SimDM Implementation

Publication in ISMDB

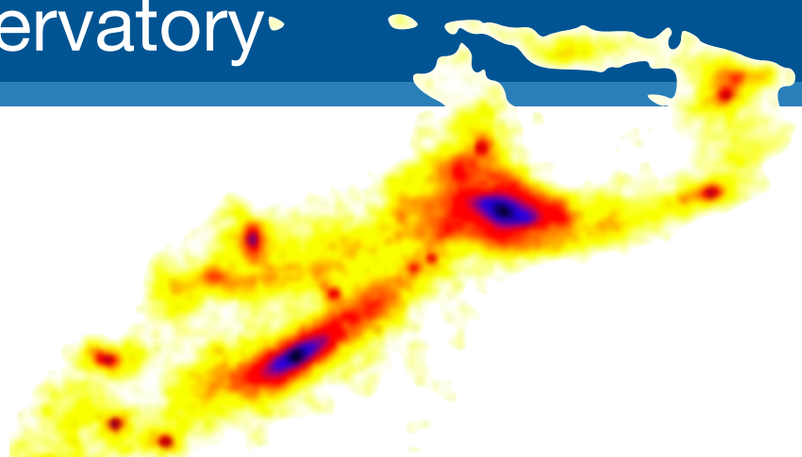
http://ism.obspm.fr

Franck Le Petit, David Languignon, Nicolas Moreau, Emeric Bron

InterOp - 4th November 2021

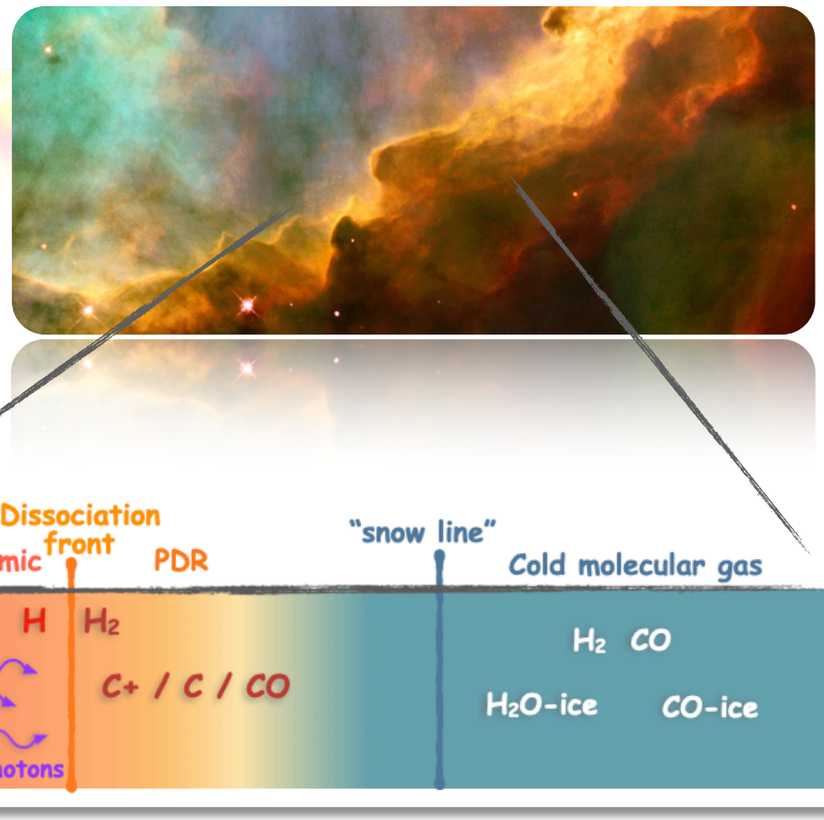
SimDM implementations at Paris Observatory

Implementations at Paris Observatory



Past implementations (pre-SimDM) :

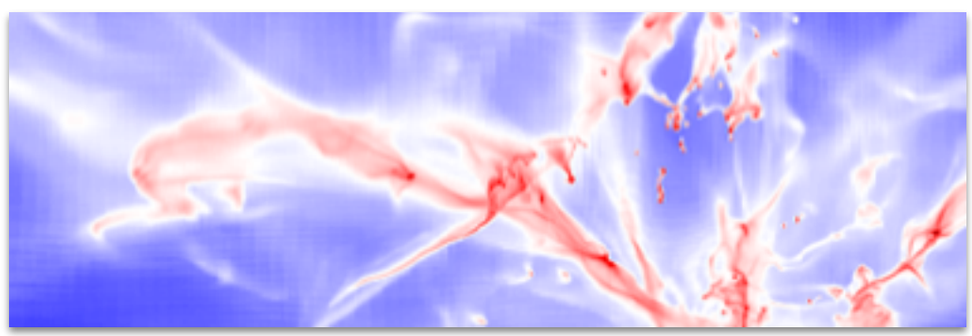
- **DEUVO** Cosmological simulations
- **GALMER** Galaxies mergers simulations
  http://galmer.obspm.fr



2 implementations with SimDM

- **ISMDB** PDR models (1D stationary code)

- **Starformat** 3D+time MHD simulations
  => becomes **Galactica** developed at CEA
  http://www.galactica-simulations.eu/db/

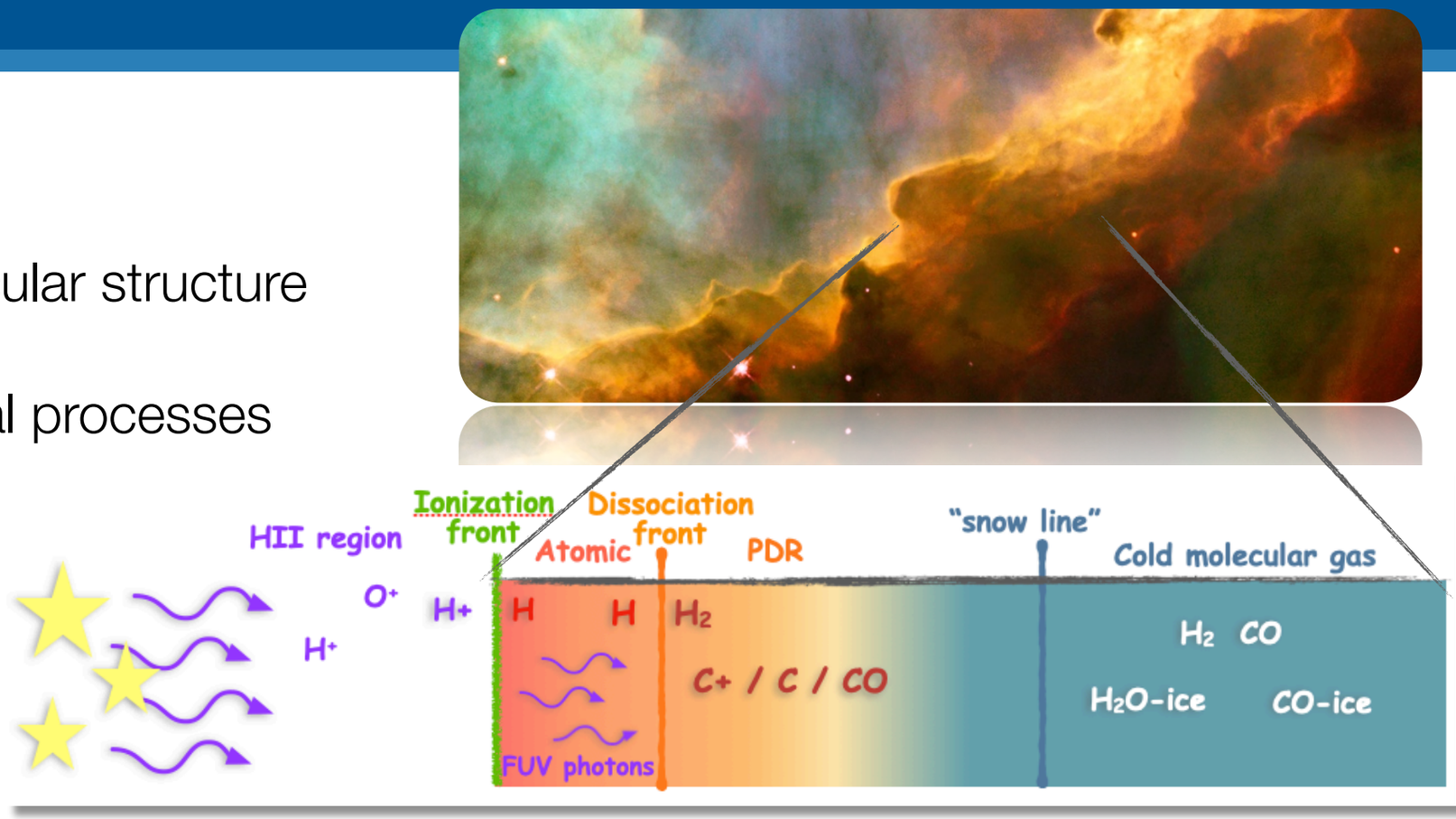


PDR code

PDR code:

- computes the atomic and molecular structure of interstellar clouds.
- analysis of physical and chemical processes

"Cloudy"-like code



Main parameters:

- UV intensity radiation field
- density, pressure
- metallicity and elemental abundances
- cosmic ray ionisation rate

Atomic & molecular data

- lines
- quantum levels
- collision rates
- chemical reaction rates
- photo-reaction cross sections

PDR code

Read inputs

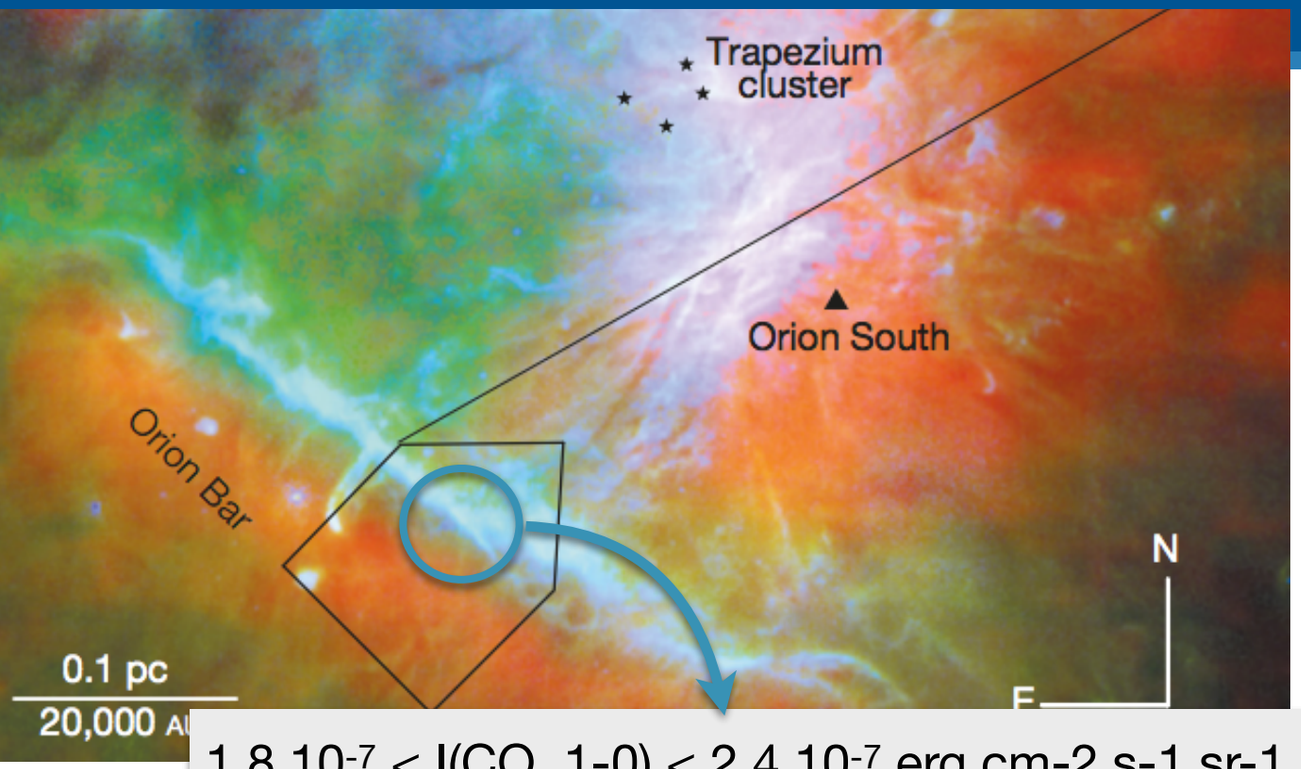
Solve the Physics

Write output files

- Chemical **density profiles**
- **Gas temperature profile**
- Grain temperature profile
- Population in quantum levels
- Chemical reaction rates

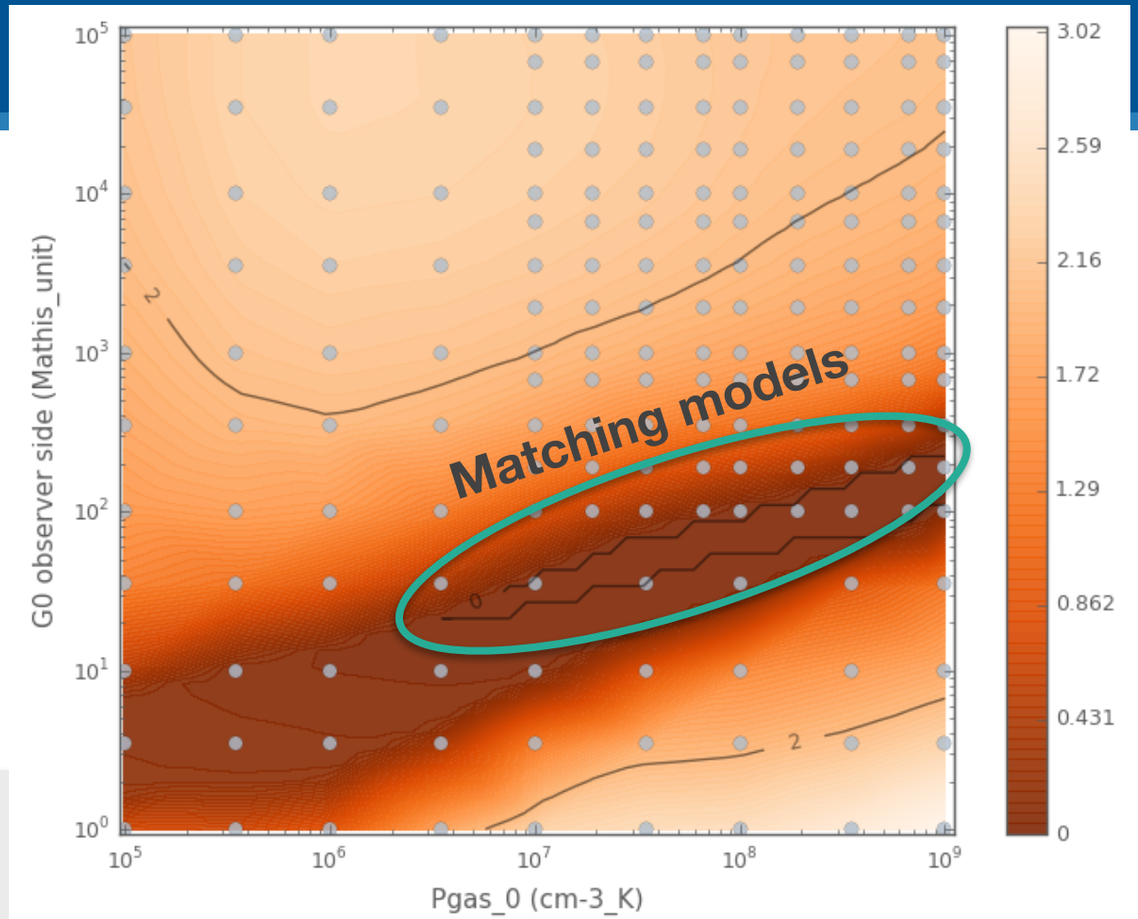
- **Line intensities**
- Column densities

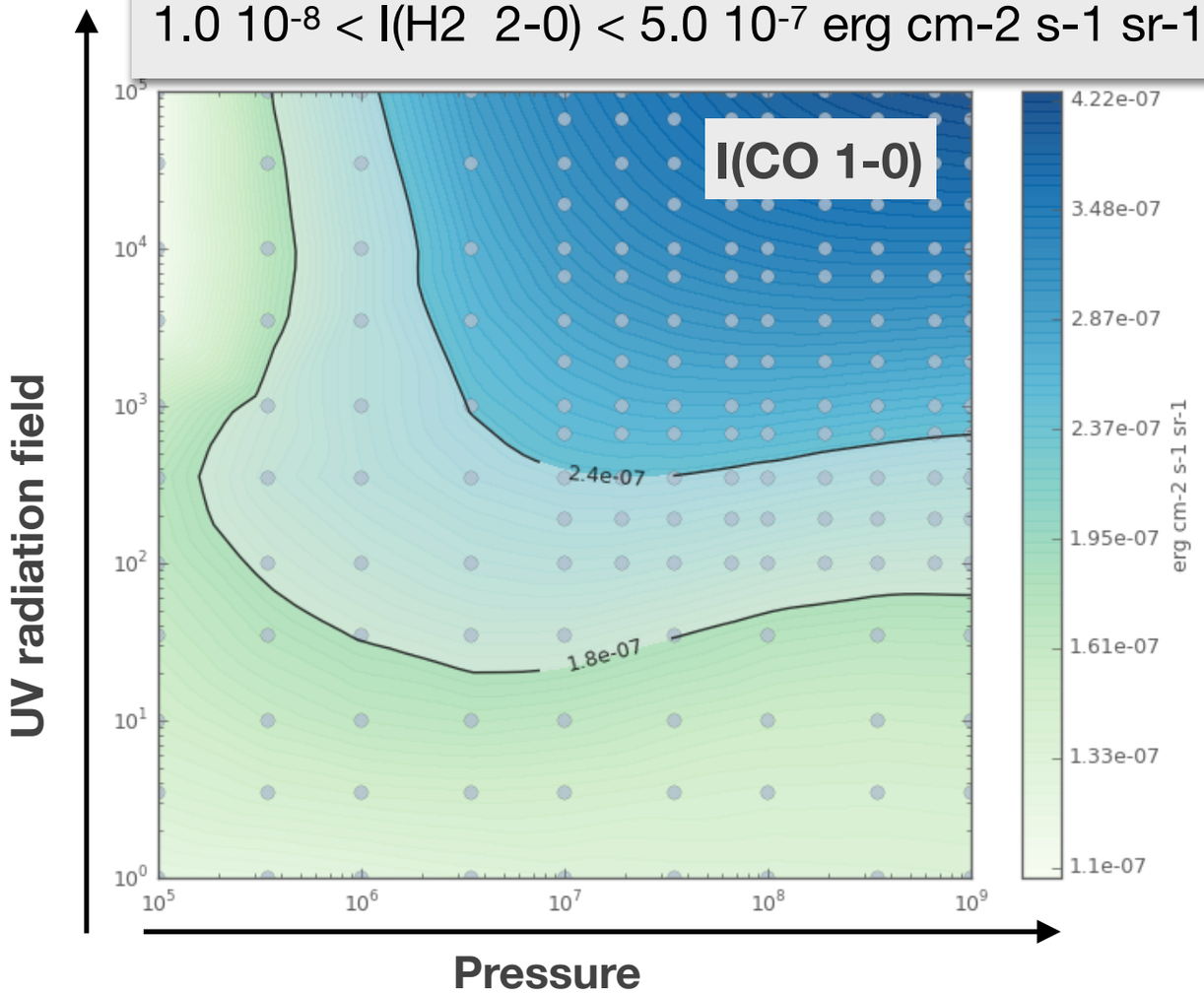
- **Spectra**

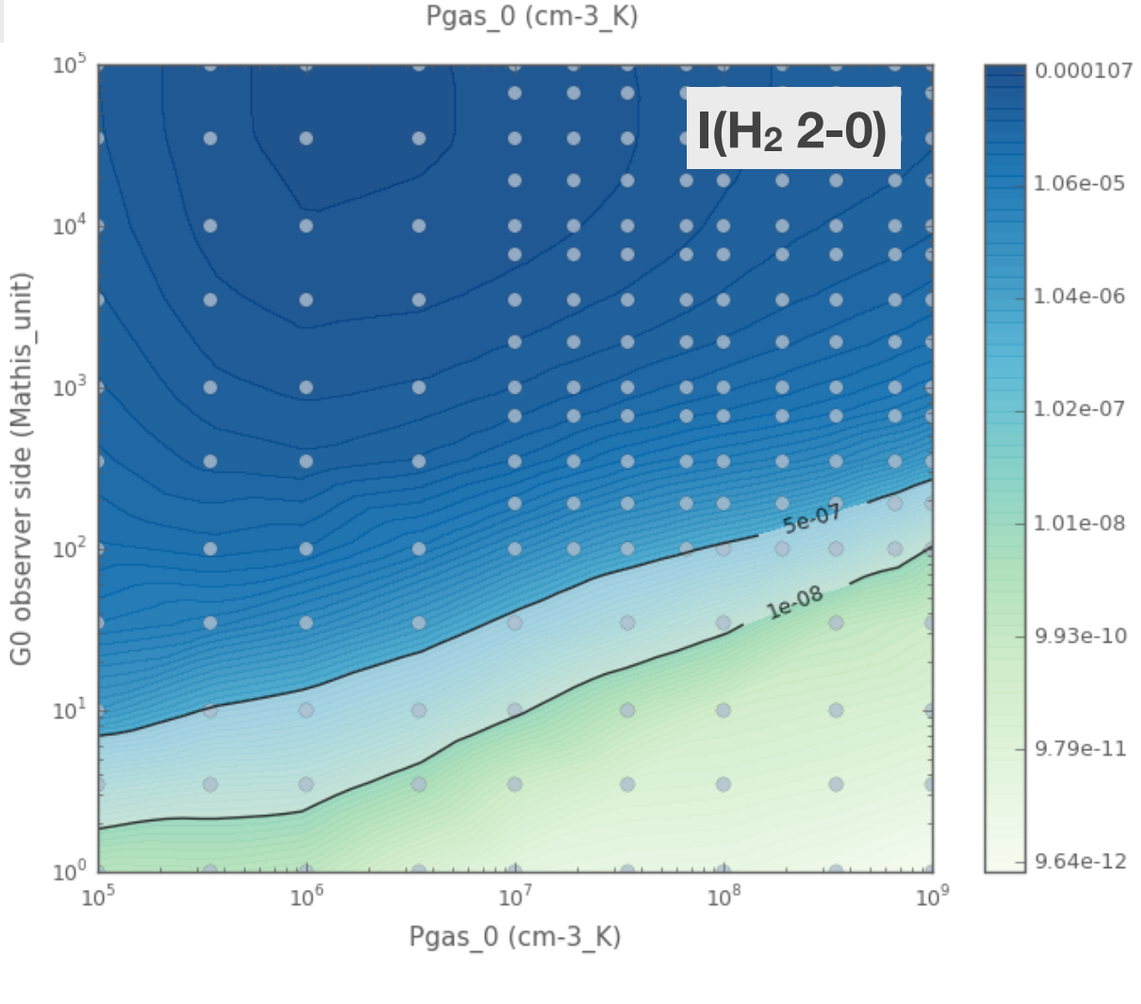


1.8 10^{-7} < I(CO 1-0) < 2.4 10^{-7} erg cm-2 s-1 sr-1

1.0 10^{-8} < I(H2 2-0) < 5.0 10^{-7} erg cm-2 s-1 sr-1



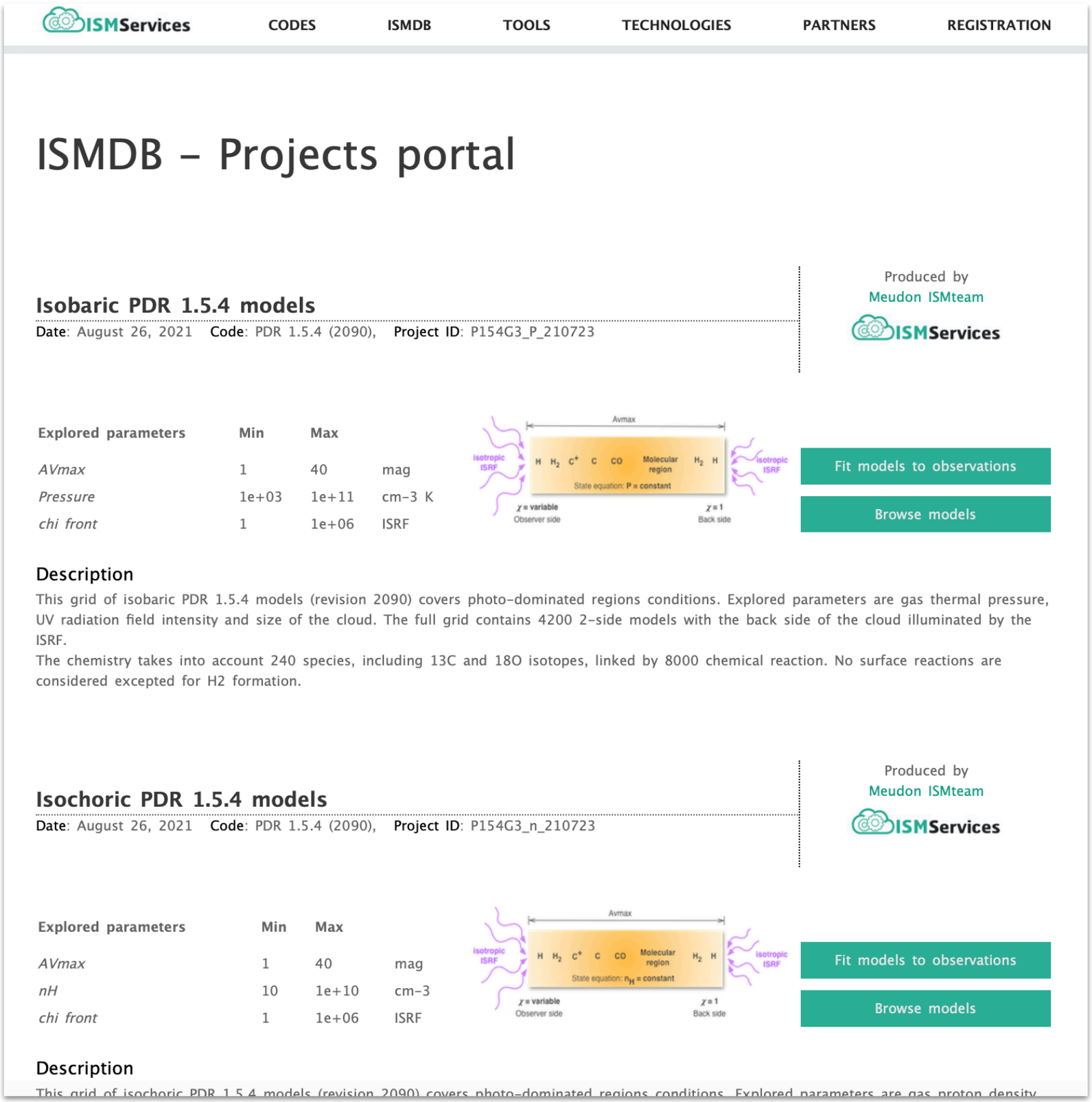




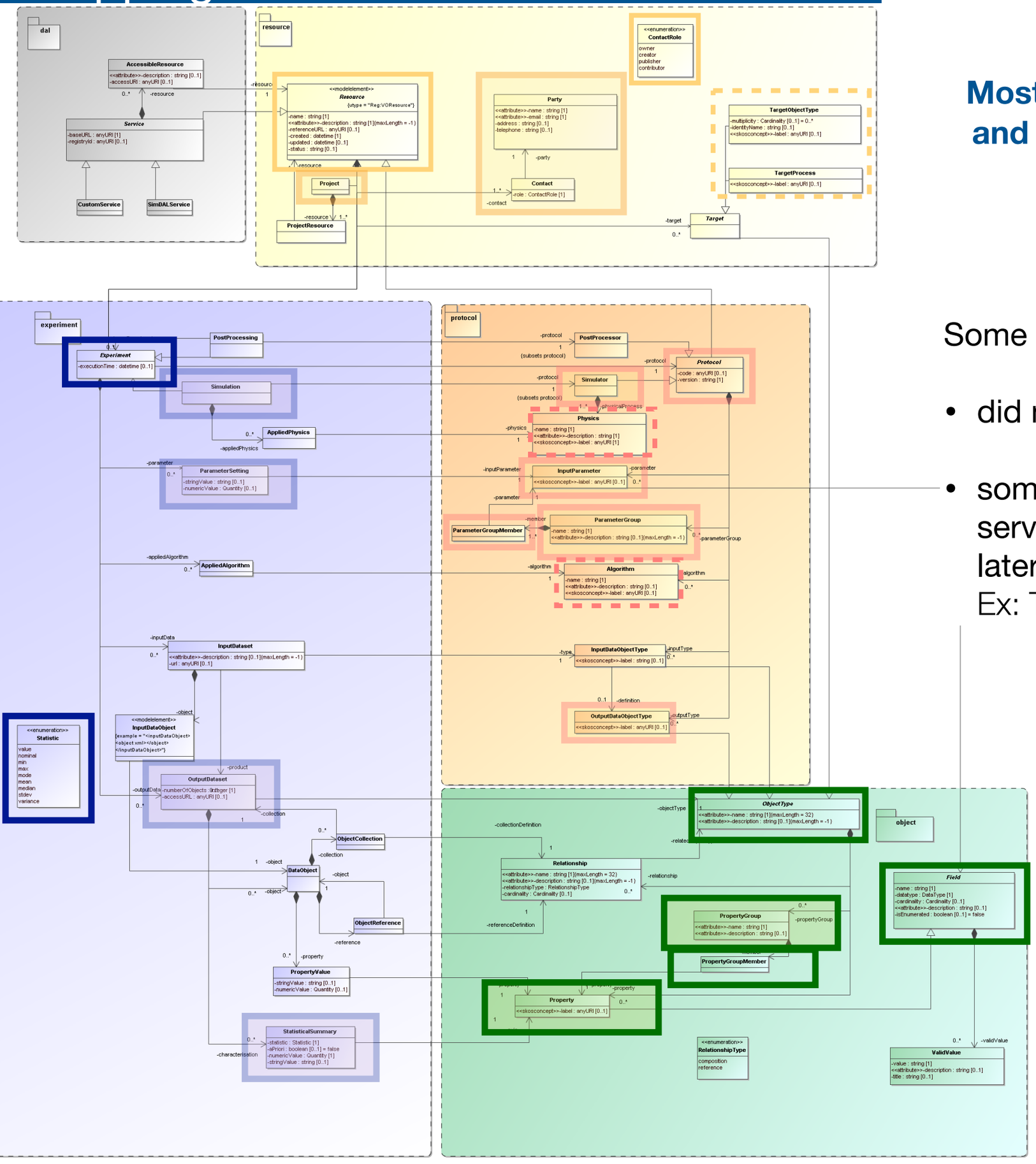
Mapping of the PDR code and its results on SimDM

ISMDB : Interstellar Medium Database
http://ismdb.obspm.fr

One grid of simulations
~ 1E3 - 1E4 PDR models



Mapping of the PDR code and its results on SimDM



Most of key SimDM classes are used and required for our implementation

Some classes not implemented yet

- did not exist when we implemented SimDM
- some metadata are not a priority for our service presently but would be easy to add later

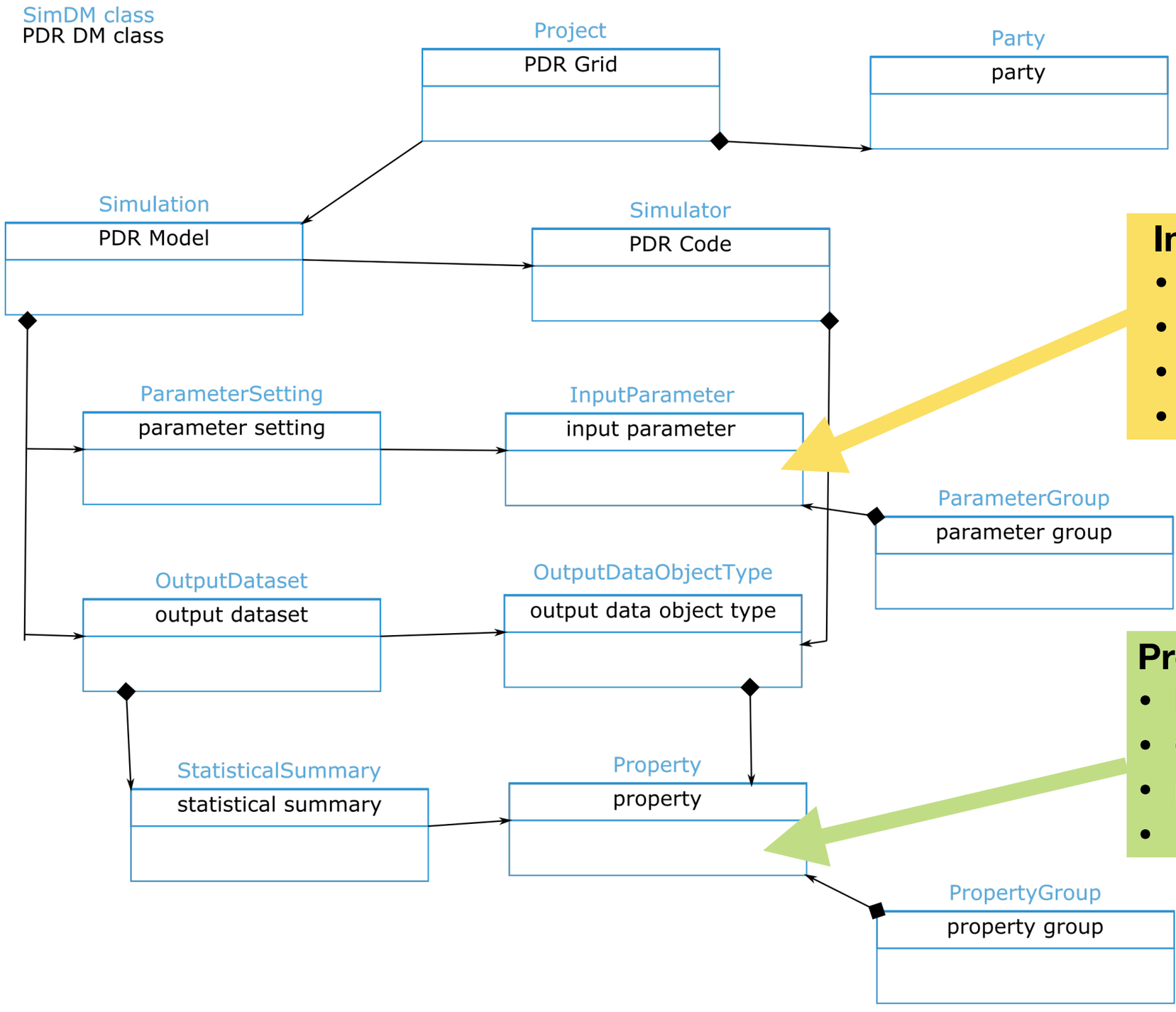
Ex: TargetObjectType, Physics, Algorithms

Mapping of the PDR code and its results on SimDM

SimDM Simulations Meta Model

instance of

PDR Simulations Data Model



Input parameters
- **density**
- **UV radiation field**
- **size of cloud**
- **…**

About 20 metadata

Properties
- **Line intensities**
- **Column densities**
- **max/min temperature**
- **…**

About 150 000 metadata

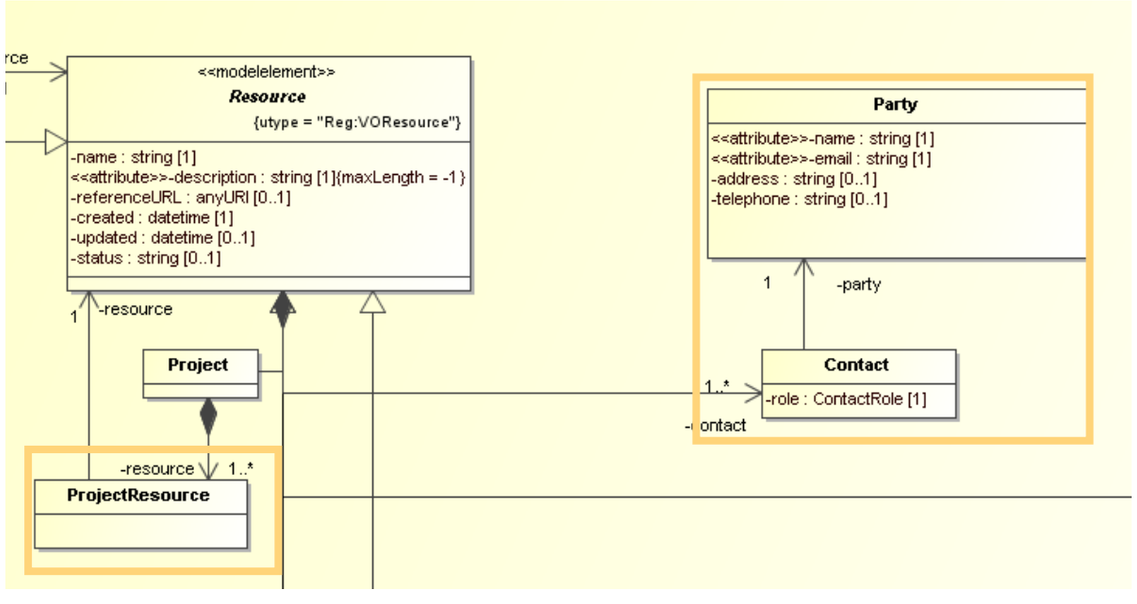
Mapping of the PDR code and its results on SimDM

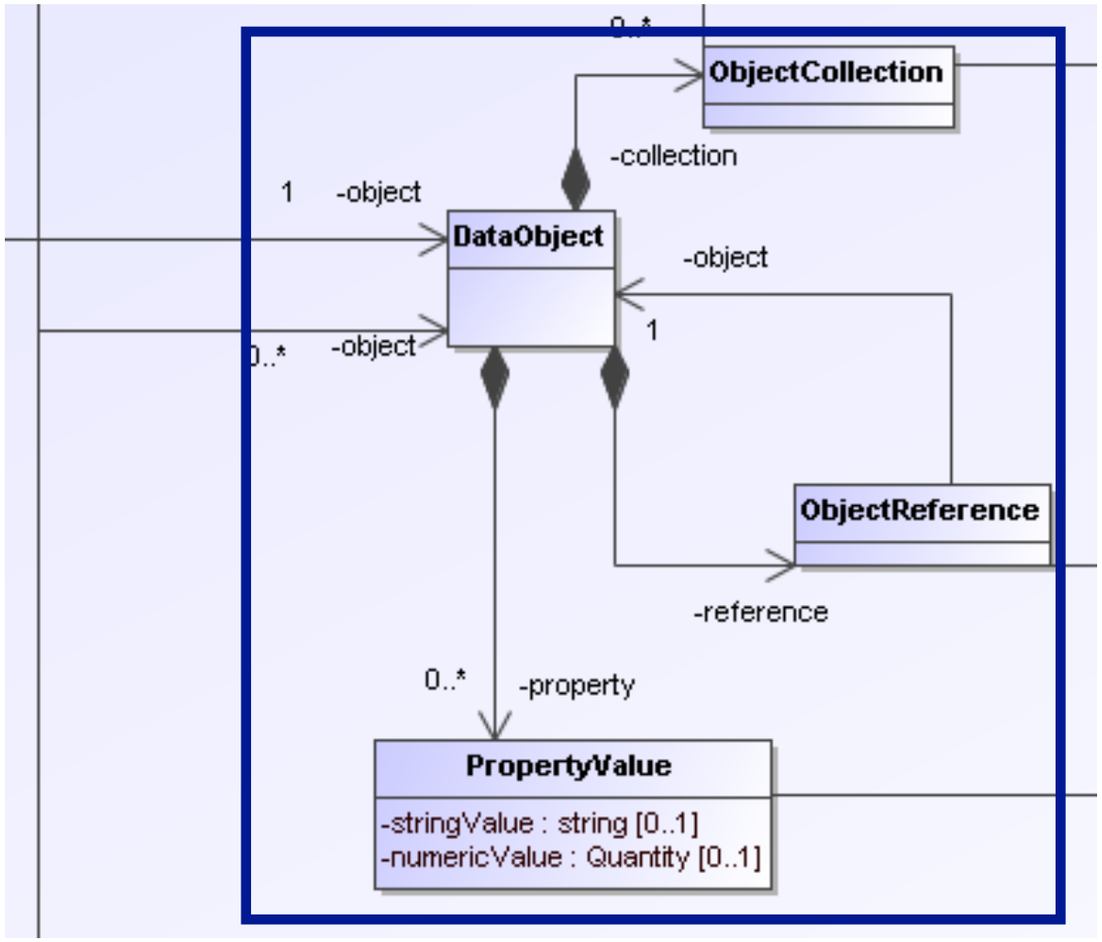
Some freedoms with SimDM

Resource part

1. We merged Party and Contact (seems was the case in old SimDM anyway)

2. We skipped the ProjectResource interface (also not sure was present in old SimDM)





Experiment part

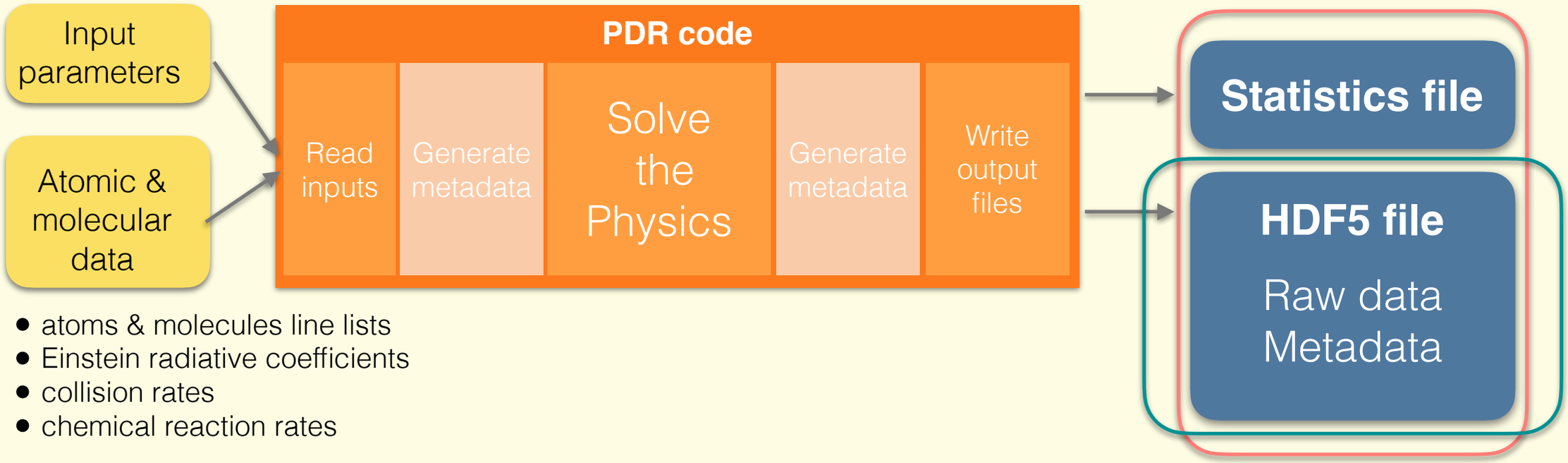
The most critical point introduced in recent SimDM allow for nested datasets.

1. For our single object "molecular cloud" we had to do the trick of defining datasets of 1 object and define "StatisticalSummary" with "statistic"="value"

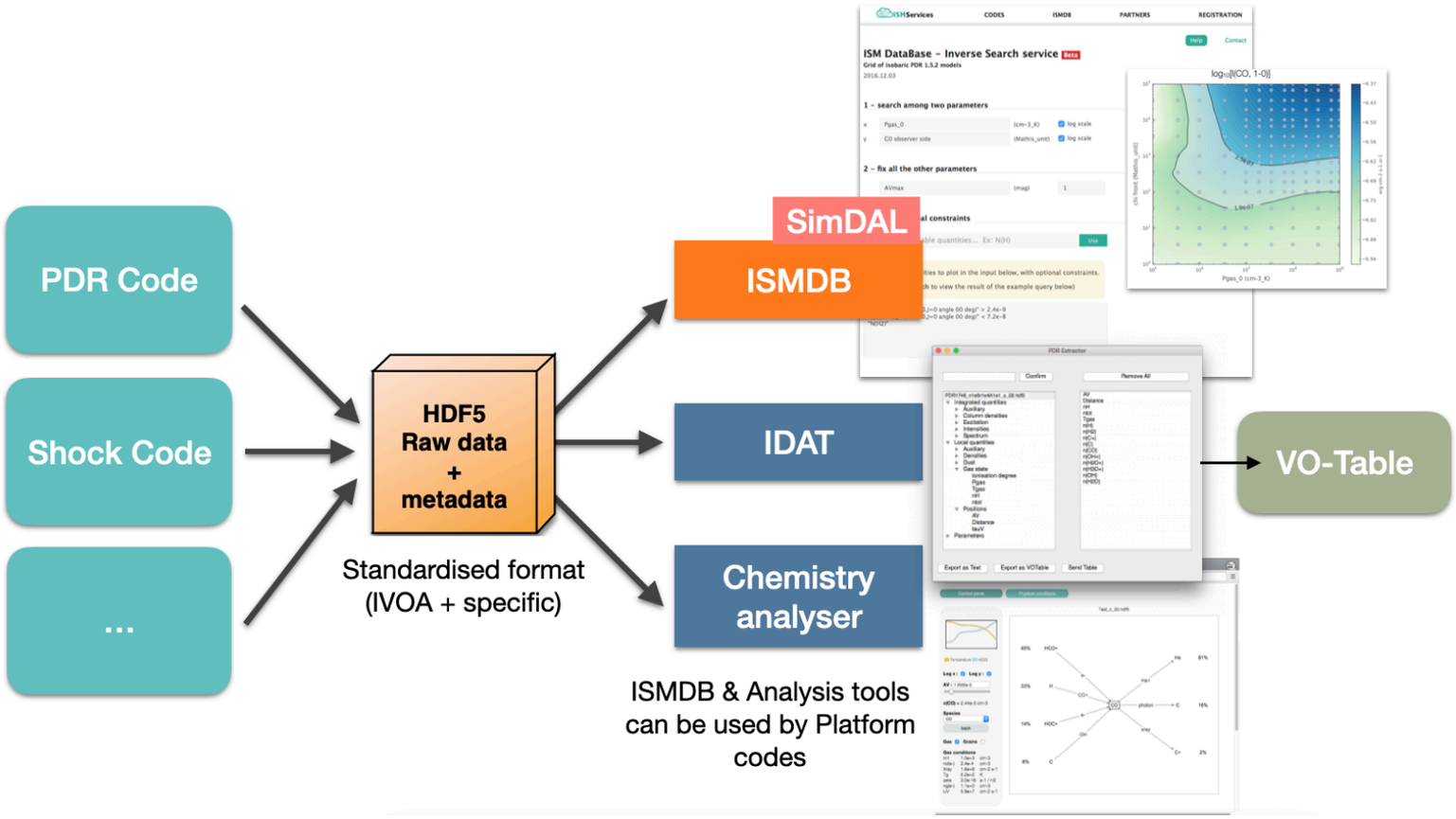
2. Is very powerful but also can bring real complexity for a client when encoded in an exchange format (VOTable in SimDAL, see VODML) due to the nesting.

Production of metadata

The production of metadata for SimDM is **integrated inside the code**

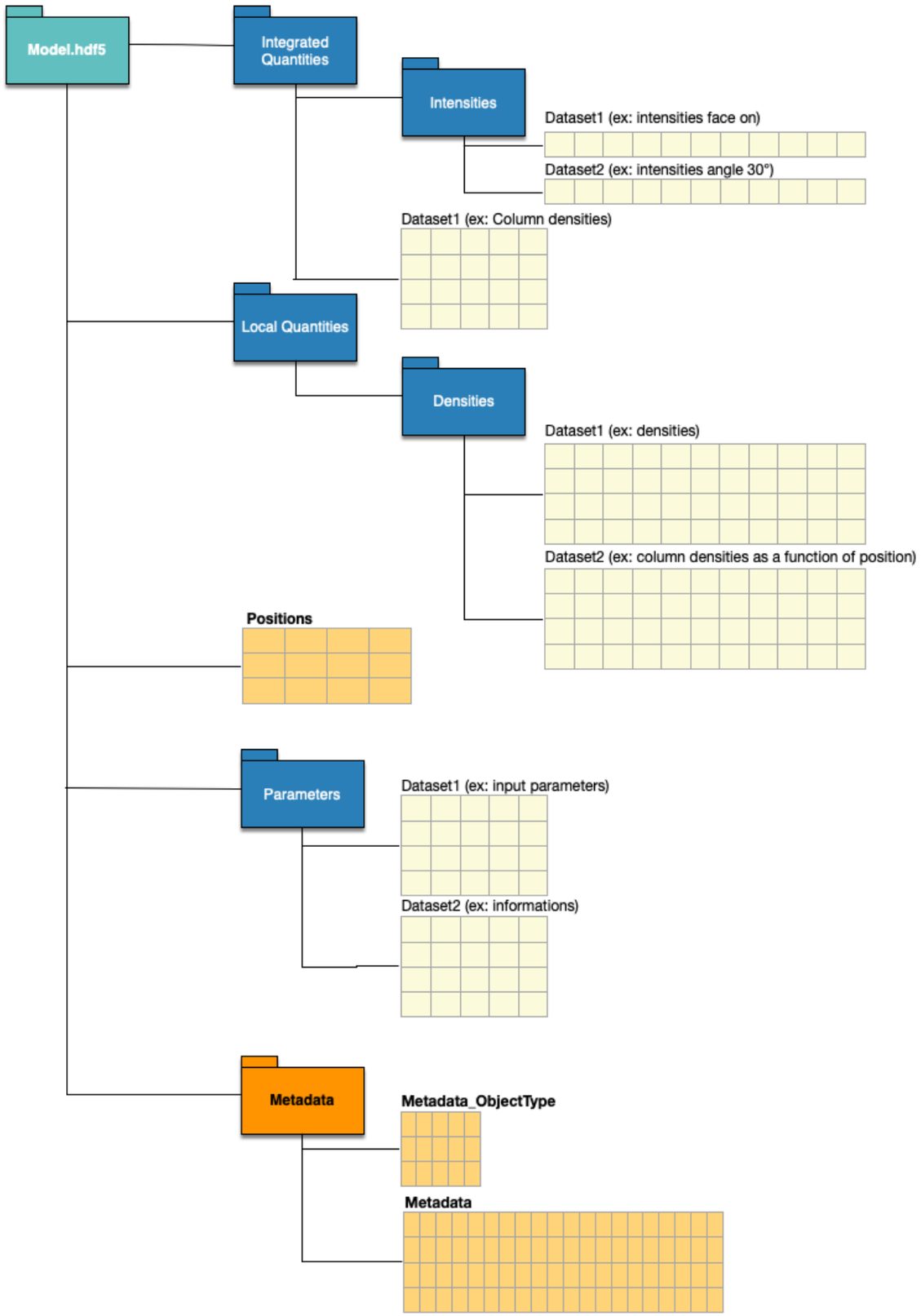


- atoms & molecules line lists
- Einstein radiative coefficients
- collision rates
- chemical reaction rates



Production of metadata : dataset

Most of SimDM metadata are stored in HDF5 files



Data - Code results

- **HDF5 with a metadata table**
  - raw data (all quantities computed by the code)
  - metadata : datatype, utype, unit, SKOS concept, …
- statistic file

Metadata tables that contains :

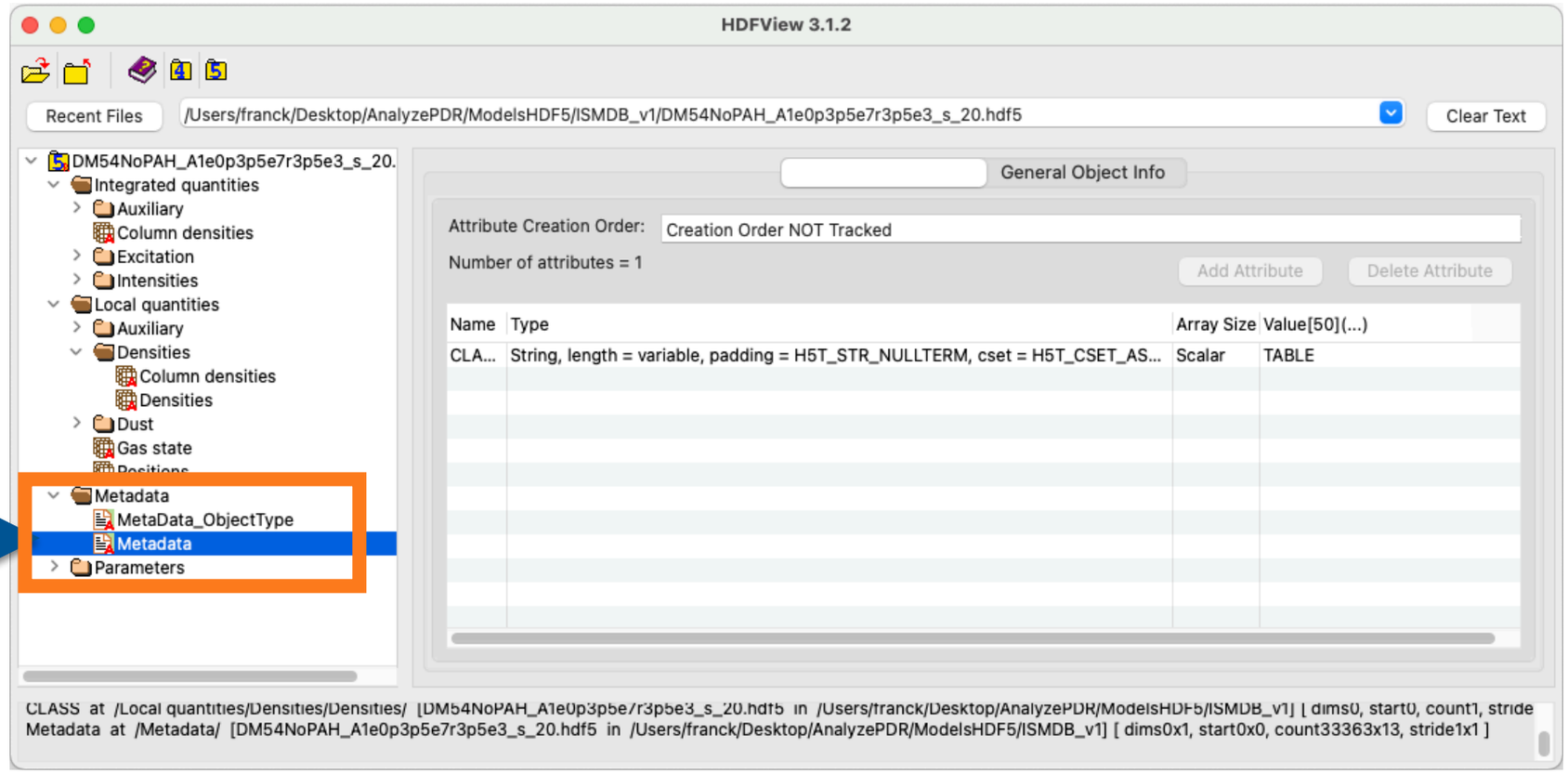
- List of object types
- SimDM attributes for all quantities computed by the code & Input parameters, …

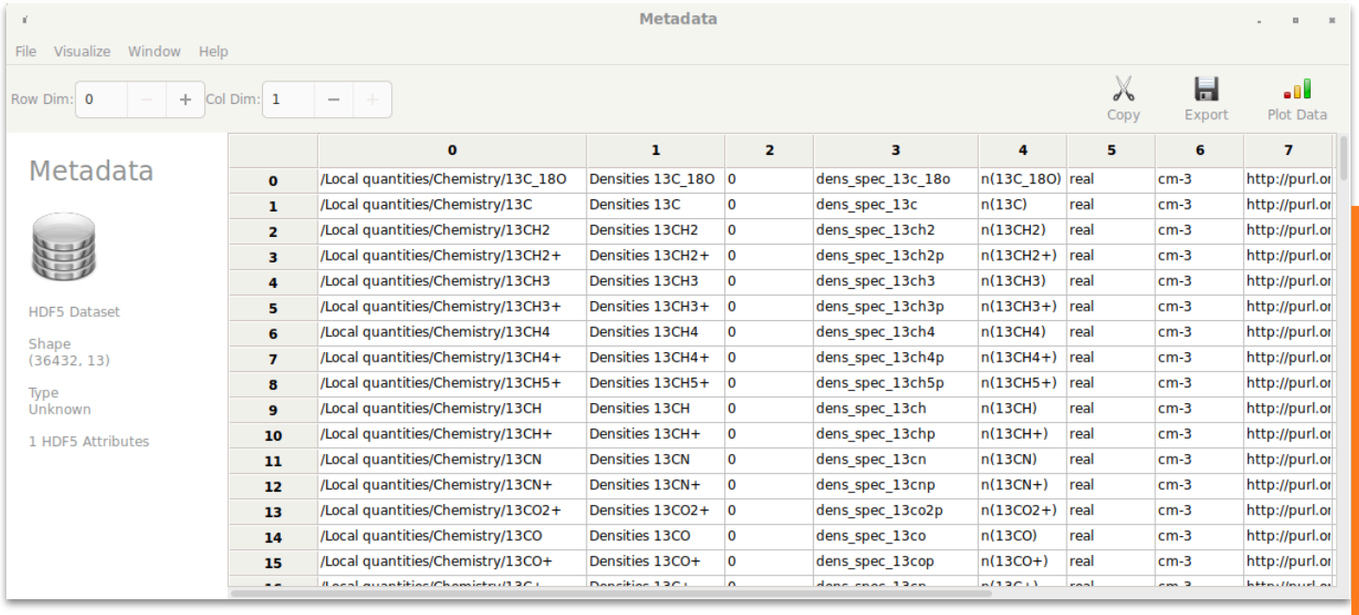
Production of metadata : dataset

How SimDM attributes are stored in HDF5 files

Content of the HDF5 file with HDFView

SimDM attributes
- Properties, Input Parameters
- ObjectTypes





SimDM attributes

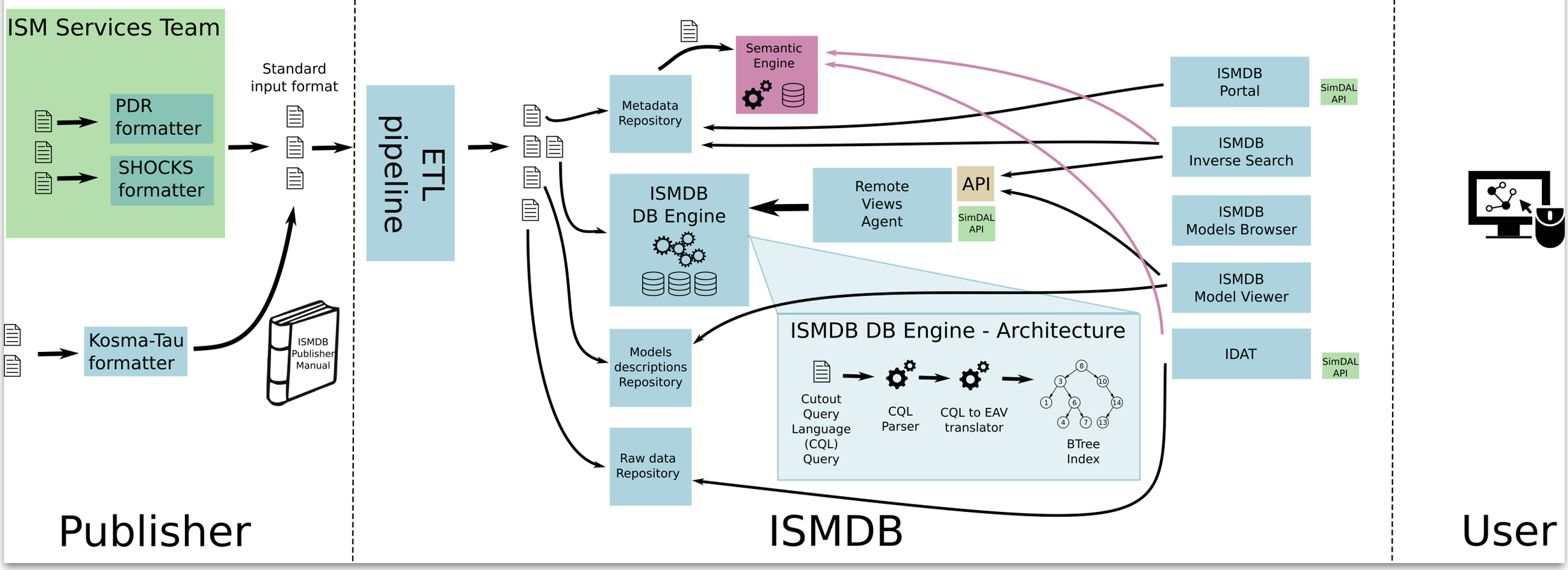
Col.	Content	Description
0	Path towards the dataset	
1	Name of the dataset	
2	Column index	
3	PubID	ID of the quantity. These IDs must be unique in the HDF5 file.
4	Hname (Human readable name)	Human readable name of the quantity. These strings are presented in User Interfaces and some must not be too long. Each quantity has also a description (column 10 of the metadata table).
5	DataType	float, integer, string
6	Unit	unit
7	SKOS	SKOS concept of the quantity
8	UCD	UCD (united content descriptor) of the quantity
9	utype	utype in the Simulation Data Model of the quantity. Often it is either Property or InputParameter
10	Description	Description of the quantity
11	ParentObject	ID of the parent of the quantity (ObjectType or ParentObject). Those are described in another specific dataset : `MetaData_ObjectType`. For example, in the PDR code, we have 3 ParentObjects : Cloud, MeshCell et Grain
12	Group	ID of the group to which the quantity is associated to.

SimDM attributes

Ingestion of metadata in ISMDB

Developing ISMDB software architecture is complex and needs time

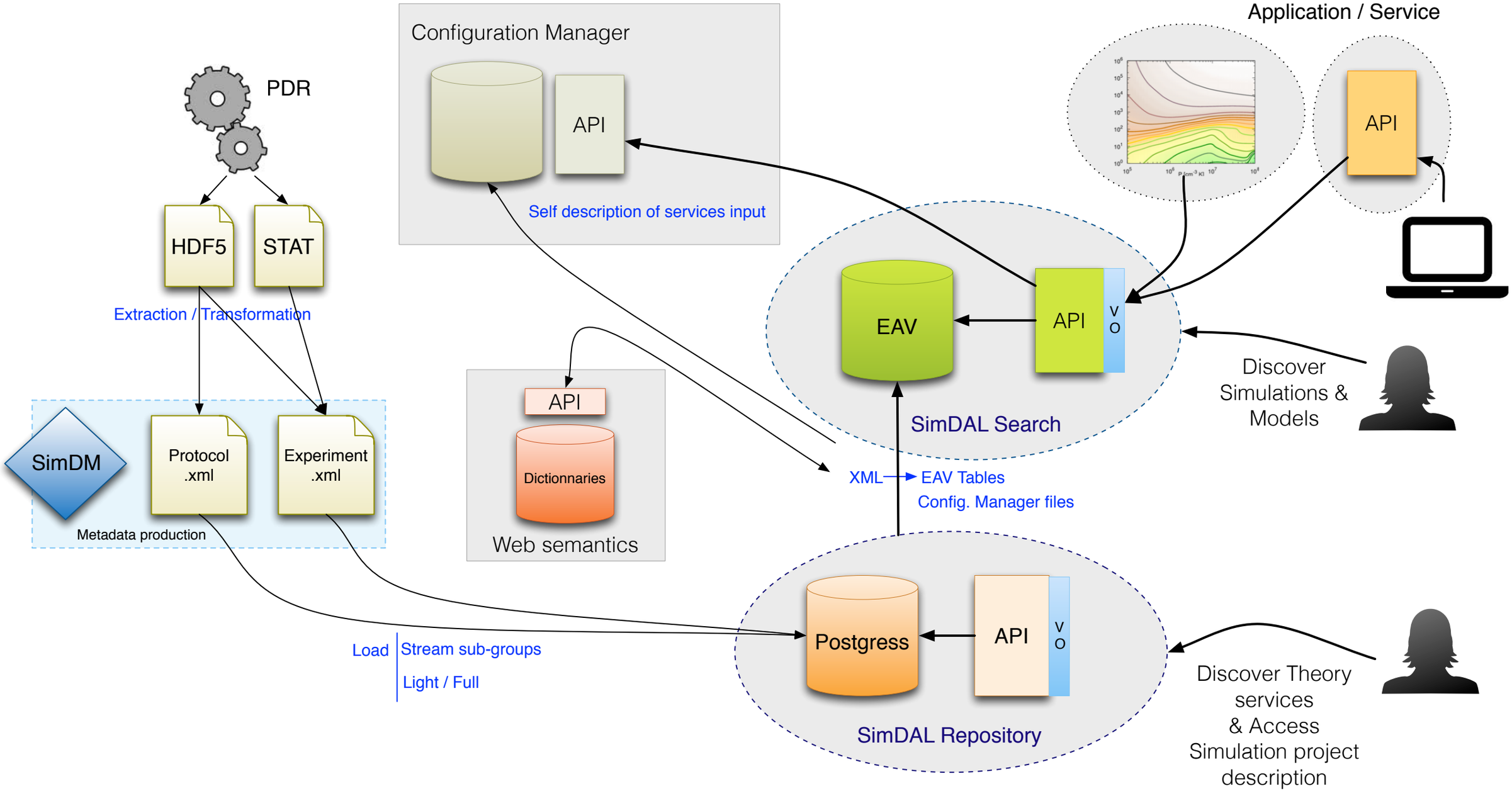
- Not any similar enough existing solution that could be re-used
- We has to **develop our own DB engine**
- **Use cases are very dynamic** over time: this is research
- The software is often complex & the output of the codes are complex, we have to **make it simple for the users**
    - This is by far the hardest part



Ingestion of metadata in ISMDB

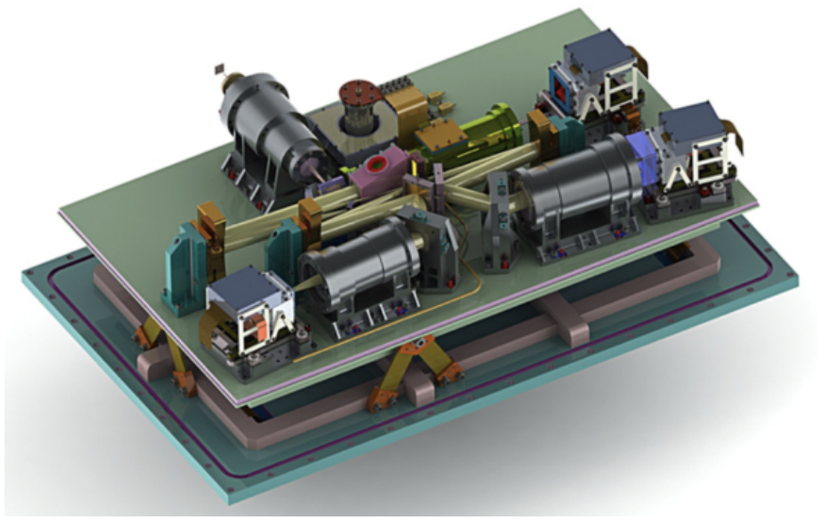
ISMDB infrastructure development

- **modular**
- based on **robust & mature technologies**
- **generic**: can integrate models from any similar code than the PDR code



Example of application

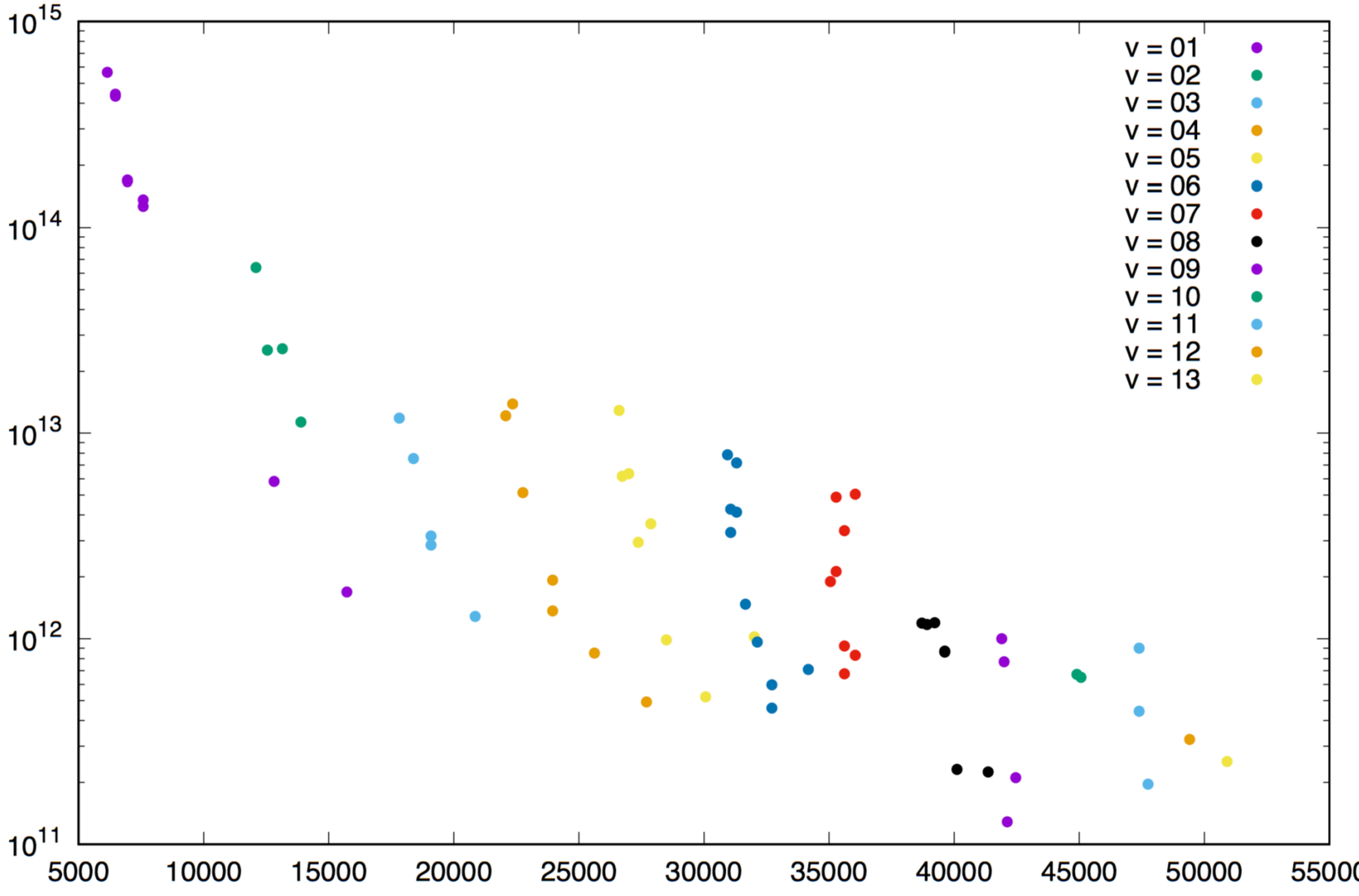
IGRINS observations
Instrument - Univ. Texas, Austin
Mc Donald observatory
Bands: H et K (1.5 to 2.5 microns)
R = 45 000



Observations of NGC 7023
(Le et al. - 2016 / ArXiV)

- Detection of **70 H_2 lines in NGC 7023**
- Conclude to a clumpy medium

H_2 excitation diagram



Column density in H_2 levels / level degeneresccy [cm^{-2}]

Example of application

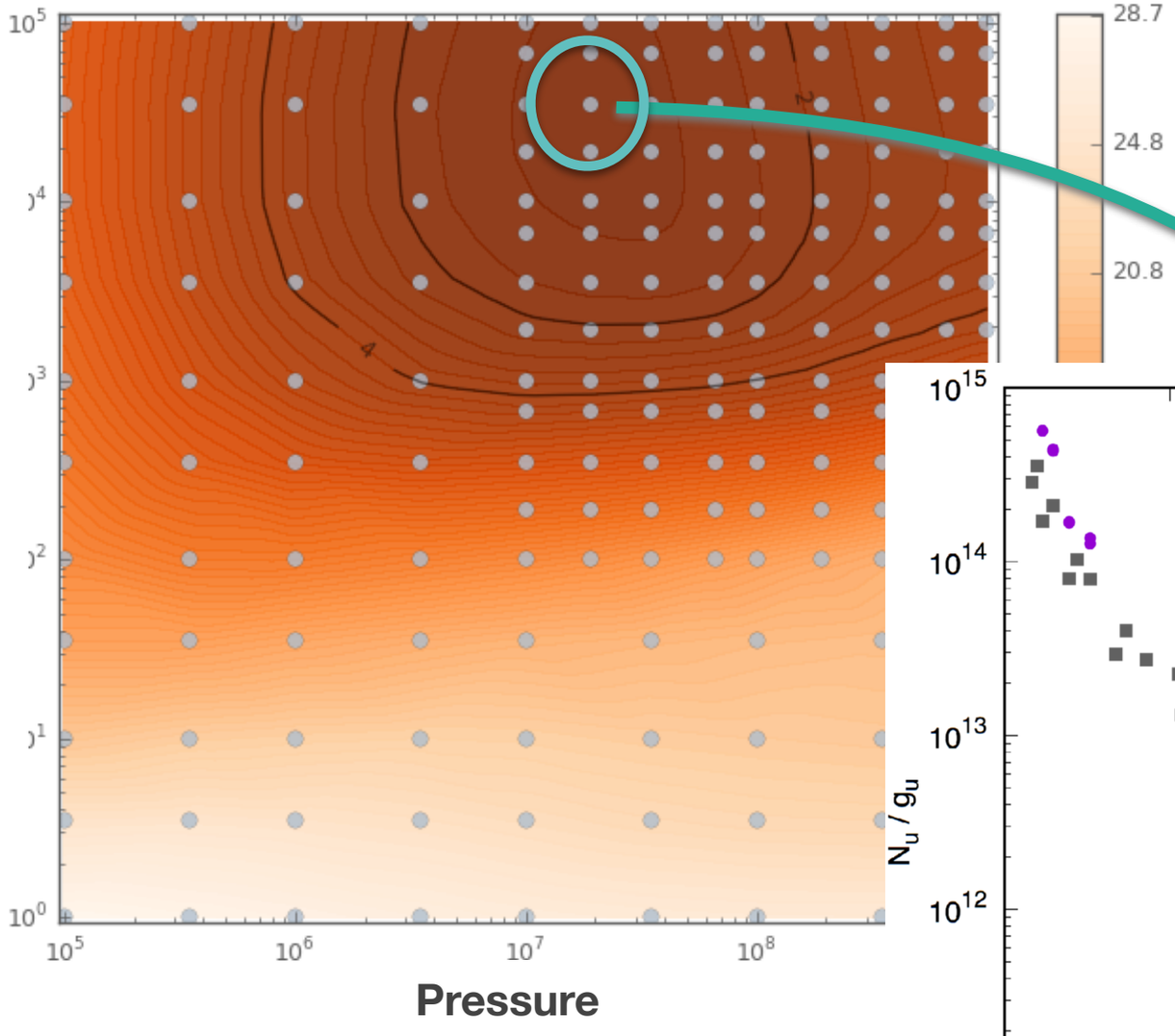
Build the query for the 70 H_2 lines (140 constraints)

```
"I(H2 v=4,J=9->v=2,J=9 angle 60 deg)" < 5.408e-06
"I(H2 v=4,J=9->v=2,J=9 angle 60 deg)" > 2.912e-06
"I(H2 v=4,J=1->v=2,J=3 angle 60 deg)" < 2.665e-05
"I(H2 v=4,J=1->v=2,J=3 angle 60 deg)" > 1.435e-05
"I(H2 v=5,J=4->v=3,J=4 angle 60 deg)" < 9.659e-06
"I(H2 v=5,J=4->v=3,J=4 angle 60 deg)" > 5.201e-06
"I(H2 v=3,J=3->v=1,J=5 angle 60 deg)" < 1.2532e-05
"I(H2 v=3,J=3->v=1,J=5 angle 60 deg)" > 6.748e-06
"I(H2 v=5,J=5->v=3,J=5 angle 60 deg)" < 9.607e-06
"I(H2 v=5,J=5->v=3,J=5 angle 60 deg)" > 5.173e-06
"I(H2 v=6,J=2->v=4,J=0 angle 60 deg)" < 7.67e-06
"I(H2 v=6,J=2->v=4,J=0 angle 60 deg)" > 4.13e-06
"I(H2 v=10,J=1->v=7,J=3 angle 60 deg)" < 4.121e-06
"I(H2 v=10,J=1->v=7,J=3 angle 60 deg)" > 2.219e-06
"I(H2 v=5,J=0->v=3,J=2 angle 60 deg)" < 1.118e-05
"I(H2 v=5,J=0->v=3,J=2 angle 60 deg)" > 6.02e-06
"I(H2 v=5,J=7->v=3,J=7 angle 60 deg)" < 6.955e-06
"I(H2 v=5,J=7->v=3,J=7 angle 60 deg)" > 3.745e-06
"I(H2 v=4,J=2->v=2,J=4 angle 60 deg)" < 1.1531e-05
"I(H2 v=4,J=2->v=2,J=4 angle 60 deg)" > 6.209e-06
"I(H2 v=7,J=4->v=5,J=2 angle 60 deg)" < 5.109e-06
"I(H2 v=7,J=4->v=5,J=2 angle 60 deg)" > 2.751e-06
"I(H2 v=6,J=1->v=4,J=1 angle 60 deg)" < 1.846e-05
"I(H2 v=6,J=1->v=4,J=1 angle 60 deg)" > 9.94e-06
"I(H2 v=6,J=2->v=4,J=2 angle 60 deg)" < 1.599e-05
"I(H2 v=6,J=2->v=4,J=2 angle 60 deg)" > 8.61e-06
"I(H2 v=5,J=9->v=3,J=9 angle 60 deg)" < 1.729e-05
"I(H2 v=5,J=9->v=3,J=9 angle 60 deg)" > 9.31e-06
"I(H2 v=5,J=1->v=3,J=3 angle 60 deg)" < 2.379e-05
"I(H2 v=5,J=1->v=3,J=3 angle 60 deg)" > 1.281e-05
"I(H2 v=13,J=1->v=9,J=1 angle 60 deg)" < 9.334e-07
"I(H2 v=13,J=1->v=9,J=1 angle 60 deg)" > 5.026e-07
"I(H2 v=6,J=3->v=4,J=3 angle 60 deg)" < 1.287e-05
"I(H2 v=6,J=3->v=4,J=3 angle 60 deg)" > 6.93e-06
```

```
I(H2 v=7,J=3->v=5,J=1 angle 60 deg)" < 1.1531e-05
"I(H2 v=7,J=3->v=5,J=1 angle 60 deg)" > 6.209e-06
"I(H2 v=4,J=3->v=2,J=5 angle 60 deg)" < 1.2961e-05
"I(H2 v=4,J=3->v=2,J=5 angle 60 deg)" > 6.979e-06
"I(H2 v=6,J=4->v=4,J=4 angle 60 deg)" < 3.523e-06
"I(H2 v=6,J=4->v=4,J=4 angle 60 deg)" > 1.897e-06
"I(H2 v=6,J=5->v=4,J=5 angle 60 deg)" < 7.878e-06
"I(H2 v=6,J=5->v=4,J=5 angle 60 deg)" > 4.242e-06
"I(H2 v=3,J=5->v=1,J=7 angle 60 deg)" < 2.457e-06
"I(H2 v=3,J=5->v=1,J=7 angle 60 deg)" > 1.323e-06
"I(H2 v=11,J=1->v=8,J=1 angle 60 deg)" < 2.899e-06
"I(H2 v=11,J=1->v=8,J=1 angle 60 deg)" > 1.561e-06
"I(H2 v=7,J=2->v=5,J=0 angle 60 deg)" < 4.628e-06
"I(H2 v=7,J=2->v=5,J=0 angle 60 deg)" > 2.492e-06
"I(H2 v=8,J=7->v=6,J=5 angle 60 deg)" < 5.824e-06
"I(H2 v=8,J=7->v=6,J=5 angle 60 deg)" > 3.136e-06
"I(H2 v=5,J=2->v=3,J=4 angle 60 deg)" < 9.425e-06
"I(H2 v=5,J=2->v=3,J=4 angle 60 deg)" > 5.075e-06
"I(H2 v=6,J=0->v=4,J=2 angle 60 deg)" < 9.776e-06
"I(H2 v=6,J=0->v=4,J=2 angle 60 deg)" > 5.264e-06
"I(H2 v=6,J=7->v=4,J=7 angle 60 deg)" < 1.2532e-05
"I(H2 v=6,J=7->v=4,J=7 angle 60 deg)" > 6.748e-06
"I(H2 v=11,J=3->v=8,J=3 angle 60 deg)" < 1.924e-06
"I(H2 v=11,J=3->v=8,J=3 angle 60 deg)" > 1.036e-06
"I(H2 v=1,J=11->v=0,J=9 angle 60 deg)" < 4.407e-06
"I(H2 v=1,J=11->v=0,J=9 angle 60 deg)" > 2.373e-06
"I(H2 v=8,J=5->v=6,J=3 angle 60 deg)" < 5.122e-06
"I(H2 v=8,J=5->v=6,J=3 angle 60 deg)" > 2.758e-06
"I(H2 v=7,J=1->v=5,J=1 angle 60 deg)" < 1.2922e-05
"I(H2 v=7,J=1->v=5,J=1 angle 60 deg)" > 6.958e-06
"I(H2 v=8,J=4->v=6,J=2 angle 60 deg)" < 5.109e-06
"I(H2 v=8,J=4->v=6,J=2 angle 60 deg)" > 2.751e-06
"I(H2 v=6,J=1->v=4,J=3 angle 60 deg)" < 2.405e-05
…
…
…
…
```

Example of application

ISMDB results:

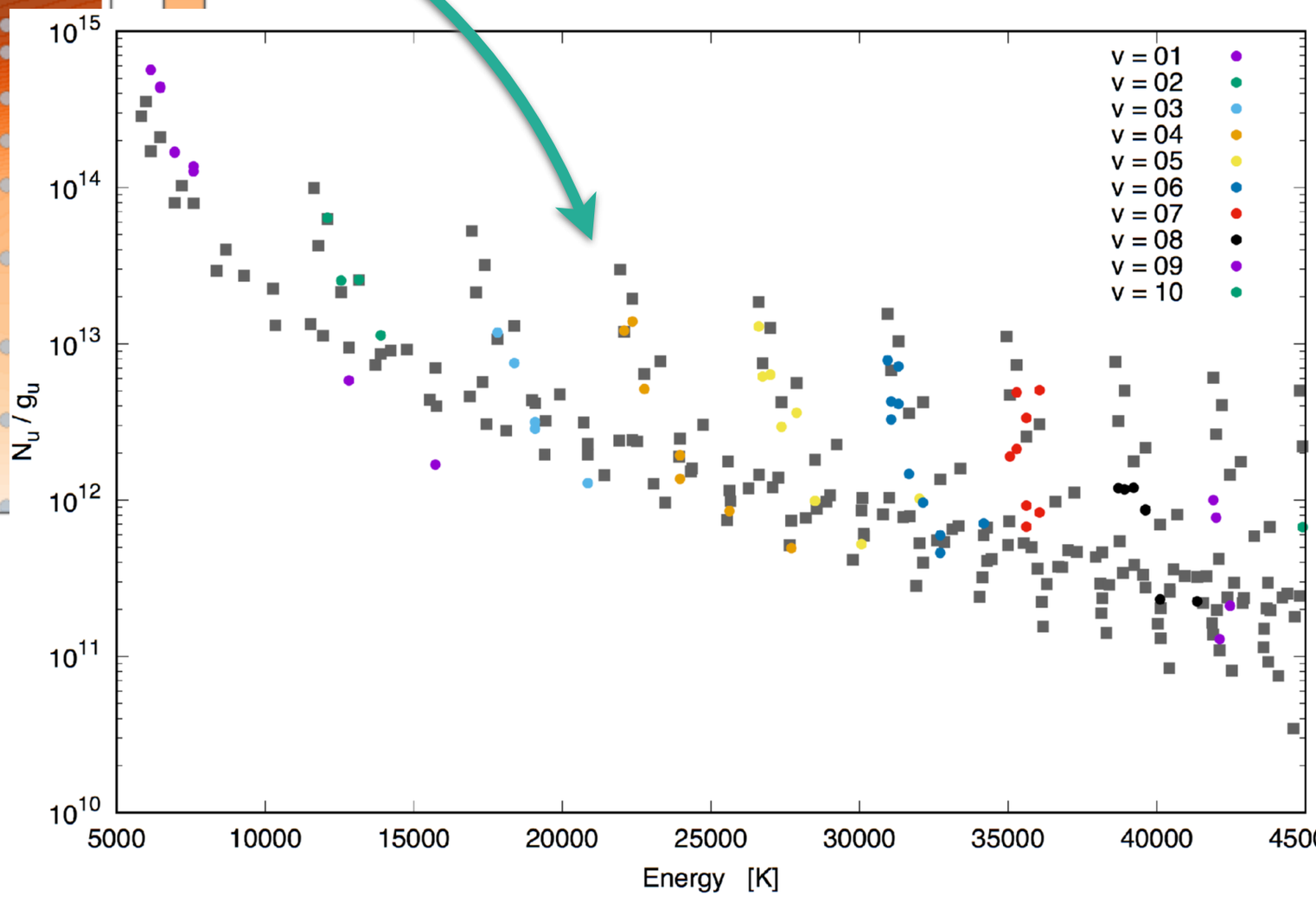


First order solution:

P = 10^7 - 10^8 K cm^{-3}

UV field ~ 10^4 mean interstellar UV radiation field

Download models



Help to the interpretation of data in a few minutes instead of weeks of work

Publication of simulations of other groups

How to convince other groups to publish in the VO using SimDM ?

- Propose them to publish in existing architecture
- Explain them how to without using SimDM !

➔ we wrote a documentation to explain to other scientists how to publish their simulations in ISMDB

ISMServices

DOCUMENTATION

ISMDB Publisher Manual

19 August 2020

Table of content

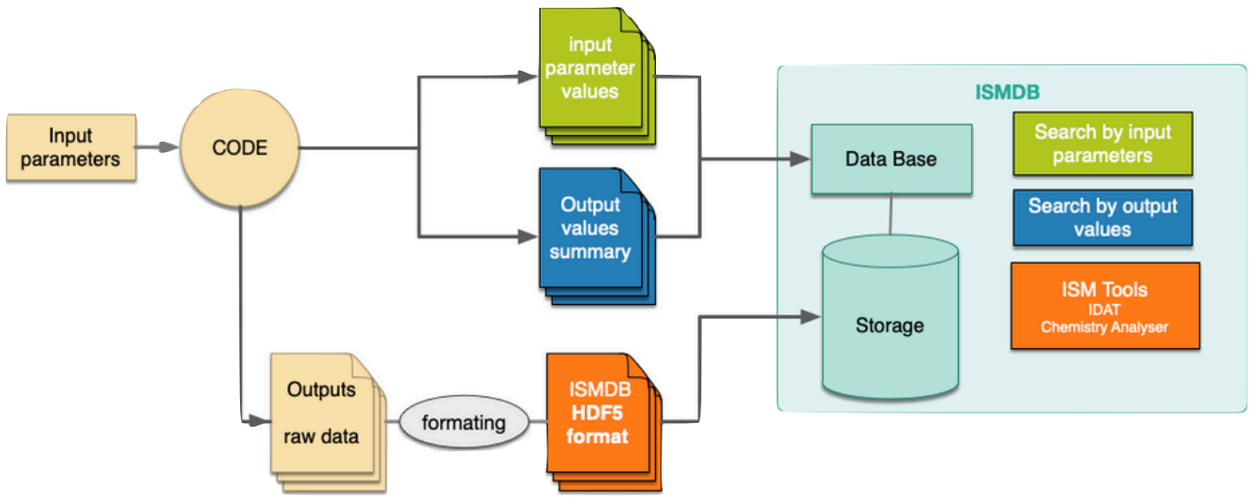
The code description file
The project description file
The input paramters files
The output files

This documentation has been formatted as an HTML page meant to be read in a web browser. Printing it will result in cutting out part of the tables and example files.

This note explains how to publish grids of models in the Interstellar Medium Data Base, ISMDB. ISMDB is designed to publish astrochemical model grids produced by various types of codes (radiative transfer codes, time dependent astrochemical codes, PDR codes, shock codes, etc.). Compared to other codes in astrophysics, such models are characterised by the large number of quantities they can compute, such as the densities of hundreds of species or the integrated line intensities of tens of thousands of lines. The structure of ISMDCB has been designed specifically to handle such high dimensional datasets. This ability to store and manipulate high dimensional data allows some of the advanced usages proposed in ISMDB.

- Structure of the HDF5 files
- Additional files : SimDM statistics, resource, …

Case	Input parameters values	Output values summary	output raw data in ISMDB HDF5 format	ISMDB feature
1	✔			Search by input parameters
2	✔	✔		Search by input parameters + inverse search
3	✔	✔	✔	Search by input parameters + inverse search + ISM Tools (IDAT, ...)
4	✔		✔	Search by input parameters + ISM Tools (IDAT, ...)

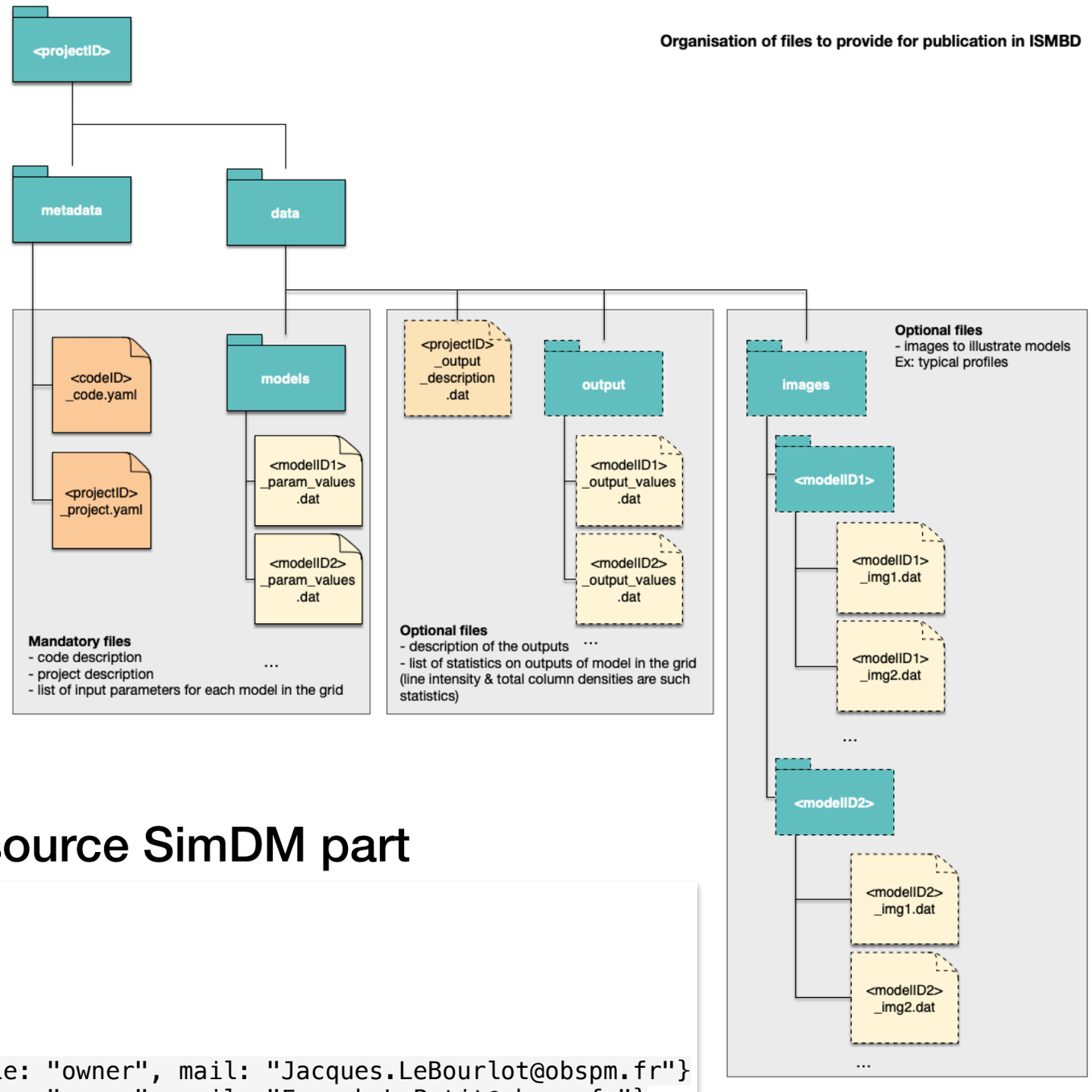


Publication of simulations of other groups in ISMDB

What scientists have to provide :

- a few **yaml files** that contains some SimDM required attributes (plus a few other informations)
- specific organisation in directories
- produce datasets in our HDF5 format

Simple to produce
No need to understand SimDM



Example of yaml file to provide for Resource SimDM part

```
code/ident:         pdr154
code/name:          PDR 1.5.4
code/family:        meudon_pdr154
code/version:       1579
code/created:       20200530
code/contacts:
                    - {name: "Jacques Le Bourlot", role: "owner", mail: "Jacques.LeBourlot@obspm.fr"}
                    - {name: "Franck Le Petit",    role: "owner", mail: "Franck.LePetit@obspm.fr"}
code/url:           https://pdr.obspm.fr
code/doc: >
                    Few lines to describe the code.
                    A full documentation could be on the web site above.
code/publisher:     ism.obspm
```

Publication of simulations of other groups in ISMDB

Simulations published in ISMDB (or coming soon)

- **Paris Observatory** - ISM Group (F. Le Petit, J. Le Bourlot, E. Bron)
  - Meudon PDR code
  - CHIMES (0D time dependent astrochemistry code)

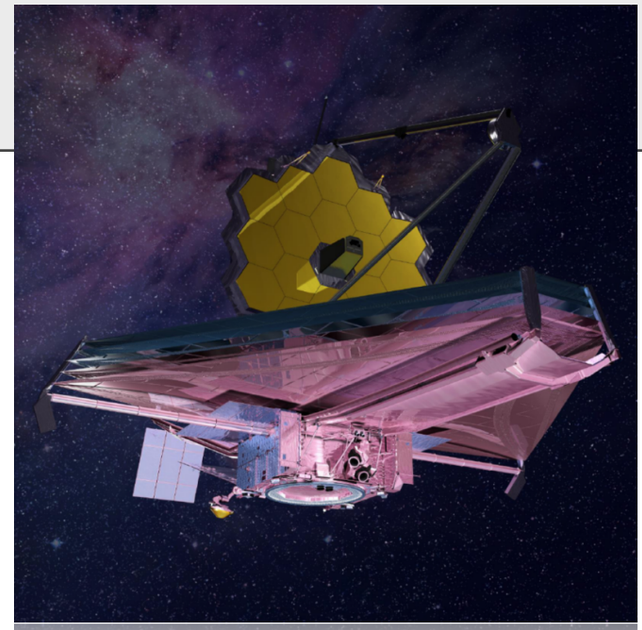
- **Köln University** - Kosma Group (M. Röllig & V. Ossenkopf)
  - Kosma-tau PDR models

- **Ecole Normale Supérieure** - Astrophysics department (A. Gusdorf, B. Goddard)
  - Shock models

- **Univ. of Maryland** - Astronomy department
  - PDR models

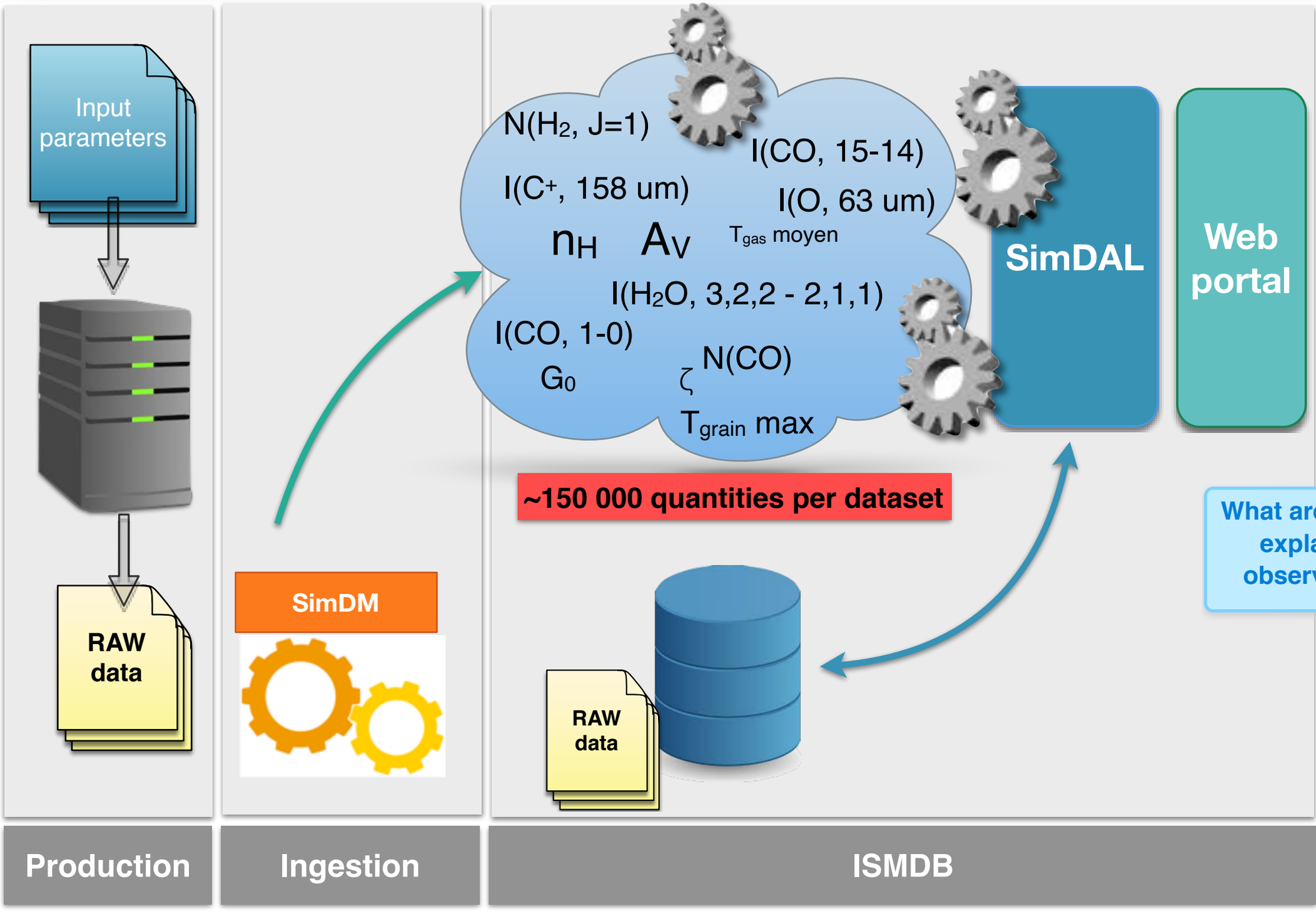
Feedback :
Easy to produce the metadata to publish simulations
Easy to be SimDM compatible without knowing SimDM !

Motivation : JWST
Need models to interpret observations
ERS JWST programs -> make tools publicly available
ISMDB + services are part of these tools



SimDM Implementation : Semantics

- not only a classical database to find pre-computed models
- but **also a tool that can *interpret* observations**



SimDM's semantics : usage

VLA archive Interface

NRAO Science Data Archive : Advanced Search Tool

Historical VLA, Jansky VLA, VLBA and GBT Data Products

Submit Query | Check Query | Clear Form

Output Control Parameters :

Choose Query Return Type :
- (•) Download Archive Data Files
- ( ) VLA Observations Summary
- ( ) List of Observation Scans
- ( ) List of Projects

Output Tbl Format: HTML
Max Output Tbl Rows: NO LIMIT
Sort Order Column 1: Starttime Asc
Sort Order Column 2: Starttime Asc

General Search Parameters :

Telescopes: [x] All [ ] Jansky VLA [ ] Historical VLA [ ] VLBA [ ] GBT

Project Code (GBT: AGBT12A_055, JVLA: 12A-256)
Project Session
Dates From
Archive File ID (partial strings allowed)
To (2010-06-21 14:20:30)
Observer Name

Position Search :

Target Name
Search Type: SIMBAD or NED
Min. Exposure (secs)
RA or Longitude (04h33m11.1s or 68.29d)
DEC or Latitude (05d21'15.5" or 5.352d)
Equinox: J2000
Search Radius: 1.0' (1d00'00" or 0.2d)
- OR - [ ] Check for automatic VLA field-of-view, freq. dependent.??

Observing Configurations Search :

Telescope Config: [x] All [ ] A [ ] AB [ ] BnA [ ] B [ ] BC [ ] CnB / [ ] C [ ] CD [ ] DnC [ ] D [ ] DA
Sub_array: [x] All [ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5
Polarization: ALL
Data Type: ALL
Observing Bands: [x] All [ ] 4 [ ] P [ ] L [ ] S [ ] C [ ] X [ ] U [ ] K [ ] Ka [ ] Q [ ] W
Frequency Range (In MHz : 1665.401 - 1720.500)

Enter Locked Project Access key :

Unique keywords may be used to unlock proprietary data from individual observing projects. Contact the NRAO Data Analysts for project access keys.

Submit Query | Check Query | Clear Form

23 paramètres de recherche
Interface complexe

ISMDB
150 000 parameters !

SimDM's semantics : usage

ISMServices CODES ISMDB

ISM DataBase – Inverse Search service Beta

Grid of isobaric PDR 1.5.2 models

2016.12.03

1 – search among two parameters

x Pgas_0 (cm-3_K) log scale

y G0 observer side (Mathis_unit) log scale

2 – fix all the other parameters

AVmax (mag) 10

3 – observational constraints

Search for available quantities... Ex: N(H) Use

"I(CO v=0,J=1->v=0,J=0 angle 00 deg)" > 1.8E-7
"I(CO v=0,J=1->v=0,J=0 angle 00 deg)" < 2.4E-7
"I(H2 v=0,J=2->v=0,J=0 angle 60 deg)" > 1E-8
"I(H2 v=0,J=2->v=0,J=0 angle 60 deg)" < 5E-7

Search

3 – observational constraints

Search for available quantities... Ex: N(H) Use

Semantics interpreter

SimDM's semantics : usage

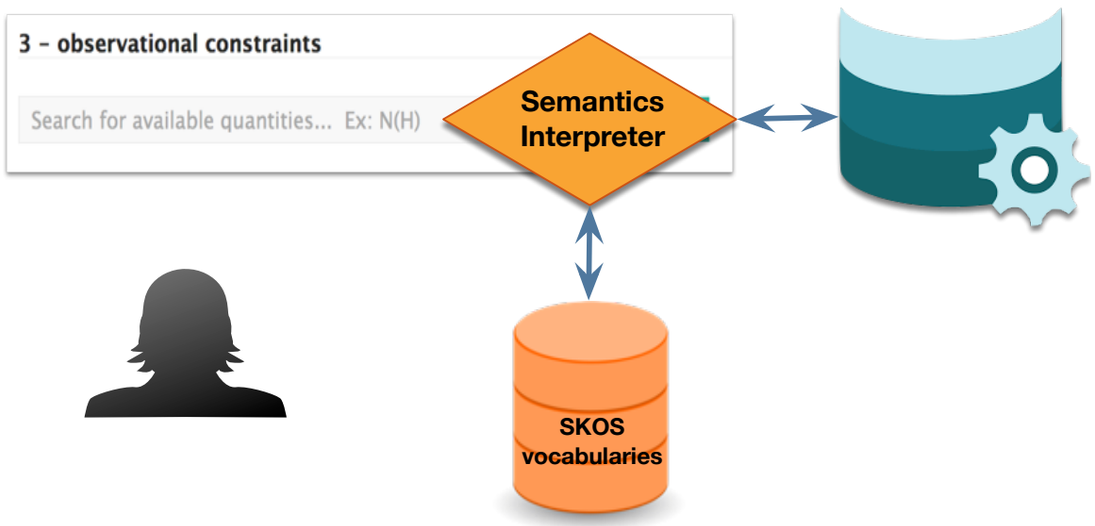
Each metadata is taggued by:

- ID
- name
- unit
- utype
- description
- **label (UCD / SKOS)**
- …

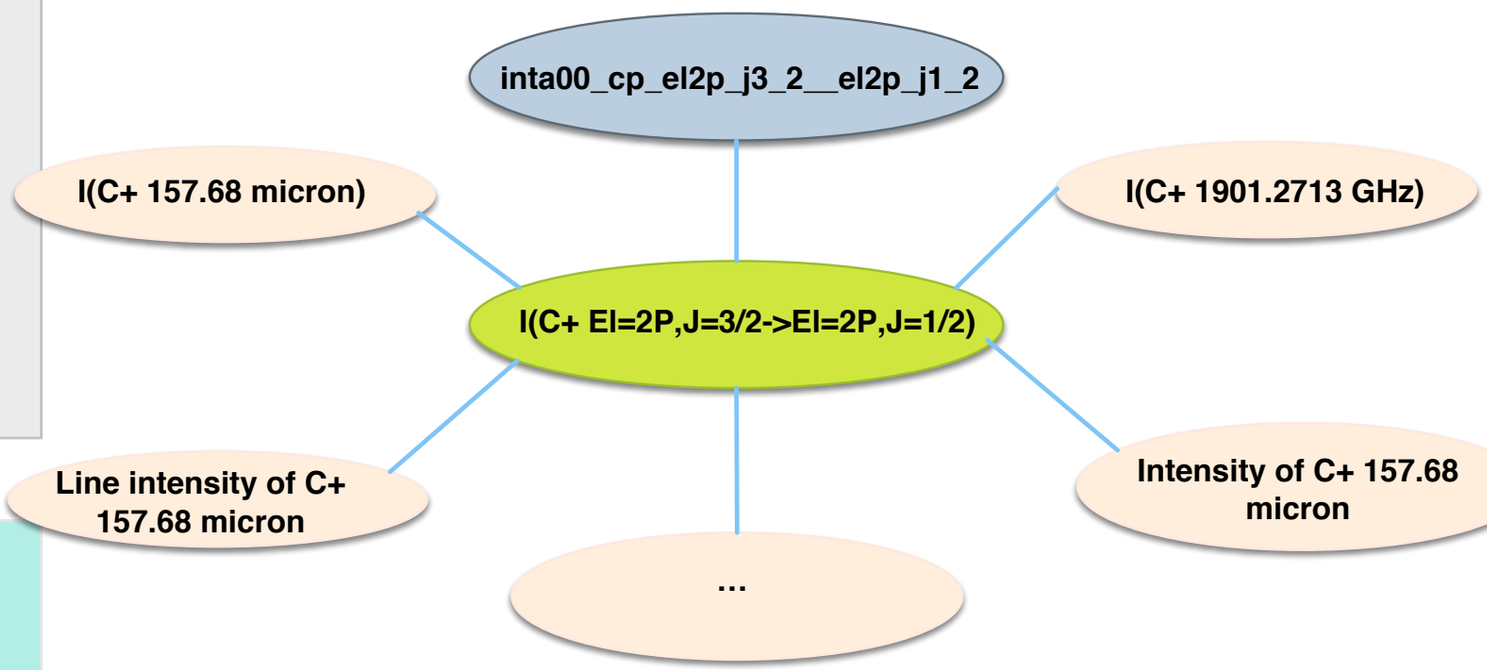
Simulation DataModel

SKOS vocabulary
For each quantity several synonyms (name, units, …)

~ 300 000 terms for the PDR code

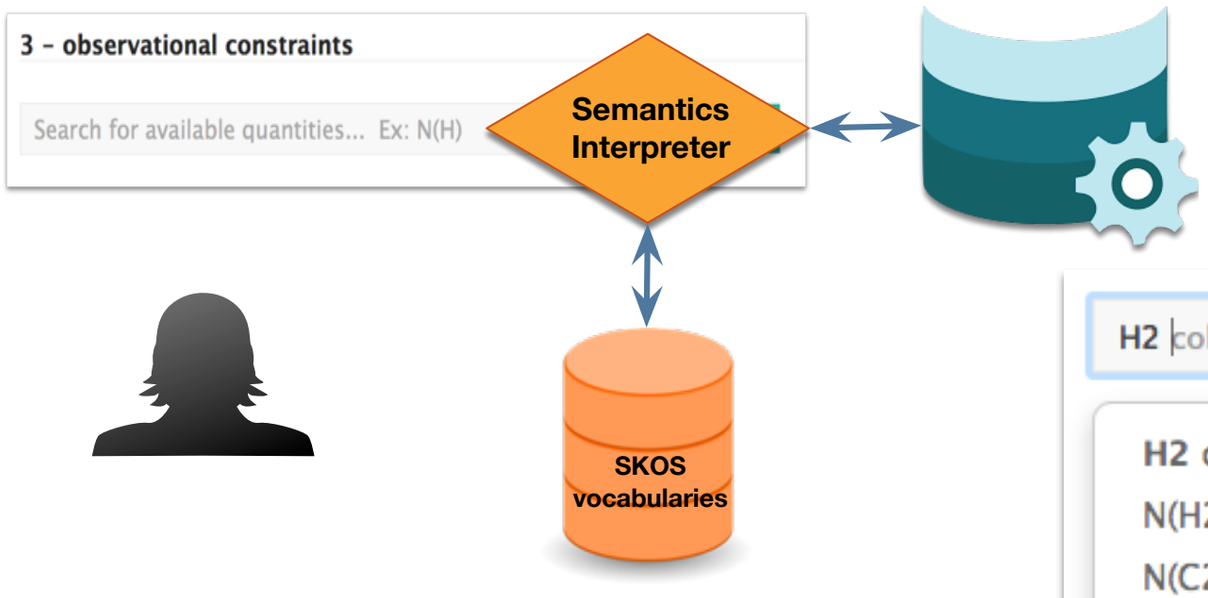


Example of the 157.7 micron C+ line intensity



ID: inta00_cp_el2p_j3_2__el2p_j1_2
PREF: I(C+ El=2P,J=3/2->El=2P,J=1/2)
ALT: I(C+ El=2P,J=3/2->El=2P,J=1/2) face on
ALT: I(C+ 157.68 micron) face on
ALT: Intensity of C+ 157.68 micron face on
ALT: Line intensity of C+ 157.68 micron face on
ALT: I(C+ 1901.2713 GHz) face on
ALT: Intensity of C+ 1901.2713 GHz face on
ALT: Line intensity of C+ 1901.2713 GHz face on
…

SimDM's semantics : usage

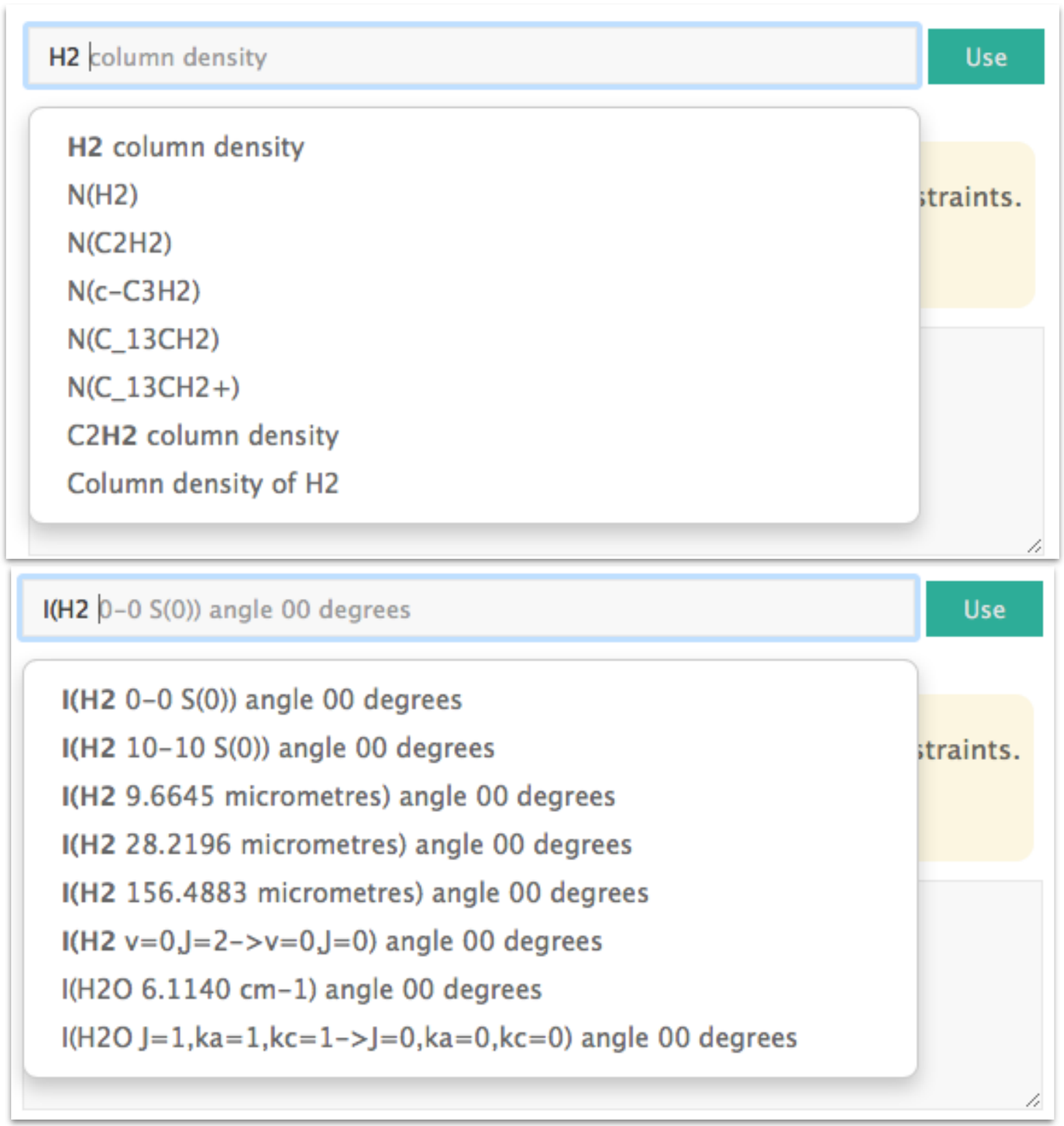


Semantics Interpreter

Semantics
SKOS: PREF + ALT
➔ synonyms

+

Ranking system
(learn from users)



Conclusion

Exemple of one implementation of SimDM

SimDM is a very powerful DM that covers many kind of simulations :
- cosmology
- MHD
- physics & chemistry of the Interstellar medium
- …

In particular, in this latter case allows to do inverse search (solve directly scientific problems) !

SimDM <u>may</u> seems to be a complex DM
BUT : **if instantiation of SimDM are presented to scientists they immediately understand it** !
➔ for scientists : no need to understand SimDM and the VO way to publish data to be VO-compliant !

We need a few ‘databases’ services (per domain) SimDM-compatible to allow scientists to publish simulations

To go further :
Datasets :
- Need to standardise datasets
- In our case we defined an HDF5 format with a specific metadata table

Semantics
- Need a VO-Theory I.G. standard for vocabularies because Vocab 2.0 is too limited, is not enough ambitious and is not suited for simulated data (contrary to Vocab 1)