# Theory & the VO

Franck Le Petit

# Theory in the VO

Theory I.G. : 2004

Many evolutions of the goals

- particle simulations
- cosmological simulations
- all kind of simulations

**Whitepaper Euro-VO DCA**

Editors: G. Lemson, H. Wozniak, J. Zuther

Date: 2008

EURO VO DATA CENTRE ALLIANCE

RI031675

EuroVO-DCA

The European Virtual Observatory Data Centre Alliance

COORDINATION ACTION

RESEARCH INFRASTRUCTURE

COMMUNICATION NETWORK DEVELOPMENT

**D11 – TEG REPORT: FRAMEWORK FOR THE INCLUSION OF THEORY DATA AND SERVICES IN THE VOBS**

Due date of deliverable: 31/10/2008

Actual submission date: 16/12/2008

Start date of project: 01/09/2006 Duration: 28 month

MPG Final version

| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) | | |
|---|---|---|
| **Dissemination Level** | | |
| **PU** | Public | X |
| **PP** | Restricted to other programme participants (including the Commission Services) | |
| **RE** | Restricted to a group specified by the consortium (including the Commission Services) | |
| **CO** | Confidential, only for members of the consortium (including the Commission Services) | |

# Why to publish simulations in the Virtual Observatory ?

## Why to *publish theoretical data* ?

- Access to the data described in your publications, so readers can extend the work ➔ **increase the impact of papers**
- readers can verify the results ➔ **increase the quality of papers**
- allows to benchmark other works with your results / **reproductibility of results**
- incresingly **mandated by funding agencies**
- **showcases** for future proposals
- …

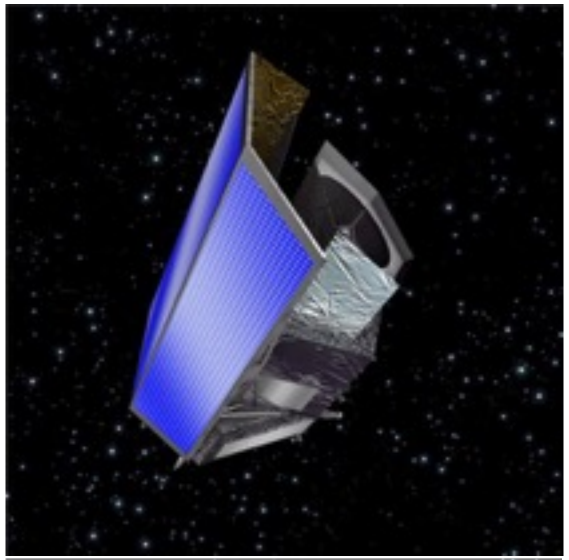
## Why to publish theoretical data *in the Virtual Observatory* ?

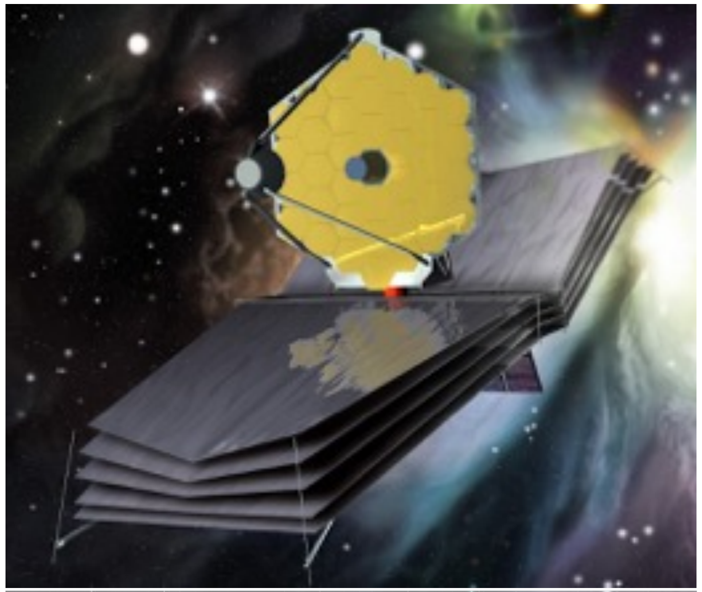
- makes results available in a **standard way**
- forces you to think carefully about your results and **improves re-usability for you**
- *maybe not give obvious benefit to you but you may agree that if others do it, you would have an easier job using their data*
- facilitates **comparison models - observations**

# Large projects

**Large projects require simulations** to prepare and interpret the observations

Ex: JWST, EUCLID, LSST, SKA, PLATO, …



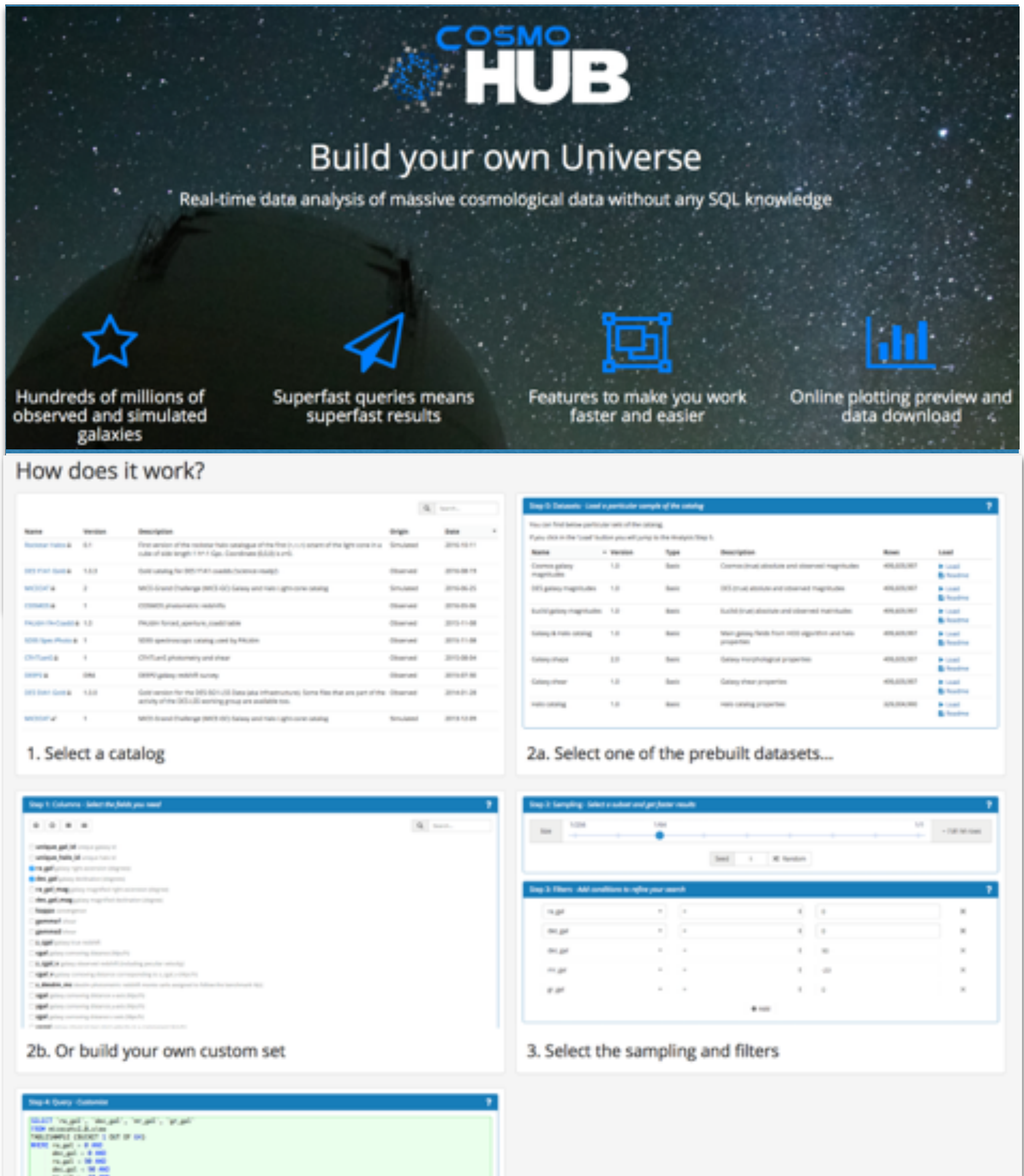


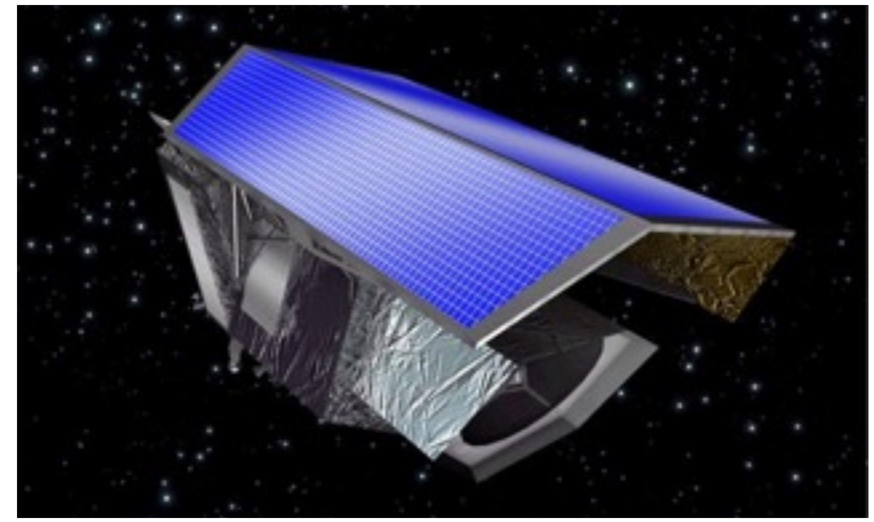
**InterOp 2016, Stellenbosch**
Presentation of several large projects: SKA, LSST, …
One general comment:
**Need to publish & access theoretical data via the Virtual Observatory**

# Large projects

## COSMOHUB / MICE

https://cosmohub.pic.es/#/home



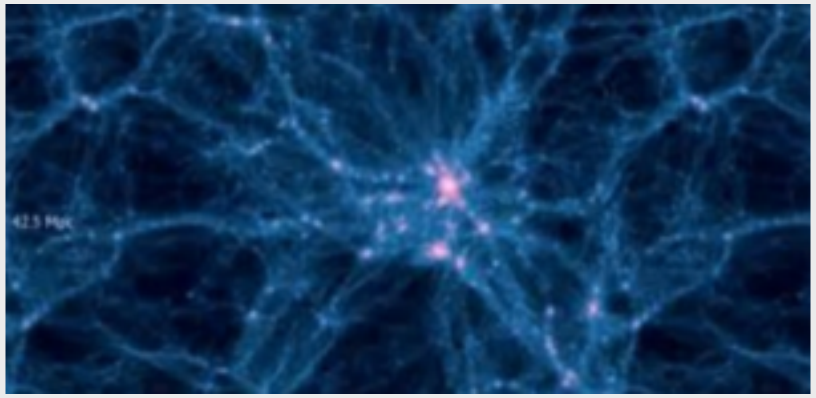



- Cosmological simulations: MICE / **Euclid**
- Access to post-processed data:
  - Halo catalog, BAO, …

**But no VO compatibility**

# Numerical simulations

Numerical simulations goes from:
- models that can be run on **simple computers**
- large simulations that require the largest **super-computers**
  Astronomy is a major driver of computational developments



State-of-the-art simulations represent **large investments**

- Grand challenge - large simulations done one supercomputer
  ~10 people teams required for massive simulations

  Exemple: Illustris simulation (2014) - 19 million CPU hours (2000 years)
  ~ 10 authors

- State-of-the-art codes require years of development by a team with several expertise
  ➔ need to share results of these codes

- Simulations are so "rich" that the developing team cannot exploit fully the simulations
  ➔ third party teams can use the simulations to do further analysis

**Many teams want to publish and share their simulations**

# Publication of theoretical data in astrophysics

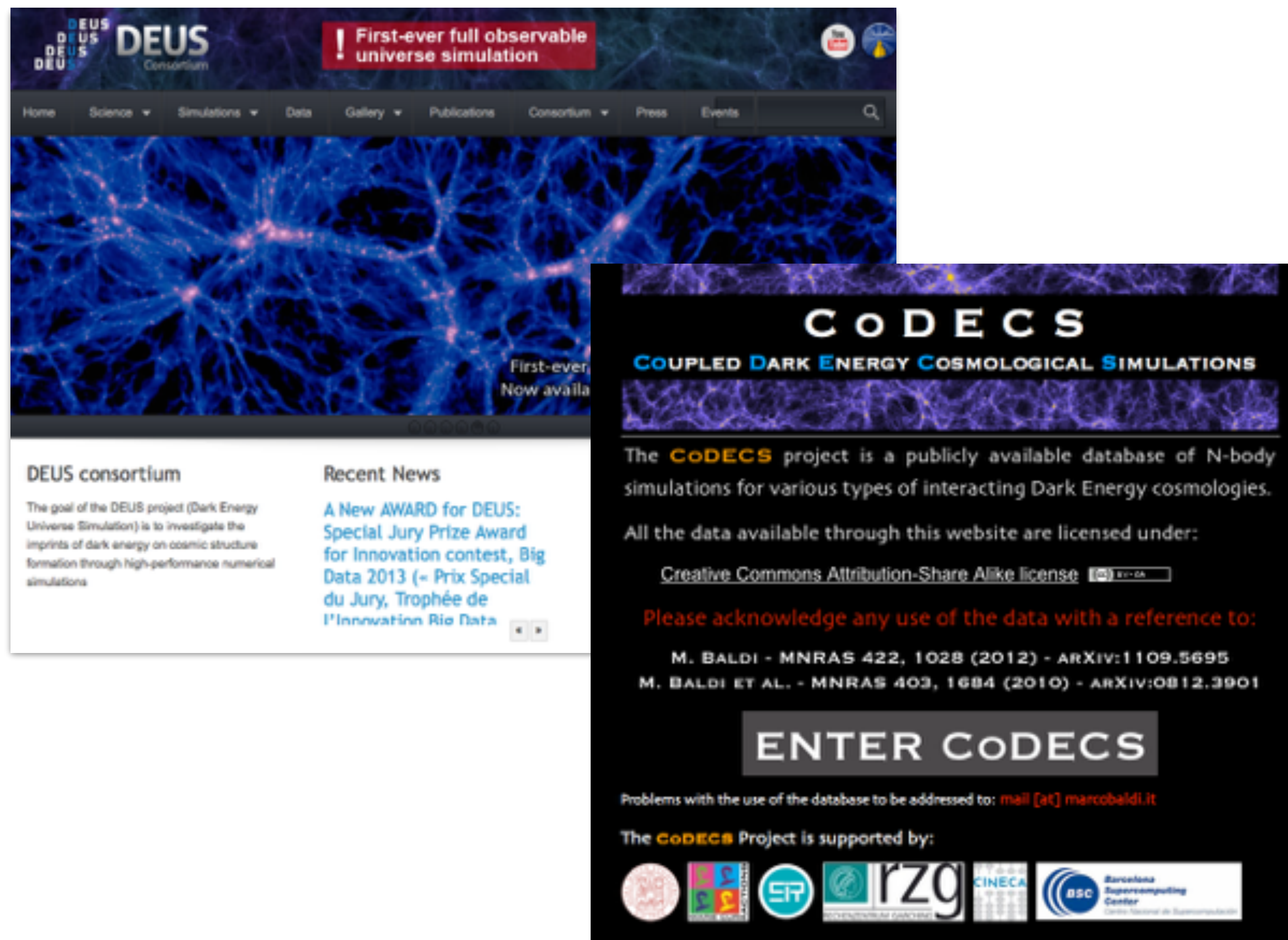
Many services to publish theoretical data have been developed in the past years

Simple search on Google and in the litterature:
These last years:

- large number of services that publish public theoretical data
- several tens of services publishing simulations

Many fields are concerned:

- Cosmology & Galaxies
- Stellar physics
- Planetary science
- Interstellar medium
- …

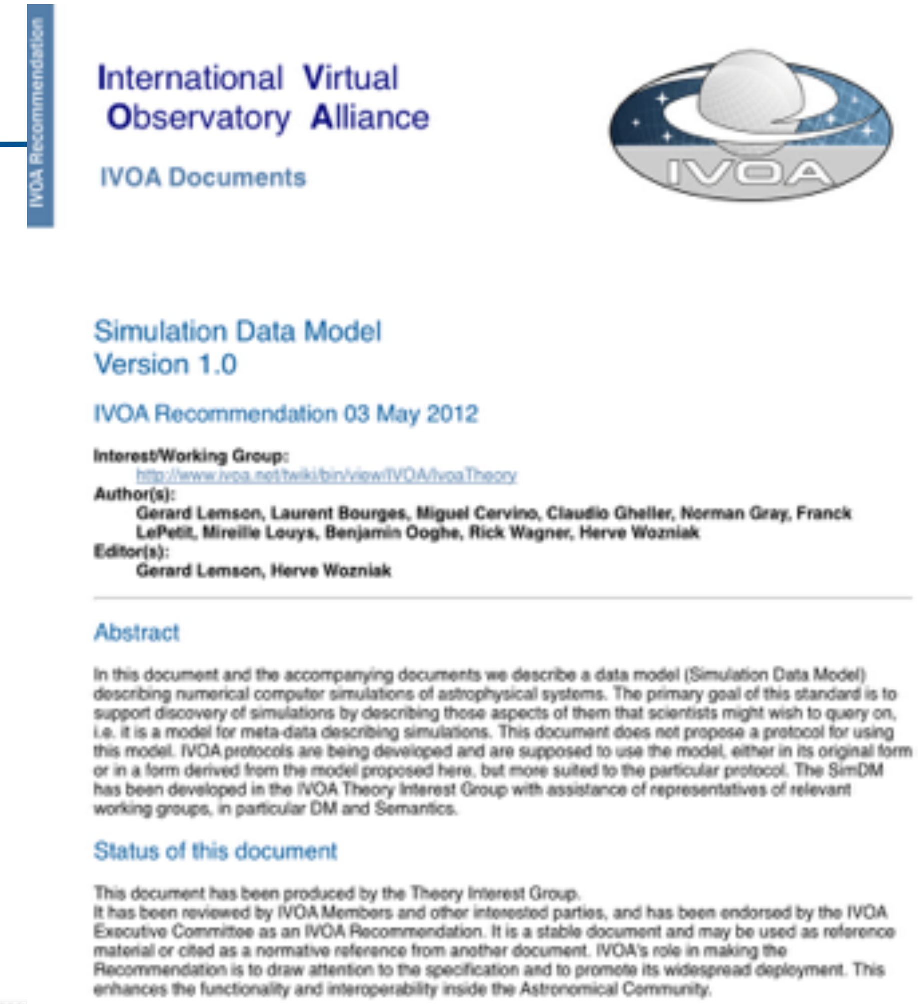



Most of these services are not VO-compliant

# IVOA Standards for Theory

## ① Simulation Data Model (SimDM)

- Description of simulations
- Meta-model
- Designed to describe all (?) simulations


IVOA Recommendation

International Virtual Observatory Alliance

IVOA Documents

Simulation Data Model
Version 1.0

IVOA Recommendation 03 May 2012

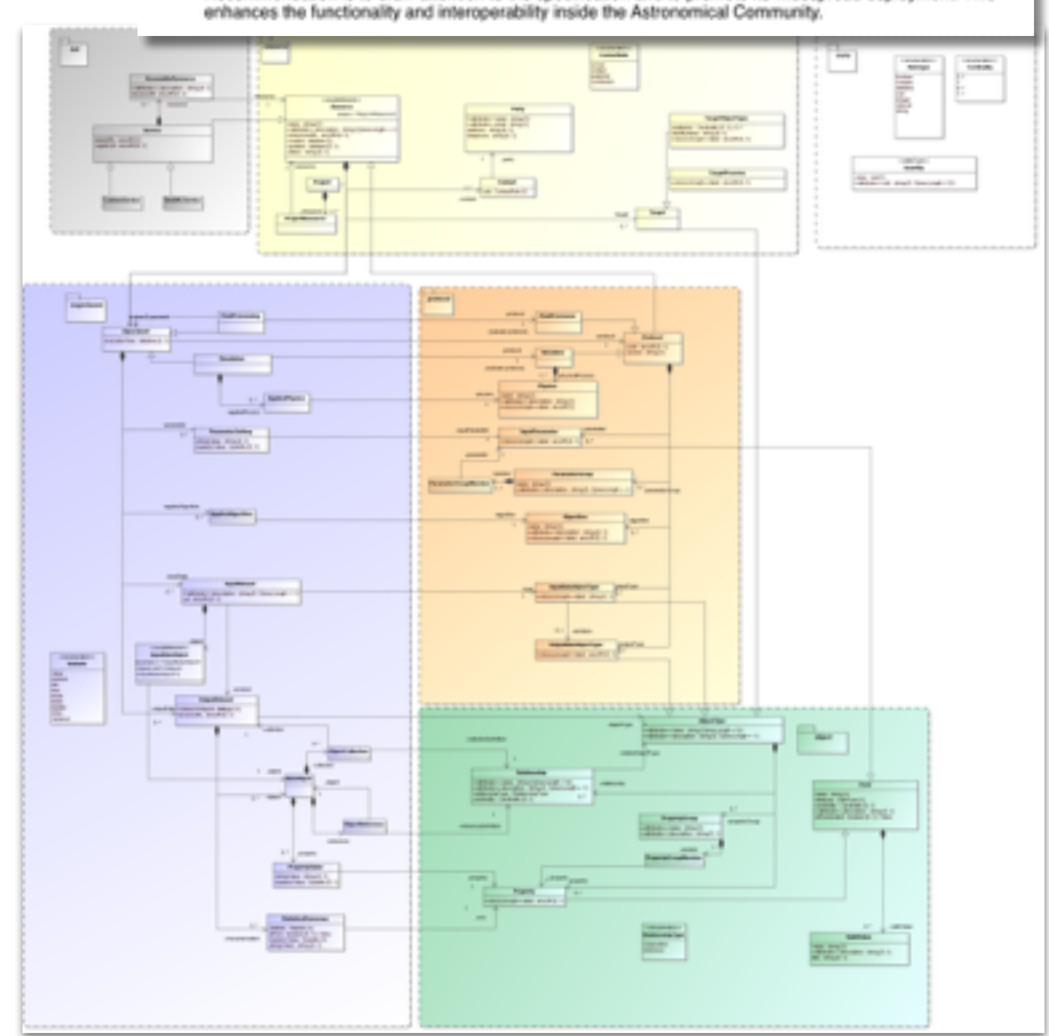
**Interest/Working Group:**
http://www.ivoa.net/twiki/bin/view/IVOA/IvoaTheory
**Author(s):**
Gerard Lemson, Laurent Bourges, Miguel Cervino, Claudio Gheller, Norman Gray, Franck LePetit, Mireille Louys, Benjamin Ooghe, Rick Wagner, Herve Wozniak
**Editor(s):**
Gerard Lemson, Herve Wozniak

Abstract

In this document and the accompanying documents we describe a data model (Simulation Data Model) describing numerical computer simulations of astrophysical systems. The primary goal of this standard is to support discovery of simulations by describing those aspects of them that scientists might wish to query on, i.e. it is a model for meta-data describing simulations. This document does not propose a protocol for using this model. IVOA protocols are being developed and are supposed to use the model, either in its original form or in a form derived from the model proposed here, but more suited to the particular protocol. The SimDM has been developed in the IVOA Theory Interest Group with assistance of representatives of relevant working groups, in particular DM and Semantics.

Status of this document

This document has been produced by the Theory Interest Group.
It has been reviewed by IVOA Members and other interested parties, and has been endorsed by the IVOA Executive Committee as an IVOA Recommendation. It is a stable document and may be used as reference material or cited as a normative reference from another document. IVOA's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability inside the Astronomical Community.



Implemented on:

- cosmological simulations
- MHD simulations
- interstellar medium micro-physics models (PDRs)
- …

➔ proof of the versatility

# IVOA Standards for Theory

## ② Simulation Data Access protocol (SimDAL)

- SimDAL Repository
- SimDAL Search
- SimDAL Cutout

allow to discover & retrieve simulations

## ③ Semantics / SKOS vocabularies

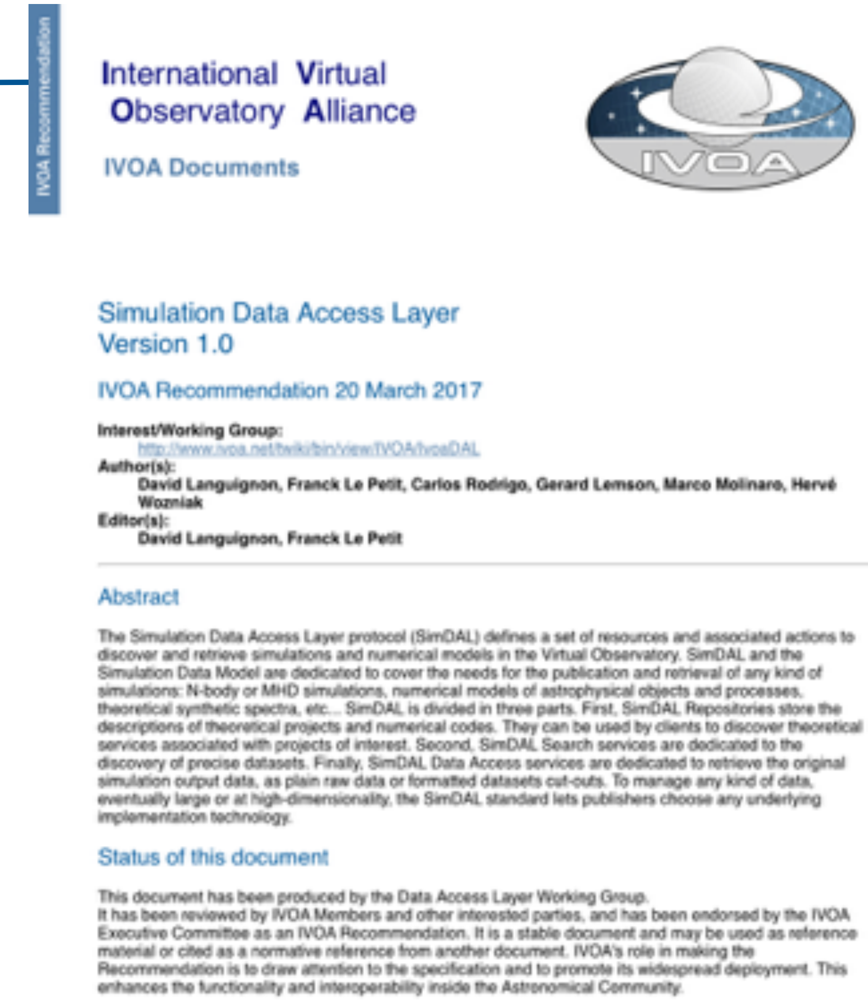
- Algorithms
- AstronomicalObjects
- DataObjectTypes ~ 700 concepts
- PhysicalProcesses
- PhysicalQuantities

## ⊗ But no common raw data format

No standard raw data format because of the **heterogeneity** (no FITS for example)

➔ a limitation to use the VO

Recommendation to use VO-Table / FITS whenever possible



IVOA Recommendation

International Virtual Observatory Alliance

IVOA Documents

Simulation Data Access Layer
Version 1.0

IVOA Recommendation 20 March 2017

**Interest/Working Group:**
http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDAL
**Author(s):**
**David Languignon, Franck Le Petit, Carlos Rodrigo, Gerard Lemson, Marco Molinaro, Hervé Wozniak**
**Editor(s):**
**David Languignon, Franck Le Petit**

Abstract

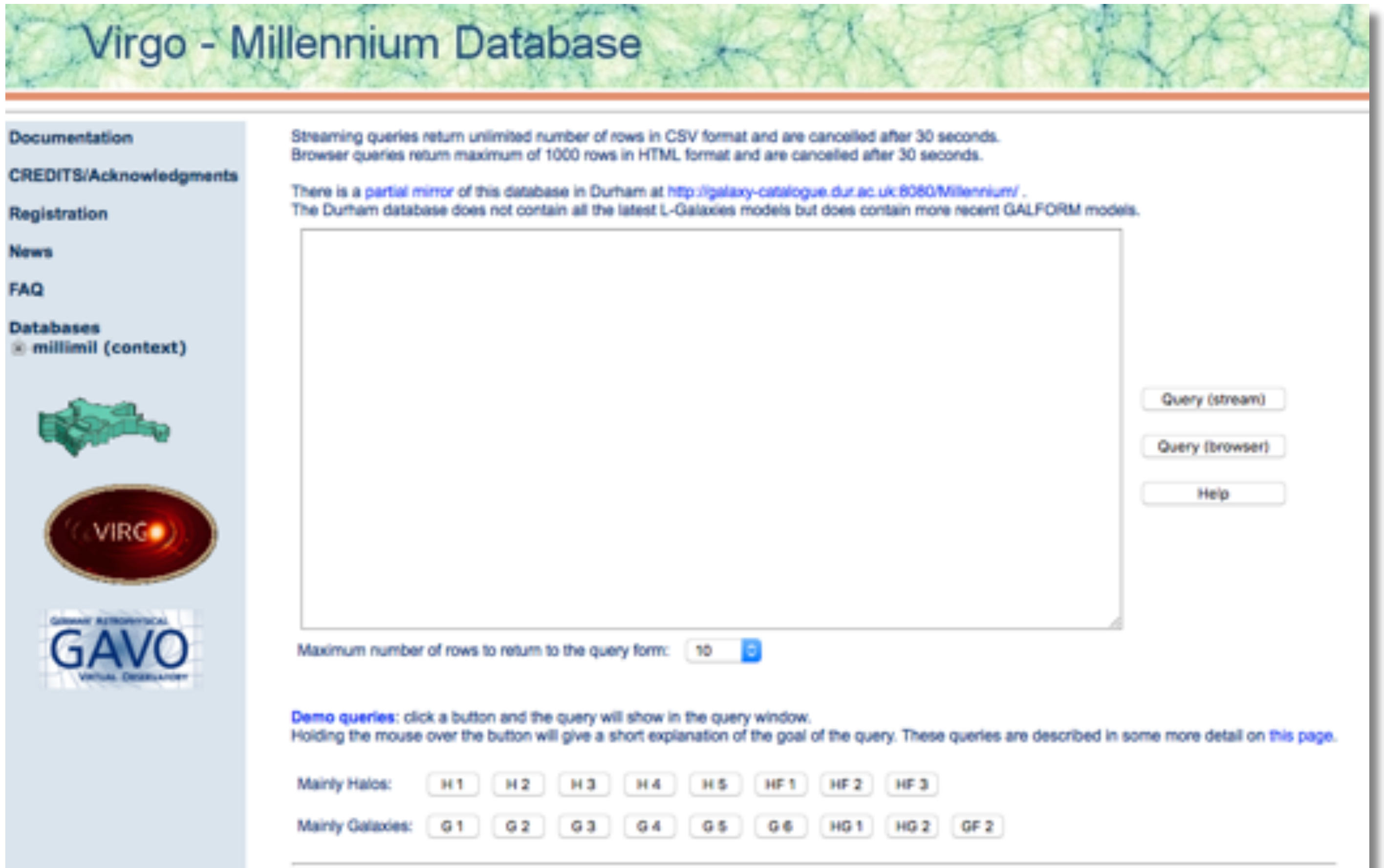
The Simulation Data Access Layer protocol (SimDAL) defines a set of resources and associated actions to discover and retrieve simulations and numerical models in the Virtual Observatory. SimDAL and the Simulation Data Model are dedicated to cover the needs for the publication and retrieval of any kind of simulations: N-body or MHD simulations, numerical models of astrophysical objects and processes, theoretical synthetic spectra, etc... SimDAL is divided in three parts. First, SimDAL Repositories store the descriptions of theoretical projects and numerical codes. They can be used by clients to discover theoretical services associated with projects of interest. Second, SimDAL Search services are dedicated to the discovery of precise datasets. Finally, SimDAL Data Access services are dedicated to retrieve the original simulation output data, as plain raw data or formatted datasets cut-outs. To manage any kind of data, eventually large or at high-dimensionality, the SimDAL standard lets publishers choose any underlying implementation technology.

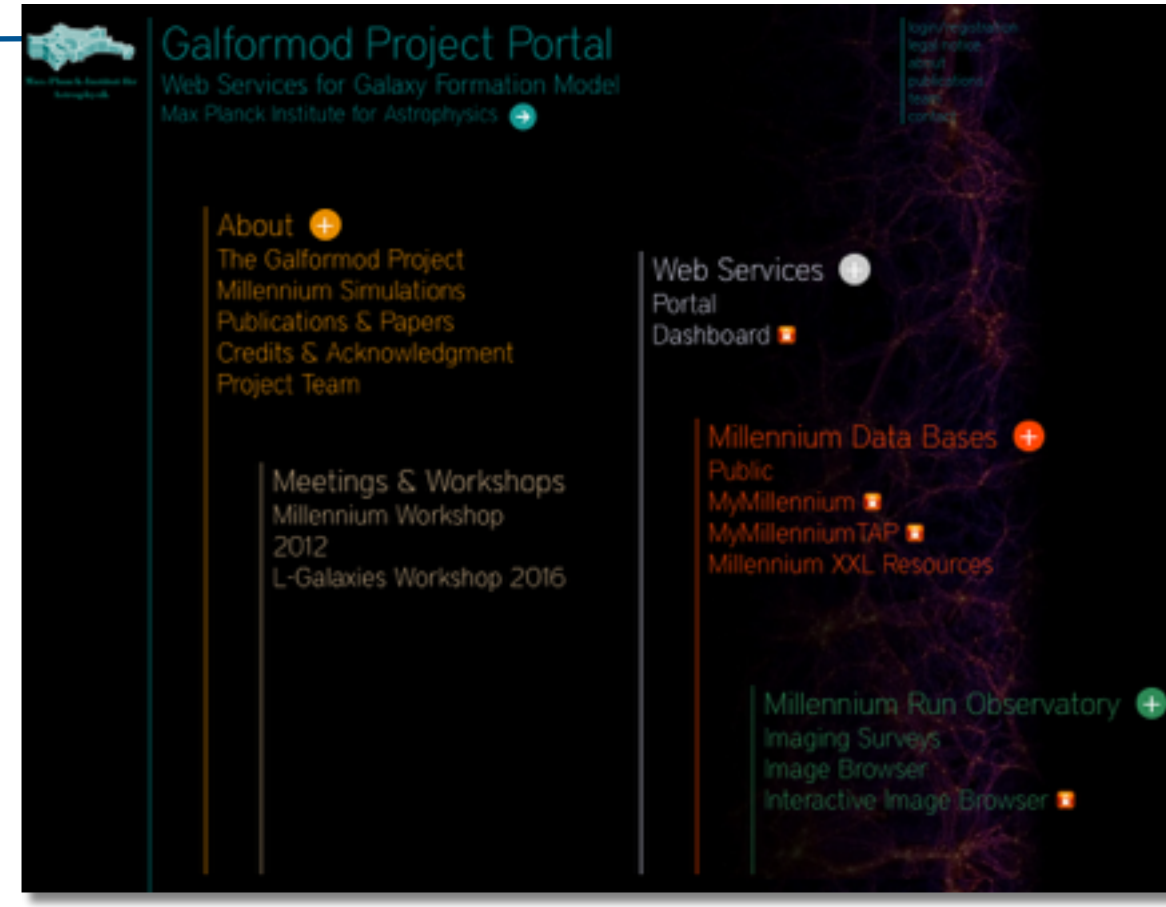
Status of this document

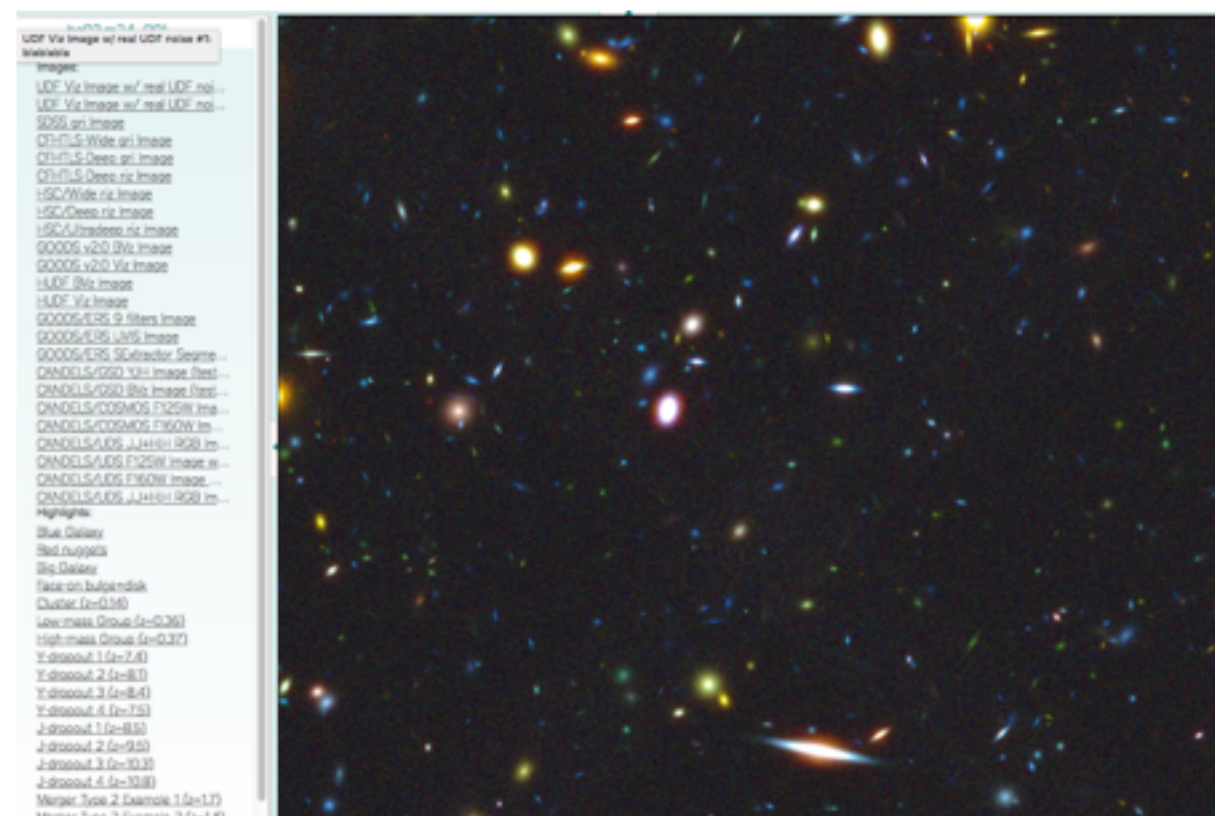
This document has been produced by the Data Access Layer Working Group.
It has been reviewed by IVOA Members and other interested parties, and has been endorsed by the IVOA Executive Committee as an IVOA Recommendation. It is a stable document and may be used as reference material or cited as a normative reference from another document. IVOA's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability inside the Astronomical Community.

# Example: Cosmology & Galaxies

## Millenium simulation



- Access to Millenium simulations
- **Many services developed above the data**
  - access to simulations
  - halo catalogs / Lightcones
  - synthetic images
  - …
- **Prototype for SimDM**
- **VO compatibility for some data (TAP / Images)**

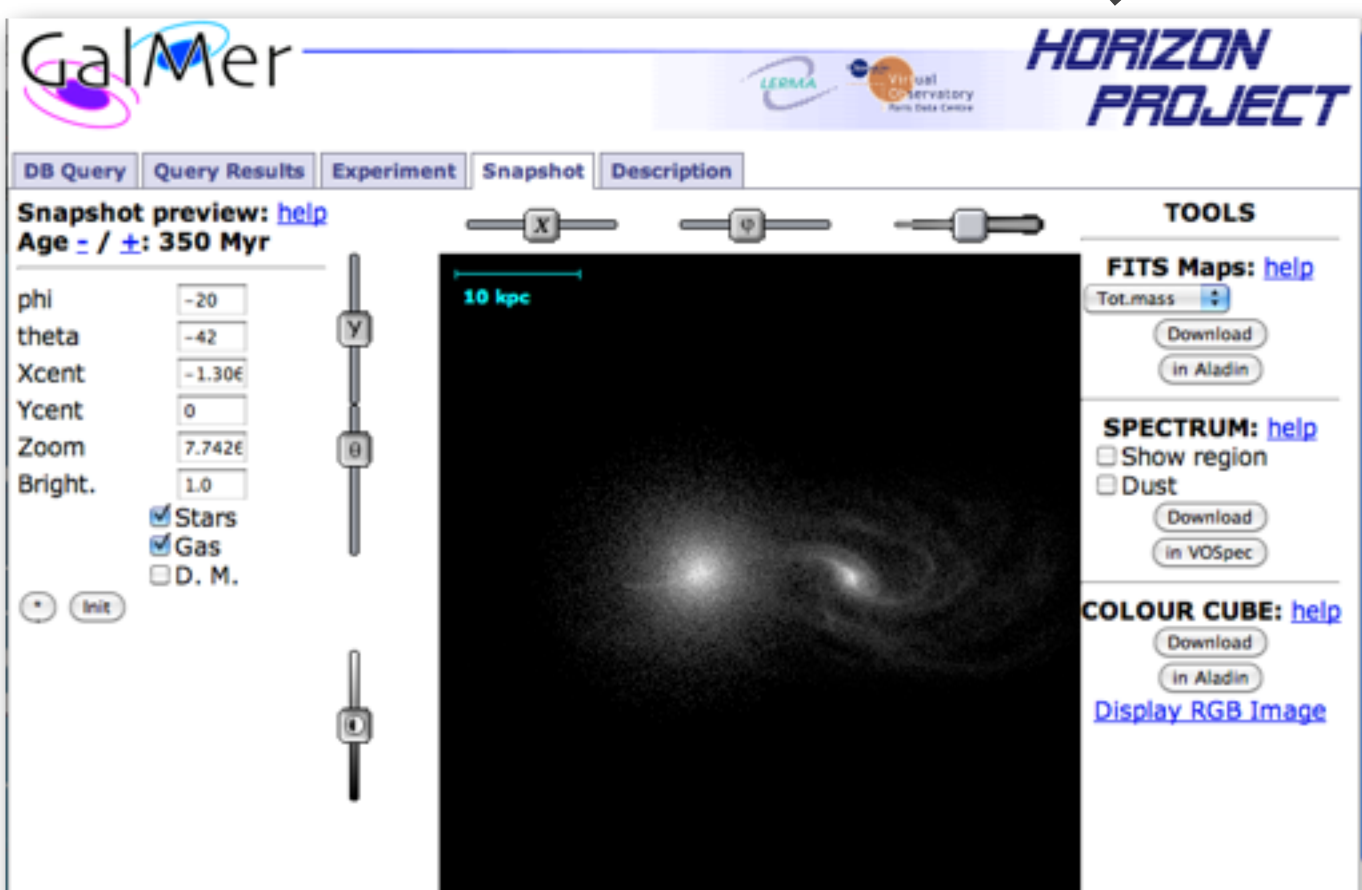
# Example: Mergers of galaxies

**GalMer**: Database of simulations of mergers of galaxies
http://galmer.obspm.fr

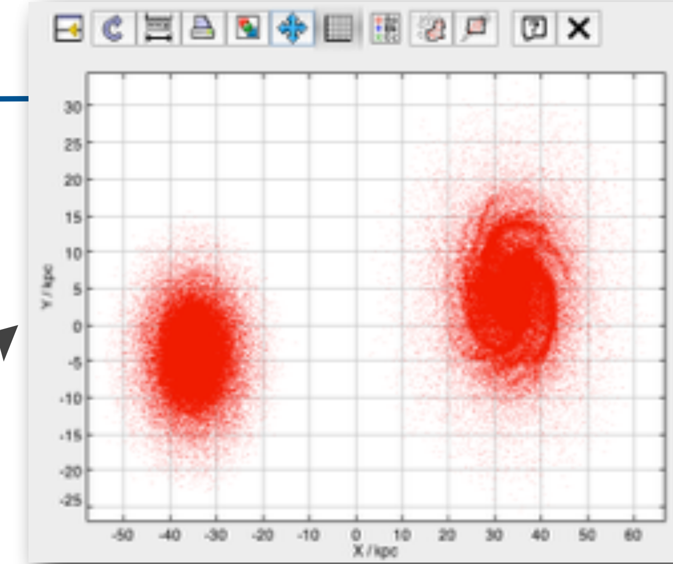
Access to:

- position of particles
- velocity field
- SED of dust
- …

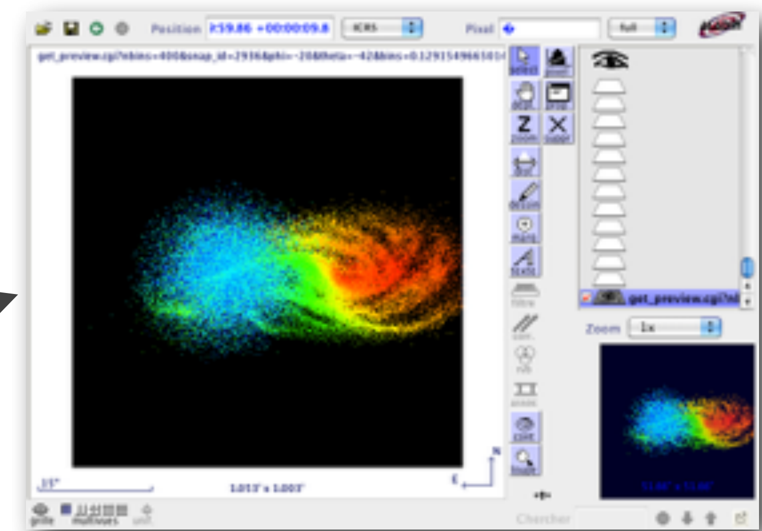
→ **Do not use VO-Theory standards**
→ **Outputs are VO-compatible**



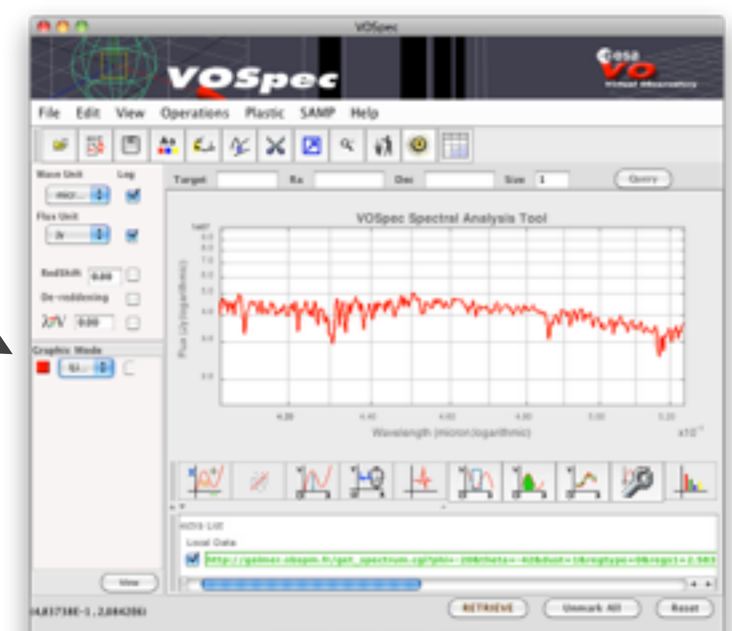

Particles




Velocities



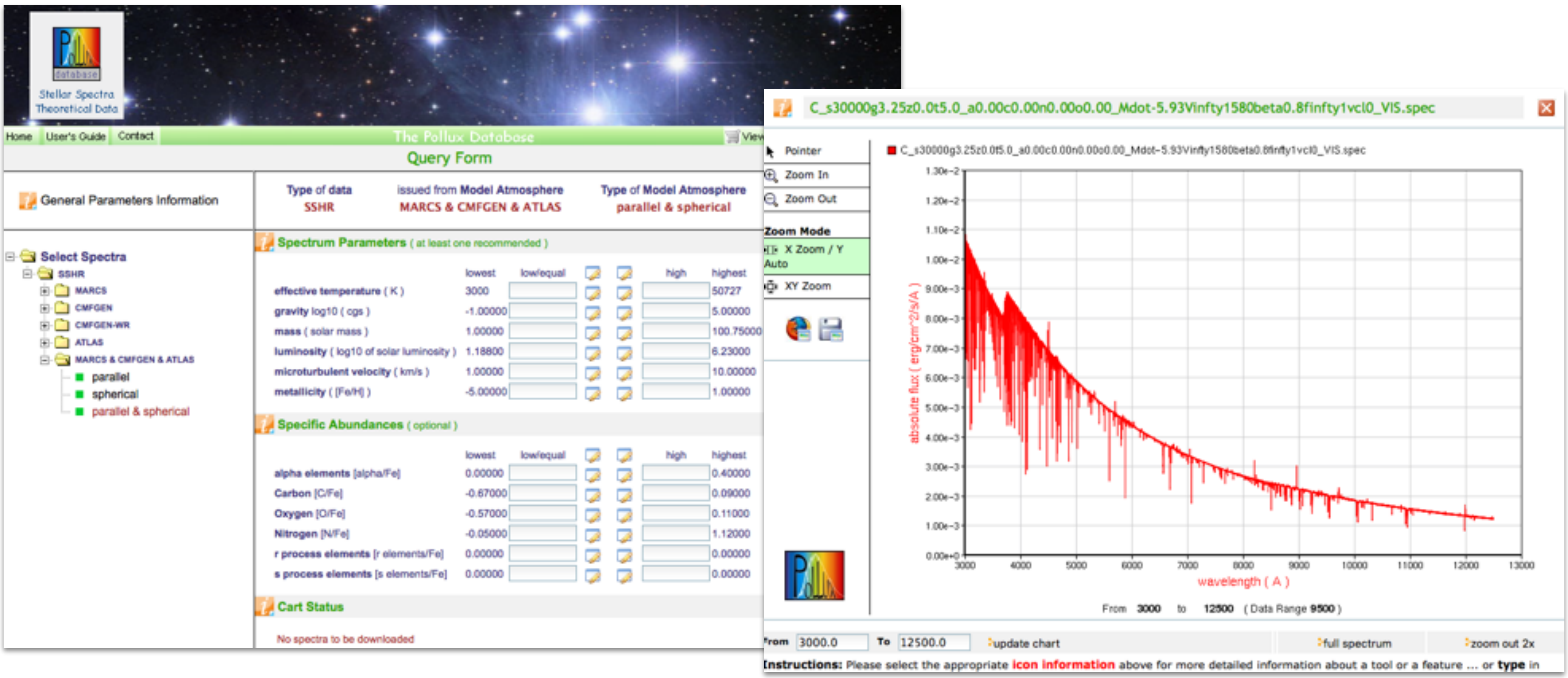
Dust Spectrum



# Example: Stellar synthetic spectra

**POLLUX database:** synthetic stellar spectra

http://pollux.graal.univ-montp2.fr




- Access to synthetic stellar spectra
- Fully VO-Compatible
  - SSA used of theoretical spectra
  - VO-Tools: CASSIS, VO-Spec, …

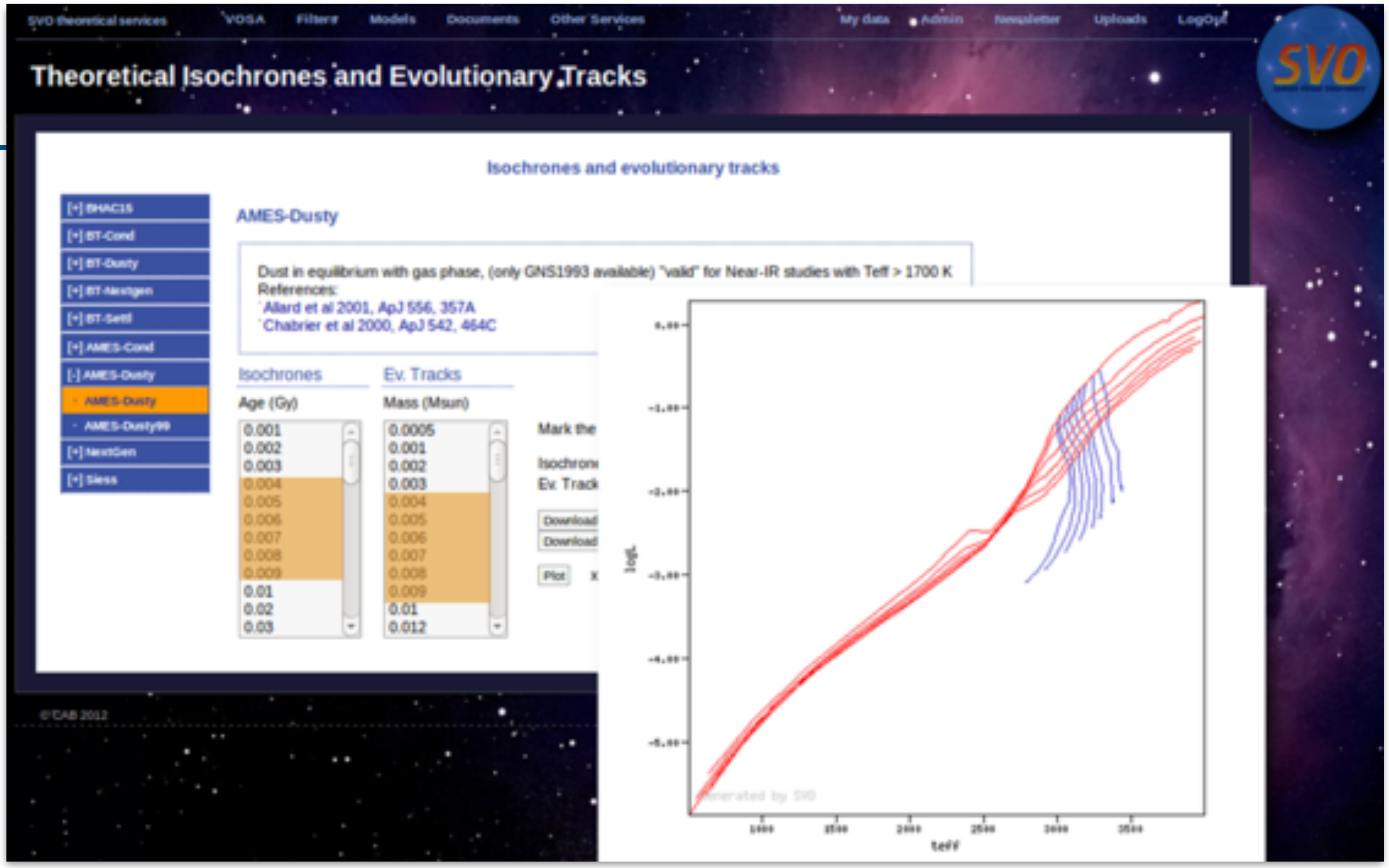
# Example: Stellar physics

## SVO Theoretical services

http://svo2.cab.inta-csic.es/theory/iso3/

Provide stellar isochrones

➔ Validation case for SimDAL

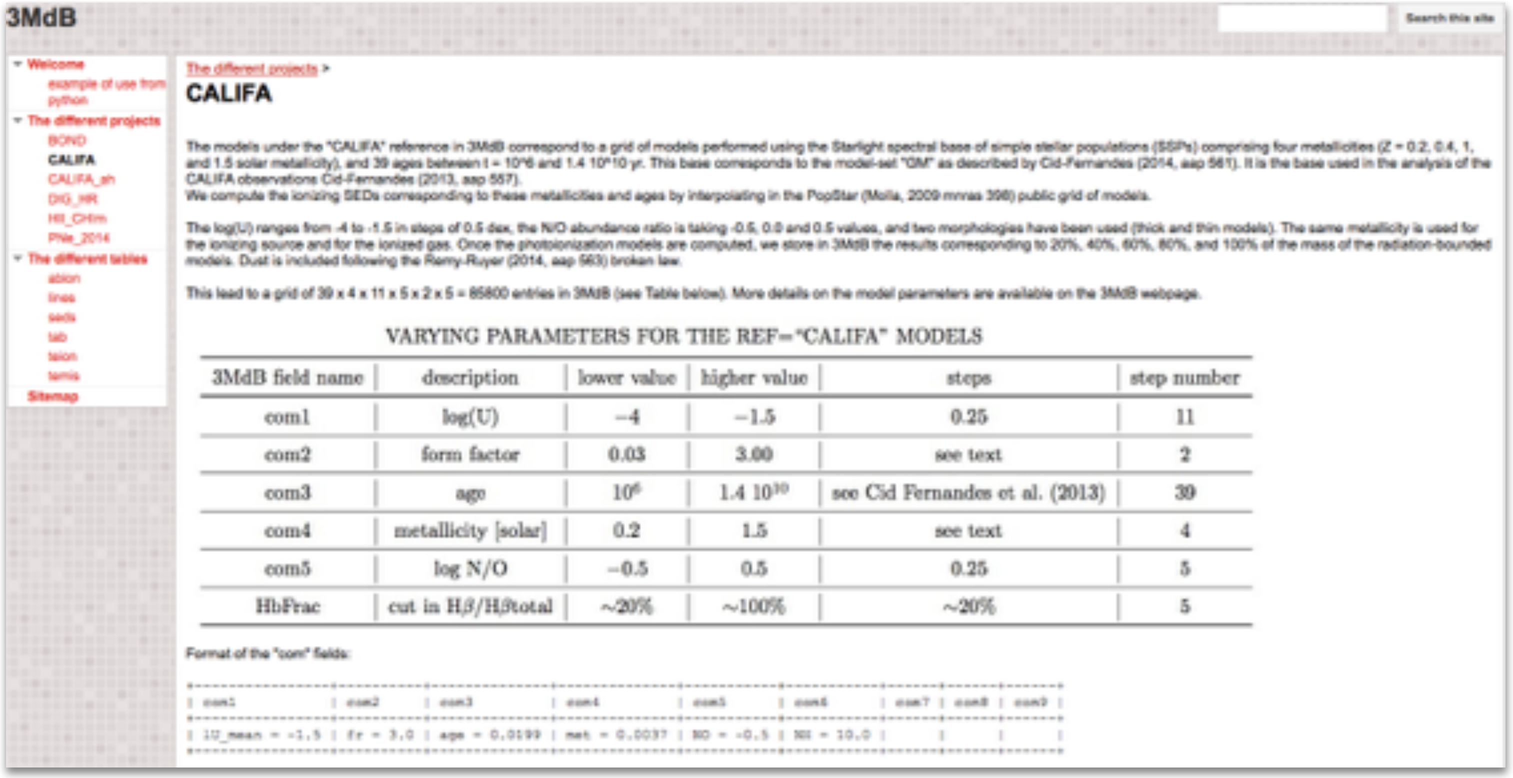



## 3MdB: cloudy models

https://sites.google.com/site/mexicanmillionmodels/

Models of chemical composition of H II regions

➔ Implementation of SimDAL underway



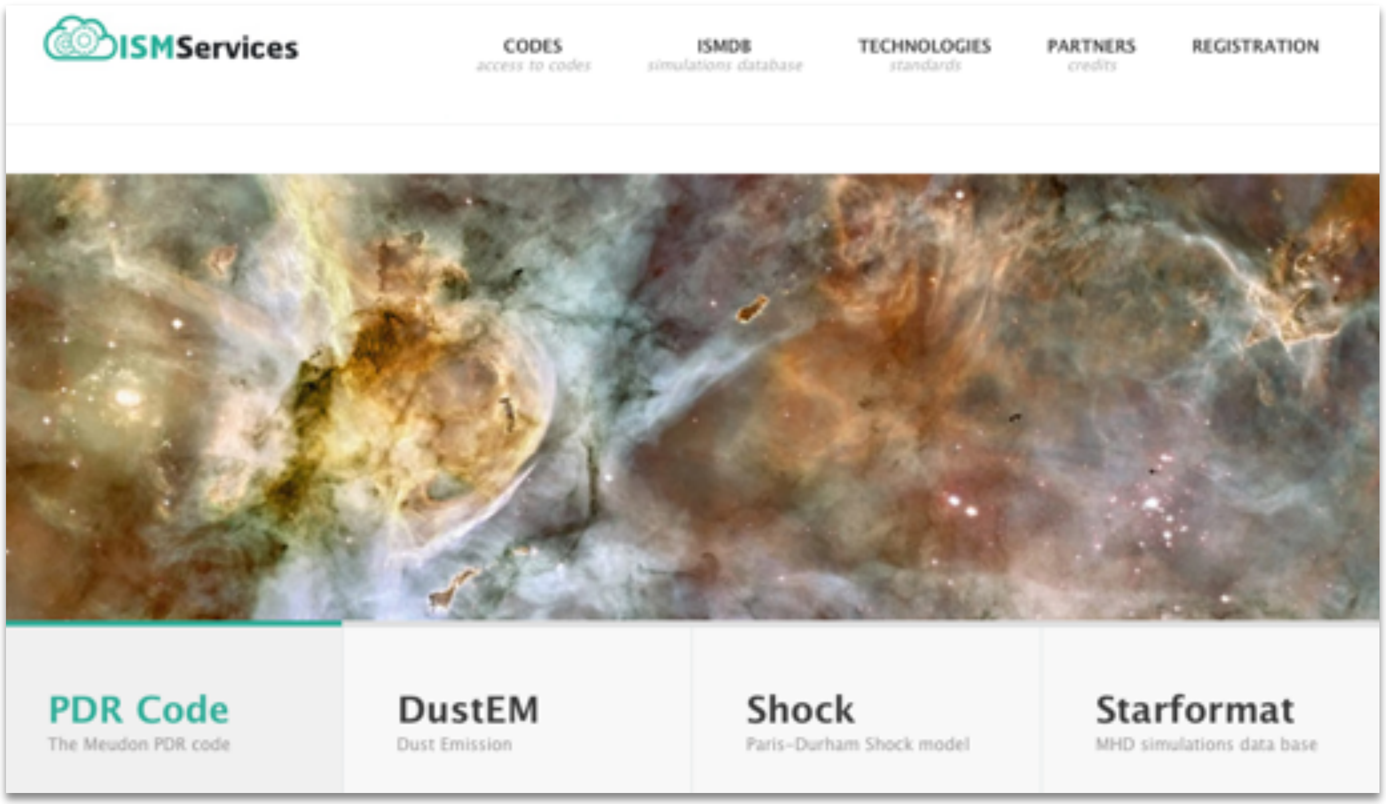

| 3MdB field name | description | lower value | higher value | steps | step number |
|---|---|---|---|---|---|
| com1 | log(U) | −4 | −1.5 | 0.25 | 11 |
| com2 | form factor | 0.03 | 3.00 | see text | 2 |
| com3 | age | $10^6$ | 1.4 $10^{10}$ | see Cid Fernandes et al. (2013) | 39 |
| com4 | metallicity [solar] | 0.2 | 1.5 | see text | 4 |
| com5 | log N/O | −0.5 | 0.5 | 0.25 | 5 |
| HbFrac | cut in Hβ/Hβtotal | ∼20% | ∼100% | ∼20% | 5 |

Format of the "com" fields:

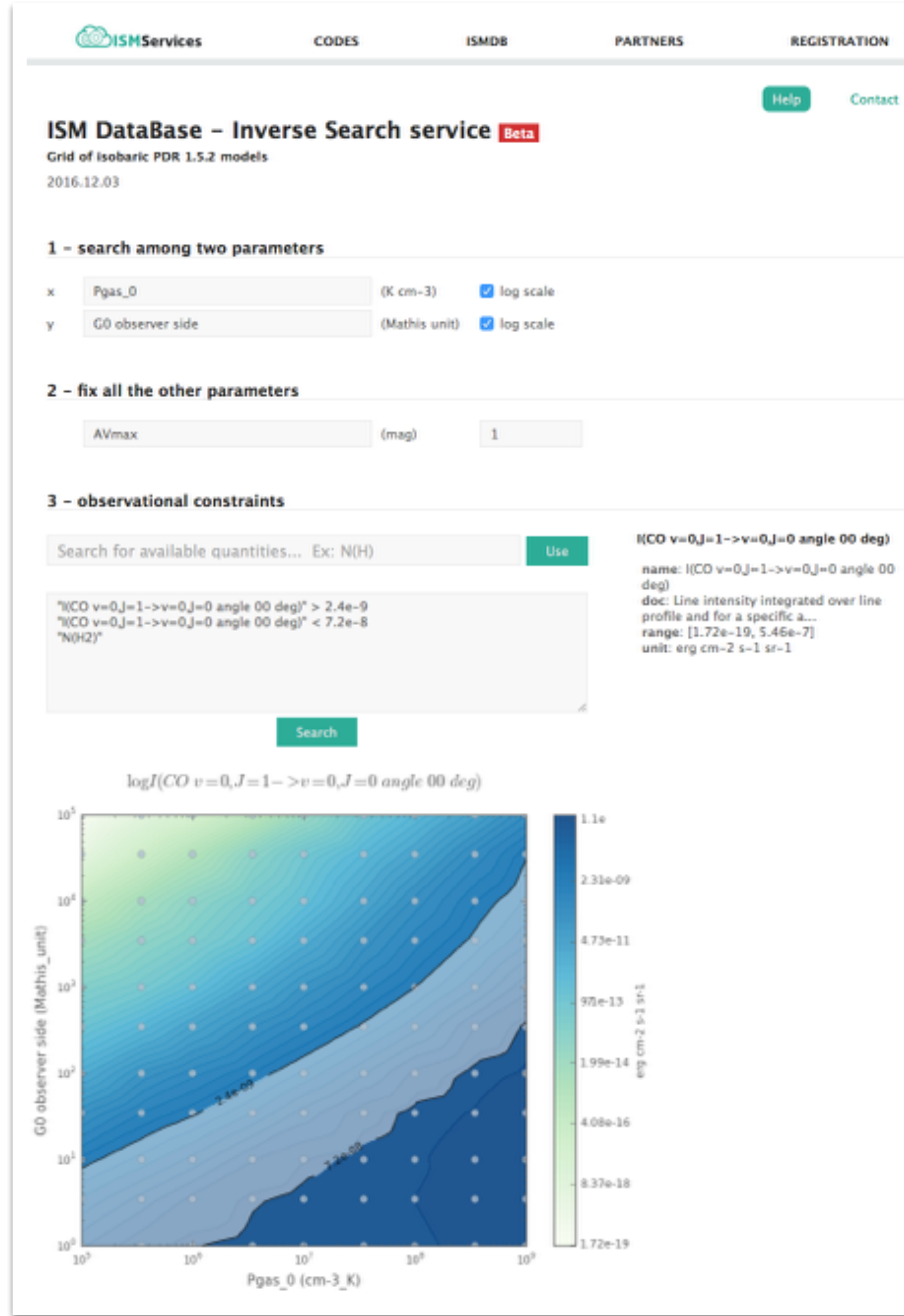
# Example: Interstellar medium

**ISM Services:** models in interstellar clouds

http://ism.obspm.fr




- Access to simulations of the Interstellar Medium
- Prediction of line intensities
- Fully VO-Theory compatible
  - ISMDB: **SimDM + SimDAL + Semantics**
  - Online codes: **PDL**
  - **SAMP** -> Topcat



# Why no more projects integrates their theoretical data in the VO ?

**Are simulations more difficult to publish in the VO than observations ?**

Simulations main specificities:

- more heterogeneous
- complex: anything that can be thought can be simulated
- no standard raw data format as FITS

Heterogeneity & complexity are solved with SimDM

**Now VO-Theory standards exist, simulations are less difficult to publish**

**Possible reasons**

- VO-Theory standards come lately for some projects
- A question of culture
  - Data centers publishing observational data exist for a long time
  - Publishing of theoretical data is more recent and less organized

- Data centers are focused on observational data: **few help for theorists**
- Requires **time**, **procedures, experience** & **manpower** to publish efficiently data
- Requires more time to publish in the VO
  ➔ **temptation to choose quick and dirty solutions that fulfill immediate needs**

# Why no more projects integrates their data in the VO ?

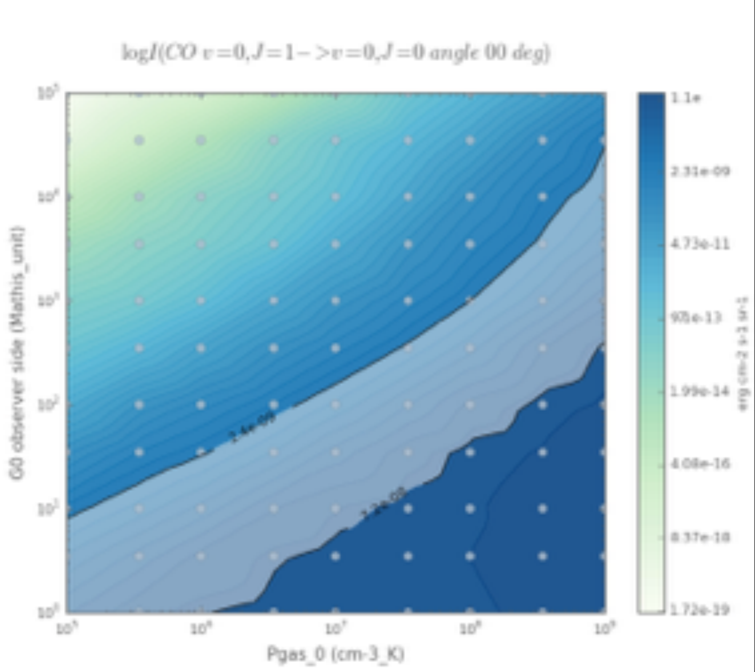
Theoretical data publishers may not see the immediate benefit

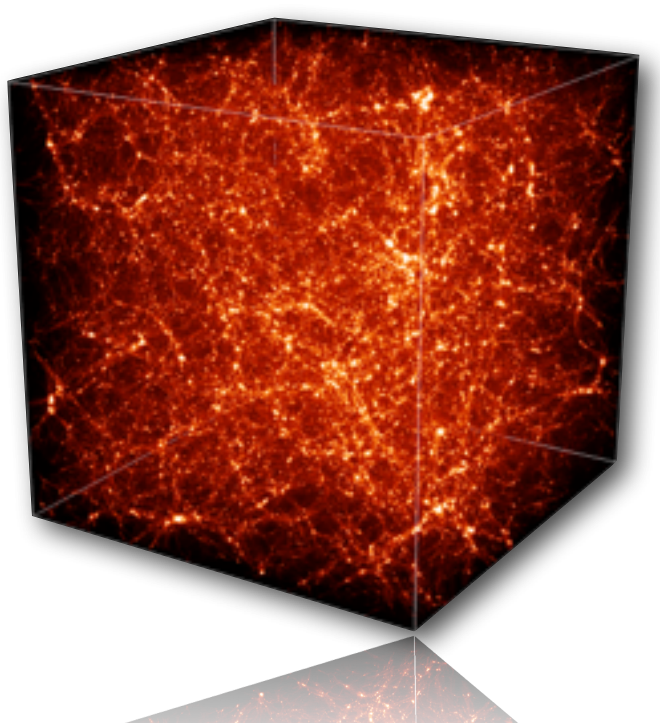
A strong motivation to publish observationnal data: **VO-Tools**

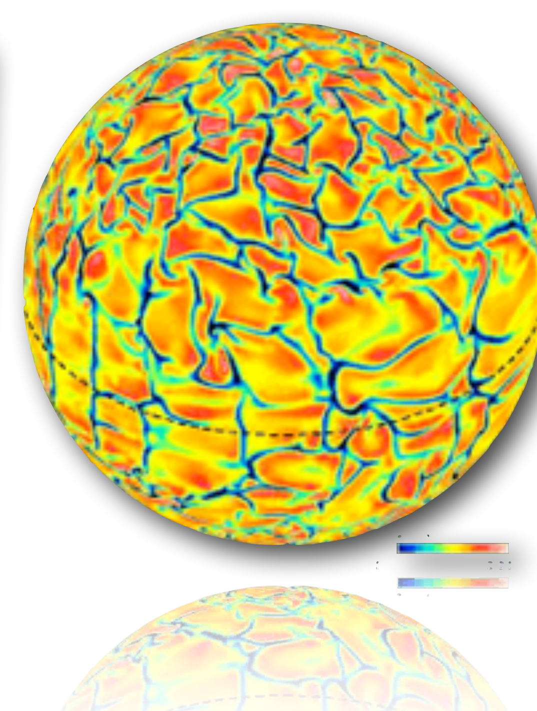
| | Tools | Observations | Theory |
|---|---|---|---|
| Catalogs / data | **Topcat**, … | ✔ | ✔ |
| Spectra | **CASSIS**, **SPLAT**, **VOSpec** | ✔ | ✔ |
| Images | **Aladin**, … | ✔ | ✔ |

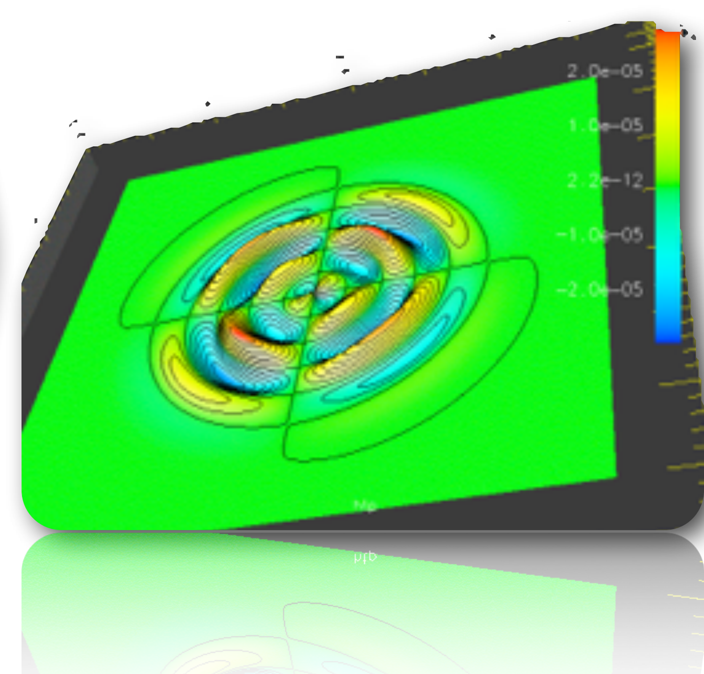
But :

- some problems for Theoretical data: positions & units, …
- no VO-Tools to plot some 2D graphs, 3D data, time-dependent data:









# What could be the strategy ?

① Do a real SimDAL Repository
➔ Gives visilibity to simulations in the VO

② Publishers of theoretical data **may wish short term benefit**
- to implement VO standards may take more time than developing a home-made solution
- VO integration has mid-term interest

Short term benefit the VO can bring: VO-Tools

➔ Theory group: Provide implementation notes to explain how to use VO standards
➔ Focus on theoretical data for which VO-Tools are ready
Ex: catalogs, spectra, images, …

Gives also the benefit of **interoperability: Observations - Theoretical data**

③ For large simulations: need to do the **link between VO and large projects**
*Example*: ASTERICS is a success to bring the VO in large projects as CTA

➔ Try to do the same for theoretical data