GhoSST VAMDC 2010 annual report

- **<u>CNRS/LPG contribution</u> (team leader: B. Schmitt; technical manager: Damien Albert)**

- **Development of GhoSST database:**

  - *Hiring of Database engineer* by CNRS (Damien Albert) in January 2010 who will act as project manager for the development of GhoSST database at LPG

  - Development of a *first prototype of the query interface* for the GhoSST database (SSDM v 0.2).
  - Development of a *program to ingest XML transmission spectroscopic data/metadata* in the database (SSDM v 0.2).

  - *Validation and preparation of data and metadata* for the GhoSST spectroscopic database.
  Data and metadata of a series of transmission spectra and optical constants of ices have been retrieved and inserted in XML files (SSDM v 0.2) ready to be ingested in the database.

  - Writing of a first draft (Documentation / description of key-words and general structure) of the *dictionary of the current solid spectroscopy datamodel* (SSDM v 0.2).
  - Delivery of this document.

  - *Survey of solid spectroscopy lab data producers and databases* (web and relational): EU+world

  - Preparation of a *first meeting (13/01/2010) of an expert group* composed of European laboratory spectroscopic data providers (16 people from 4 countries attended).
  The aim of this expert group is to advise the Europlanet and VAMDC teams on the necessary (and optional) improvements and extensions of the solid spectroscopy data model that we intend to propose to be, at least, an European standard for this field, and possibly an international one.

Dictionary of the current solid spectroscopy datamodel (SSDM v 0.3).

Constituent (possibly also for use for mineral elements (atoms) description)

Key-word	Type	Level	Table	Exp	Unit	Description
constituent_index	int(10)	B	Const	F	--	Automatic incremental number given to new constituent
Species mixing						
constituent_number_species	int(10)	S	Const	F	No	Number of different species (molecules or element) of the constituent ➔ calculated from list_species_index (note: =1 for material_type = "pure" or "raw")
list: [species_family]	List: [enum(text)]	U	Const	F	--	Table of the family of species of the constituents ("molecule", "mineral naturel") for constituents for a complex "organic material": maybe "complex molecular solid" ? (note: all these families refer to specific sub-databases where the "constituents" are stored)
List:[species_index]	List: [int(11)]	U	Const	F	--	Table of the "species_index" of the different species composing the constituent (will be "molecule_index" or "element_index" depending on the "species_family")
List:[species_is_precursor]	List: [?]	U	Const	F	--	Table of flags telling for each species if they are "precursor species" or "actuel species" composing the constituent (for processed synthetic materials)
List:[species_mole_fraction_precursor]	List: [Float]	U	Const	F	No	Table of mole fraction of the different precursor species in the constituent (processed synthetic materials only)
List:[species_mole_fraction_actual]	List: [Float]	U	Const	F	No	Table of mole fraction of the different actual species in the constituent (synthetic materials only)
List:[species_mass_fraction_precursor]	List: [Float]	U	Const	F	No	Table of mass fraction of the different precursor species in the constituent (processed synthetic materials only)
List:[species_mass_fraction_actual]	List: [Float]	U	Const	F	No	Table of mass fraction of the different actual species in the constituent (synthetic materials only)

Data model extensions
towards SSDM v1.0

- to describe :
  - additionnal instrument techniques
    - Key-words to describe
      - Bidirectional / Biconic reflectance spectroscopy
      - Emission spectroscopy
      - IR + Raman + fluorescence micro-spectroscopies
  - new sample types
  - new spectra types and spectral products

- additional types of materials
    - Minerals / Rocks
      - Which classification to use ?
      - Which attributes to completely characterize one mineral ?
    - Complex organics
      - How to characterize a complex synthetic or natural organic solid ?

Expert working group on solid spectroscopy data model

– 1st meeting –

Wednesday 13 January 2010 – LPG, Grenoble

in the frame of VAMDC & Europlanet RI programs
(also supported by ASOV – PCMI)

1st Expert working group meeting on solid spectroscopy data model

List of participants (16)

- Damien Albert, LPG, Grenoble, France
- Donia Baklouti, IAS, Orsay, France
- Pierre Beck, LPG, Grenoble, France
- Mario D'Amore, DLR, Berlin, Germany
- Yves Daydou, DTP, Toulouse, France
- P. Duvernay, PIIM, Marseille, France
- Joern Helbert , DLR, Berlin, Germany
- Karine Demyk, CESR, Toulouse, France
- Erwan Le Menn, LPGN, Nantes, France
- Giuseppe Leto, Catania Astrophys. Obs., Italy
- Patrick Pinet, DTP, Toulouse, France
- Eric Quirico, LPG, Grenoble, France
- Bernard Schmitt, LPG, Grenoble, France
- Bhala Sivaraman, Open University, Great Britain
- Patrice Theulé, PIIM, Marseille, France
- Razvan Caracas, ENS-Lyon, France

Other people interested (10):

- Gabriele Arnold, Univ. Münster, Germany
- Emmanuel Dartois, IAS, Orsay, France
- Cornelia Jäger, Institut für Festkörperphysik, Jena, Germany
- Stéphane Le Mouelic, LPGN, Nantes, France
- Nigel J. Mason, Open University, Great Britain
- Gilles Montagnac, ENS-Lyon, France (TBC)
- Harald Mutschke, AIU, Jena, Germany
- Robert Georges, IPR, Rennes, France
- Ella Sciamma O'Brien, LATMOS, Verrières, France
- …

Aim of expert group

- **Advise** the VAMDC and Europlanet teams on the necessary (and optional) improvements and extensions of the solid spectroscopy data model schema

- **Help to build** some parts of data model schema
  - minerals and rocks, organic matter,
  - sample preparation protocols …

- **Define** (all)
  - the query key-words
  - the information key-words
  - the units

- **Review** the new data model schema (SSDM v0.3+), dictionary and documentation
  - April-May 2010

- **Test** (with end-users) the implemented database query interface
  - End 2010 ?

Aims of Data model

- Development of an ***European Solid Spectroscopy Data Model*** (ESSDaM) that we intend to propose to be, at least, an European standard for this field, and possibly an international one.

- Include in the data model schema, up to some limited complexity level the requirements of most of the *European solid spectroscopy data producers*

Schedule:

- “completed” solid spectroscopy data model by April 2010
- Dictionary and documented (v0.3/v0.4?) by May-June 2010

How to proceed

- start from the GhoSST core data model schema (v.0.2) and expand it with some changes.

- extend mostly on:
  - new types of spectroscopy techniques (reflection and emission spectroscopy, Raman, fluorescence …)
  - other types of solids (minerals/rocks, organic matter, …)

- Complete, improve, generalize, homogeneize key-words

Limitations of the evolution of SSDM data model towards v1.0

- to keep this version 1.0 simple although as complete as possible for the most fundamental data.

- We favour **wide spectroscopic data types and material types** coverage in v1.0

  to the detriment of complex samples and processes.

SSDM Data model extensions

In **bold** are “dimensions” that are already implemented in SSDM v0.2

- materials:
  - **molecular solids**
  - minerals, rocks
  - complex organic compounds

- samples:
  - **Solids or liquids,** molecular aggregates
  - **1 (pure)** or “n” material components in any type of mixture
  - **Molecular mixing, granular mixing, layering (2 layers)**
  - Core/mantle grains

- ranges:
  - **UV**
  - **Visible**
  - **Near, mid and far-IR**
  - sub-mm

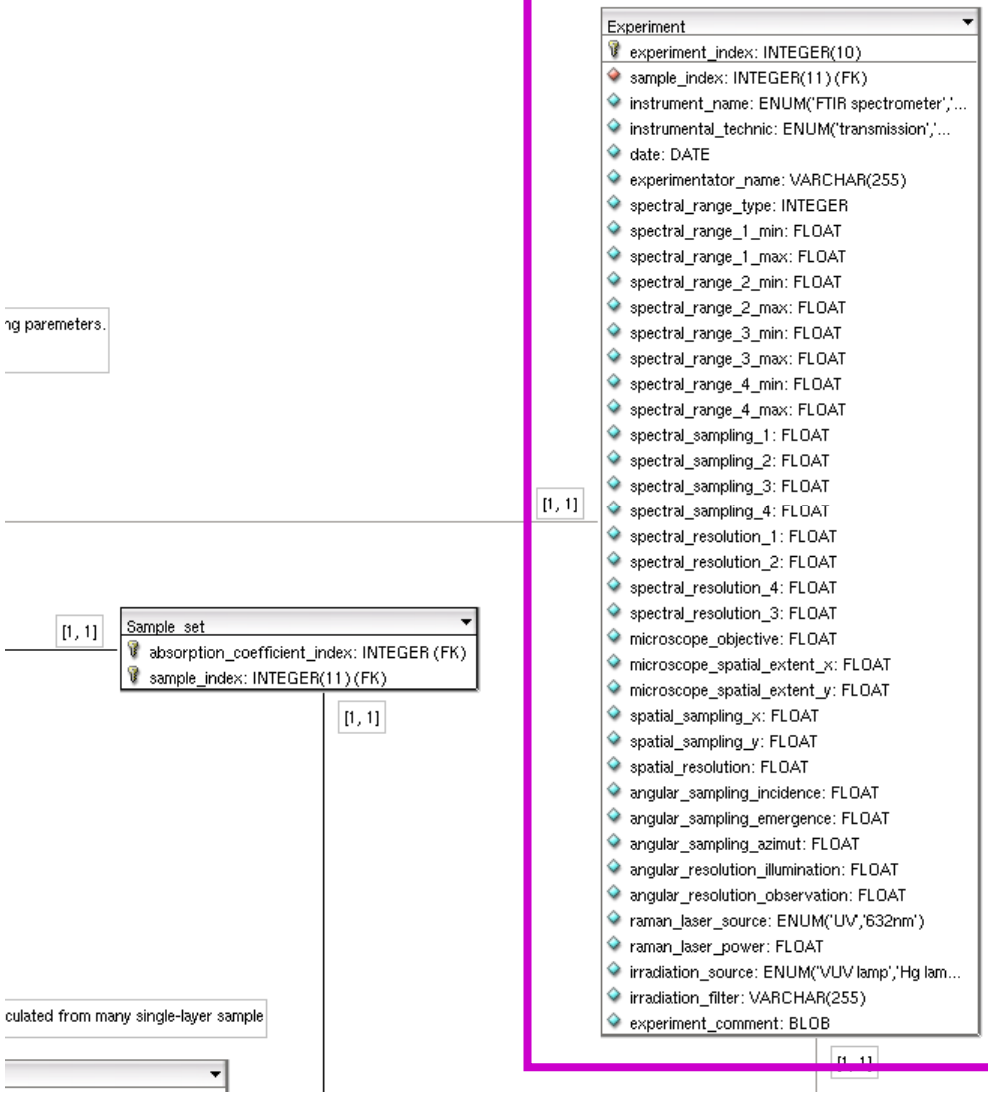
SSDM Data model extensions

- sample processes:
  - **Temperature cycling**
  - Simple irradiations (source)
  - gas interactions (adsorption …)

- spectroscopic techniques:
  - **Transmission**
  - Reflection (bidirectional, biconic, …)
  - Emission
  - Microscopy
  - Raman
  - Fluorescence

- data and products:
  - **Spectra (raw and transformed)**
  - **Absorption coefficient and Optical constants**
  - Spectral products from the other techniques
  - **Band list for molecular solids** (+ adsorbed in/on minerals)

Instrument parameters



Measurement technics

+ reflectance

+ emission

Spectral / Angular / Spatial

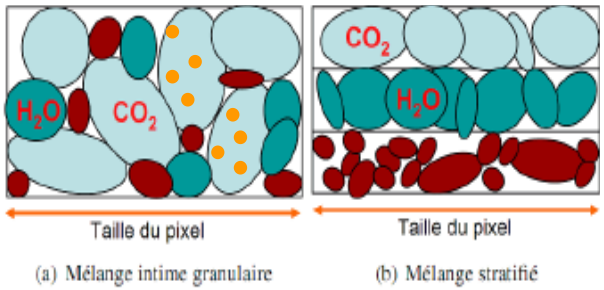
- range
- resolution
- sampling

+ polarization, …

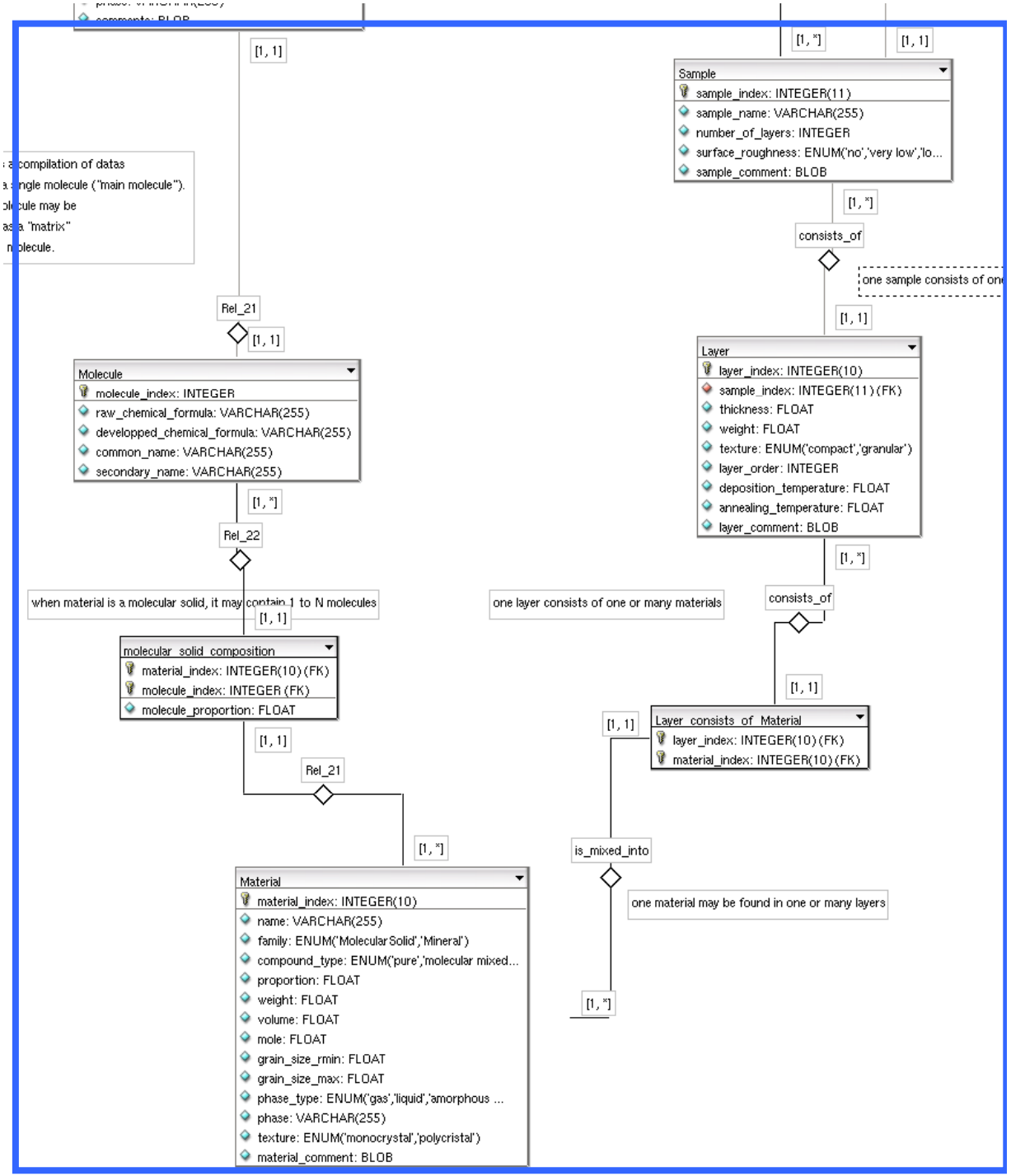
Sample

- Layers
+ core/mantle
- Materials
- Constituents :
  Molecules
  + minerals

+ formation protocol
+ irradiations



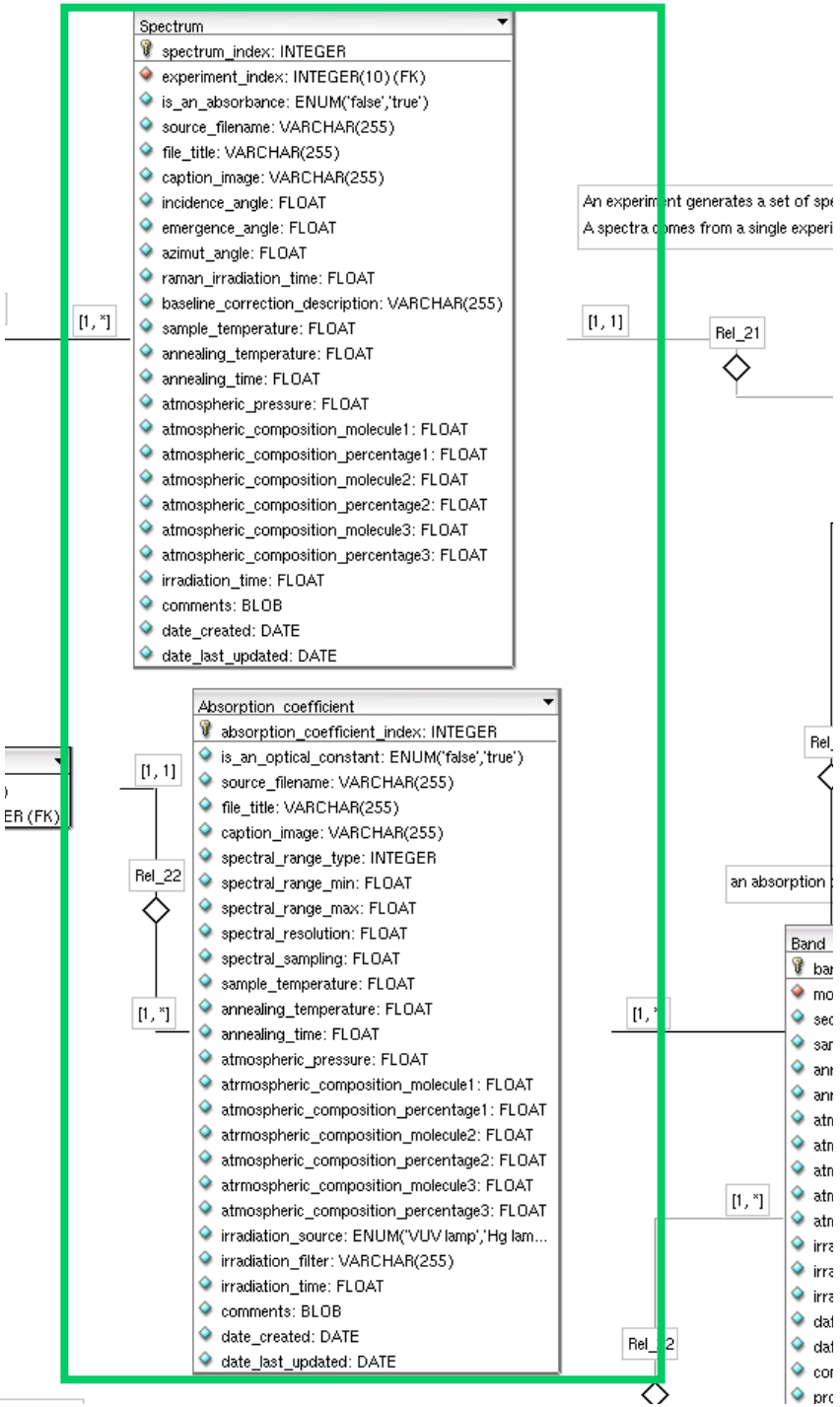
(a) Mélange intime granulaire (b) Mélange stratifié



Spectra

- Transmission spectra
- Normalized absorbance
- Absorption coefficients
- Optical constants

+ reflection / emission spectra
+ emissivity, BRDF, …
+ polarized spectra

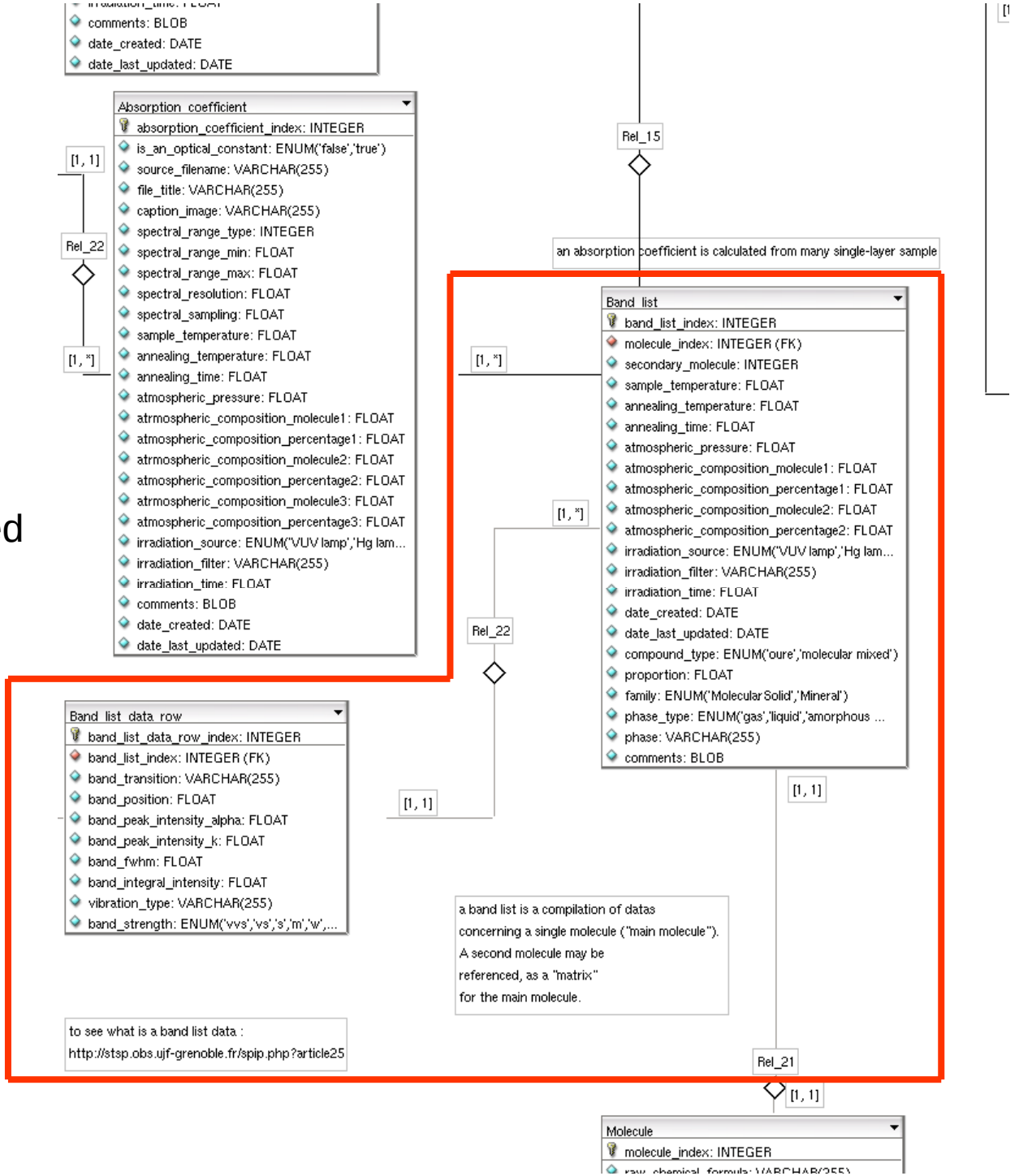


Band list

- Molecular solids
+ adsorbed molecules

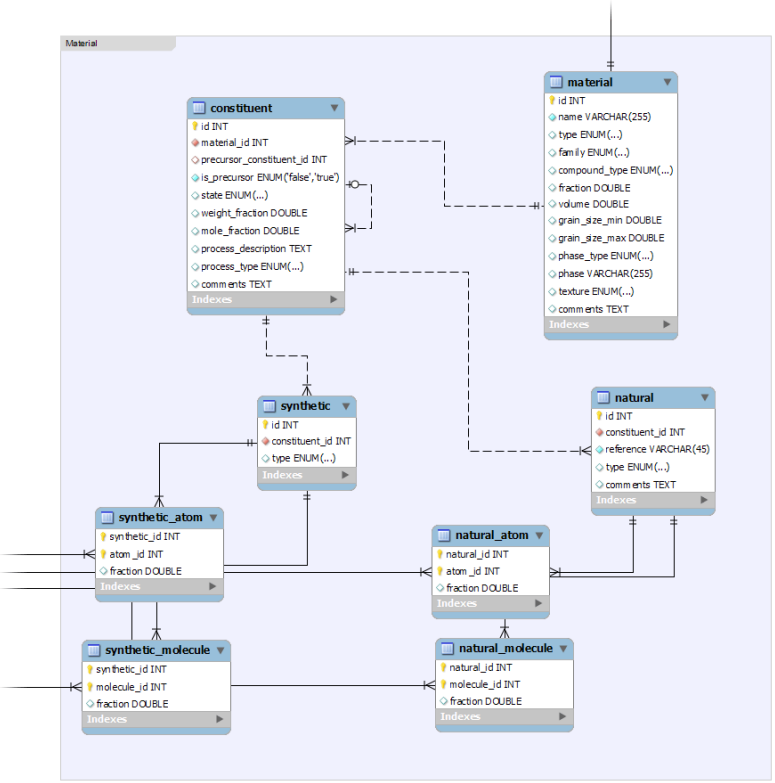
- Positions,
- Width,
- Intensities: peak, integrated
- Vibration modes (+ type)
- References

+ errors



Solid Spectroscopy Data Model

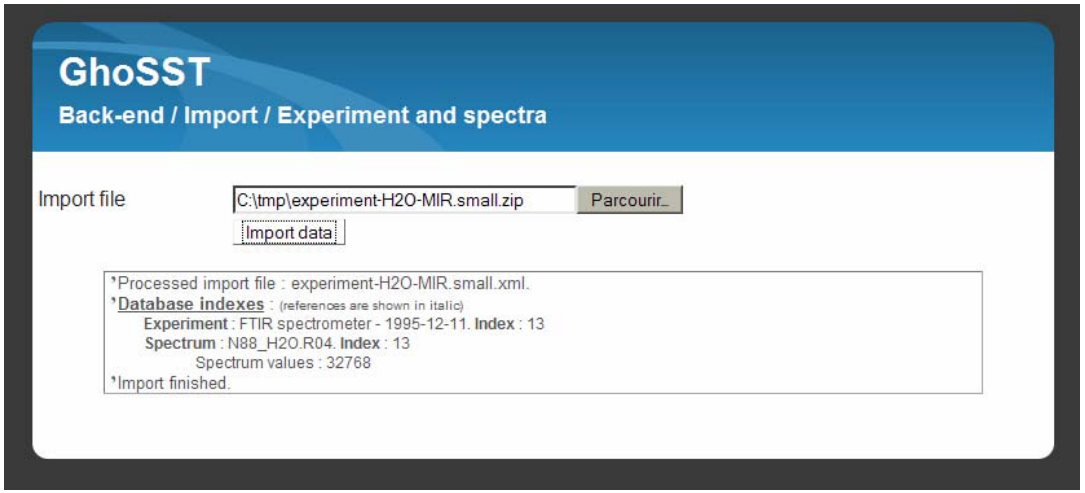
- Relational database advantages
  - Permits to store a wide range of parameters variations in experiments.
  - Experiment samples can handle complex structures (multiple layers / materials / constituents / precursors).
  - Samples and experiments can have a relation to a parent sample/experiment, enabling full history for those.
  - At term for an advanced use every parameters could be searchable.
- Molecule database identification, two standard available in SSDM :
  - InChIKey (International Chemical Identifier).
  - CAS Registry Number (Chemical Abstracts Service).



Database example : materials & constituents structure

Data Ingestion Interface Prototype

- Simple web tool permitting the import of experiment / samples / spectra.
- Based on XML input files with server-side XSD validation
- For additional data files an archive containing the import XML file and associated data files can be uploaded (multiple format accepted).
- Instant feedback to the user :
  - For data validation : displays full XSD errors to ease location of potential problem.
  - After import : report of activity done, ensuring the data has been correctly handled.
- Expertise in XML validation to be used when interconnecting whith VO in XSAMS.



Interface prototype feedback example

PLANNING 2010 « Spectro-database »

- Solid Spectroscopy Data Model (SSDM) extension (with VAMDC)

  (+ reflection & emission spectroscopy + minerals)

  - expert group meeting
  01/2010 Done

  - new SSDM + draft dictionary document
  04/2010 in progress

  - documented SSDM
  05/2010 in progress

- Implementation of spectro Database on new

PLANNING 2010-12 « Spectro-database »

- Data ingestion tools (XML files)
  - Transmission data
    07/2010 in progress
  - Band list data
    12/2010
  - Reflection / Emission data
    2011

- Feeding of databases :
  - transmission spectro (LPG) :
    07-12/2010
  - band list database (LPG)