# THEORY-1

Simulation Data Model and Semantics

# Goals of the session

- clarify/specify the recommendation process (adopted by IVOA at Baltimore INTEROP)

- discussion on comments/questions received on theory@ and [dm@ivoa.net](mailto:dm@ivoa.net) lists

- report on Vocabularies progresses and demos.

# Recommendation process (1/2)

- after DM WG agreement, it is suggested that
  - the full Note is transformed into a WD
  - part of Appendix (general DM interest) could be turn into a separated Note
  - the WD will mention all related documents (XML, XSD, PNG, ...)
- the abstract should also link to all related documents

# Recommendation process (2/2)

- continue WD discussions on theory@ and dm@ list
- DM WG chair starts RFC in January 2011

# Comments received so far : Format of this talk

- Comment: comment or question received (thanks to Franck and Miguel)
- Answer: received (thanks to Gerard) or my own view ☺ ! should be changed to TIG one <u>after</u> discussion
- Solution or reason: in any. Received or my own view ☺ (again…)
- Action: have to agreed on during this session

- *Therefore, since this represents mostly my view in order start the discussion, the definitive agreed answers can only emerge after this discussion*
- *They will be send on the mailing lists after the meeting*

# Simulation DM or Simulation Database DM?

- *Comment: "the description of the document quoted by Mireille "a Note about the data model defined for simulation data" and the description of the posted document looks to be different and now it is not clear to me what is the final goal of the document."*

- Answer:
  - per se, SimDM is a VO model describing Simulation metadata. Its intent is to allow discovery of simulations according to
  - It has been extended for inclusion of all known kind of simulations. Unknown ones (at the date of today) can be integrated by updating SimDM if necessary.
  - SimDM will be used not only in SimDB but also (in part and w/ other representation) in SimDAL services (see. Sect 3.1)

- Solution: fixe the name « Simulation Data Model » (i.e. SimDM) instead of SimDB/DM or Simulation Data Base model or whatever…

- Actions:
  - update the text and rename the document (HW)
  - don't use anymore the name 'SimDB/DM'

# VOTables vs XML files

- *Comment: "If the DataModel is intended to provide a description of theoretical data, i.e. a datamodel for theory and not just a datamodel for a database of theoretical results (…) some examples are needed in the document, in particular VOTables of final theoretical products ."*

- Answer: "*From this model we can (and do) derive different representations that can be useful in particular circumstances. For this proposal the XML schema and UTYPE representations are required. So we have a schema that defines the format of XML documents containing such descriptions. We have examples of such documents, and these should go into the document prepared by Franck. They are not VOTables. The metadata is too hierarchical to easily fit in the flat table structure. We do have a relational database mapping, useful for example in a TAP context.*"

- Solution: N/A

- Actions:
  - none

# Spatial vs non-spatial simulations

- Comment: too few examples of simulation not-in-space wrt to simulation in 3D+1 space
- Answer: Right. A comment in the Note points out this in Sect. 3.3 (Domain analysis) but also the case at various places. This is due to an historical bias (starts with SNAP…) and previous Note versions,
- Solution: rephrase some part of the text.
- Action:
  - Gerard and myself: editing
  - all: send use cases, examples (named one, not "I know someone who…") etc.
  - received from Miguel (for Sect. 3.3):

  - How the results are parametrized?
  - Can I access grids of models? can I access individual results?
  - Which are the inputs ingredients (usually, which data collections are used?)
  - How I can run a simulation? Can I do it on-the-fly?
  - Can include my simulations in the VO in a easy way?, What I should do?
  - Can i compare different simulations? Can I compare the simulation with my data?
  - Which simulations provide diagnostic tools? (i.e. distance/extinction/quasi-scale free quantities)
  - Can I combine the results of different simulations in a single file adapted for my needs (e.j. own code)?

# Simulator / postprocessor

- Comment: do we need to make the difference between *simulator* and *postprocessor* ?
- Answer:
  - *postprocessor* (in the context of 'theory') works on *simulated data* without adding physics but makes use of *algorithms*
    - A step-by-step problem-solving procedure. In the context of mathematics, computer science and related field, it is an established, recursive computational procedure for solving a problem in a finite number of steps.
  - *simulator* makes use of *algorithms* to solve *physical equations* and create *simulated data*
- Reason:
  - initially, the difference has been made because some published products are not *raw* simulated data (such as particle pos/vel) but only *results* from *transformation* or *recombination* of original simulated data (such as halo pos/vel in dark matter simulations).
- Action:
  - need more use cases and examples
  - can be discussed during the RFC period

# Histogram or not histogram?

- Comment: histogram or not histogram
- Answer:
  - as the SimDM main use if for discovery, histogram has been leaved out the specification.
- Reason:
  - There was no use case for this feature.
  - from a formal pov, not part of a statistical summary.
- Action:
  - could be included in a future upgrade of the model
  - but need additional use cases to assess how much important it is

# SimDM and Access to the data

- *Comment: "the model includes some access data fields (Sect 4.7 and 4.8 in particular), Following the SSAP it is the access protocol which includes its own data model for access spectra."*

- Answer: Right. But SimDM will be used by SimDAL. See Gerard's talk on Friday session (SimTAP)

- Reasons:
  - there is no agreement in the 'theory' community (as a whole or even astronomical part of it) for a common data format (no real equivalent to FITS adopted by IAU)
  - need to specify what the user will find when he/she will access the data through a SimDAL service

- Actions:
  - none

# Editing issues (1/2)

- Use of word « *simulation* » for various concepts:
  - simulation code, simulation result(s), simulation class (of the model) etc.
  - suggested solution: "be more specific in each case specially for simulation code and simulation result (and, for instance, insist that Simulation class is always in bold and Uppercase)"
- Answer :
  - End-of-page note # 1 : "We will use the term *simulations* for the running of a simulation code as well as for their results. And we will often include post-processing codes and their results as well."
  - check all the document to make clearer the distinction between "result", "code", "action of running the code" where it is necessary to clarify
  - **<u>Simulation</u>** class is already in blue and bolded and underlined (cannot be uppercased to be consistent w/ the rest of the document)

# Editing issues (2/2)

- Use of word « *protocol* » for various concepts:
  - IVOA *protocol* vs experimental *protocol*
  - suggested solution: "replace protocol by e.g. procedure"
- Answer:
  - both uses are (unfortunately) correct
- Solution:
  - replace *protocol* by *procedure*?
  - add *experimental* to *protocol* where relevant (in the text, not the Class)?
    - According to Oregon State University, an *experimental protocol* is a detailed plan of a scientific experiment that specifies experimental methods, data collection and sampling schedules.
  - ask to semantics WG?
- Action:
  - in the text, change *protocol* into *experimental protocol*
  - **Protocol** class doesn't change.

# Border line cases?　　　　　　　　　　(1/2)

- Comment: can a collection of empirical/observed/theoretical tracks/spectra be described by SimDM?

- Answer: depends if the intent is to publish a 'theoretical' library or uses it as an input parameter for further modelling.
  Example: http://www.iac.es/proyecto/miles/

  - Miles and CaT libraries are made of observed spectra

  - SSP models (Vazdekis et al. 2009) based on theses libraries are covered by SimDM

  - inputparameter (in protocol class) and/or inputdataset (in experiment class):

    - IMF : Unimodal, bimodal, Kroupa universal, Kroupa revisited (up to 7 various slopes…)

    - theoretical stellar isochrones (calibrated with stellar photometric libraries for magn. and colors)

    - [M/H] and/or Z : 7 values

    - Stellar libraries to compute SED and Lick indices (or so).