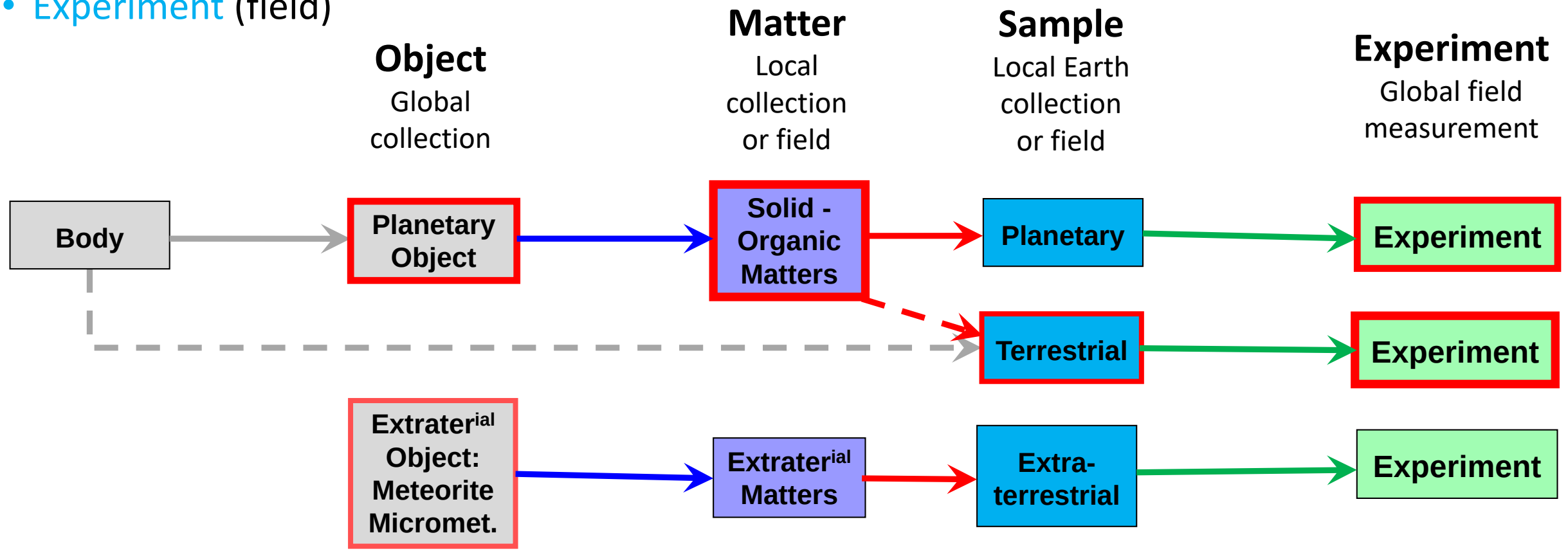
# Links


BANDLIST
Bands
States
SAMPLE
Layers
Materials
Constituents
Sample processings
Experiment
Body
Planetary Object
Extraterial Object
SPECTRA
Experimentalists
Extraterial Matters
Solid - Organic Matters
Publis
Instruments
Database
Laboratory
Molecules
Solids Minerals Liquids
Phases
Species
Chemical functions
Crystal sites
Bonds
Atoms
links Exp/spectra
links Sample
links Bandlist
links Instrument
links Matters
links Species
links to Publication

# Geolocation: collection and field measurements (v 0.9.0)

- Objects (except IDPs)
- Matters (except extraterrestrial)
- Sample (simplified case for Earth)
- Experiment (field)

- Body or object, Place, country
- Coordinates
  - System, type
  - Lat/long, altitude



# Description and import of multi-angle spectral data (v 0.9.0)

- **Surface spectra with 3D angles (illumination – observation – azimuth) :**

  - « spectrum_type »: physical  measurement type
  - « spectral_observation_mode » : spectrum, multi-wavelength, …
  - « angle_observation_geometry »: direct, bidirectional, directional-hemispheric, …

4 import options depending on dataset organization in original file(s):

- unique file with full spectro-photometric data set
- series of spectrum files at single geometry in a multi-angle dataset
- series of photometric data files at single wavelength in a multispectral dataset
- series of spectro-photometric data files with a single data type in each file

➔**all data homogeneously stored in the same 4D structure with same angle definition !**

+ read and import the 3 specific formats from Bern, IPAG and IRAP

+ selection of spectra within spectro-photometric data for experiment preview

# Better versions management (v 0.9.0)

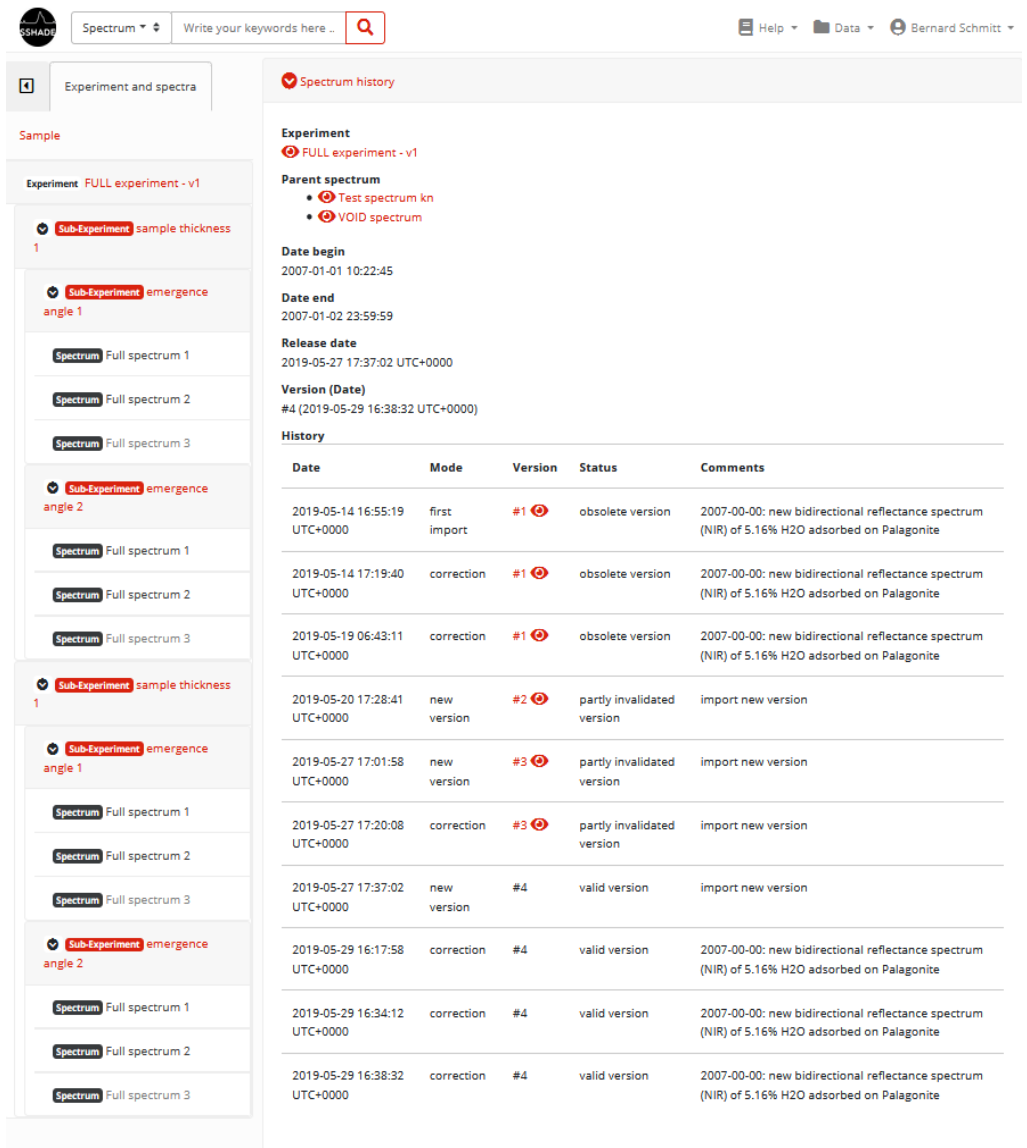
- Spectrum versions:

- 2 options
  - New version:
    - Explanation what is new
    - Import new spectrum
    - Keep old version(s) with metadata

  - Invalidate:
    - Explanation why it is invalidated
    - Keep invalidated version(s) with metadata
    - Possible link to another spectrum with equivalent data

- Experiment versions:
  Detemined from spectrum version
  - New version:
    => new experiment DOI (extension .Vn) **(TBD)**

➔INTERFACE:
- ➢ Tell if obsolete/invalidated and when a new version exists
- ➢ Provide version & correction history



# SSDM – future evolutions (2020…)

- Future options to be developed (≥ 2020)

  - **Bandlists** (upgrade from GhoSST) + **band parameters**
    [Europlanet-2024 RI]

  - Implementation of 4-parameters **polarization** import and storage
    [2020 ?, when import/storage formats fully defined]
  - Implementation of n-entries **model parameters** import and storage
    (when ?)
  - Implementation of **spectro-images** import and storage
    (later …?)

  - Addition of several **fundamental physical properties** of solids (Ps, thermodynamic properties, …)
    (2021/22…?)

# Development of SSHADE interface

# Development of SSHADE interface

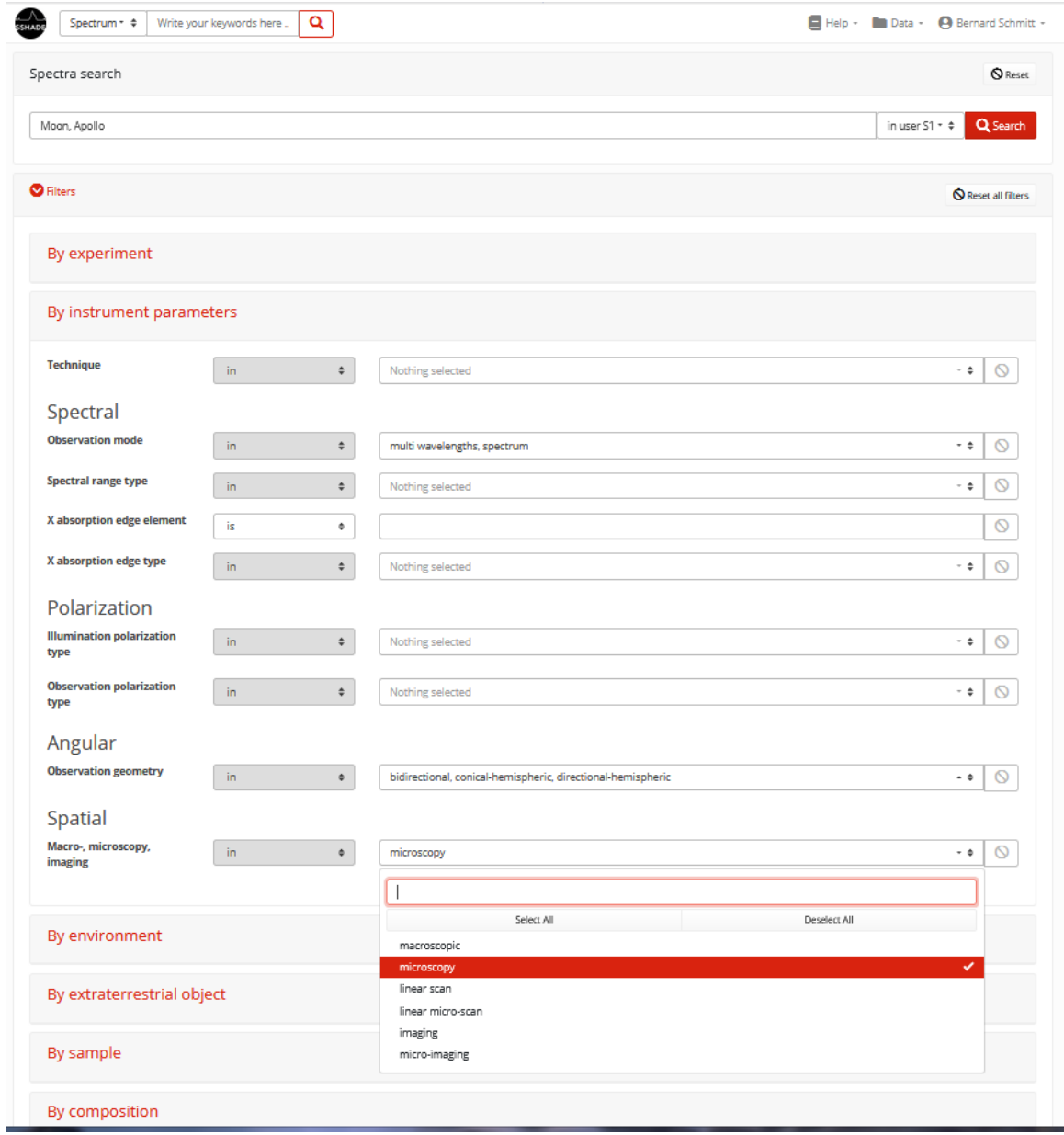
## A lot of new features and tools !!!

- For users: Login / Search / Visualization / Export / Dashboard
- For data providers: Detailed search, import, verification/publication tool
- For SSHADE managers: management of database, providers, members, ….

Better organized, More efficient, More stable, ….

# User: Search interface

- **Search tool:**
  - ✓ more efficient top search bar based on « **Elasticsearch** »
  - ✓ more filters choice
  - ✓ Reset modes: global, per keyword
  - ✓ …
- **Result list :**
  - ✓ Well working experiment grouping

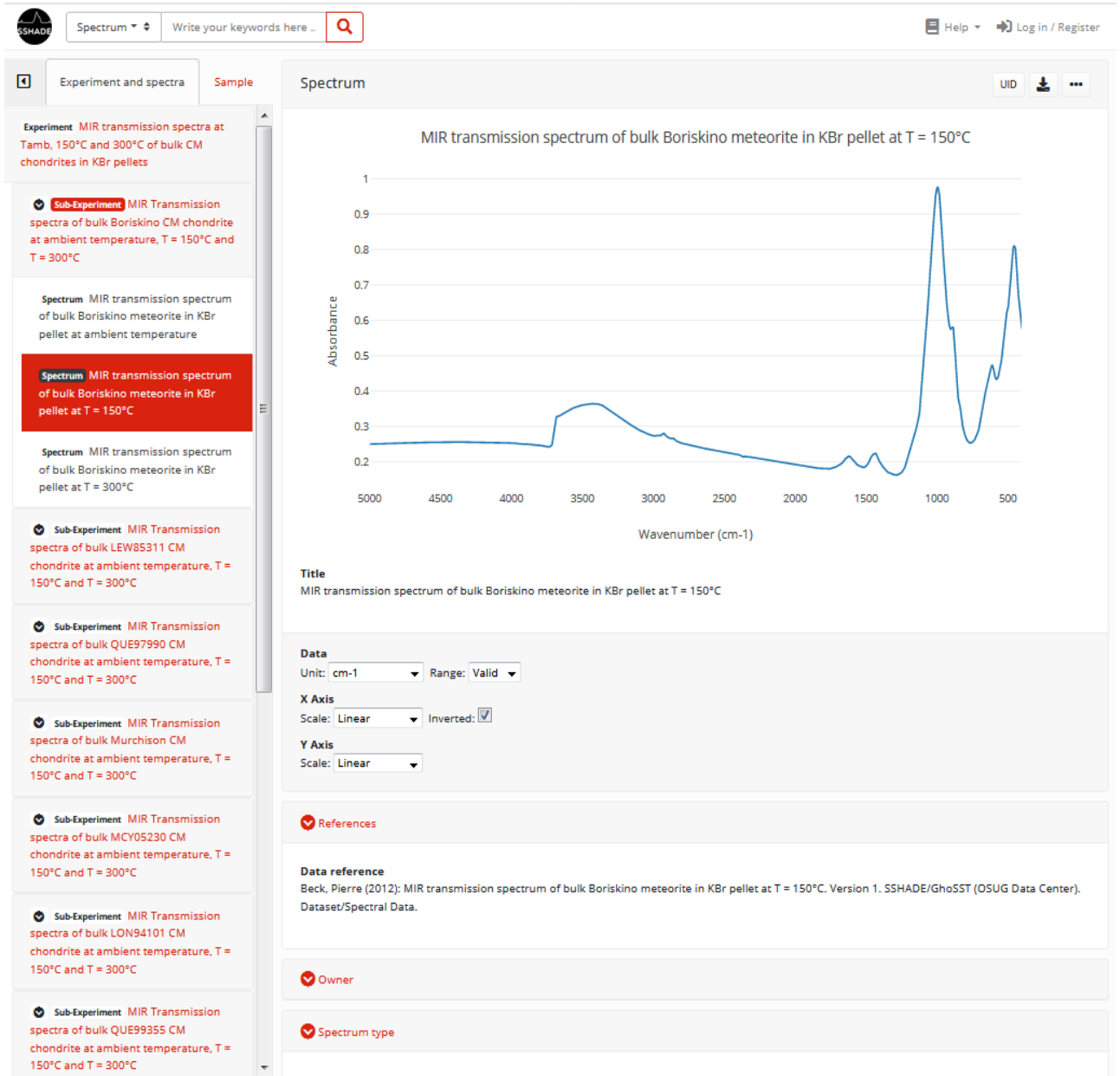


# User: Experiment-spectra / Sample data interface

- **Experiment display :**

**Better structuration**

- ✓ Exp-spectra and sample Tabs
- ✓ Highlight of displayed data
- ✓ Highlight of spectra with common sample
- ✓ …



# User: Experiment-spectra / Sample data interface

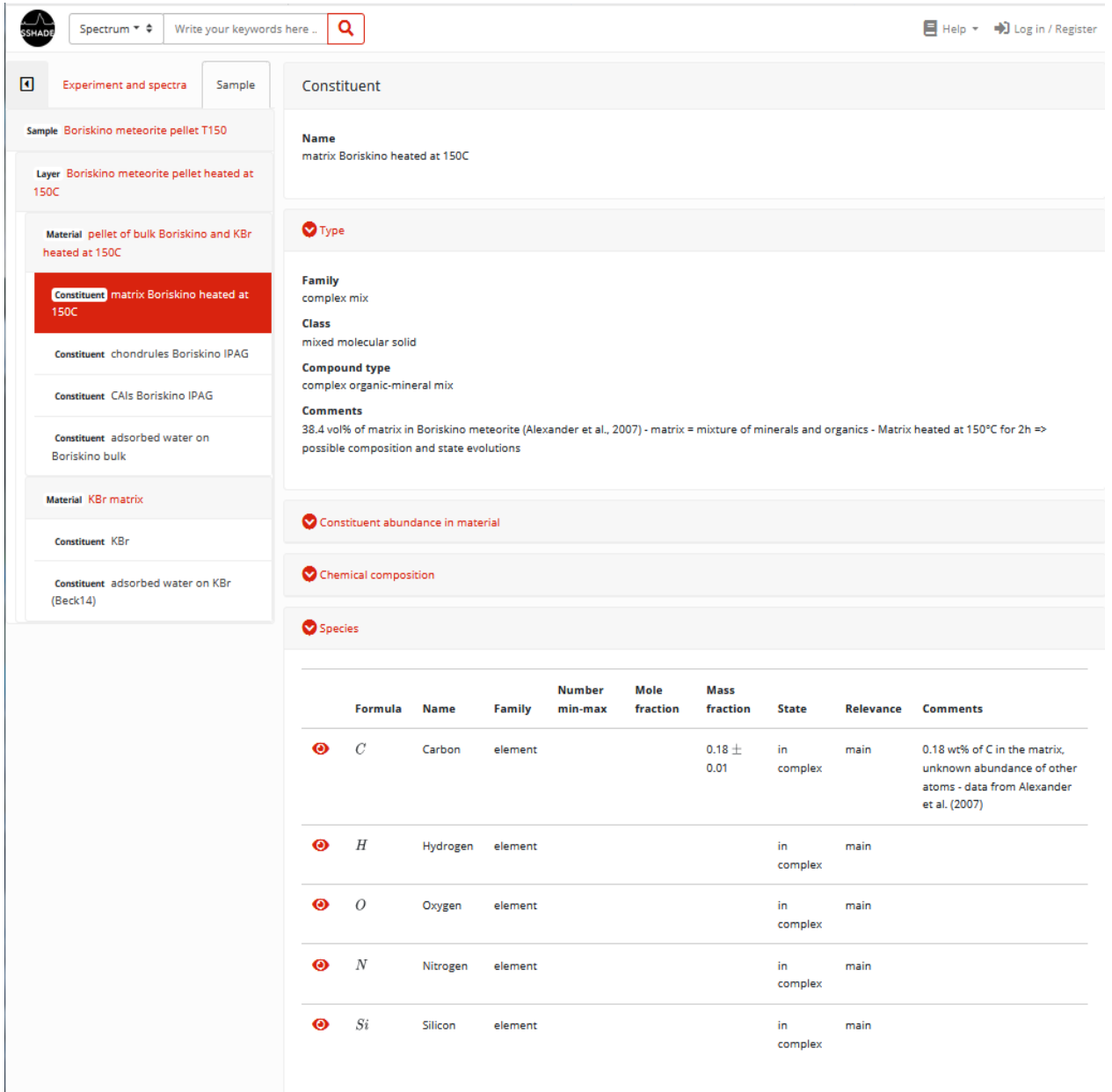
- **Experiment display :**

**Better structuration**

- ✓ Exp-spectra and sample Tabs
- ✓ Highlight of displayed data
- ✓ Highlight of spectra with common sample
- ✓ …

- **Samples :**

- ✓ Highlight of displayed sample structure level
  - ✓ Collapside structure still TBD

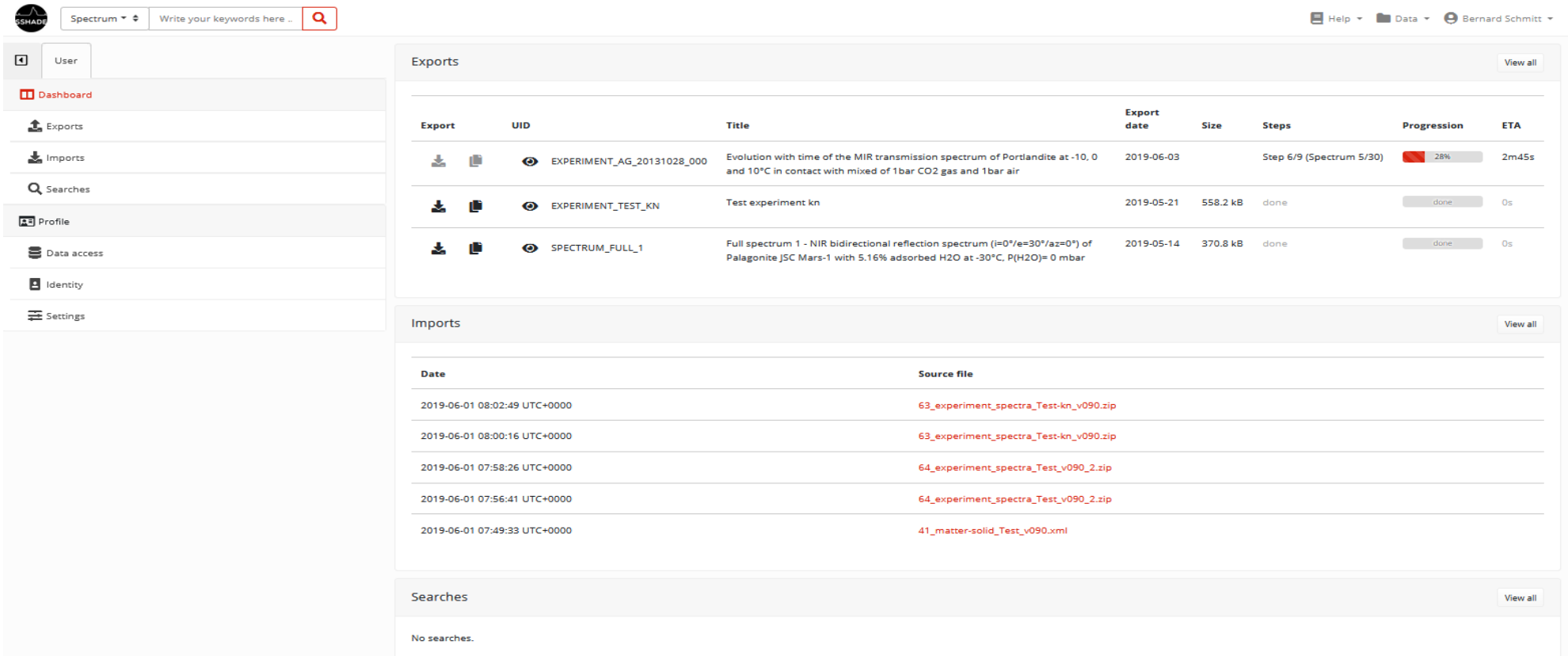


# User: Dashboard interface

- **Dashboard:**

✓ Export history (date, export progression, download, link for sharing, view, …)

✓ Import history (date, source file, download)

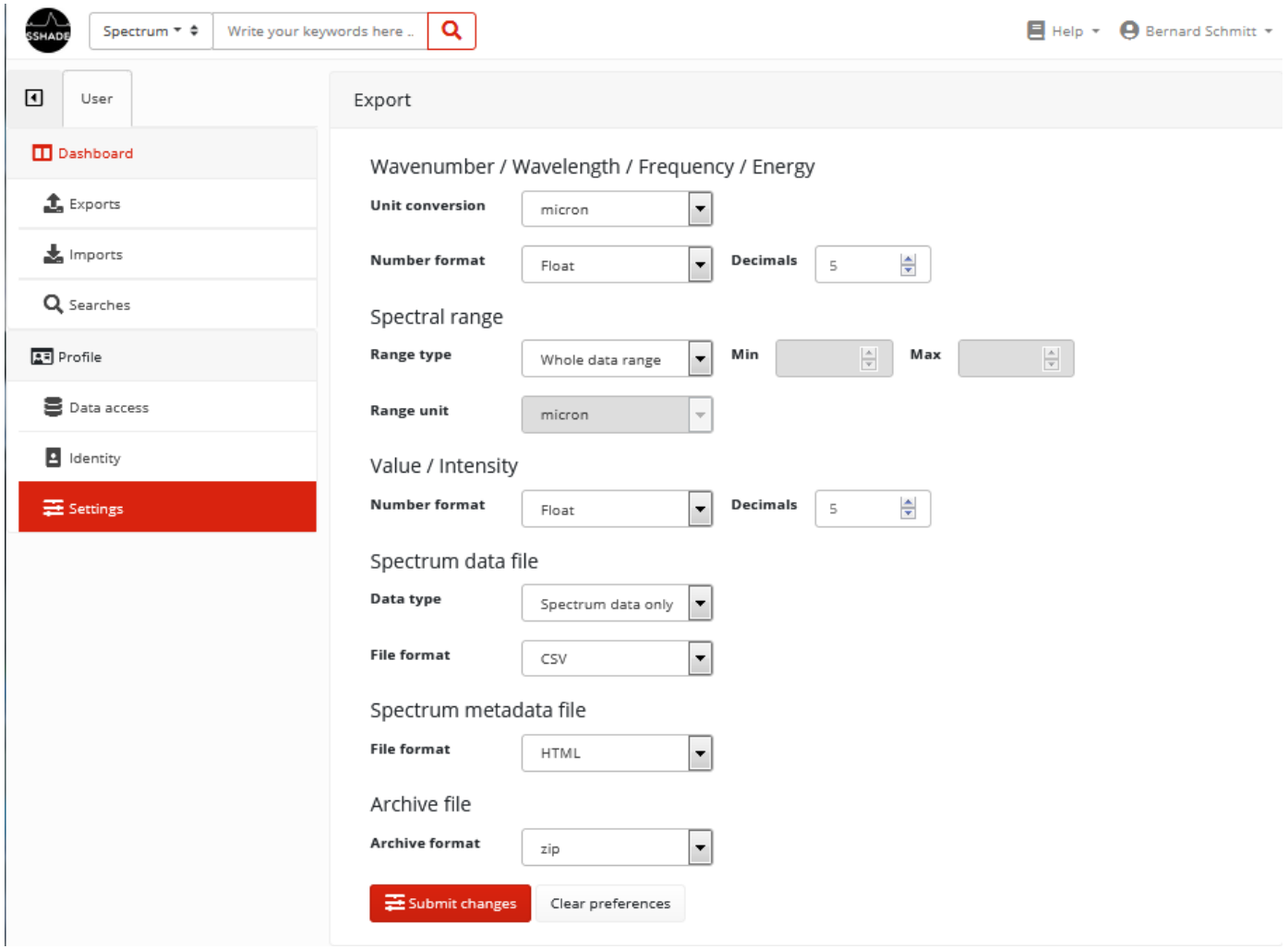


# User: Export setting interface

- **Export settings**

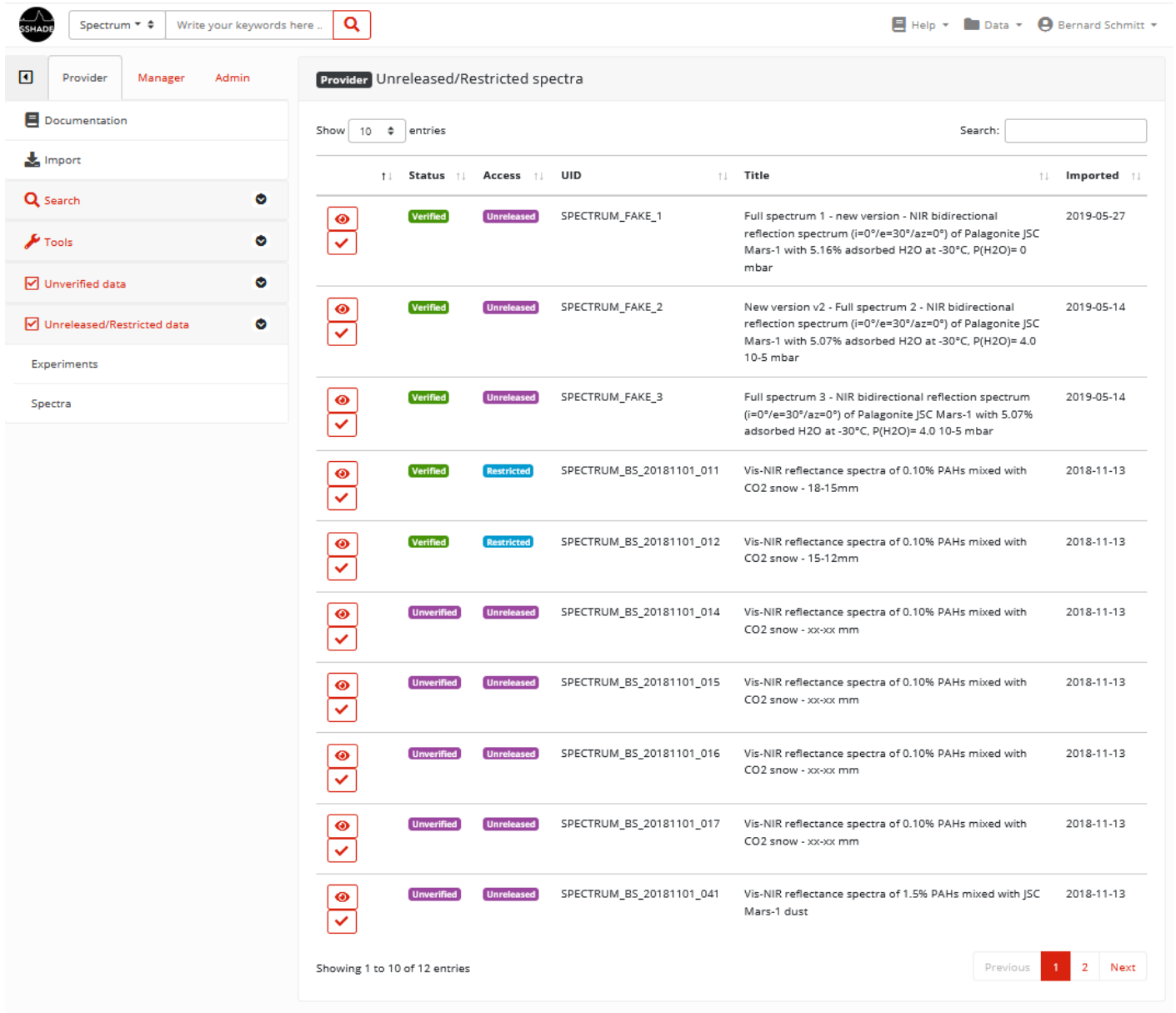
  - ✓ Unit
  - ✓ Spectral range
  - ✓ Data and metadata format
  - ✓ Export file format
  - ✓ …

  Can be set as ‘user preferences’ or at each file export



# Data provider: data verification / release interface

- **Unverified Experiments / spectra**
  - ✓ Unverified/verified status
  - ✓ Verification process
- **Unreleased/Restricted data**
  - ✓ Private / Restricted / Public status
  - ✓ Publication process
  - ✓ DOI generation
- **Public data**
  - ✓ DOI upgrade



# PID: the DOI in SSHADE

**DOI:** Unique identifier for

- SSHADE (*doi:10.26302/SSHADE*)
- each Database (ex: *doi:10.26302/SSHADE/GHOSST*)
- each Experiment (each version: TBD) (doi: 10.26302/SSHADE/EXPERIMENT_BS_20121213_002.V1)

**Automatically created (and upgraded)** using a set of KW of the datamodel:

- at database creation
- when an experiment is set to 'Public'

DOI redirects to a 'landing page'

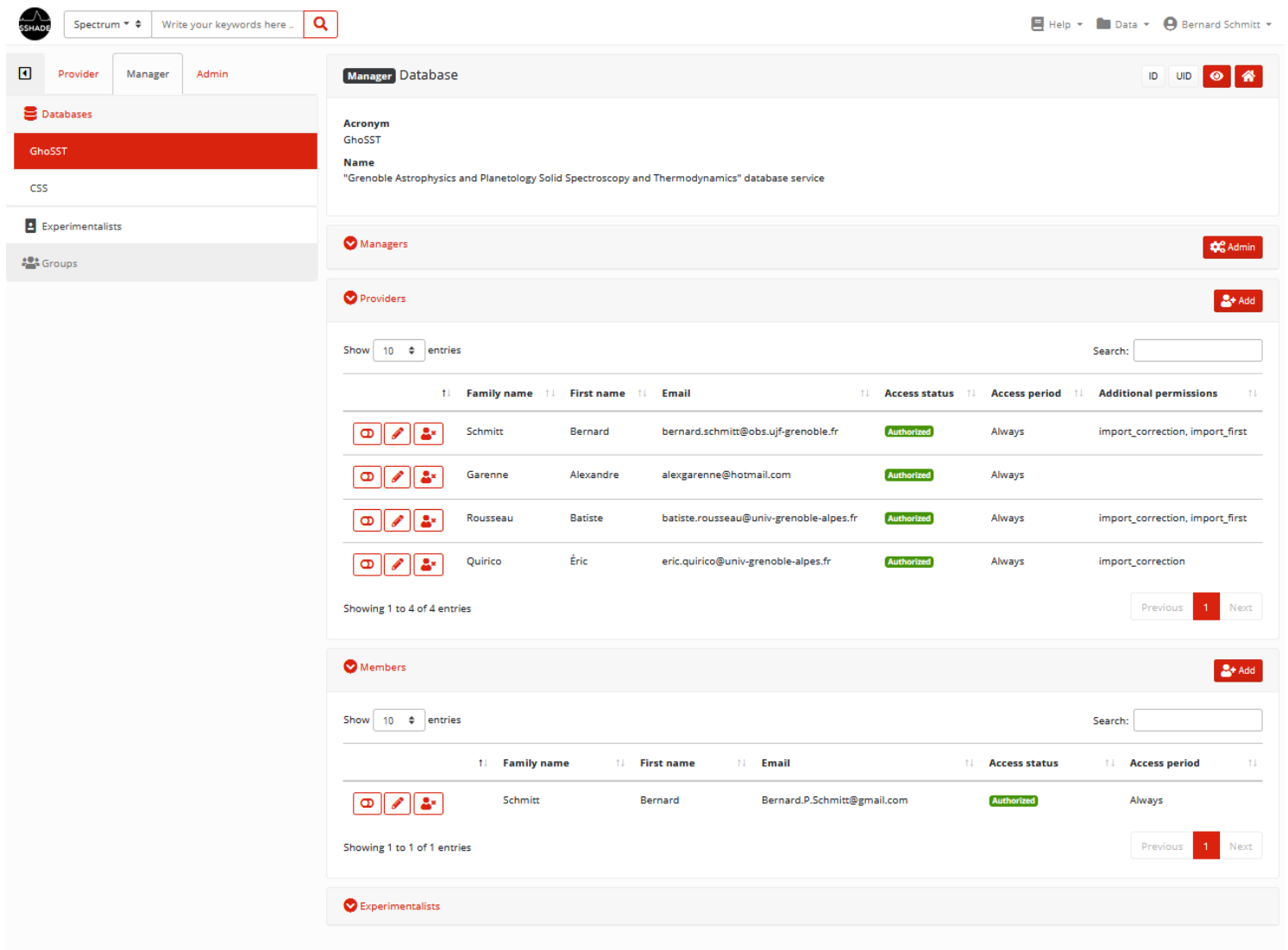
**Information in DOI metadata:**

- **(ABS) Mandatory KW**: Identifier, Creators, Title, Publisher, PublicationYear, ResourceType
- **Recommended KW**: subject, contributors (many types !), date, description, geolocation
- **Optional KW**: language, format, version, …

***Data Reference:*** Pommerol, A.; Schmitt, B. (2007): NIR bidirectional reflection spectrum of Smectite SWy-2 for different grain sizes at 298K. Version 1. SSHADE/GhoSST (OSUG Data Center). Dataset/Spectral Data. doi: 10.26302/SSHADE/EXPERIMENT_BS_20121213_002.V1

# Database manager: database mangement interface

- **Database management**
  - ✓ Providers
  - ✓ Providers import rights
  - ✓ Members
  - ✓ Link to experimentalist data



# Development of SSHADE VO

## Development of SSHADE Virtual Observatory (VO) access for VESPA

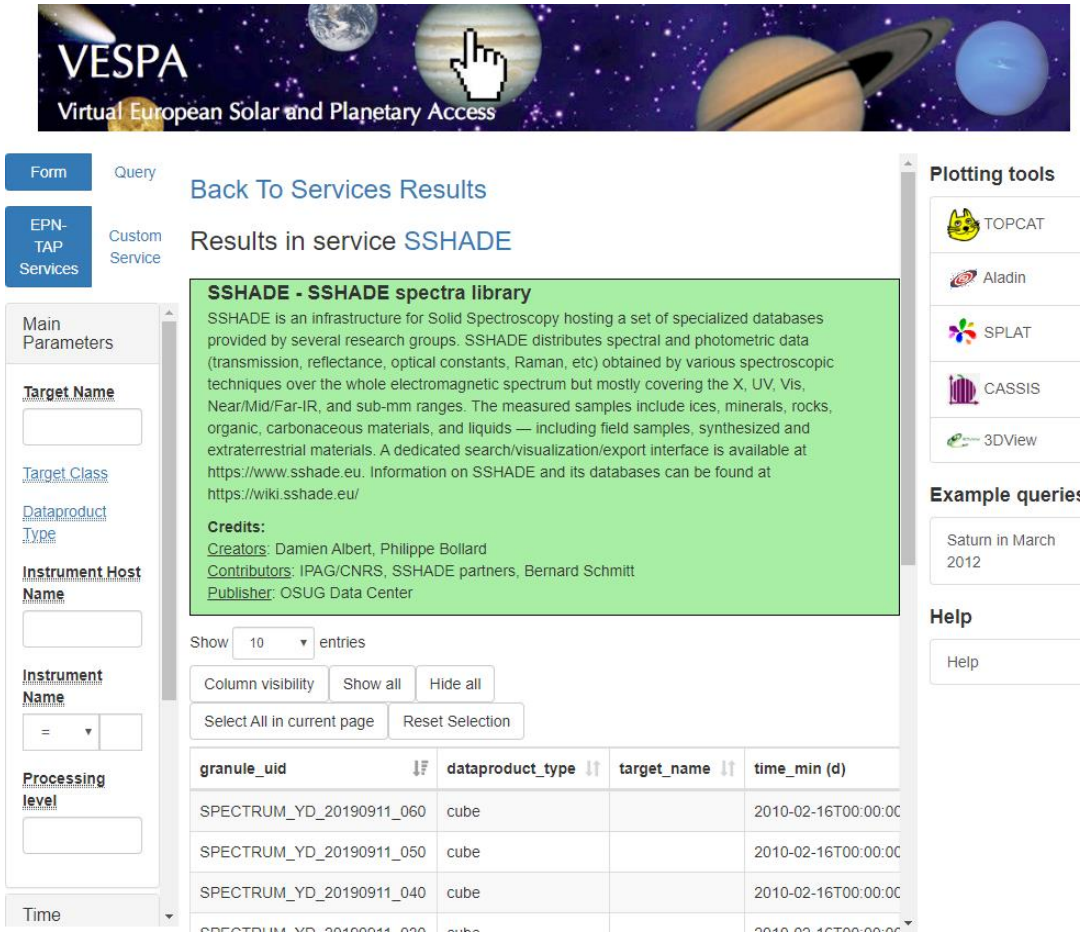
- Provide VO search on a limited number of main metadata
  - *species name/formula, compound type, object name, spectral type, T, P, grain size…*
- Allow to retrieve metadata and data for displaying in VO and associated tools/services
- Provide a link to the data in SSHADE (spectra in VOTable)

✓ Meeting *(Nov. 2018)* to define and implement a few new KW for better access and **search of laboratory data**

➔ implemented by VESPA in EPN-TAP

➔ **implemented in SSHADE**

➔ public VO completed and delivered last summer 2019



# Future SSHADE developments

# Future SSHADE developments

## Europlanet-2024 RI (2020-2023)

- Start: February 2020

- VESPA work package
  - SSHADE development: **band list of molecular solids**
    (+ extension of fundamental solids)
  - New partners => **new databases (8 - 12)**

## Other needed developments

- Multispectra visualization (internal comparison and with observations)
- BRDF : 4D multi-angle spectral and photometric visualization

# Bandlist of molecular solids

- Bandlist:

  List of band parameters and vibration modes of an isotopic molecule

  - in a simple constituent (2-3 species maxi)
  - in a defined environment (T, P, …)

  - Bands parameters
    - position (energy),
    - width, shape, …
    - intensities (peak and integrated)
    - accuracies / quality / evaluation
  - Transitions assignment
    - states QN, anharmonic coefficients, …

  ➔ link to a constituent, mostly fundamental solid phases

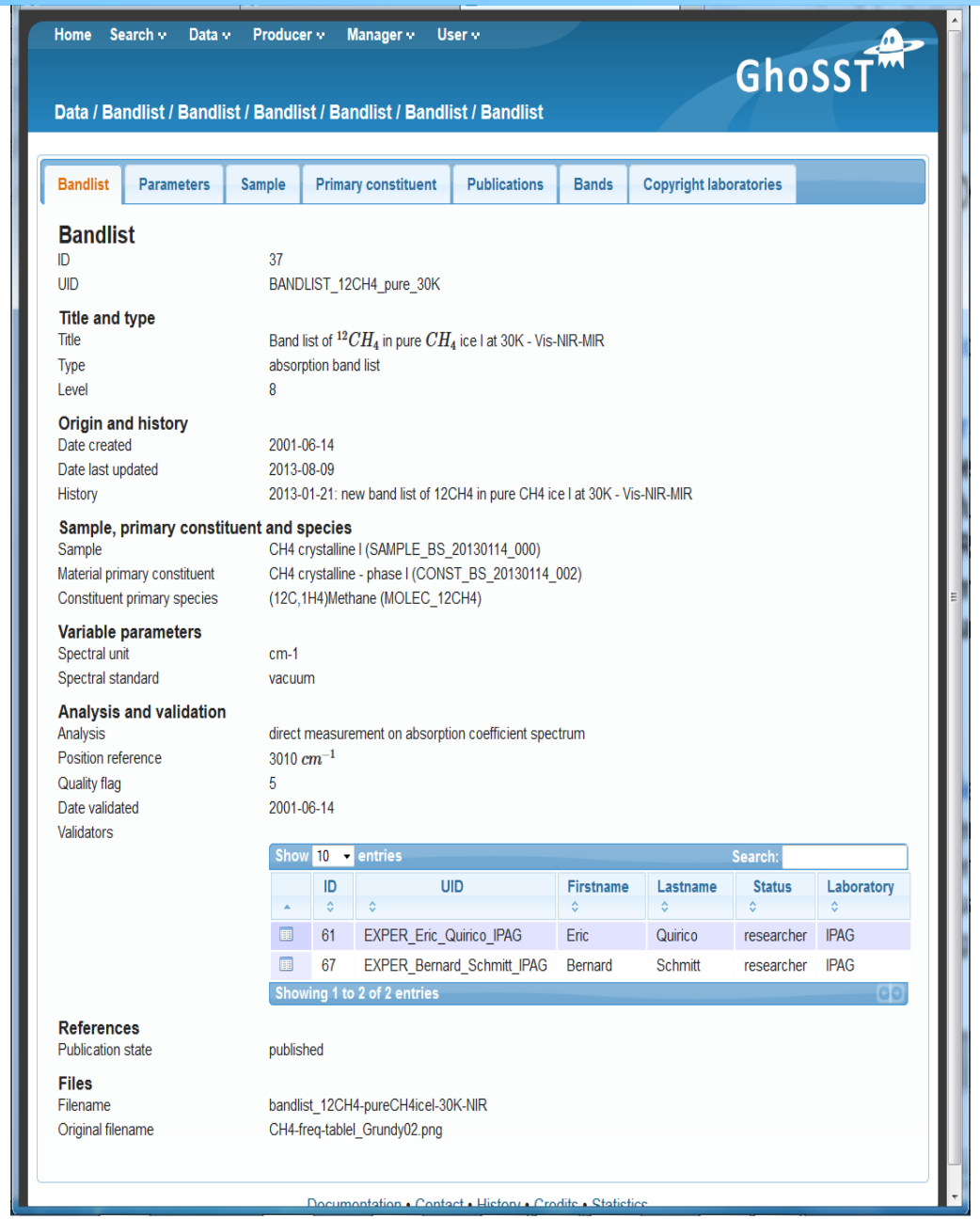
- Band parameters

  Variation of band parameters with temperature, pressure …

- Molecular vibration modes parameters

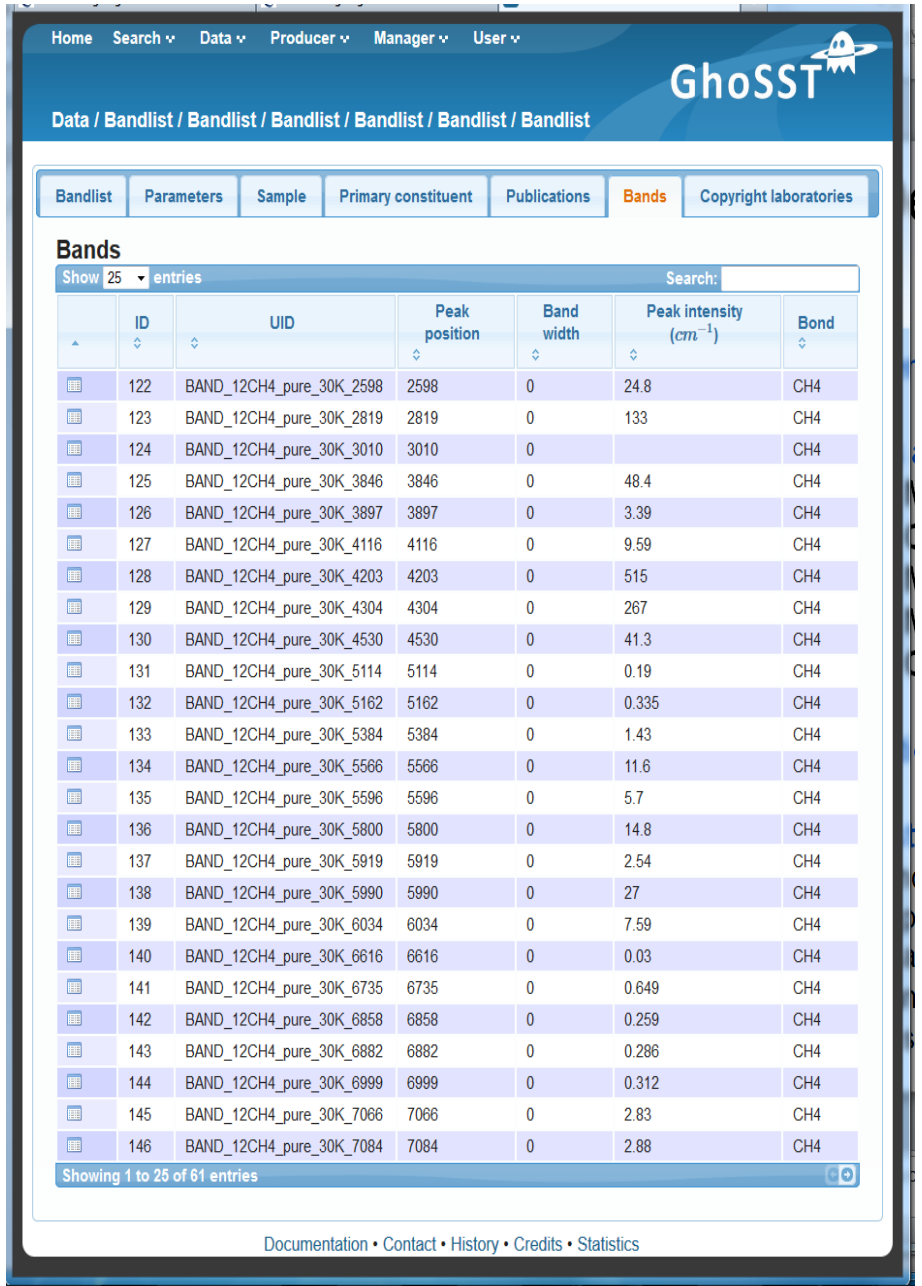
  Harmonic frequencies and anharmonic and interaction terms of molecular species in molecular solids

# Band lists and Bands




'old GhoSST':

- 15 bandlists
- 167 bands



# Band list

- Development/adaptation of bandlist datamodel
  - prototype already in GhoSST (to be adapted/modified to fit v0.9.0)

- Development of:
  - Bandlist database
  - Search tool / Visualization tool / Export interface

- Filling of the database

**Review the available data for molecular solids**

- Partner's data (see examples in 'old GhoSST')
- Publications

  => critical review and selection

  => selection committee ? (➔ 2020-22)

# SSHADE is FAIR ?

# FAIR

## Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

- ✓ ***UID and DOI for a set of spectra*** *(experiment') and their metadata*
- ✓ ***UID for each spectra*** *and its metadata*

F2. Data are described with rich metadata (defined by R1 below)

- ✓ *Various sets of metadata on sample, experiments, spectra, instruments, …*

F3. Metadata clearly and explicitly include the identifier of the data they describe

- ✓ *UID and DOI are metadata of the experiment*

F4. (Meta)data are registered or indexed in a searchable resource

- ✓ *DOI registered at DataCite*

# Accessible

A1. (Meta)data are retrievable by their identifier using a standardized protocol

A1.1 The protocol is open, free, and universally implementable

- ✓ *DataCite DOI https protocol*

A1.2 The protocol allows for an authentication and authorization procedure

- ✓ *??*

A2. Metadata are accessible, even when the data are no longer available

- ✓ *Data and metadata are persistently available, even when obsolete (old version) or discarded data (invalidated version)*

# Interoperable

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

- ✓ *SSDM datamodel common to the 20 databases of the SSHADE consortium. So some a largely shared language.*
- ✓ *Also interoperability developed with some VO (VESPA, VAMDC)*

I2. (Meta)data use vocabularies that follow FAIR principles

- ✓ *??*

I3. (Meta)data include qualified references to other (meta)data

- ✓ *Reference to Publications (DOI, …), data in other databases, …*

# Reusable

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

X *Still missing … (need agreement of all SSHADE partners…)*

R1.2. (Meta)data are associated with detailed provenance

✓ *Experimentalist/validator, instrument/laboratory, publications, …*

R1.3. (Meta)data meet domain-relevant community standards

✓ *SSDM use as much as possible the standards of definition of solids, spectra, …, but in crossing communities fields some tradeoff are made.*