



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



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**
- increasingly **mandated by funding agencies**
- **showcases** for future proposals
- ...

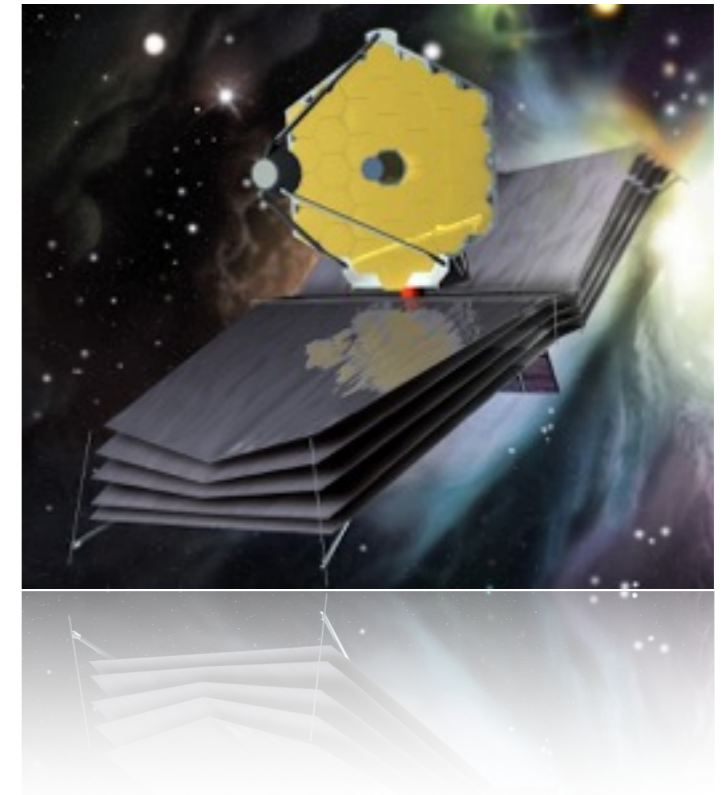
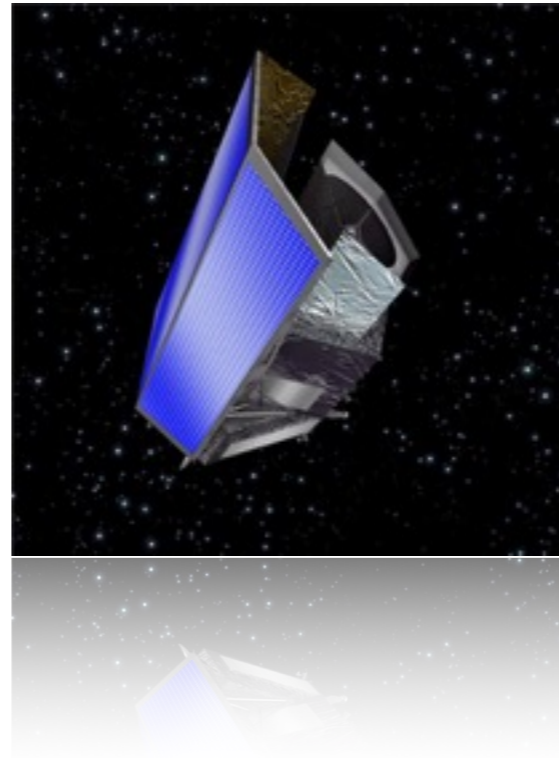
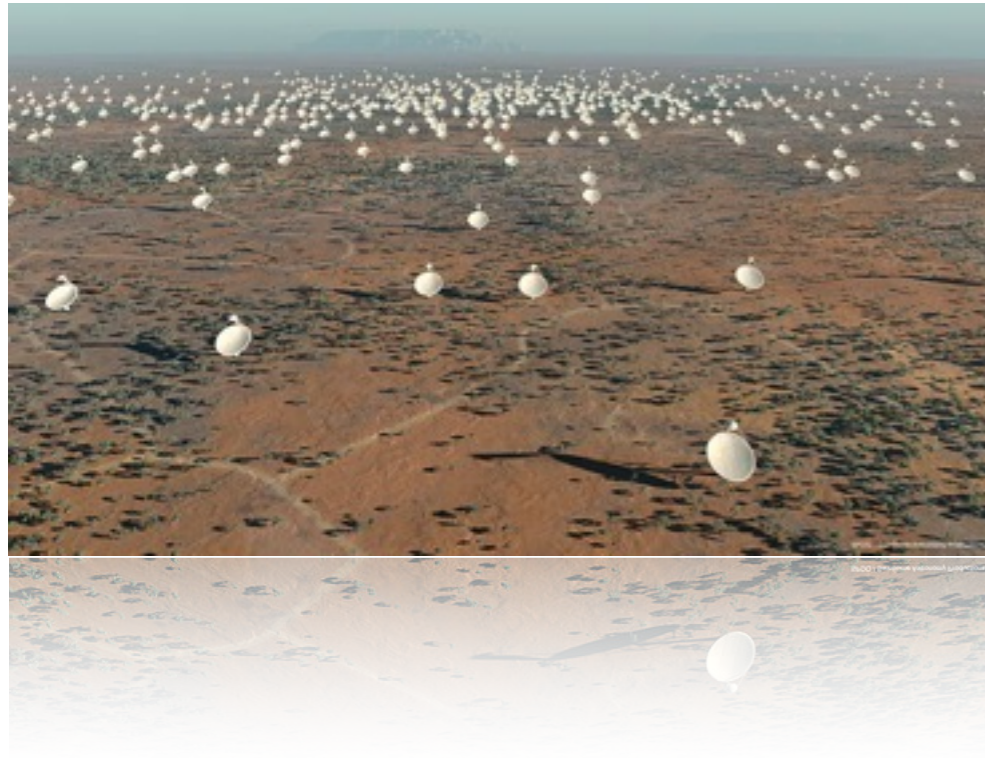
Why to publish theoretical data *in the Virtual Observatory* ?

- 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>



COSMO HUB
Build your own Universe
Real-time data analysis of massive cosmological data without any SQL knowledge

Hundreds of millions of observed and simulated galaxies
Superfast queries means superfast results
Features to make you work faster and easier
Online plotting preview and data download

How does it work?

1. Select a catalog

Name	Version	Description	Origin	Date
cosmos-hub-1.0	1.0	cosmos-hub-1.0	Observed	2019-03-11
MICE-1.0	1.0	MICE-1.0	Observed	2019-03-11
Euclid-1.0	1.0	Euclid-1.0	Observed	2019-03-11
MICE-1.0	1.0	MICE-1.0	Observed	2019-03-11
Euclid-1.0	1.0	Euclid-1.0	Observed	2019-03-11
MICE-1.0	1.0	MICE-1.0	Observed	2019-03-11
Euclid-1.0	1.0	Euclid-1.0	Observed	2019-03-11
MICE-1.0	1.0	MICE-1.0	Observed	2019-03-11
Euclid-1.0	1.0	Euclid-1.0	Observed	2019-03-11
MICE-1.0	1.0	MICE-1.0	Observed	2019-03-11

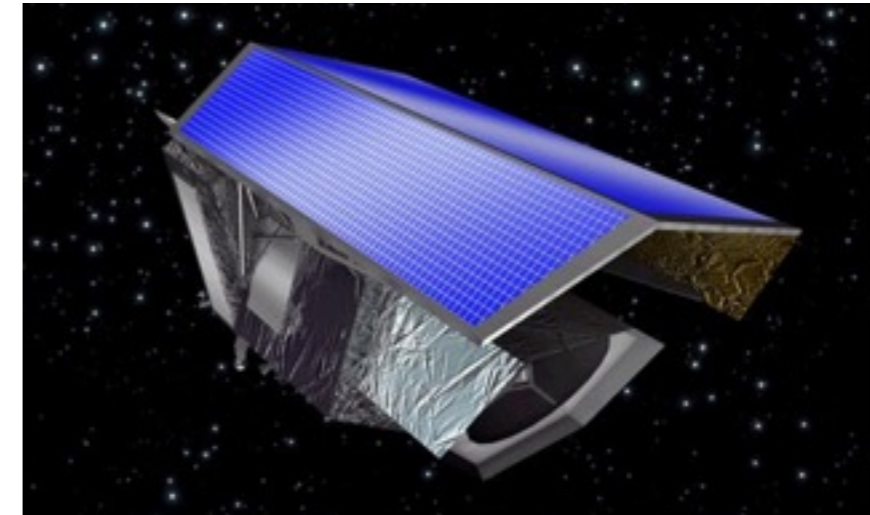
2a. Select one of the prebuilt datasets...

Name	Version	Type	Description	Size	Load
cosmos-hub-1.0	1.0	Table	cosmos-hub-1.0	400,000,000	Load
MICE-1.0	1.0	Table	MICE-1.0	400,000,000	Load
Euclid-1.0	1.0	Table	Euclid-1.0	400,000,000	Load
MICE-1.0	1.0	Table	MICE-1.0	400,000,000	Load
Euclid-1.0	1.0	Table	Euclid-1.0	400,000,000	Load
MICE-1.0	1.0	Table	MICE-1.0	400,000,000	Load
Euclid-1.0	1.0	Table	Euclid-1.0	400,000,000	Load
MICE-1.0	1.0	Table	MICE-1.0	400,000,000	Load
Euclid-1.0	1.0	Table	Euclid-1.0	400,000,000	Load
MICE-1.0	1.0	Table	MICE-1.0	400,000,000	Load

2b. Or build your own custom set

3. Select the sampling and filters

```
SELECT * FROM cosmos-hub-1.0 WHERE z < 0.1
```



- 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 on 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



① Simulation Data Model (SimDM)

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

Implemented on:

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

→ proof of the versatility

Simulation Data Model Version 1.0

IVOA Recommendation 03 May 2012

InterestWorking Group:

<http://www.ivoa.net/wiki/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):

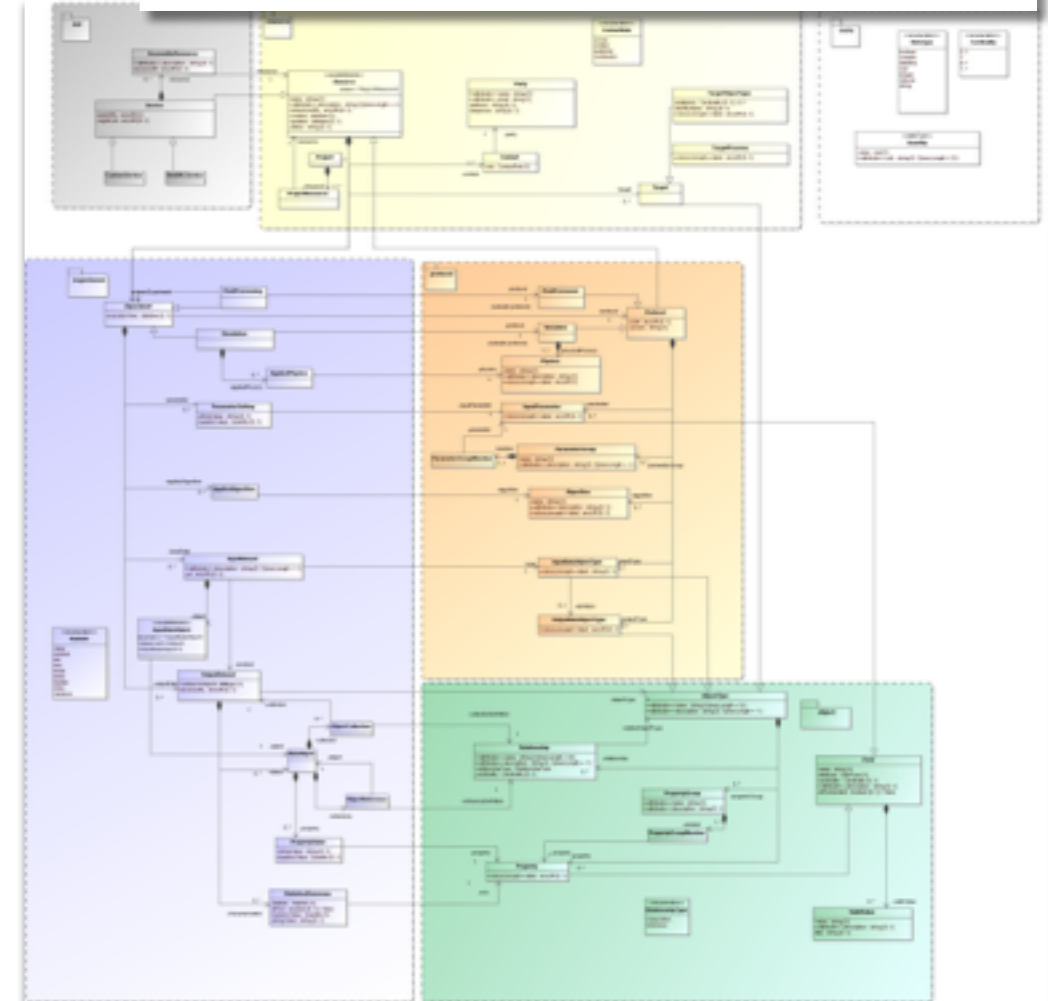
Gerard Lemson, Herve Wozniak

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.





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

Simulation Data Access Layer Version 1.0

IVOA Recommendation 20 March 2017

Interest/Working Group:

<http://www.ivoa.net/twiki/bin/view/IVOA/ivosDAL>

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.

Home Search concepts Help

This service is dedicated to scientists and VO developers who wish to publish theoretical services described by the [Simulation Data Model](#).

As described in the [IVOA](#) standard, Simulation Data Model, registrations of theoretical services, require to provide several URIs corresponding to semantics keywords describing services and simulations. VO-Theory concepts are based on SKOS description as recommended by the [IVOA Semantic Working Group](#).

Example of a VO-Theory URIs : <http://url.obo.inria.fr/oboc/Algorithms/GaussSeidel>

This website is dedicated to the discovery of these URIs. Navigate through the broader, narrower, related terms to discover the most precise concept you wish.

To suggest new concepts or corrections, contact : support.vothery@obspm.fr

Search concepts

3+1 Formalism 8-Wave Scheme Accelerated Lambda Iteration

Adaptive Mesh Refinement Advection Upstream Splitting Method

Algorithm Alternating Direction Implicit BiConjugate Gradient

BiConjugate Gradient Stabilized Block Based AMR

Bulirsch-Stoer Cell Based AMR Cell Centred

Central Difference Scheme Chebyshev Iteration

Conjugate Gradient Method Conjugate Gradient Squared Method

Constrained Transport Coupled Escaped Probability

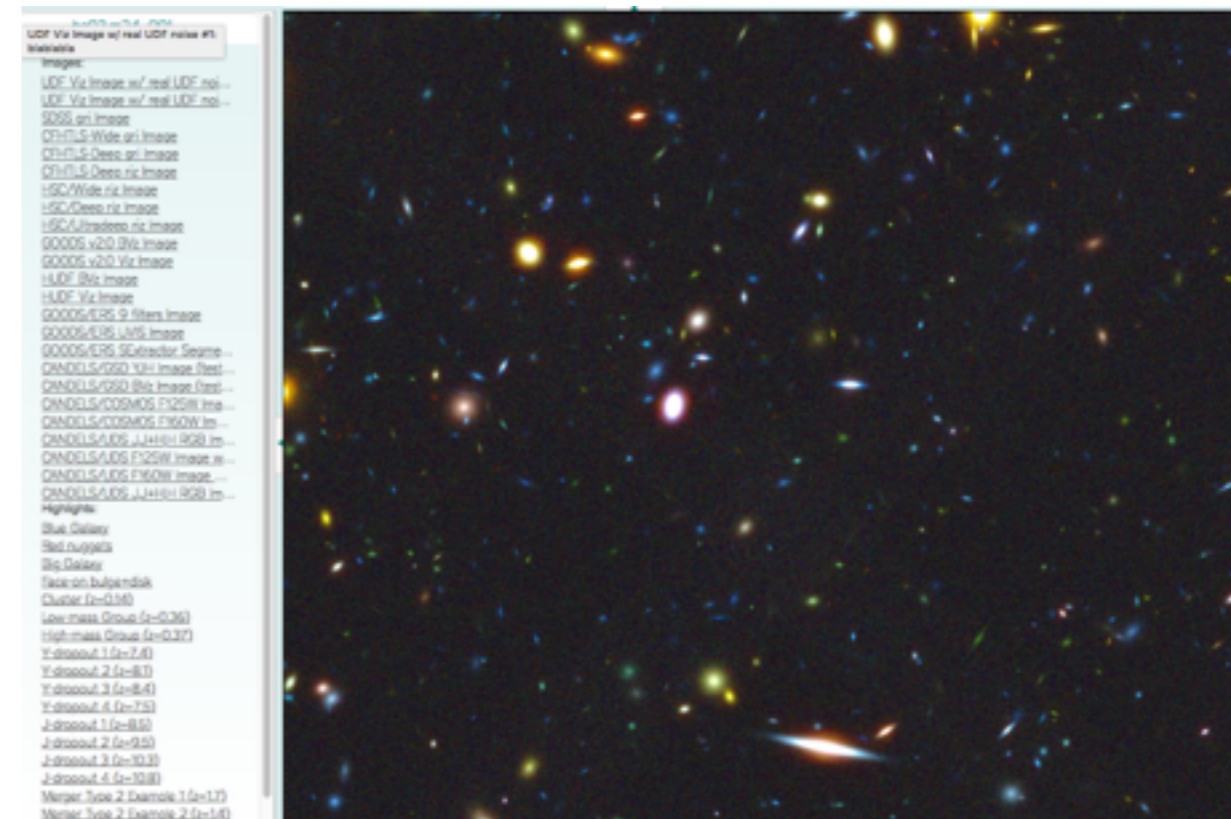
Crank-Nicolson Discontinuous Galerkin

Discontinuous Galerkin methods Escape Probability Euler

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

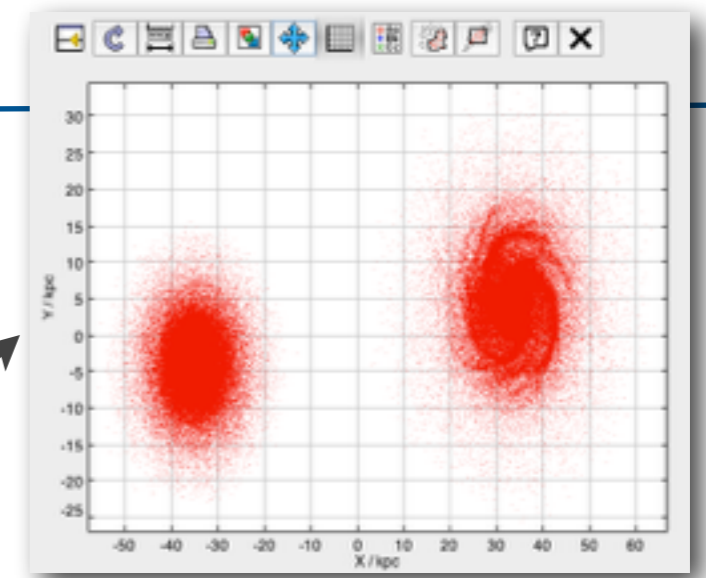
<http://galmer.obspm.fr>

Access to:

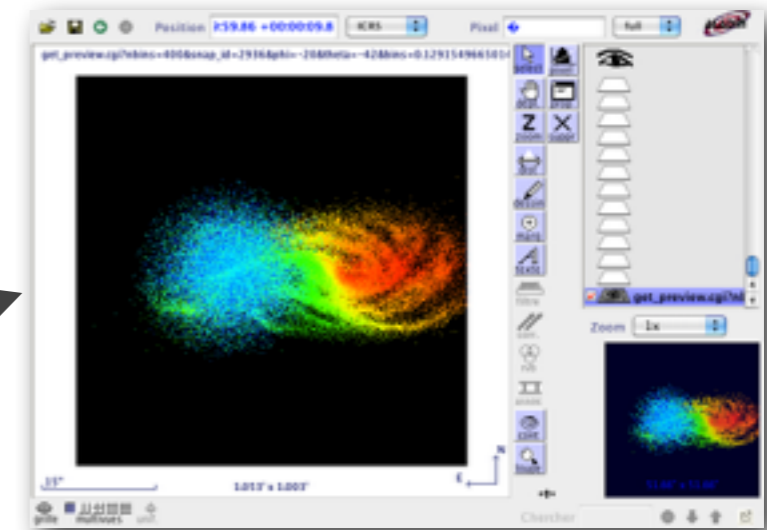
- position of particles
- velocity field
- SED of dust
- ...

→ Do not use VO-Theory standards

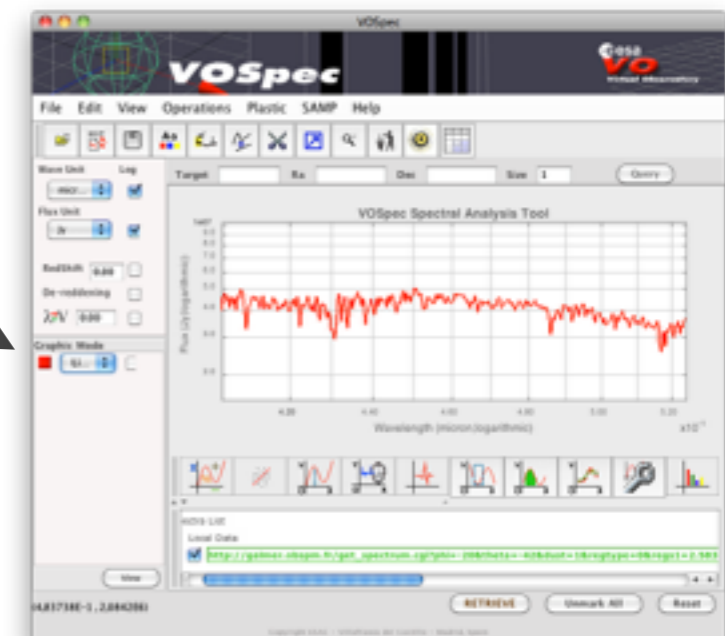
→ Outputs are VO-compatible



Particles



Velocities



Dust Spectrum

Example: Stellar synthetic spectra

POLLUX database: synthetic stellar spectra

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

The screenshot displays the POLLUX database interface. On the left, a navigation tree shows 'Select Spectra' with sub-categories: SSHR, MARCS, CMFGEN, CMFGEN-WR, ATLAS, and MARCS & CMFGEN & ATLAS (with sub-options for parallel, spherical, and parallel & spherical). The main area is a 'Query Form' with sections for 'Spectrum Parameters' and 'Specific Abundances'. The 'Spectrum Parameters' section includes fields for effective temperature (K), gravity log10 (cgs), mass (solar mass), luminosity (log10 of solar luminosity), microturbulent velocity (km/s), and metallicity ([Fe/H]). The 'Specific Abundances' section includes fields for alpha elements [alpha/Fe], Carbon [C/Fe], Oxygen [O/Fe], Nitrogen [N/Fe], r process elements [r elements/Fe], and s process elements [s elements/Fe]. A 'Cart Status' section at the bottom indicates 'No spectra to be downloaded'. On the right, a spectral plot shows 'absolute flux (erg/cm^2/s/A)' versus 'wavelength (A)'. The plot title is 'C_s30000g3.25z0.0t5.0_a0.00c0.00n0.00o0.00_Mdot-5.93Vinfy1580beta0.8finfy1vcl0_VIS.spec'. The x-axis ranges from 3000 to 13000 Angstroms, and the y-axis ranges from 0.00e+0 to 1.30e-2. The plot shows a red line representing the synthetic spectrum with numerous absorption lines. Below the plot, there are controls for 'From 3000.0 To 12500.0', 'update chart', 'full spectrum', and 'zoom out 2x'. An 'Instructions' box at the bottom reads: 'Please select the appropriate icon information above for more detailed information about a tool or a feature ... or type in'.

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

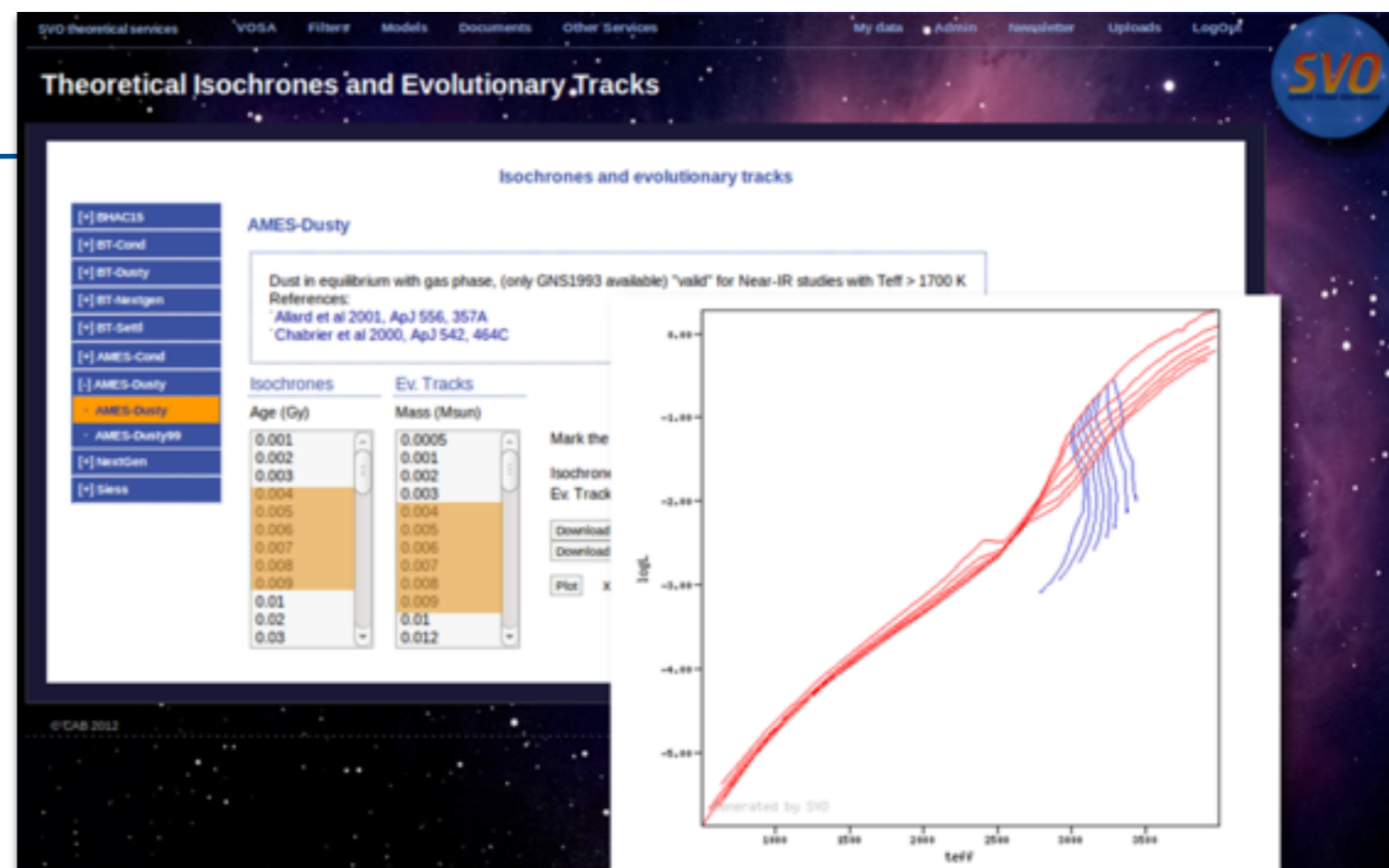
Example: Stellar physics

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

Welcome

The different projects

BOND
CALIFA
CALIFA_sh
DIG_HII
HII_Chim
PNe_2014

The different tables

abun
linc
seds
tab
talon
tunis
Sitemap

The different projects >

CALIFA

The models under the "CALIFA" reference in 3MdB correspond to a grid of models performed using the Starlight spectral base of simple stellar populations (SSPs) comprising four metallicities ($Z = 0.2, 0.4, 1,$ and 1.5 solar metallicity), and 39 ages between $t = 10^6$ and $1.4 \cdot 10^{10}$ yr. This base corresponds to the model-set "GM" as described by Cid-Fernandes (2014, aap 561). It is the base used in the analysis of the CALIFA observations Cid-Fernandes (2013, aap 557). We compute the ionizing SEDs corresponding to these metallicities and ages by interpolating in the PopStar (Molla, 2009 mras 398) public grid of models.

The $\log(U)$ ranges from -4 to -1.5 in steps of 0.5 dex, the N/O abundance ratio is taking $-0.5, 0.0$ and 0.5 values, and two morphologies have been used (thick and thin models). The same metallicity is used for the ionizing source and for the ionized gas. Once the photoionization models are computed, we store in 3MdB the results corresponding to 20%, 40%, 60%, 80%, and 100% of the mass of the radiation-bounded models. Dust is included following the Remy-Ruyer (2014, aap 563) broken law.

This lead to a grid of $39 \times 4 \times 11 \times 5 \times 2 \times 5 = 85800$ entries in 3MdB (see Table below). More details on the model parameters are available on the 3MdB webpage.

VARYING PARAMETERS FOR THE REF="CALIFA" MODELS					
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	$1.4 \cdot 10^{10}$	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 $H\beta/H\beta_{\text{total}}$	$\sim 20\%$	$\sim 100\%$	$\sim 20\%$	5

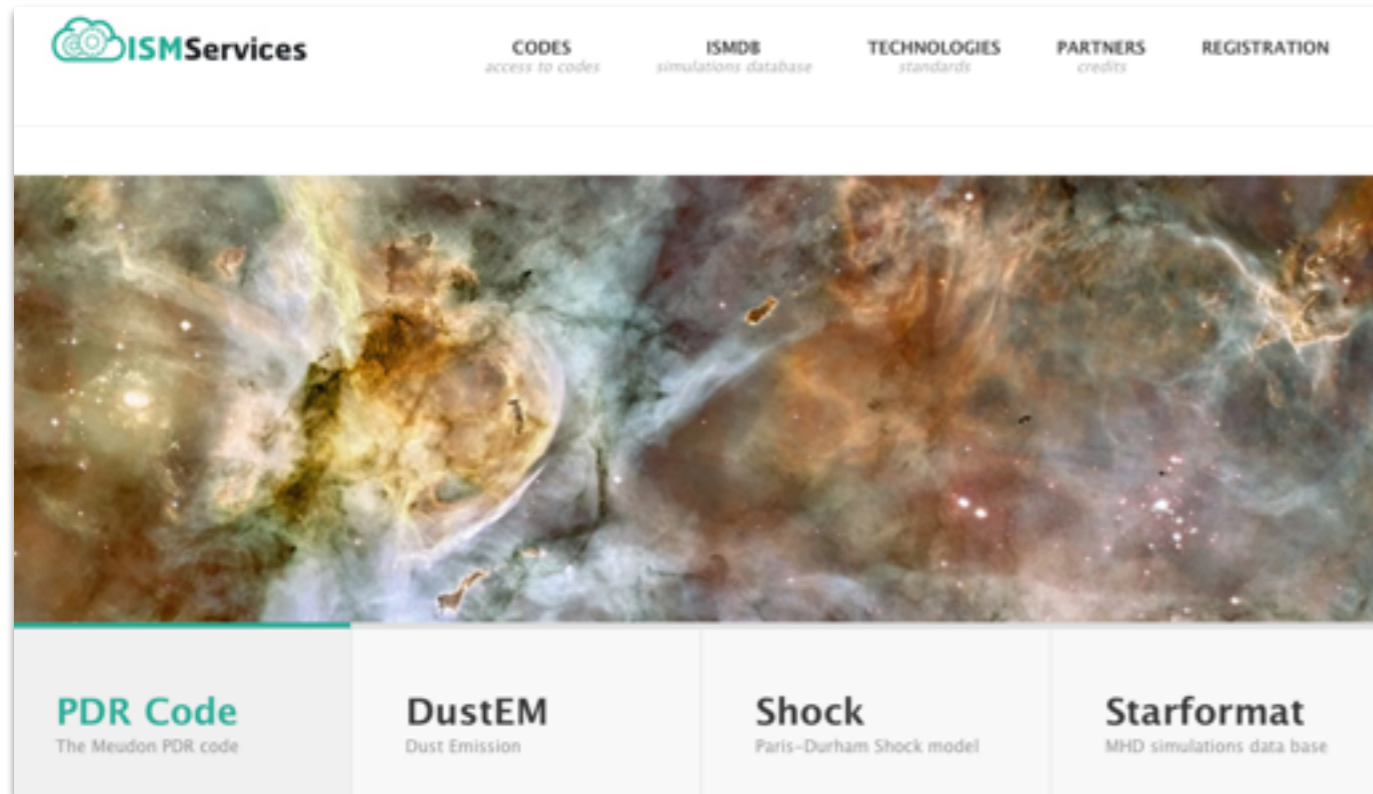
Format of the "com" fields:

```
| com1 | com2 | com3 | com4 | com5 | com6 | com7 | com8 | com9 |
| log_mean = -1.5 | fr = 3.0 | age = 0.0199 | met = 0.0037 | SO = -0.5 | NO = 10.0 | | | |
```

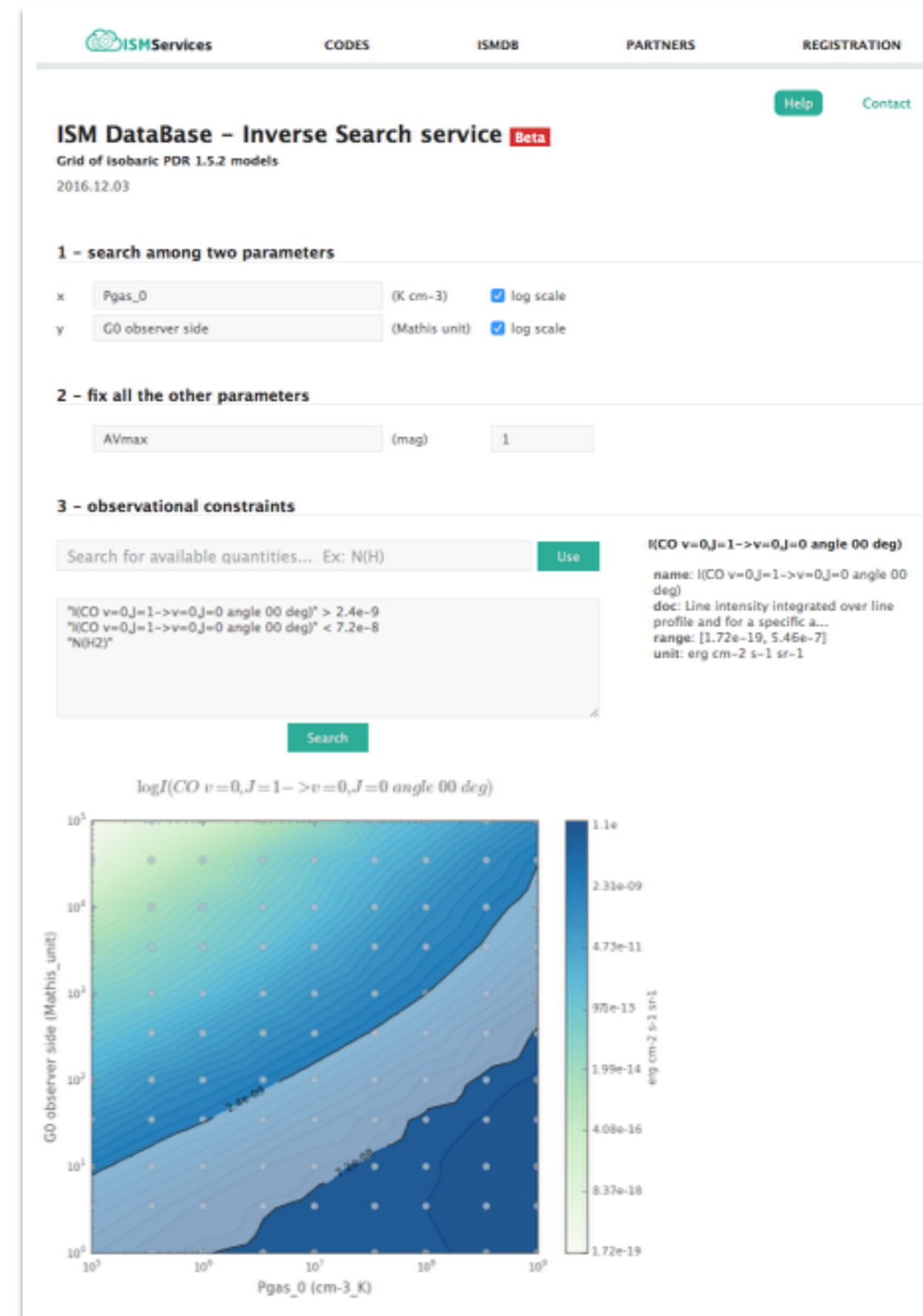
Example: Interstellar medium

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 ?

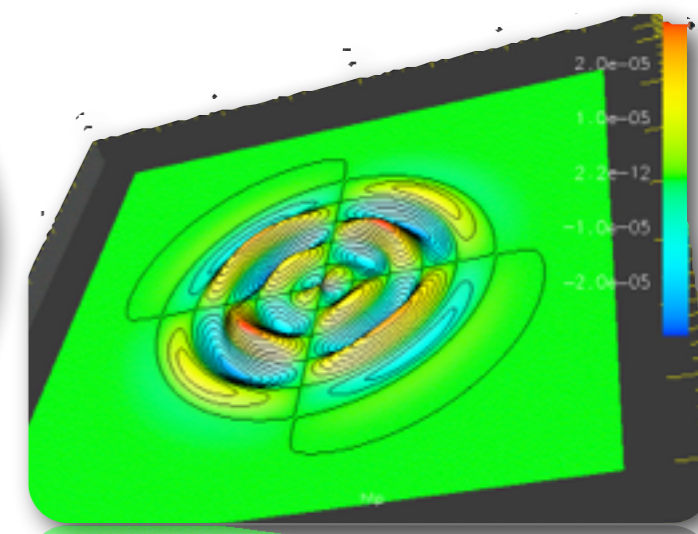
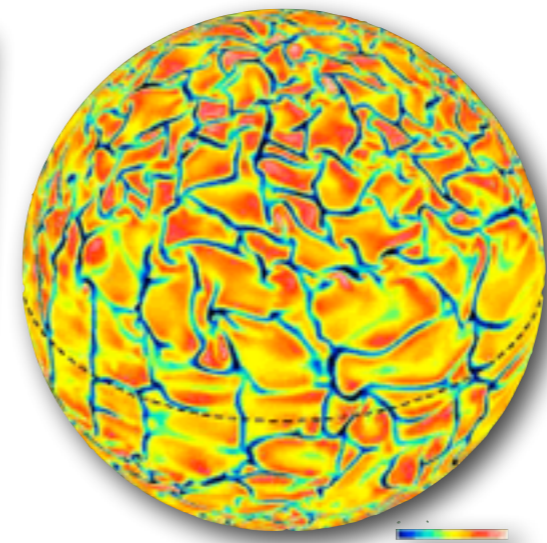
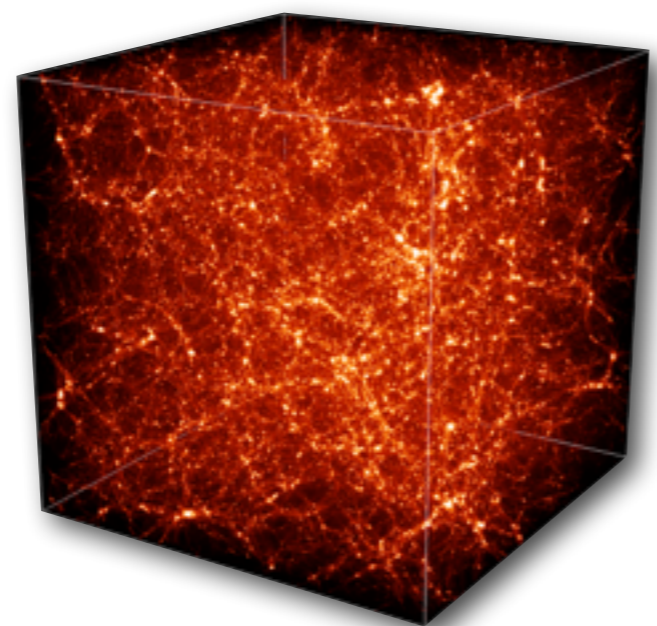
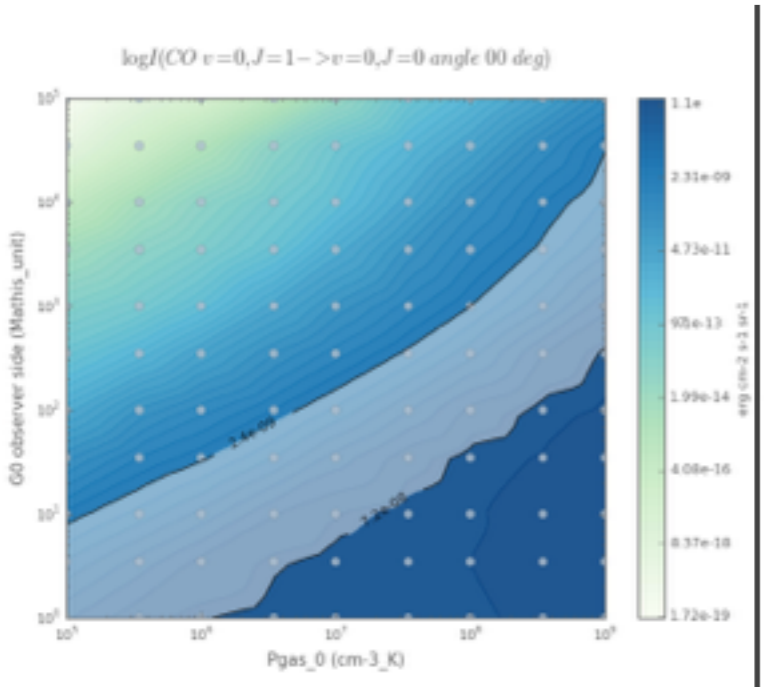
Theoretical data publishers may not see the immediate benefit

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 visibility 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

