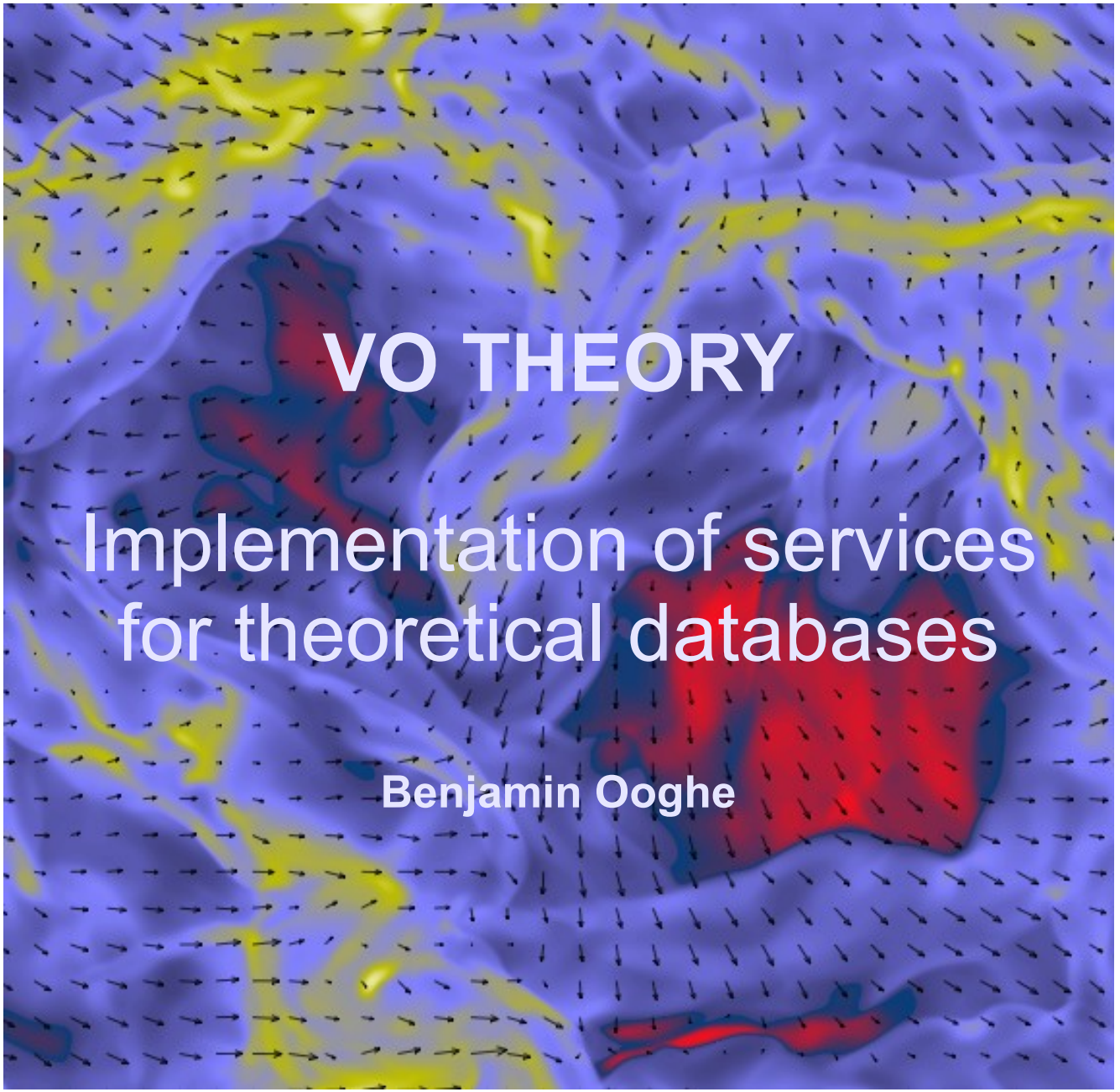


VO THEORY

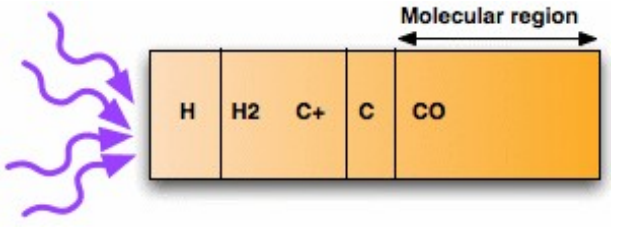
Implementation of services for theoretical databases

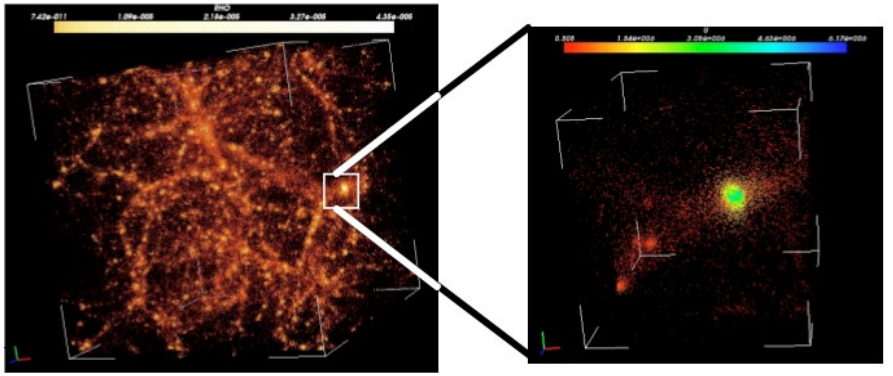
Benjamin Ooghe

l'Observatoire de Paris — LERMA
Laboratoire d'Étude du Rayonnement et de la Matière en Astrophysique
l'Observatoire de Paris — LUTH
Laboratoire Univers et Théories
EURO VO DATA CENTRE ALLIANCE
l'Observatoire de Paris Virtual Observatory Paris Data Centre

Theoretical Data

- All theoretical simulations but very different types of data
  - PDR models
    - Cloud properties, line intensities, molecules abundances, ...
  - StarFormat simulations
    - Molecular cloud boxes
    - Dense cores, clumps
  - DEUVO simulations
    - Cosmology halos



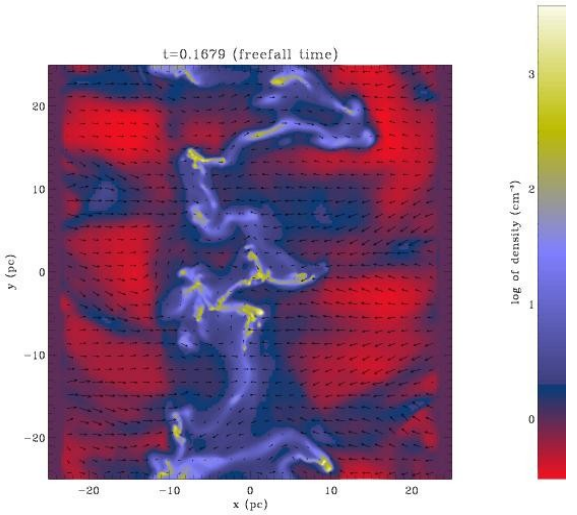


- Each kind requires multiple kinds of codes
  - PDR : geometry, time dependant or not, different chemistries, ...
  - StarFormat and DEUVO : Ramses 2 & 3, Flash, Gadget

=> How to store all of these different data consistently into the VO

VO Theory – Simulations Data Model

- Common need to describe any simulation and its results
  - codes and parameters
  - objects simulated
  - characterisation and properties
    at different time intervals → snapshots
  - postprocessings on the results



- Common need to publish and display the results
  - Retrieve data through VO protocols
  - Browse results by hierarchy and find by metadata query

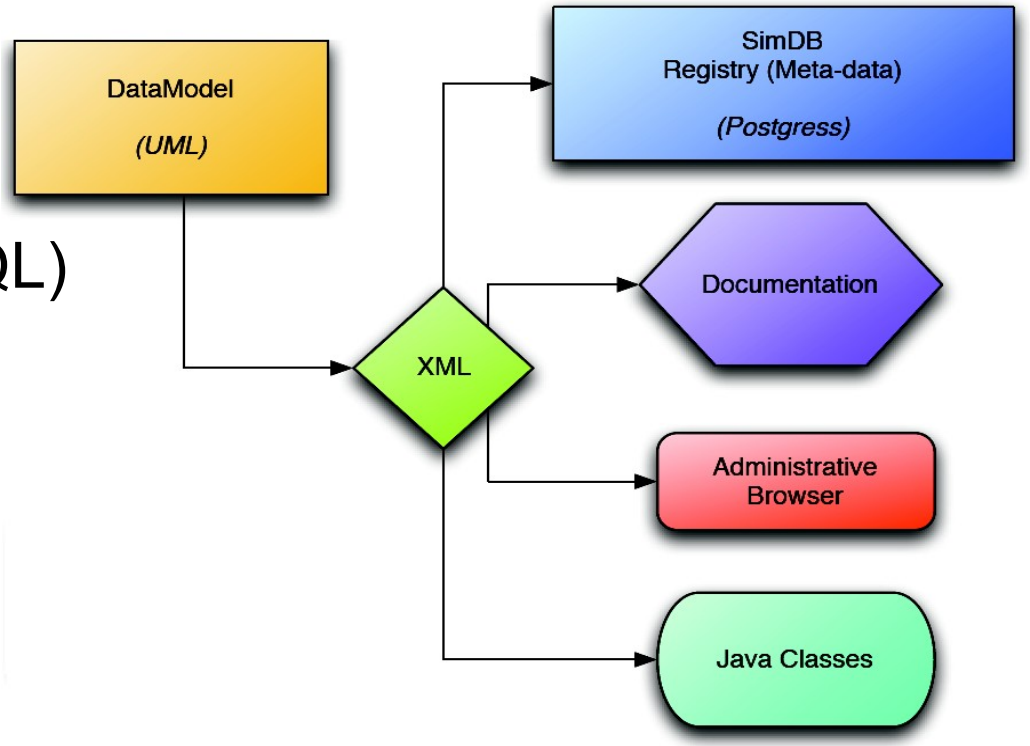
    → limits to the general simulations model to display the results

    → need of specific web interfaces for different science purposes

VO Theory – Simulations Data Model

- SNAP (Simple Numerical Access Protocol)
  - DataModel to store : SimDM
  - Data Access Protocol to retrieve : SimDAP (Web, VOQL, ...)

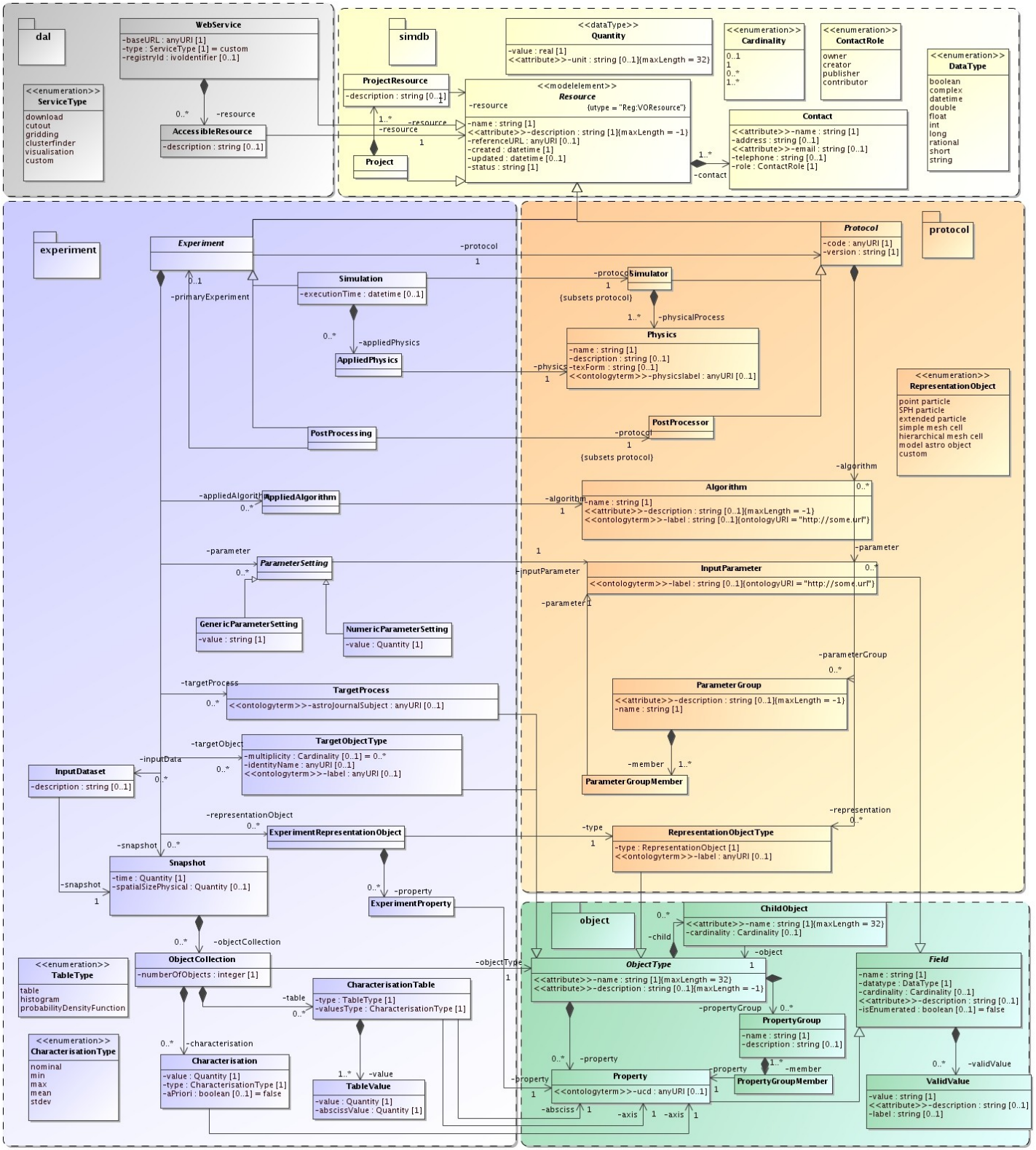
- SimDB / VO-URP : G. Lemson & L. Bourges (Euro-VO)
  - DataModel UML
  - DMTransformer
  - DB managing (PostGresSQL)
  - Java interface
  - WebServices (SimDAP)



→ http://code.google.com/p/vo-urp/

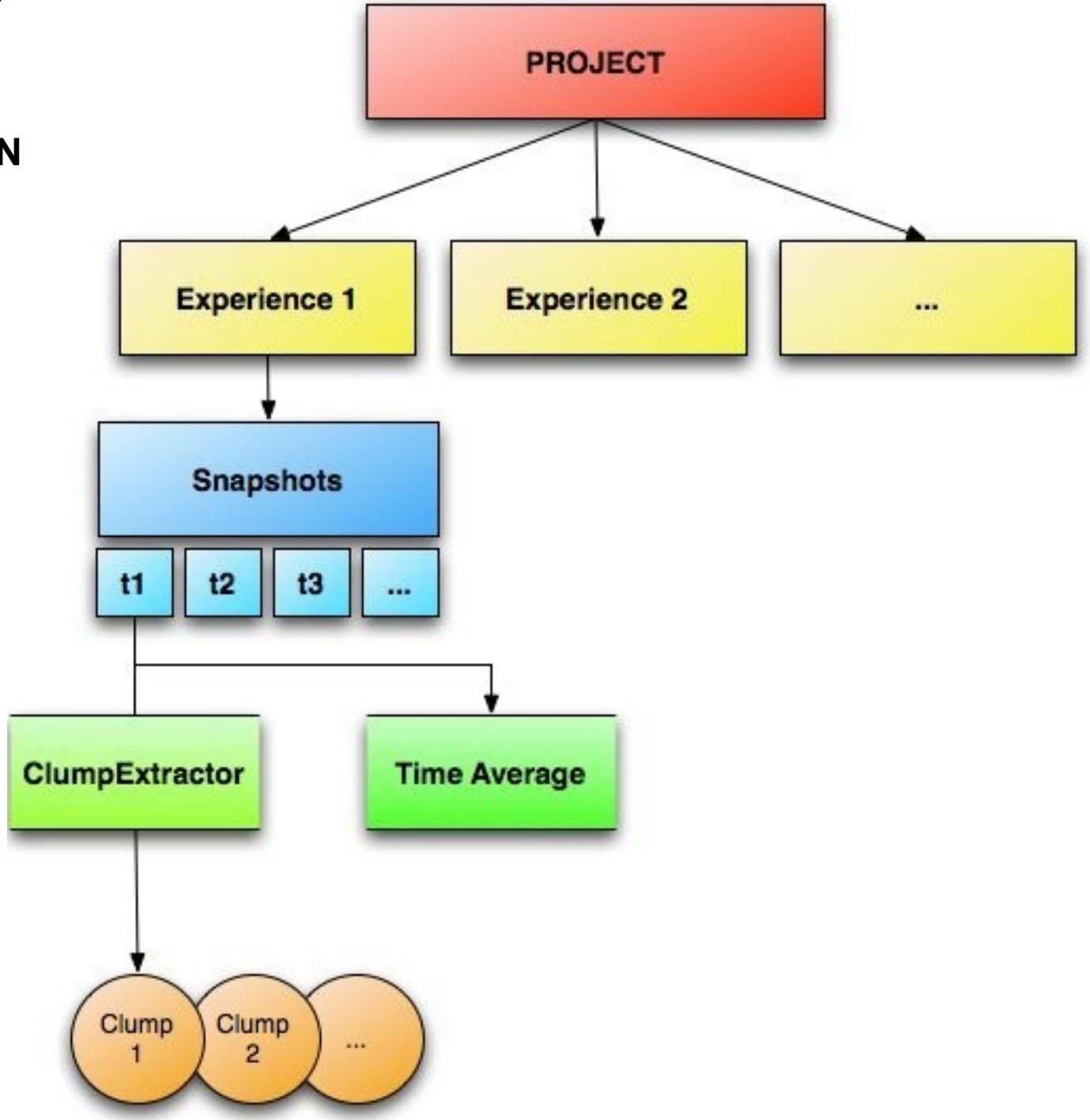
IVOA THEORY

SIMULATION DATA MODEL IMPROVED



IVOA THEORY

SIMPLIFIED MODEL ORGANIZATION

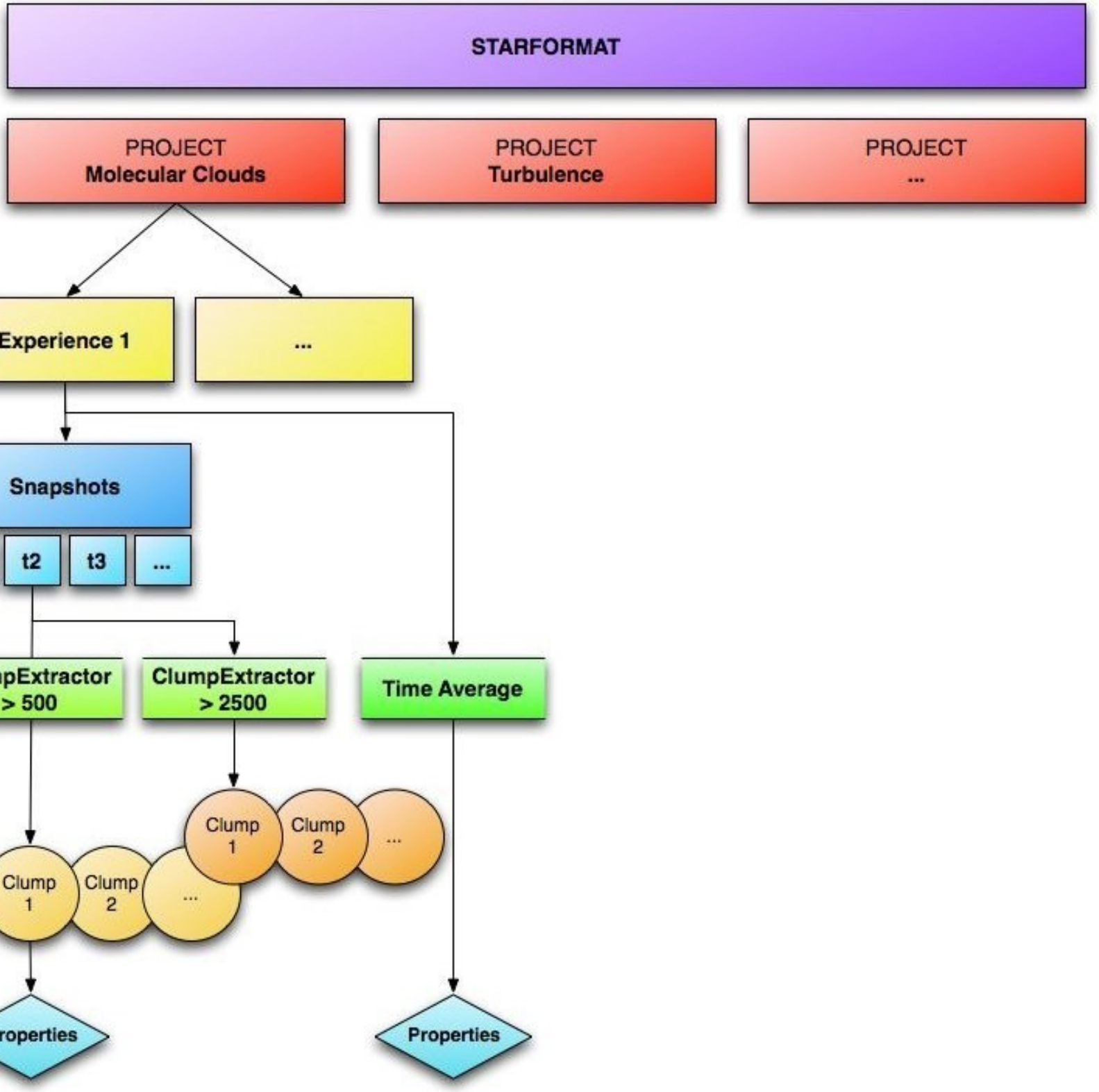


Data description for querying

- For a proper data browsing, documentation is needed :
  (for both simulations and postprocessings)
  - Simulation's description (code used, simulated objects, contacts, ...)
  - Physical processes involved (MHD, heating, gravitation, turbulence forcing, …)
  - Parameters (boundaries and initial conditions, grid definition, …)
- Description of results for each time step (snapshots) :
  - Results or statistics on the results to help observers identify data
  - Descriptive files like images, probability density functions, ...
  - Eventually raw or postprocessed data
- To provide further VO interoperability with worldwide apps
  - Descriptions must match VO semantics standards (UCDs / UTypes)

IVOA THEORY

SIMPLIFIED MODEL ORGANIZATION FOR STARFORMAT



Benjamin Ooghe

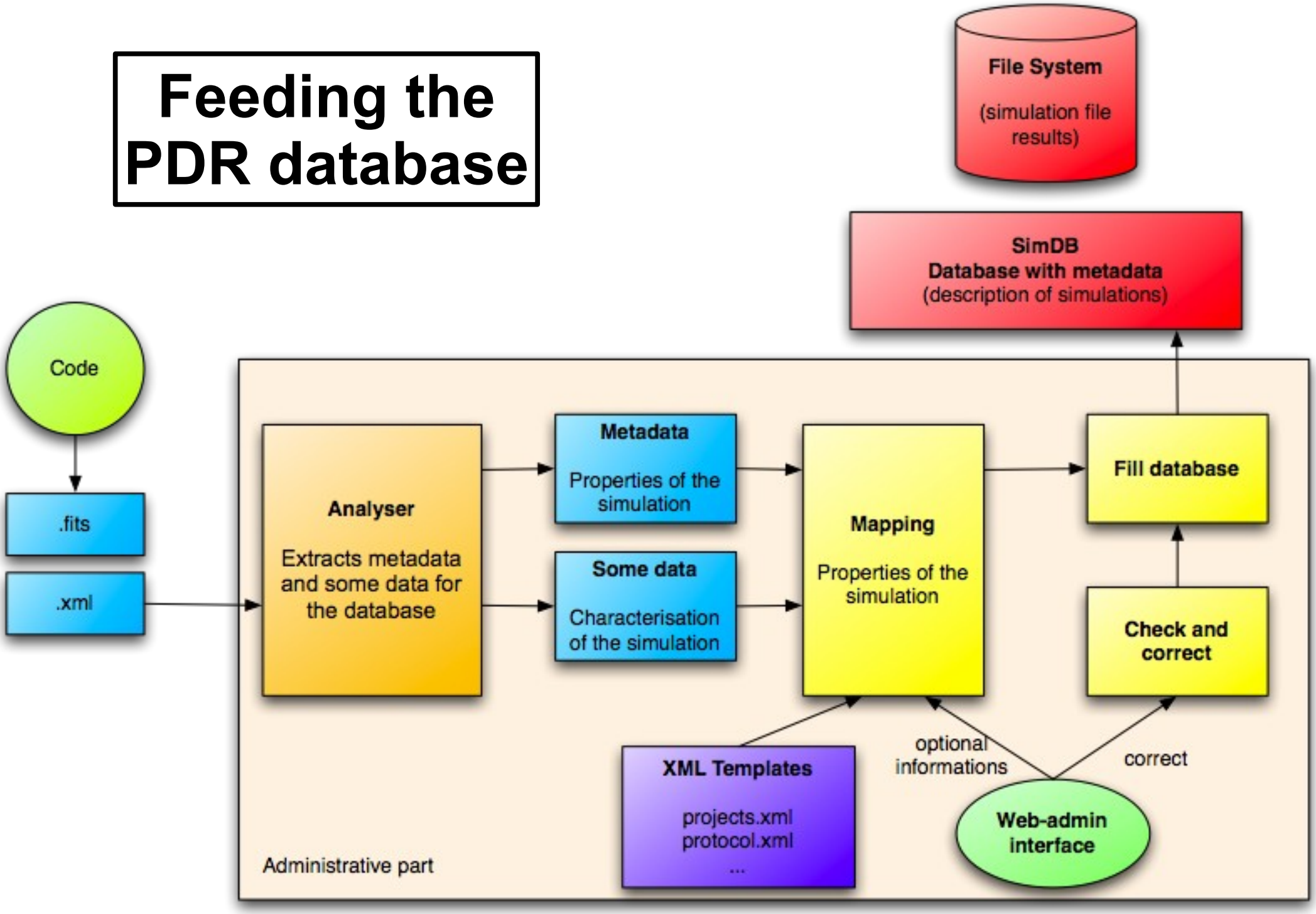
Feeding the databases

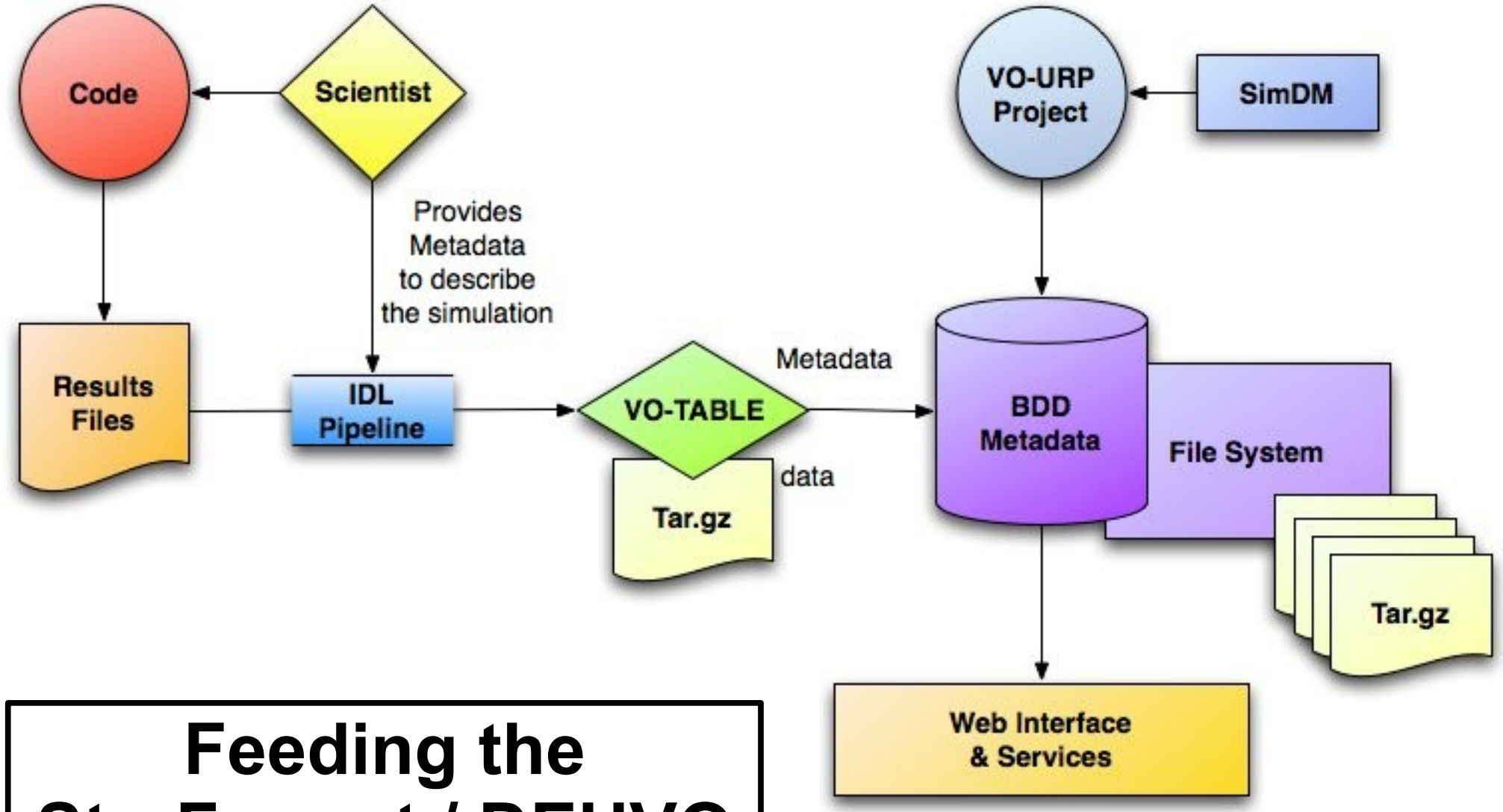
- The scientist :
  - Runs simulations, identifies useful data and collects results
  - Eventually computes useful extra results (statistics, plots, ...)
  - Provides all results as a double-components set:

    → MetaDatas as a generic XML or VOTable file

    → Raw data files like images, .fits, .tgz archives, ...
- The software :
  - Collects and reads filesets using SAVOT + JPA in VO-URP
  - Identifies and associates new, previous and modified codes with new descriptions, parameters, properties, possible targets, ...
  - Stores each snapshot as a database element with all of its descriptive elements and values
  - Associates the related files into the corresponding filesystem

Feeding the PDR database





Feeding the StarFormat / DEUVO databases

Making the ingestion XML files

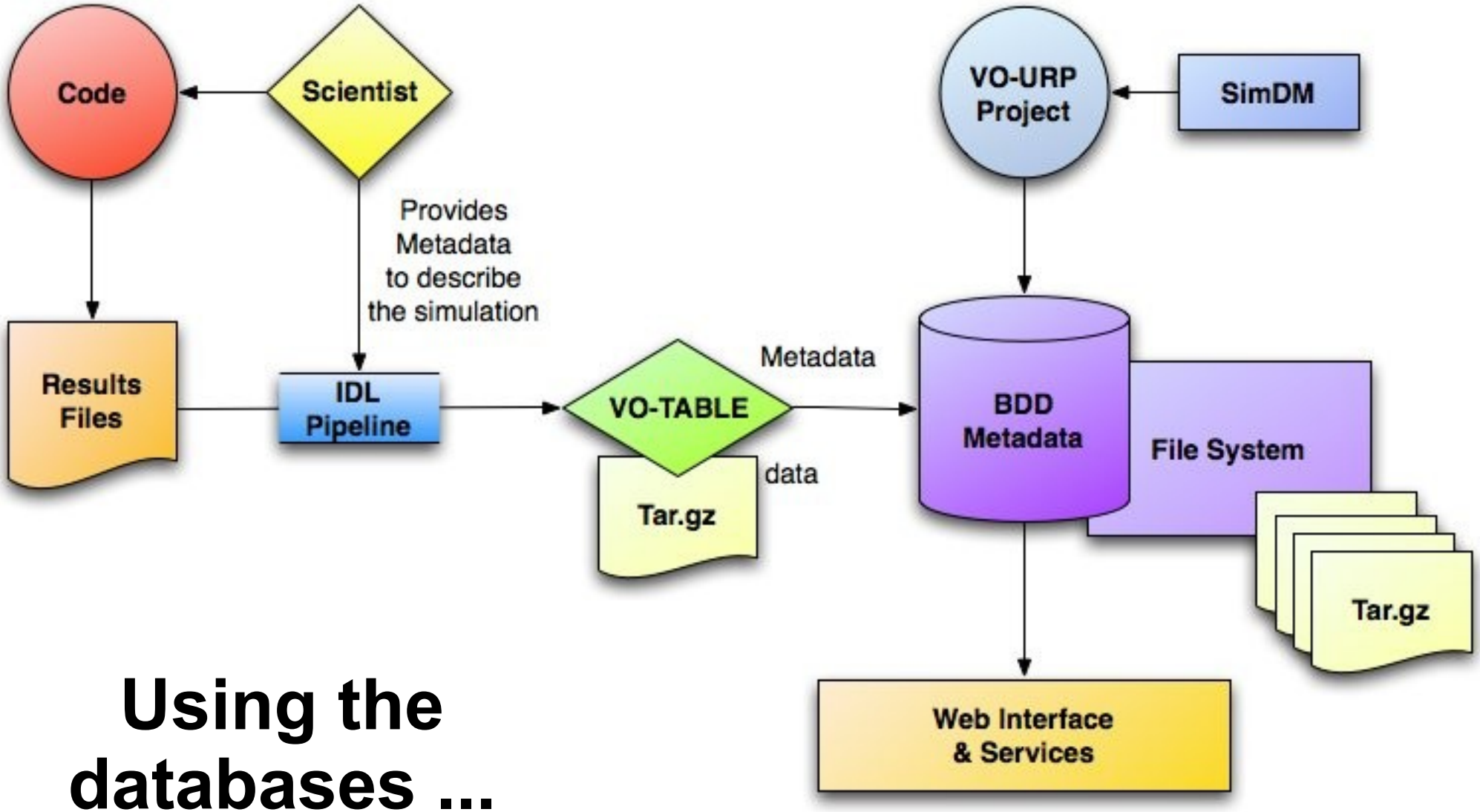
- For any scientific application need for a collaborative work between the Scientist and the IT engineer for specific cases

  → Generate organized metadatas and results to fill the database model

```
    </DATA>
  </TABLE>
</RESOURCE>
<RESOURCE name="experiment">
  <TABLE name="experiment" nrows="1">
    <FIELD name="publisherDID" datatype="char"/>
    <FIELD name="name" datatype="char"/>
    <FIELD name="description" datatype="char"/>
    <FIELD name="referenceURL" datatype="char"/>
    <FIELD name="class" datatype="char"/>
    <FIELD name="primaryExperiment" datatype="char"/>
    <DATA>
      <TABLEDATA>
        <TR>
          <TD>00080-_-clumps_10000</TD>
          <TD>Snapshot 80, Clump extraction with min density 10000</TD>
          <TD><![CDATA[]]></TD>
          <TD>http://roxxor.obspm.fr/StarFormat</TD>
          <TD>postprocessing</TD>
          <TD>FORM_MC-_-THY3D_iso_wt_wf</TD>
        </TR>
      </TABLEDATA>
    </DATA>
  </TABLE>
  <TABLE name="parameters" nrows="1">
    <FIELD ID="dens_threshold" name="Density threshold" ucd="phys.density" datatype="flo
      <DESCRIPTION>The density threshold above which the clumps from the snapshot are ar
    </FIELD>
    <DATA>
```

StarFormat example : IDL pipeline

- P. Hennebelle's RAMSES simulations :
  - IDL treatments to compute statistics and generate images
  - Extension of the IDL routines to automate feeding :
    - read results and store on model's schema in a XML
    - collect images and various files in archive
  - Need for consequent human descriptive work
  - Need to push data to the database
- Need to adapt the pipeline for different scientific apps
  - DEUVO tryouts successful, even with 500 000 halos



Using the databases ...

→ http://pdr.obspm.fr

→ **http://starformat.obspm.fr**

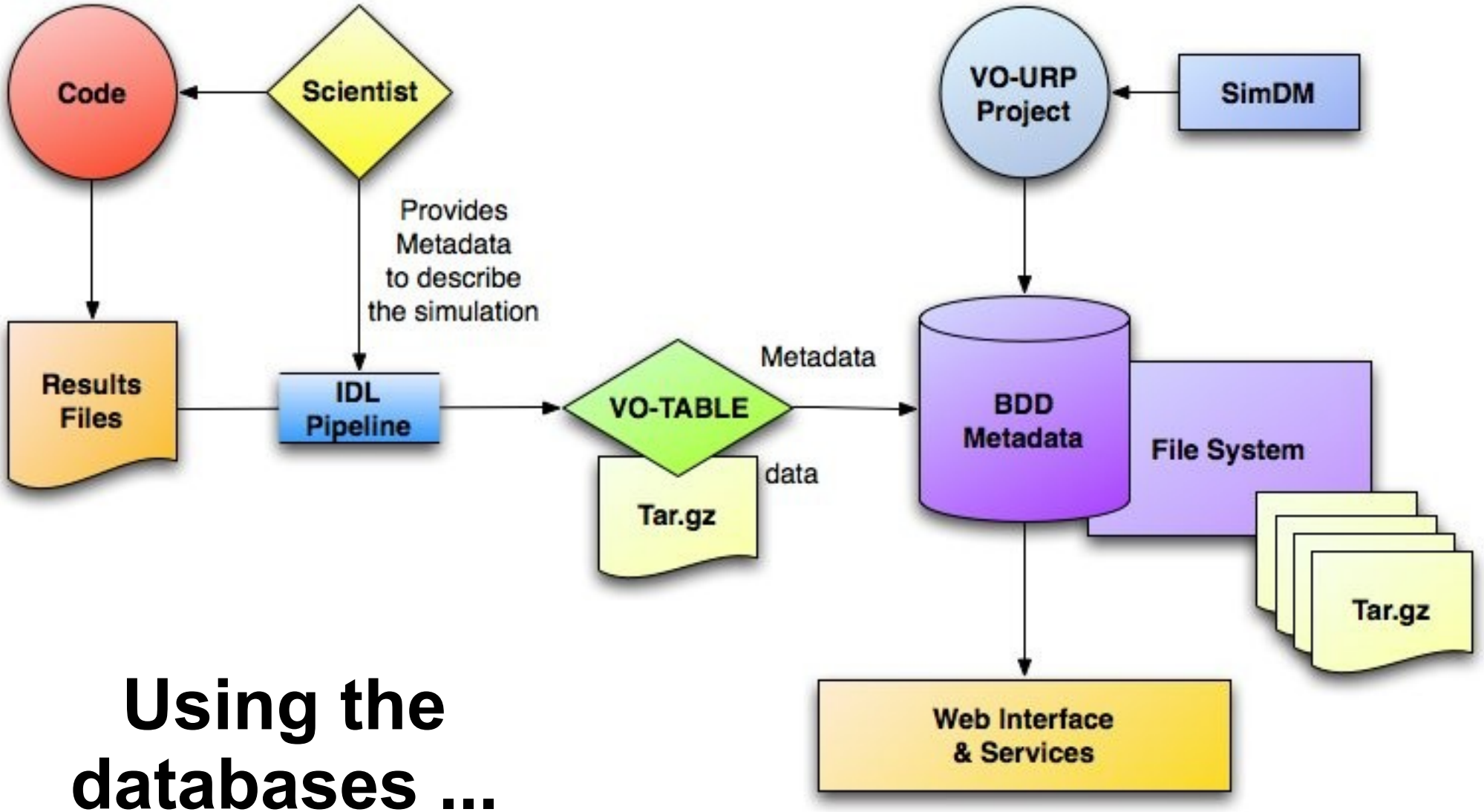
PDR web interface

- L. Bourges' work at Euro-VO + N. Moreau
- PDR code online : download, run on grid, analyzer webapp
- Database queries on all models or just on specific sets :
  - queries on input parameters
  - reverse queries on statistical results → observer's need
- Models display and webservices :
  - Parameters, Structure, Column Densities, Line intensities, ...
  - Plots on demand
  - Download data or export in VO-Table
    → interoperability with TopCat, Vo-Spectre
  - More to come...

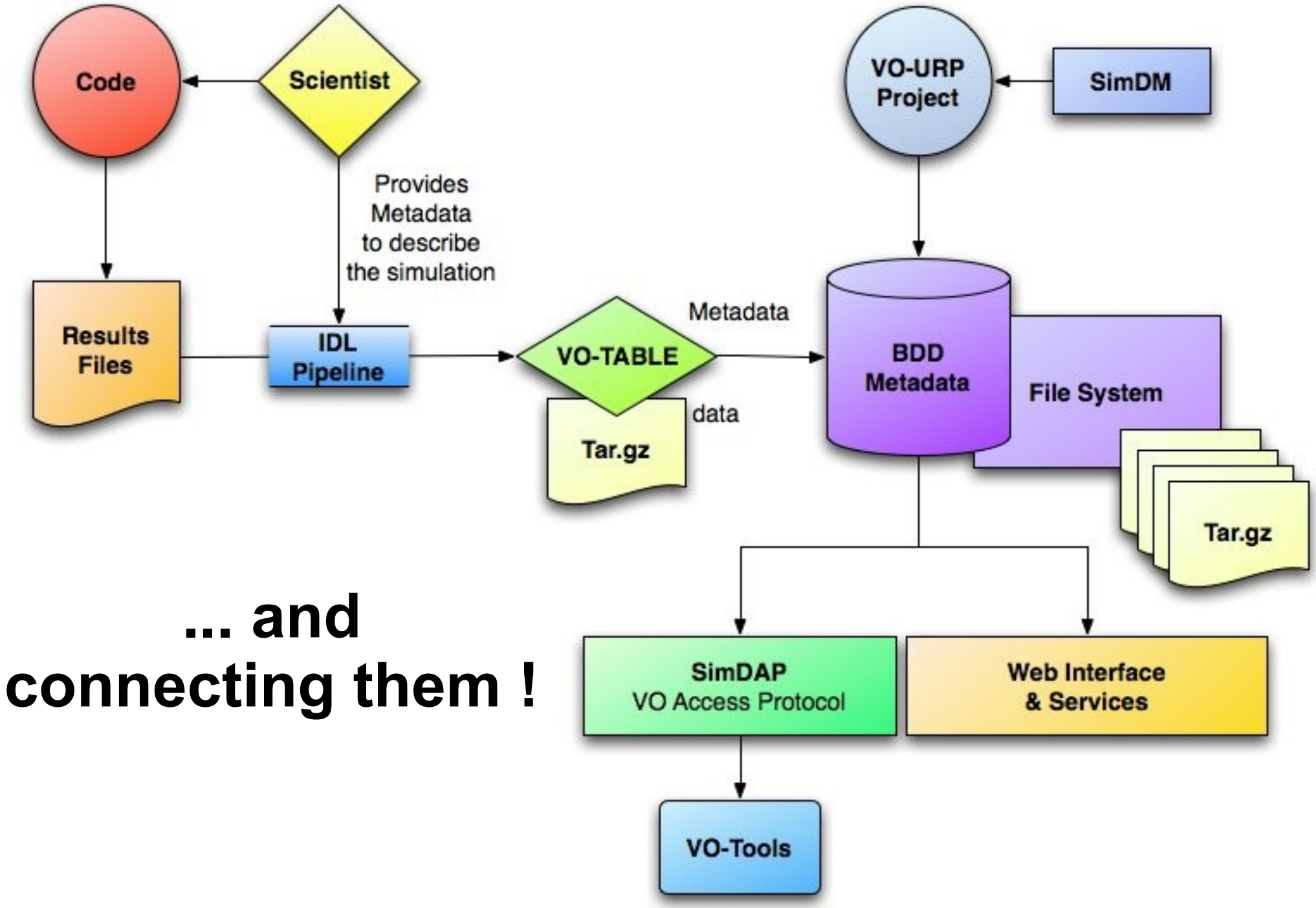
http://starformat.obspm.fr

StarFormat web interface

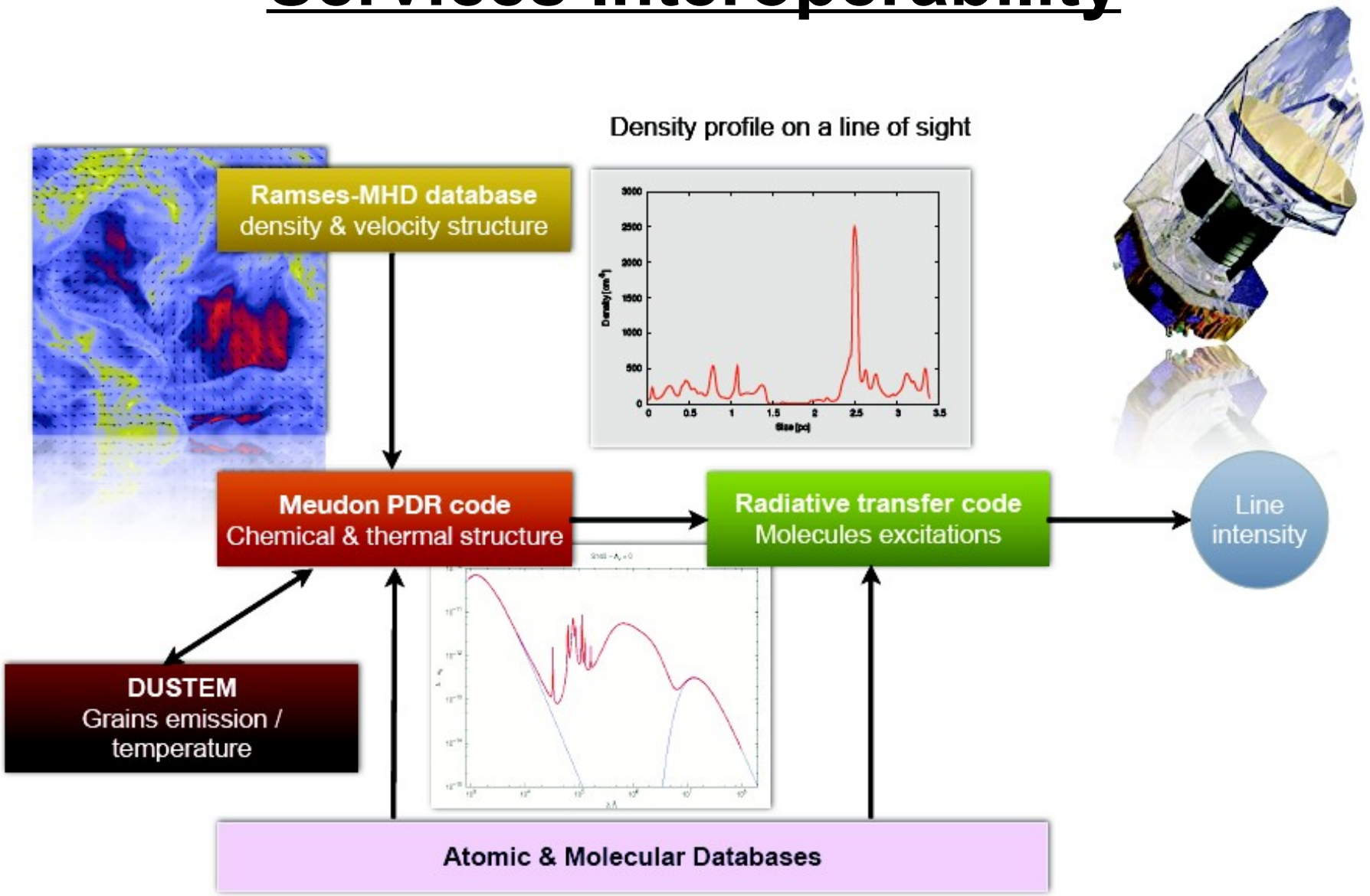
- Future Interstellar Medium services platform :
  → connect PDR and StarFormat at some point
  → need for visual and technical coherence between websites
- But different technical applications :
  - Double step queries :
    1) query simulation's parameters to identify sets of simus
    2) query postprocessing's result properties to find clumps
  - Display snippets, download images, PDF, plot clumps distribution on demand, ...
  - Cut data from raw data on demand?
- Next : Extend to other simulations (LERMA + ZAH, Germany)
  - turbulent boxes, chemistry postprocessings, ...
    → new needs to come for queries and presentation



Using the databases ...


Code
Scientist
VO-URP Project
SimDM
Provides Metadata to describe the simulation
Results Files
IDL Pipeline
VO-TABLE
Metadata
BDD Metadata
File System
Tar.gz
data
Tar.gz
... and connecting them !
SimDAP
VO Access Protocol
Web Interface & Services
VO-Tools

Services Interoperability



VO Theory usage report

- Abstraction of the model allows good genericity !
  - Very different datasets can fit in the same registry
  - Easy multiple and parallel maintenance
- Depth of the model allows good adaptability !
  - Theory projects evolve constantly and can be updated easily
  - Very different kind of queries can be used for each science app
- But still need of individual human work per project :
  - Adapt ingestion process by mapping data to the model
  - Adapt web interface to display results in accordance with specificities of each scientific project