ChemFlow, chemometrics using Galaxy

Virginie Rossard, Jean Claude Boulet, Fabien Gogé, Eric Latrille, Jean-Michel Roger

ChemFlow

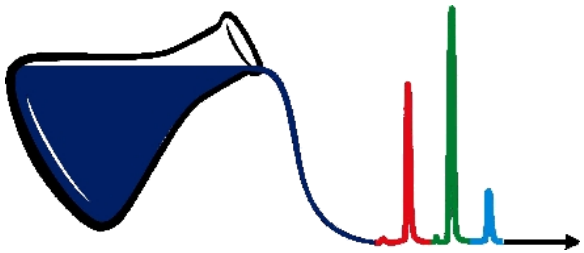
ChemFlow, chemometrics using Galaxy

V. Rossard*, J.C. Boulet*, F. Gogé**, E. Latrille*, J.M. Roger**

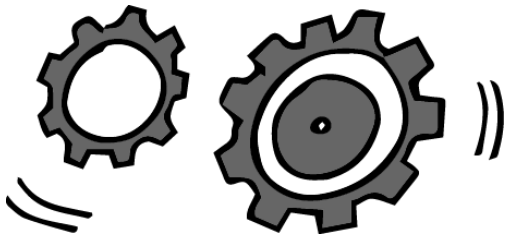
*INRA, Institut National de la Recherche Agronomique

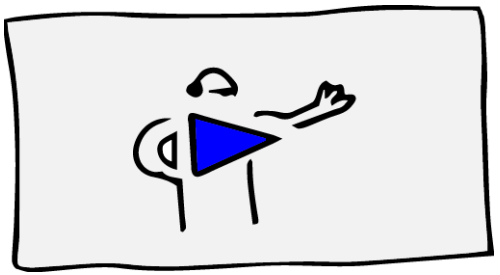
**IRSTEA, Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture

1. Chemometrics
2. The ChemFlow Tool
3. The MOOC

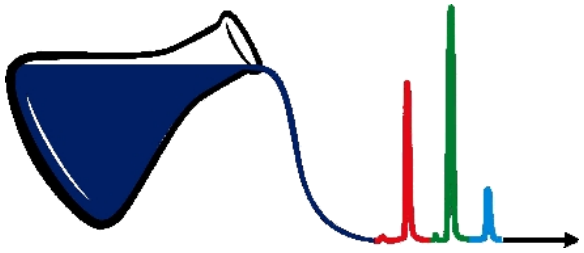


By courtesy of the Division of chemometrics,
Norwegian Chemical Society



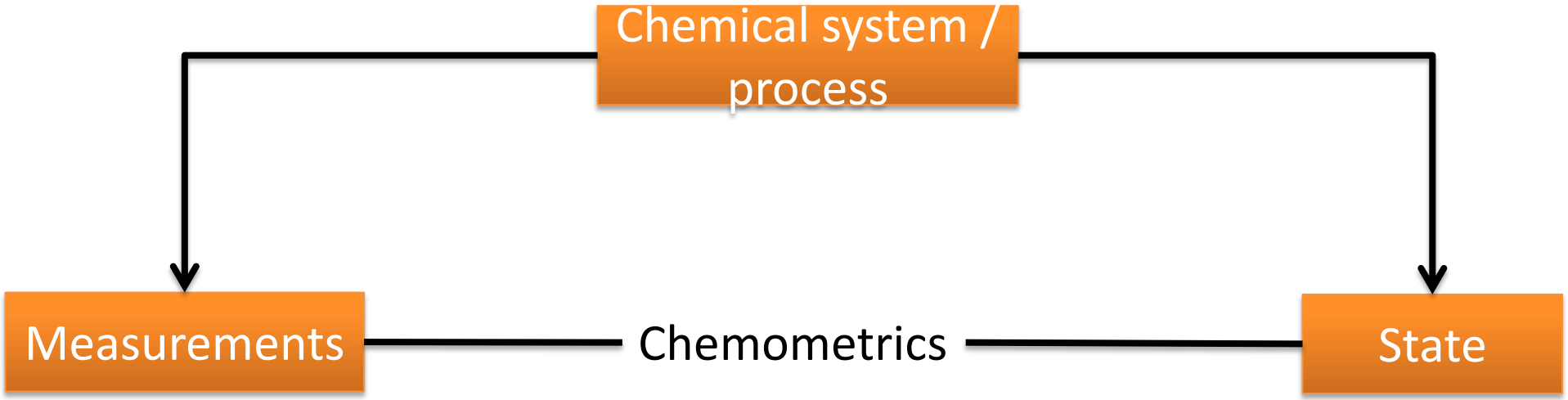


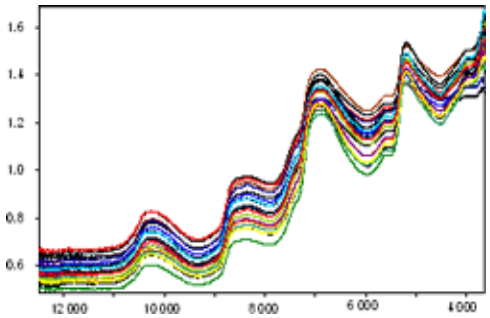
1. Chemometrics
2. The ChemFlow Tool
3. The MOOC



By courtesy of the Division of chemometrics,
Norwegian Chemical Society

Definition (International Chemometrics Society):
Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods.

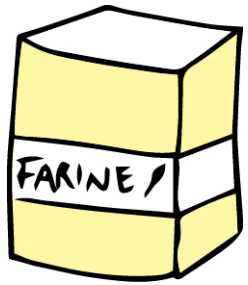


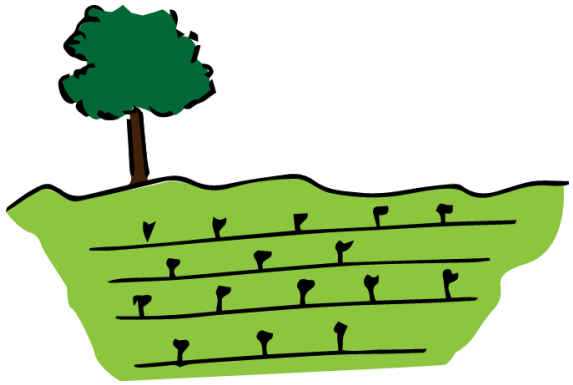


Infrared spectra

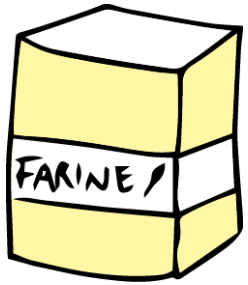
- Multivariate model
- Mathematics
- Statistics

- Gluten content of flour
- Year of production



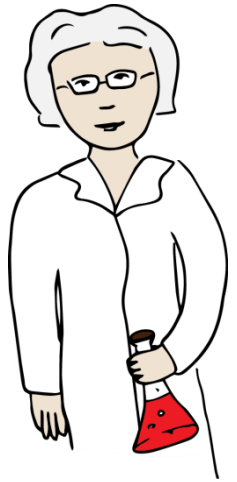


Application domain



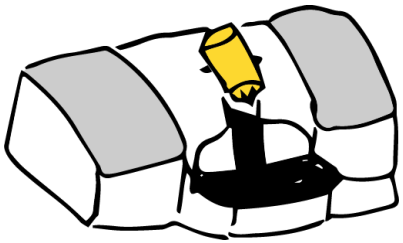
Product

Sample



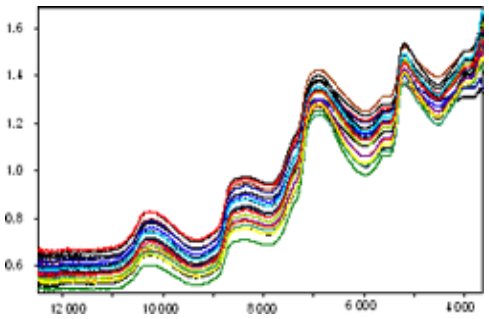
Operator

Spectrometer



- Gluten content of flour
- Year of production

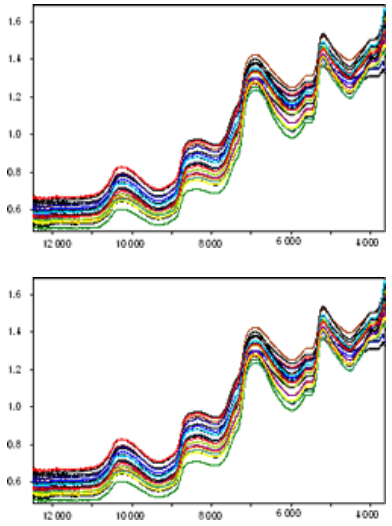
Measurements



Infrared spectra

Chemometrics

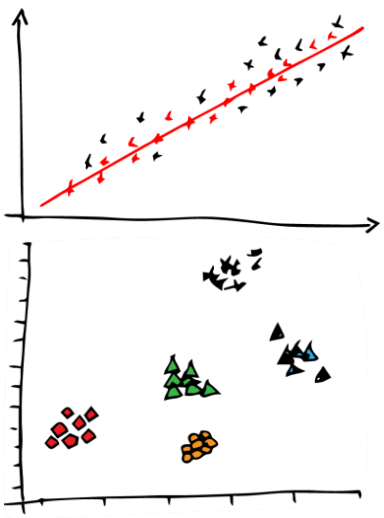
```
LOAD('MY DATA.CSV');
PLOT(X');
X=DETREND(X)
RESPLS=PLS(X,Y,10,20,1)
PLOT(RESPLS...
```

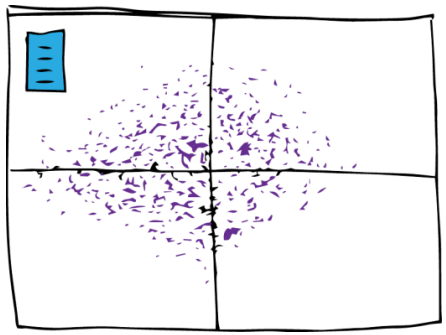


Infrared spectra

To

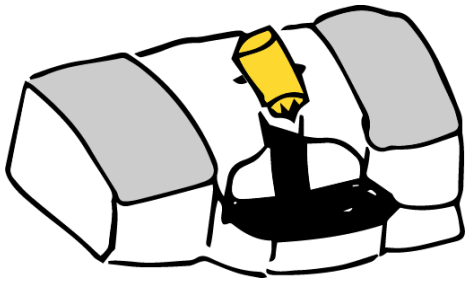
```
LOAD('MY DATA.CSV');
PLOT(X');
X=DETREND(X)
RESPLS=PLS(X,Y,10,20,1)
PLOT(RESPLS...
```

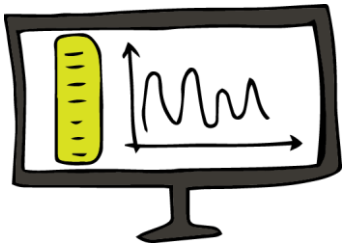




Predict biochemical content using a model

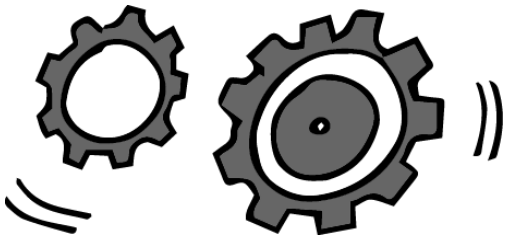
Laboratory





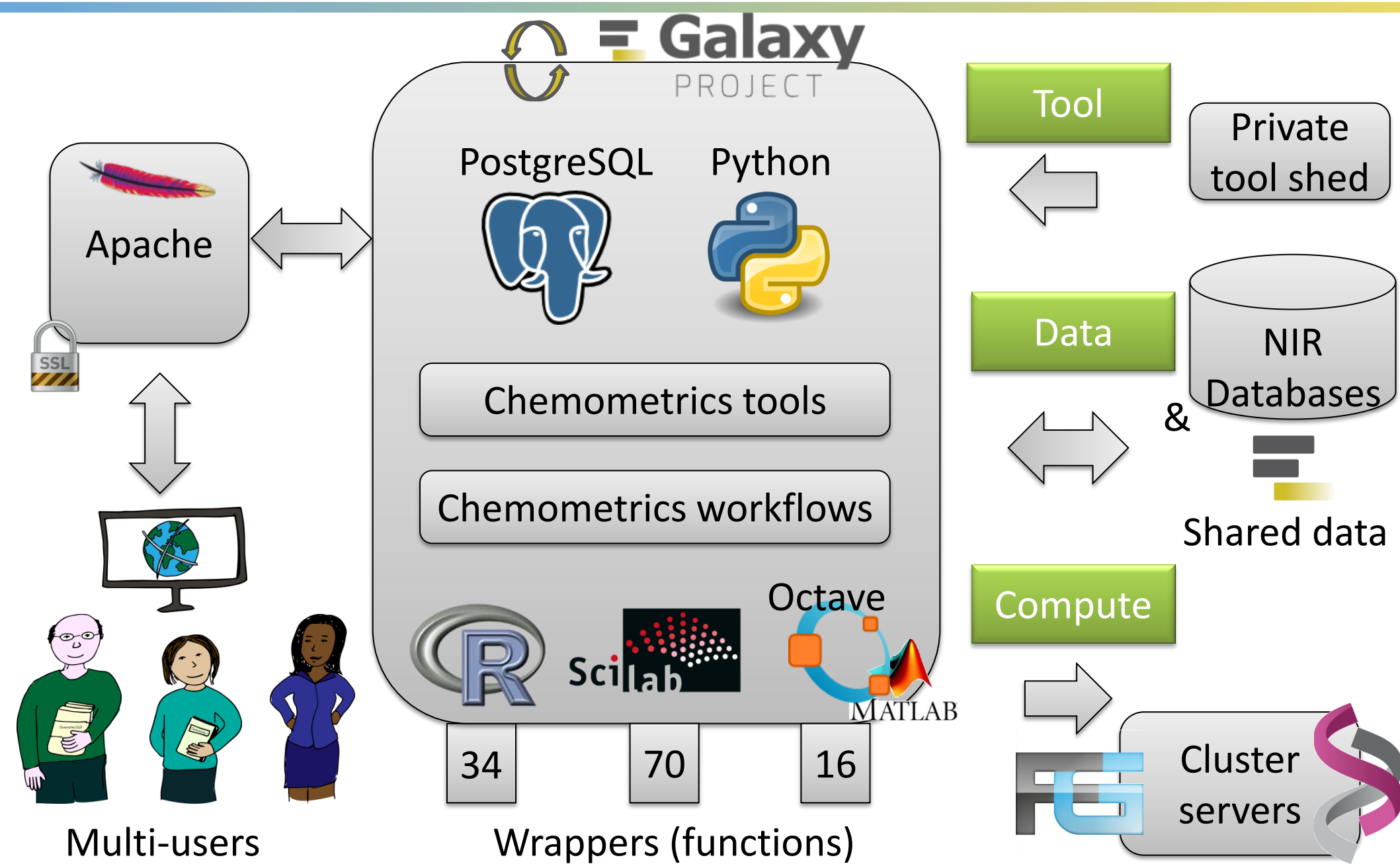
Inexpensive, faster, preserve sample

1. Chemometrics
2. The ChemFlow Tool
3. The MOOC



Common practice	Specificities and added values of ChemFlow
One software for each type of spectrometer	**Accessibility, web service**
The collaborators develop tools in different langages	Reinforce the chemometrics community by **collaborative**, **multi-users and shared** tools
Developed in various langages   	**Interoperability**
Difficulties to reproduce data treatment chains	Reproducibility and traceability: **workflow**
• User friendly software: all copyright. • Free software: programing.	**User friendly** and **free** software

Common practice	Specificities and added values of ChemFlow
One software for each type of spectrometer	**Accessibility, web service**
The co … langag…	…ty by … tools
Develo…	
Difficul… chains	…traceability. **Workflow**
• User friendly software: all copyright. • Free software: programing.	**User friendly** and **free** software


Galaxy
PROJECT
Apache
SSL
PostgreSQL
Python
Chemometrics tools
Chemometrics workflows
Octave
Scilab
MATLAB
34
70
16
Tool
Private tool shed
Data
NIR Databases
&
Shared data
Compute
Cluster servers
Multi-users
Wrappers (functions)



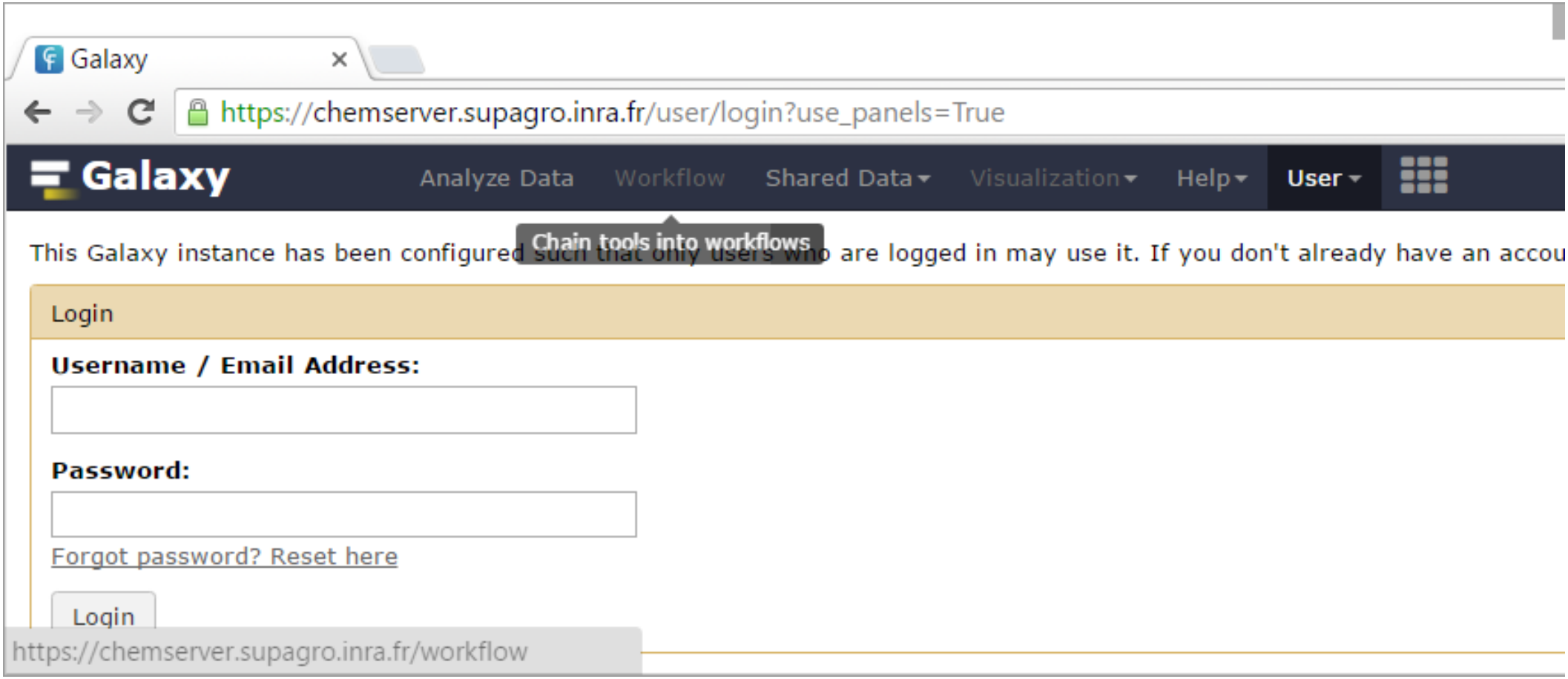

Step 1 Register an account automatically

Step 2 Connect to ChemFlow

Step 3 Upload your data OR use shared data

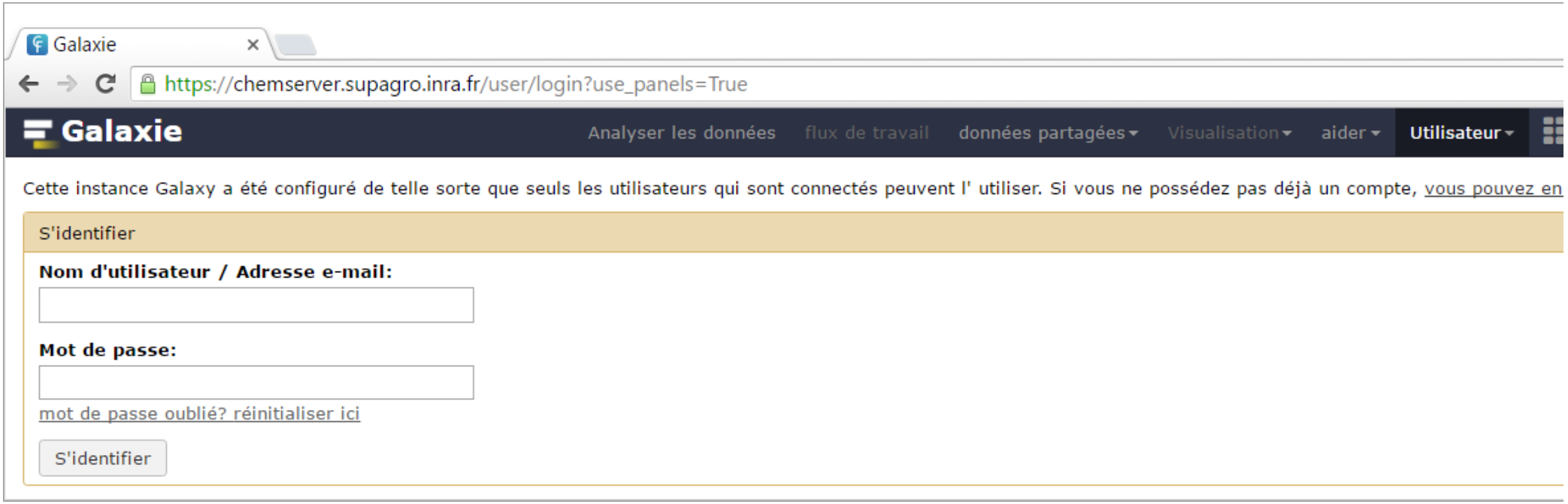
Step 4 Play chemometrics

A tool with a configuration immediately useful
CONNECT & PLAY



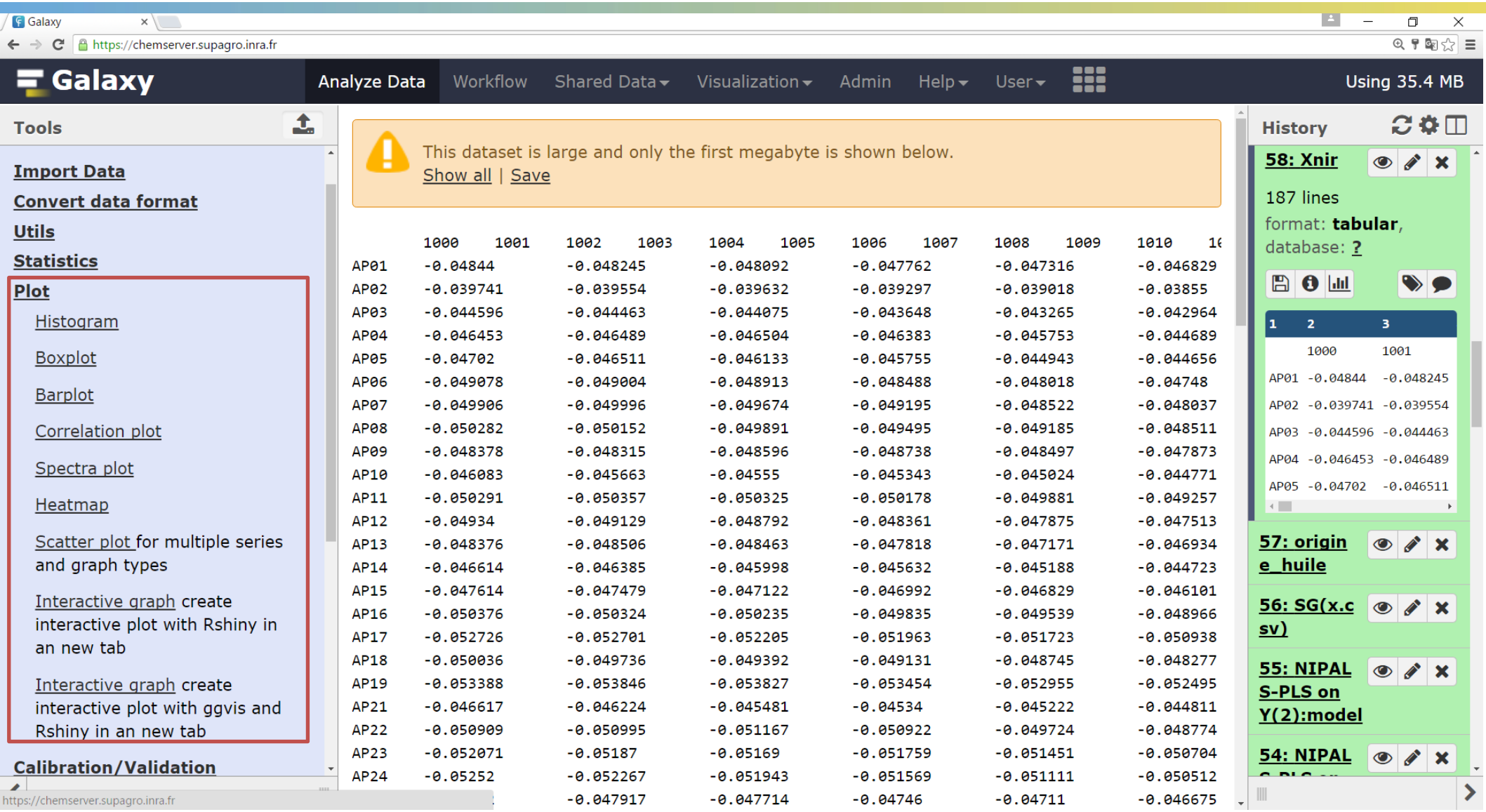
In Google Chrome, the tool is automatically translate

In Google Chrome, the tool is automatically translate

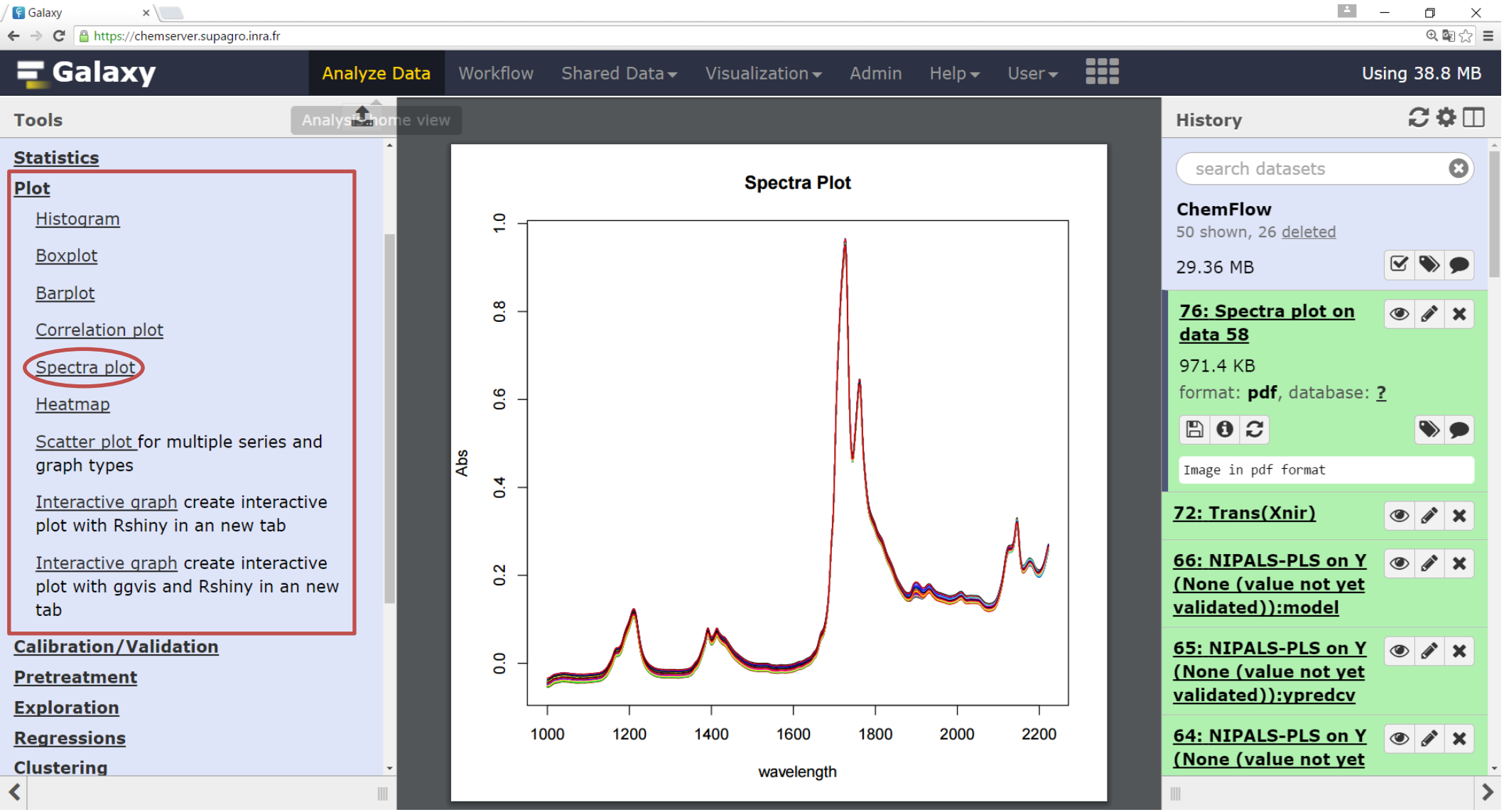


ChemFlow: home interface

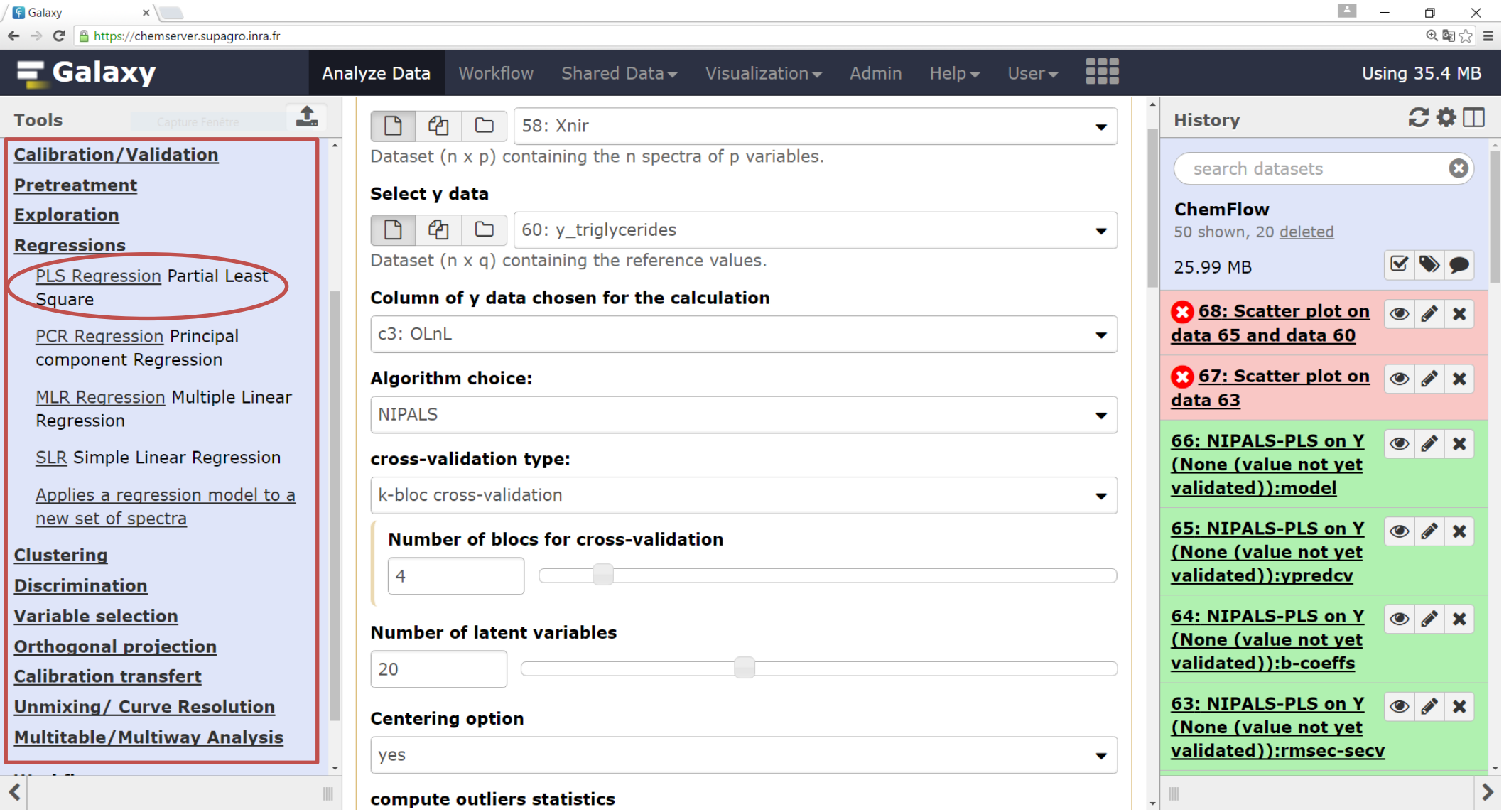
ChemFlow: plot interface

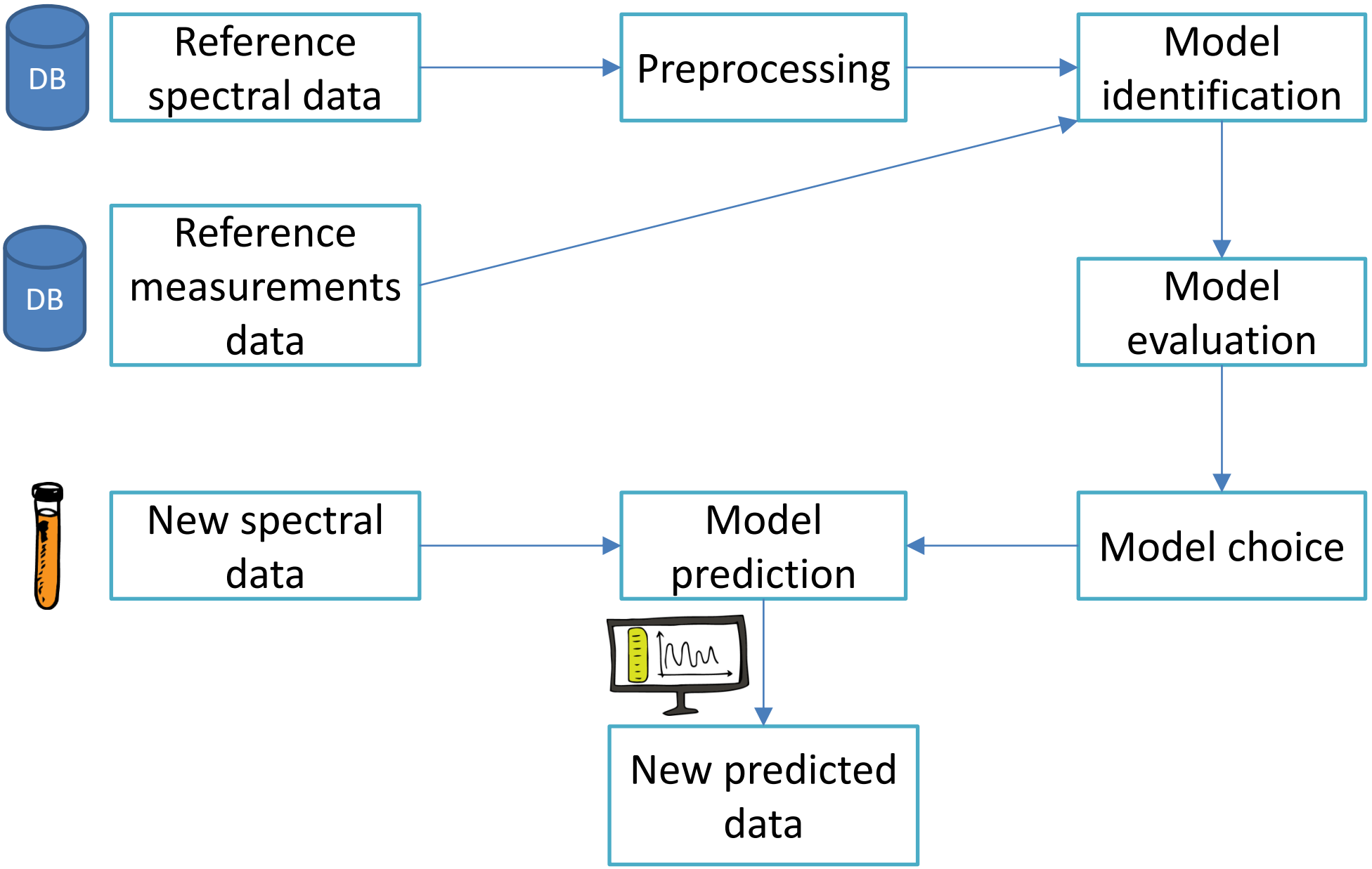


ChemFlow: plot interface

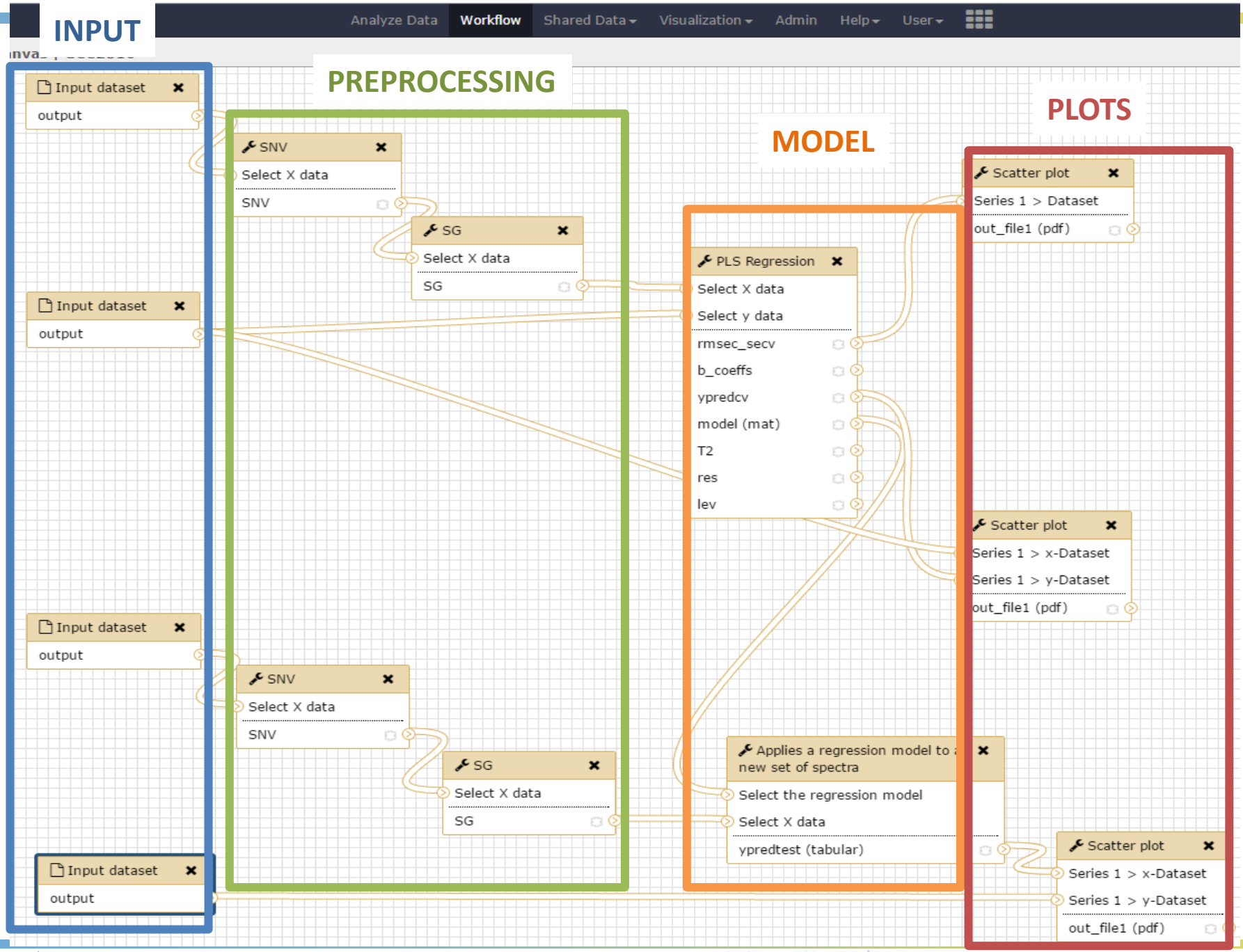


ChemFlow: chemometrics tools interface

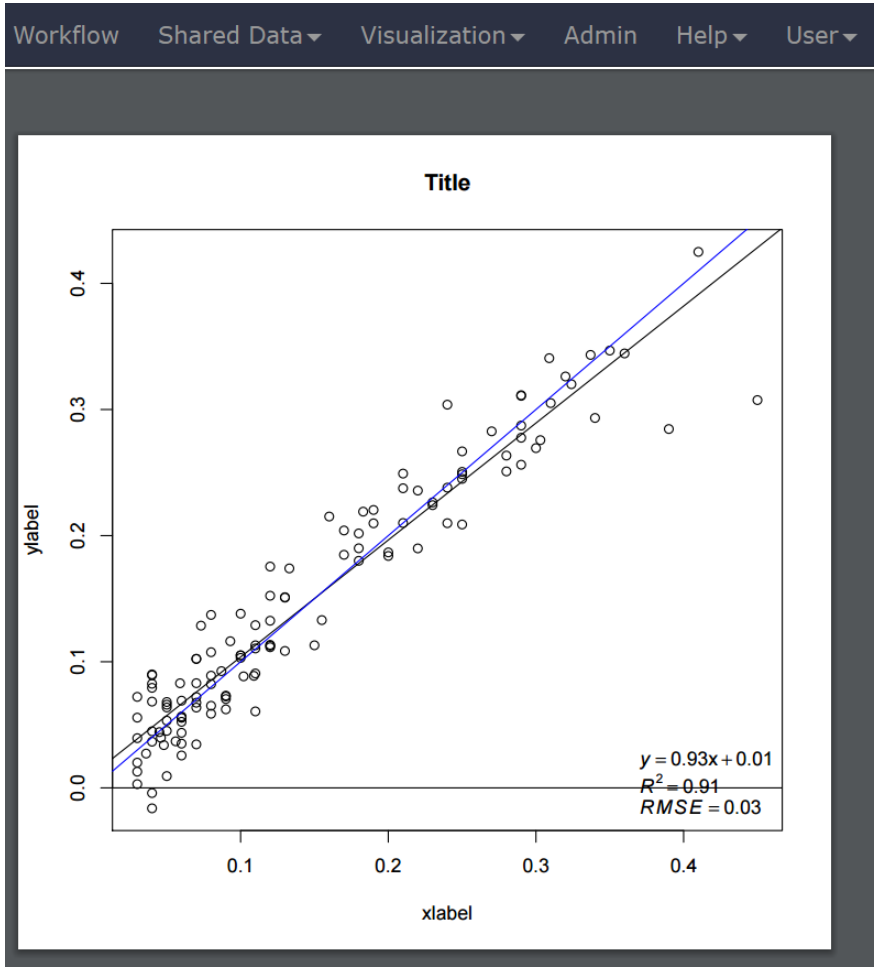


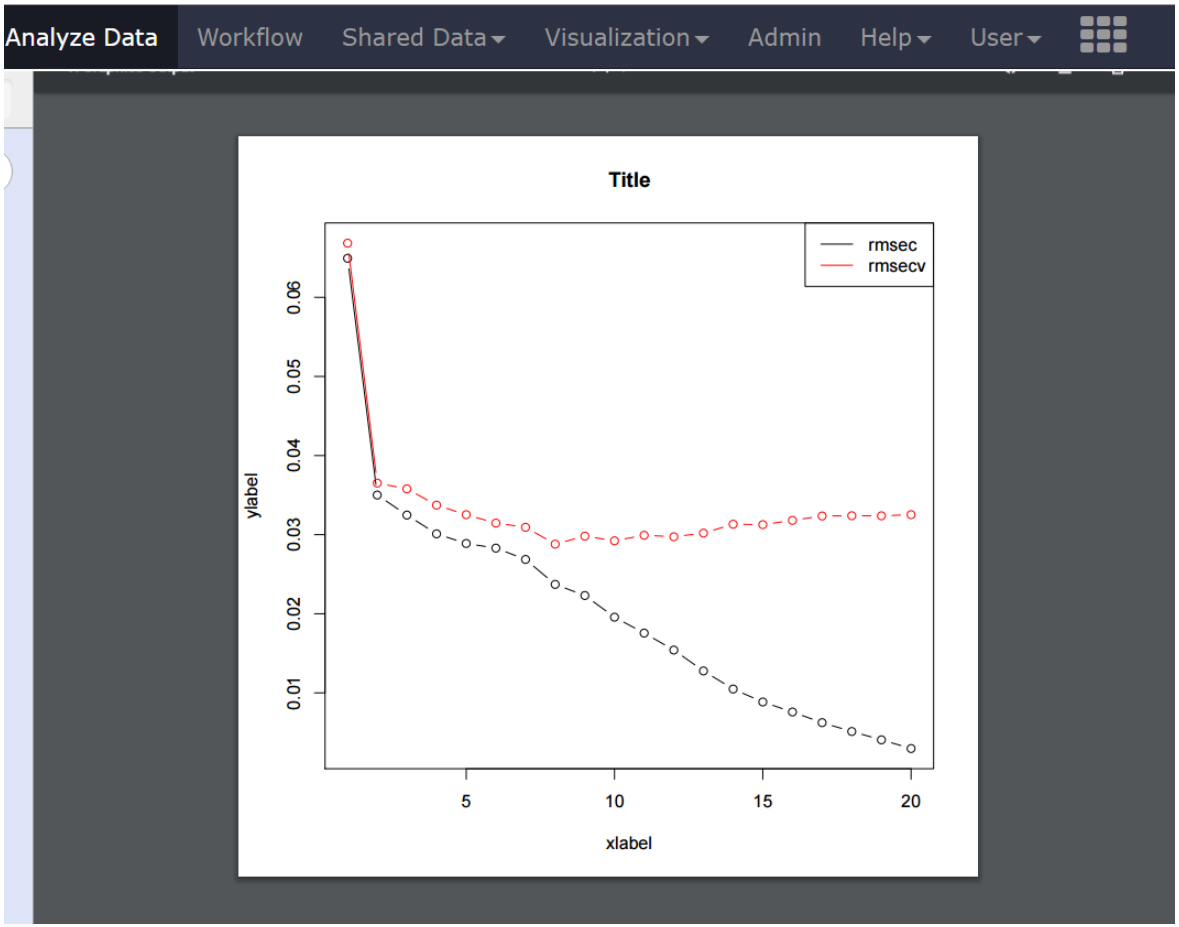

DB
Reference spectral data
Preprocessing
Model identification
DB
Reference measurements data
Model evaluation
New spectral data
Model prediction
Model choice
New predicted data

ChemFlow: chemometrics workflow

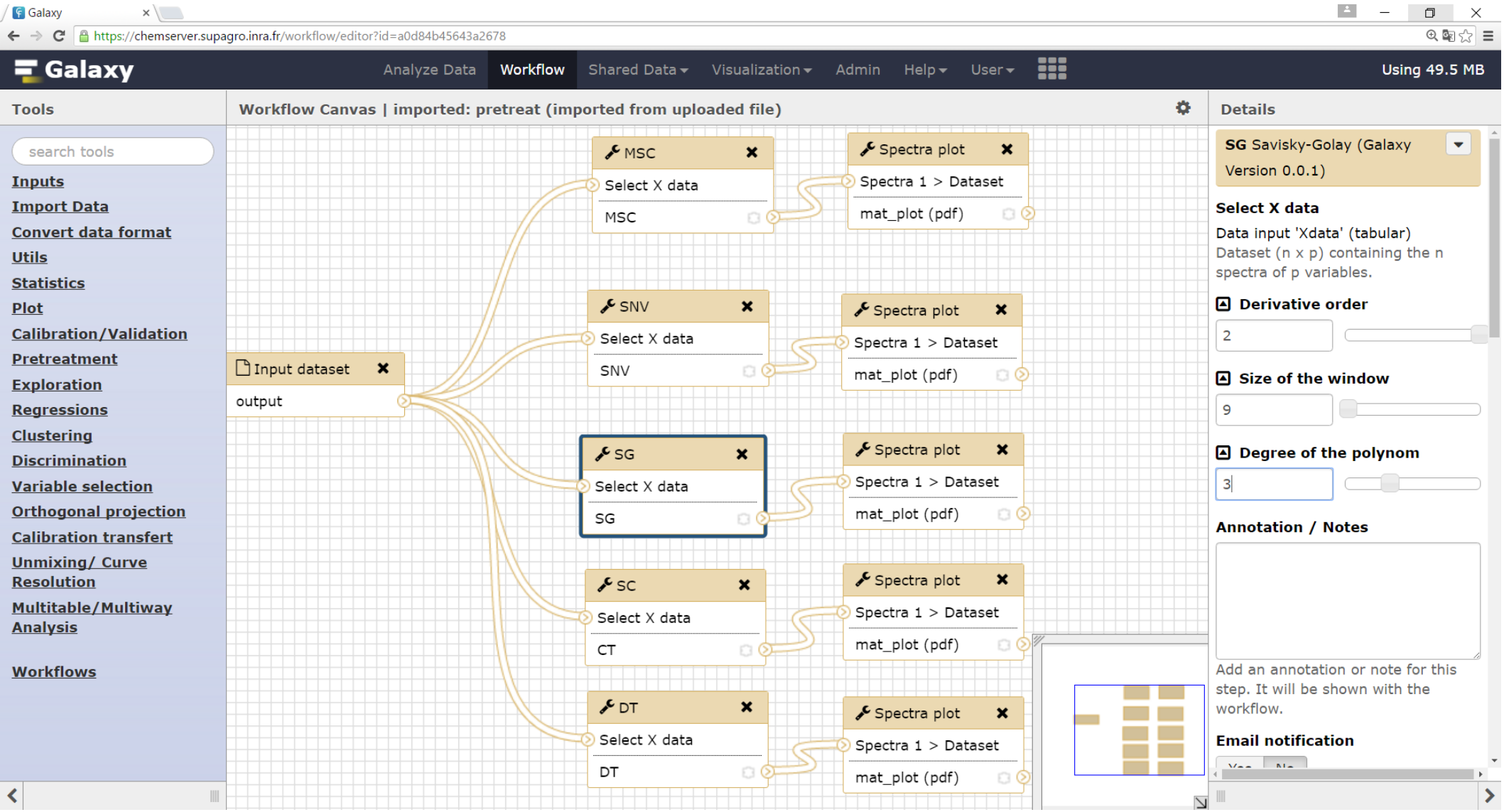


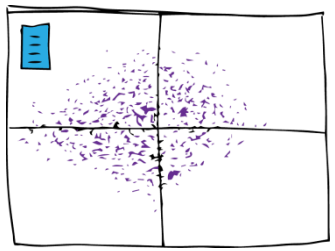
ChemFlow: graphical outputs





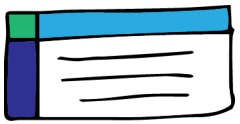
ChemFlow: functionnal test workflow





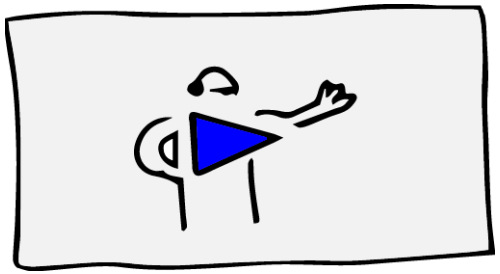
–Test and validation tools (Planemo…)

–Interactive graphics (R-Shiny)



–Connection with specific databases

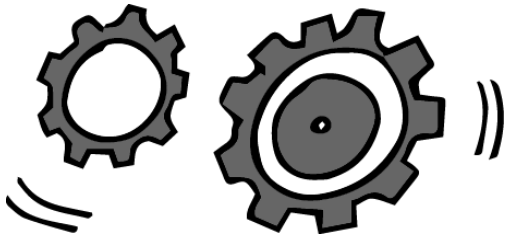
The project: the three pillars



CheMOOCS

-MOOC, Massive Open Online Course

Free and available for everybody

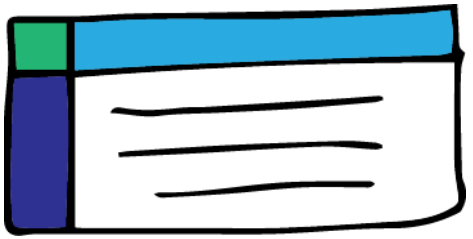


ChemFlow

- tool

ChemFlow

Using a collaborative conception

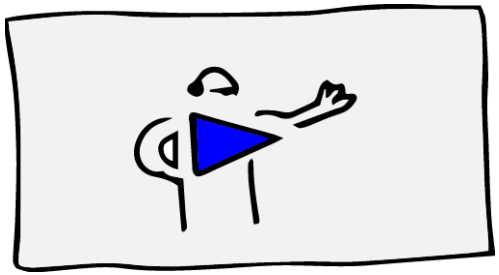


ChemData

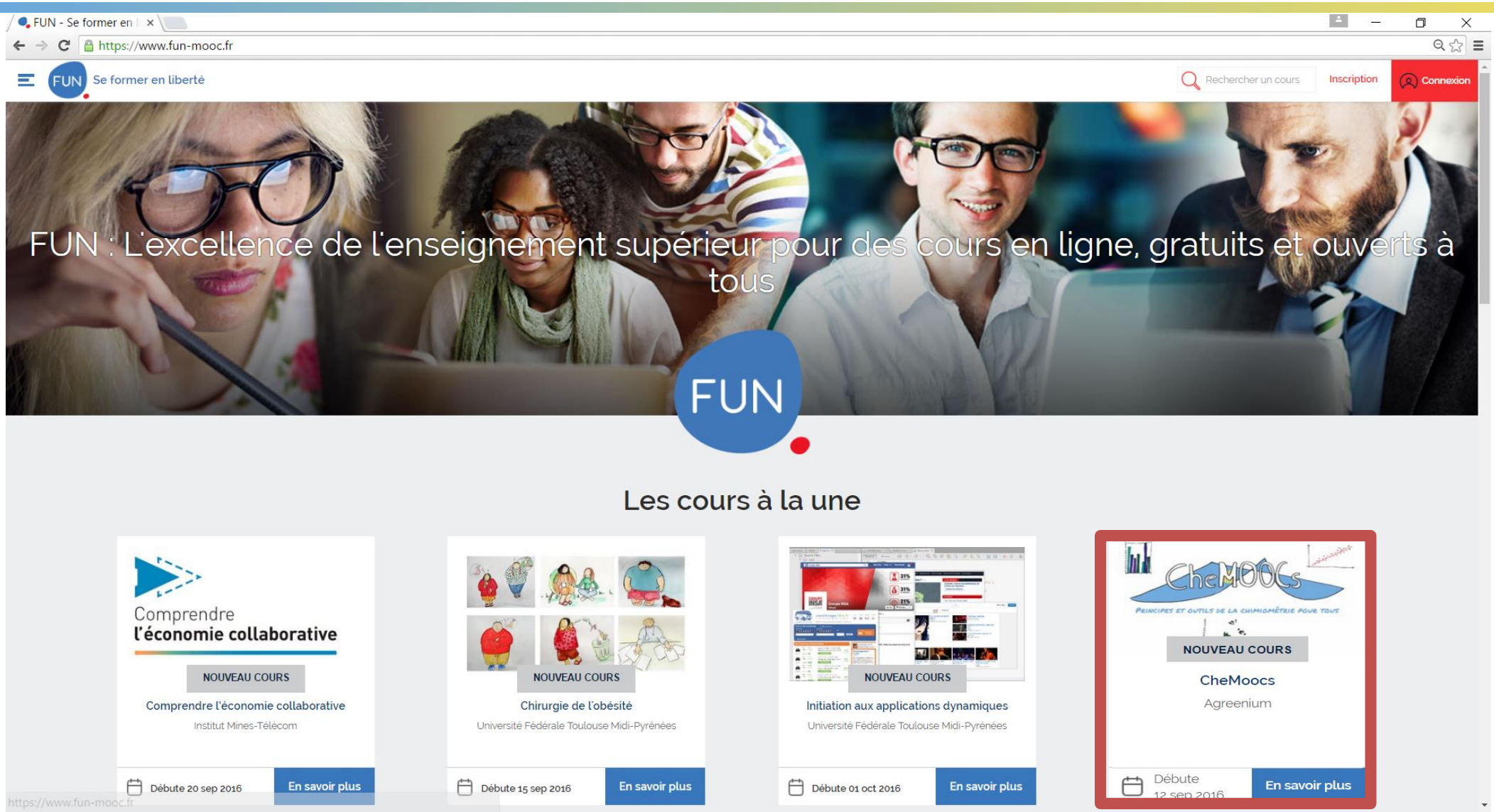
- database

Data sharing

1. Chemometrics
2. The ChemFlow Tool
3. The MOOC

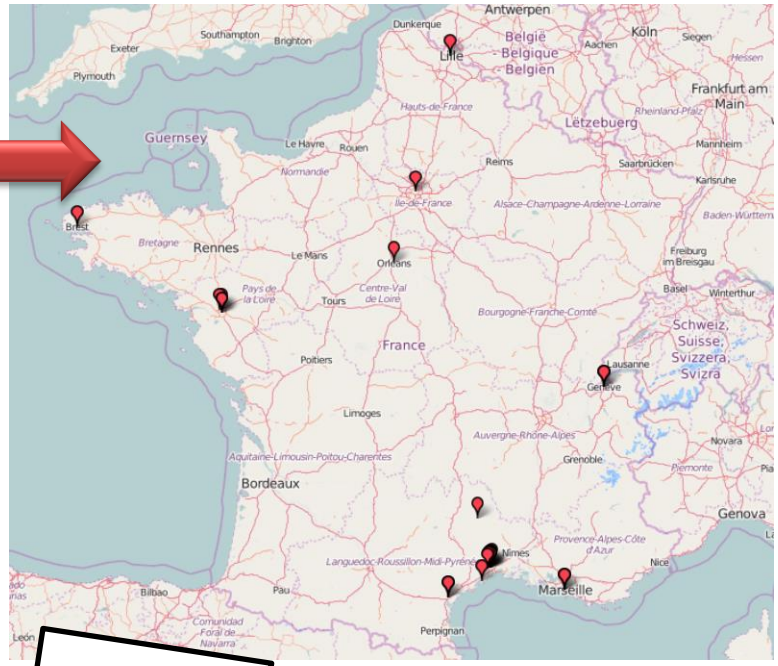


The MOOC: using FUN, a french course platform



Courses in french langage

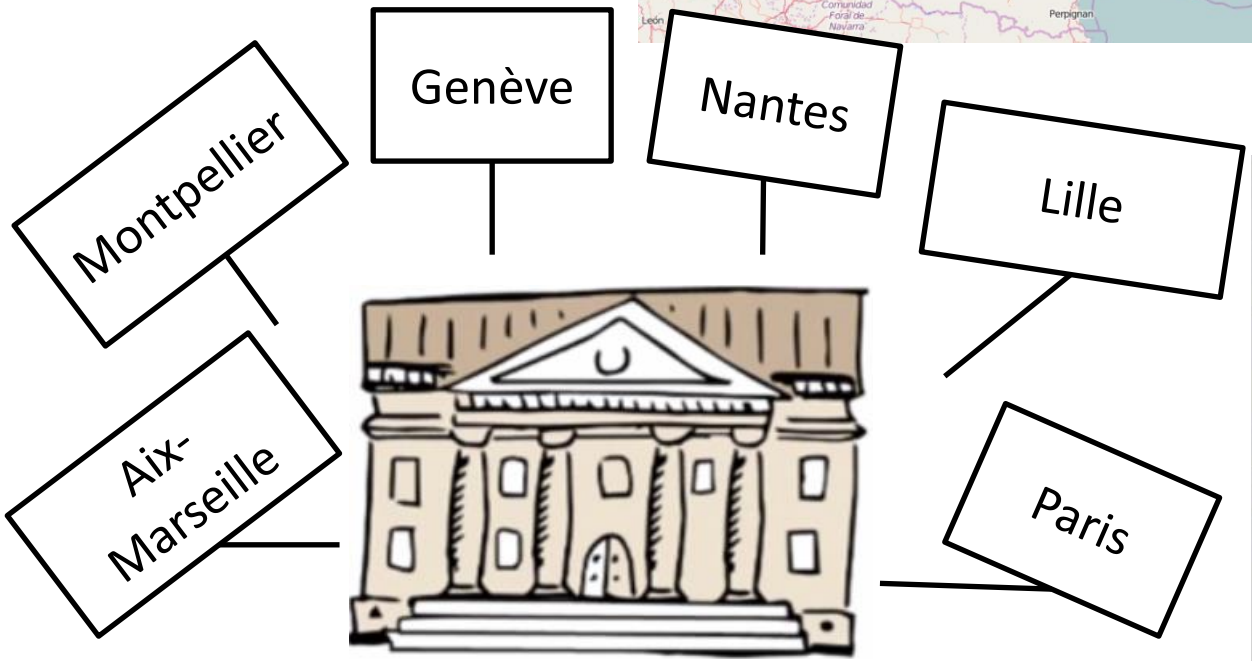
FUN, France Université numérique
France Digital University



IT team

24 collaborators

- Academic
- Industrial



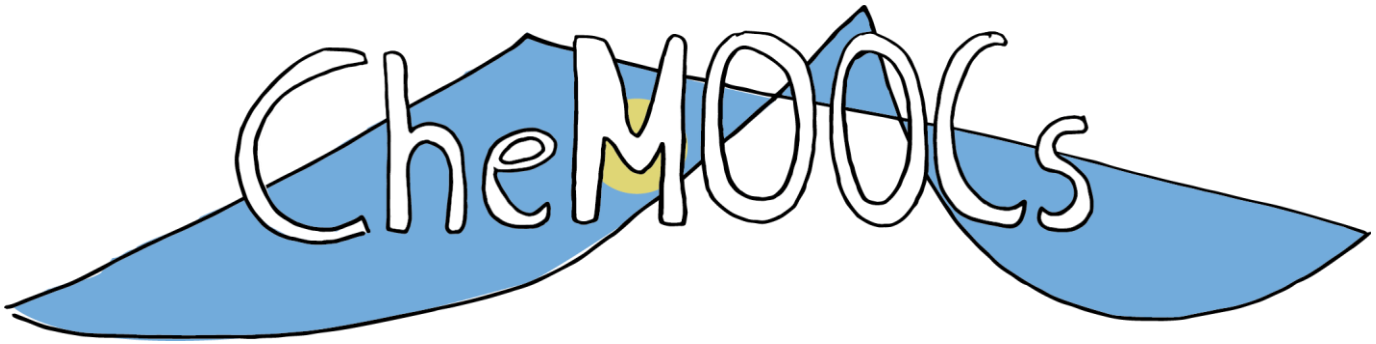
– Targeted audience

- Students: initial learning
- Professionals (industrial, teacher, researcher): training

– Aims:

- New competence
- Competence improvement

- Launch on FUN in French language: September 12th 2016
- Registration from June 15th to www.FUN-MOOC.fr
- MOOC from September 12th to November 10th



26 registrations per day

THANKS

The EIC, *Equipe d'Informatique Collective*, team in Montpellier at Supagro/INRA

Illustration et Logo (sous licence CC BY SA) par Fanny Monod (fanzyl.net) et Marc Lanssens de SupAgro Montpellier

agropolis fondation

This project was funded under the reference ID-1401-005 from the program "Investissements d'avenir" (Labex-Agro : ANR610-LABEX-0001-01)