Spectral Database: from data model to web interface

Nils Paulhe, Christophe Duperier, Daniel Jacob, Etienne Thévenot, Jean-Francois Martin, Claire Lopez, Franck Giacomoni

METABOHUB'S SPECTRAL DATABASE: FROM DATA MODEL TO WEB INTERFACE

NILS PAULHE[1], CHRISTOPHE DUPERIER[1], DANIEL JACOB[2], ÉTIENNE THÉVENOT[3], JEAN-FRANCOIS MARTIN[4], CLAIRE LOPEZ[1], FRANCK GIACOMONI[1]

[1] METABOLISM EXPLORATION PLATFORM, INRA (CLERMONT-FERRAND/THEIX)
[2] BORDEAUX METABOLOME PLATFORM, INRA (BORDEAUX)
[3] THE METABOLOMEIDF , CEA (SACLAY)
[4] METABOLOMICS AND FLUXOMICS PLATFORM, INRA (TOULOUSE)

Introduction

MetaboHUB is an infrastructure of metabolomics and fluxomics that provides tools to research teams and partners; a full workpackage focuses on bioinfomatics tools design and development, including one called "MetaboHUB's Spectral Database". This tool is created to gather and share knowledge and data from both Mass and NMR Spectra.

The core of the database called "data-model", is a computational representation of each entity involved in spectra analysis and chemical compounds annotation.

It was designed with spectrometry experts knowledge; their experience will also be required to build an interactive identification interface for spectral analysis.

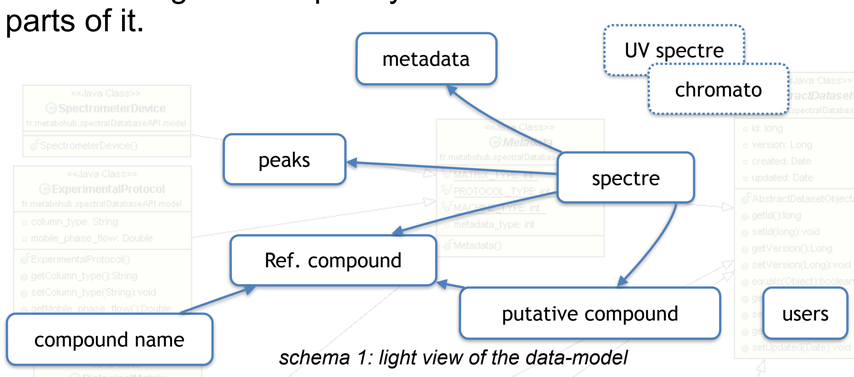
This database will be available through a web-portal and open for each partner of the MetaboHUB consortium in a first time, to the whole community later.

Methods

For the MetaboHUB's Spectral Database code architecture, we use modelling concepts and objects-objects relations to represent mass / nuclear magnetic resonance (NMR) spectrometry complexity. Manipulated entities and their links are based on data / metadata used by chemists in laboratory routine. Our work is split in three levels with data-model building, users services designing and code engineering.

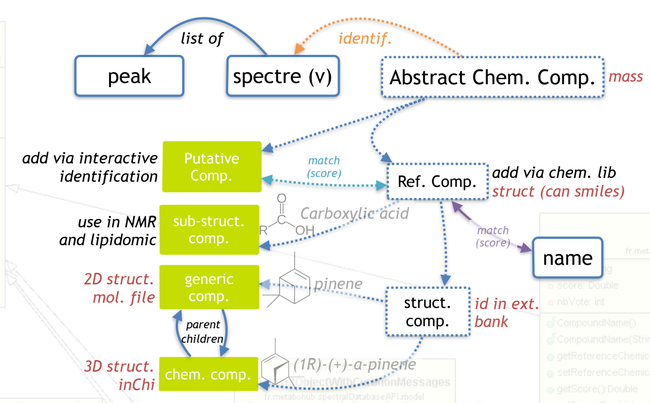
data model

Removing the complexity of the full data-model allows us a better understanding of it; the first schema proposes a light view of relation entities, others are focused on parts of it.



schema 1: light view of the data-model

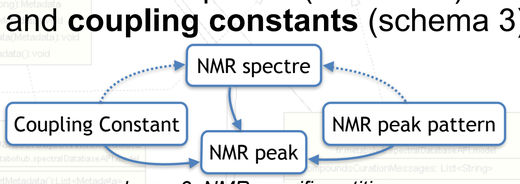
Light view of the data-model with main entities: both **reference** and **putative compounds** are linked to **spectra**. A specific entity, **metadata**, is dedicated to link spectra by their shared properties. The **user** entity is independent: it is used only for web-portal access control, not for spectra or compounds data ownership.



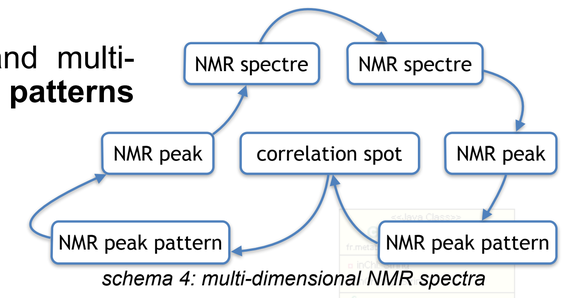
schema 2: compounds matryoshka dolls model

Link spectra to compounds: spectra match 4 types of compounds, common properties or relations given by a virtual "parent" object in the data-model. We provide **reference compounds** via chemical libraries manager. **Putative compounds** are annotated with an interactive identification tool and stored in the database. Three different entities of each reference compound allow us to match putative compounds with missing properties (like the 3D structure).

NMR spectra specificity: the data-model must support one and multi-dimensional spectra (schema 4) and specific entities like **peaks patterns** and **coupling constants** (schema 3).



schema 3: NMR specific entities



schema 4: multi-dimensional NMR spectra

*Parts of the data-model are specific for spectra technics (like **MS-MS**), others are created for futur developments (like **biological matrix spectra annotation**).*

user services

There are two ways to access the data: though a web-portal or a webservice. Each service on the web-portal gets his own/specific interface, with user control access. Additional interfaces, like spectral annotation tools, will be implemented later as plugin.

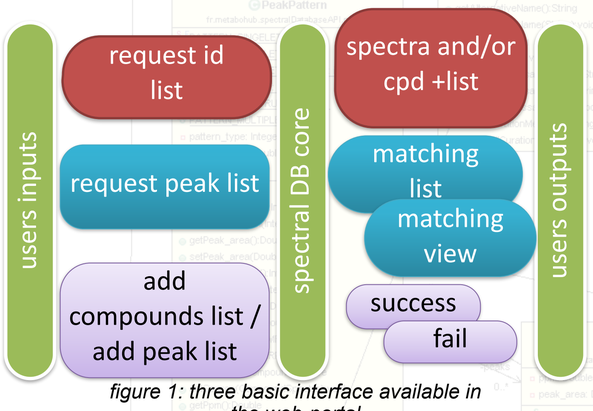
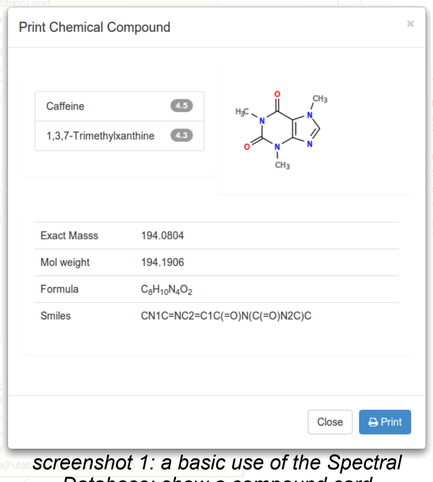


figure 1: three basic interface available in the web-portal

Interfaces and user access: some database queries are available for all users (get data), others require an authentification (add or modify elements, data curation).

"Spectral annotation tools" building is an aim of the MetaboHUB's Spectral Database. We have to create the structure to be used to host and link all data provided by chemists.

We use a user-friendly framework (twitter bootstrap) to design web interfaces.



screenshot 1: a basic use of the Spectral Database: show a compound card

computing

Code architecture and frameworks are an originality of this project. Developers use java frameworks and common generic toolbox to keep to the time scheduled and interoperability. There are three levels of architectures: the spectral database core (figure 2), implemented in both webservices and the web-portal model-view-controller (MVC) (schema 5).

Code architecture: the MVC pattern is better for code maintainability, team development and security.

Webservices are the way to open with bots data stored in the spectral database, submitting massives requests and hudge computing. We use popular technologies like REST and Json.

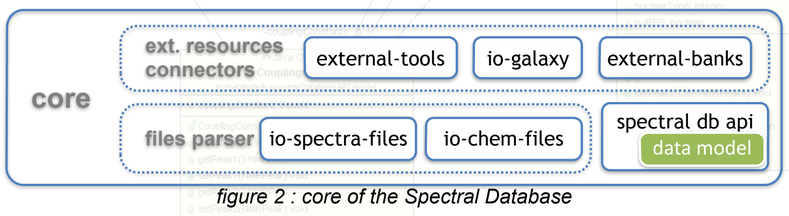
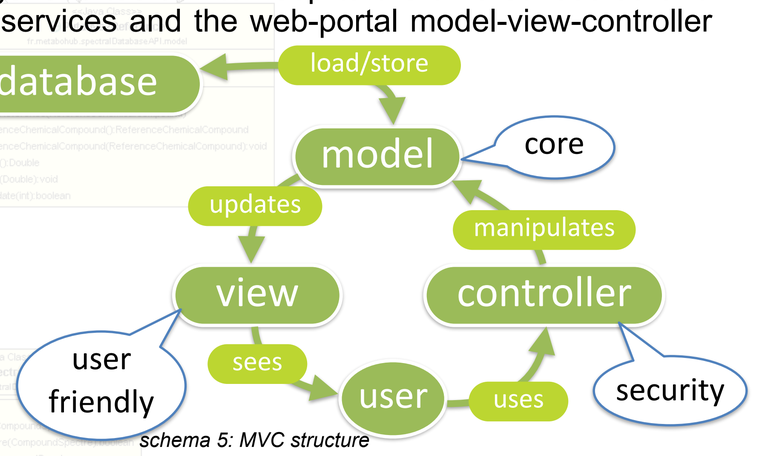


figure 2 : core of the Spectral Database

The core of the database is split in stand-alone blocks, which allow the community to re-use them in others projects. Each block is specific to one function.



schema 5: MVC structure

To store data in the database, we use the Hibernate framework. We keep the model relations and constrains, with an abstraction of technologies used for SQL requests.

Conclusion

Still in early development phase, some functional blocks are already being tested.

The first challenge was to validate this data-model and the use cases with chemists. We are currently testing the chemical library module with data provided by MetaboHUB partners.

Two milestones are coming: a first to provide a mechanism to import spectral data in the data-model and feed the database, a second to define metadata around spectral analysis.

INRA - Unite de Nutrition Humaine
Plate-Forme "Exploration du Métabolisme"
Centre de Theix
63122 Saint genes Champanelle
France

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

http://www.clermont.inra.fr/plateforme_exploration_metabolisme