

# SimDAP et S3

## Can they merge?

Rick Wagner

Laboratory for Computational Astrophysics,  
San Diego Supercomputer Center

May 25, 2009

# SimDAP Summary

- Goals

- Provide a service standard for retrieving simulation data
- Mimic other second generation DAL protocols (e.g., TAP, SSAP, etc.)
- Low technical overhead for implementation

- Important Data Model Elements

**Protocol** A piece of software (Enzo, Gadge, halo finders)

**Experiment** (i.e., Simulation) the process of running a piece of software with a set of input parameters

**Snapshot** Results of the simulation at some point in time. Could be in files or a database

# Current SimDAP Operations

Required?	Operation	Input Parameters
Yes	GetAvailability	
Yes	GetCapabilities	
Yes	ListExperiments	INPUTDATASET (O)
Yes	ListSnapshots	EXPERIMENT (R)
Yes	QueryData	EXPERIMENT (R), SNAPSHOT (O), PROPERTIES (O)
No	Cutout	EXPERIMENT (R), SNAPSHOT (O), PROPERTIES (O), VOLUME (O)
No	Preview	EXPERIMENT (R), SNAPSHOT (O), PROPERTIES (O)
No	Custom	EXPERIMENT (R), SNAPSHOT (O), PROPERTIES (O)

# VOTable Response

- My current suggestion is for nested RESOURCE elements
  - RESOURCE utype="simdb:Simulation"
    - RESOURCE utype="simdb:Snapshot"
    - RESOURCE utype="simdb:Snapshot"
- However, this is very different from other DAL services,
- and is harder for client tools to interpret.
- But I think this is the correct way to serialize the data model to XML, and for Claudio's TDFF

## Comments on S3

- Focuses user interaction, but
  - each operation should be independent, i.e., recursion is optional and automatable.
- Very similar to SIAP, SSAP with custom parameters
- Likewise, VODataService supports a complete description of tables and parameters.

# Proposal

Priority: *One* service for simulation (theory) data access (I don't care what it's called).

## Two Purposes

- A: Data access to existing data, including pre-defined operations (e.g., preview, cutout)
- B: Self-describing custom operations
  - special operations on existing data (halo finding, ionization levels)
  - parameterized queries for data ( $h_0 < 0.7$ )
  - dynamically generating data (spectra)

# Key Points/Ideas

- Service response is the same for all data queries both SimDAP-like and S3-like.
- Push descriptions of the data to the other side of the access URL (cómo S3).
- Access URL can lead to anything. . .
  - . . . image
  - . . . VOTable of spectra, particles, etc.
  - . . . file describing binary data, VOTable con un TDFF description, VisIVO binary table and data.
- Describing custom operations (S3-like) is done using GetCapabilites.

# Importance of the Data Model

- Key Point: All theory tools should talk the same language; the woorden and concepten come from the data model.
- Queries for existing data are really queries against the data model.
- Service operations which modify or create data are protocols.
- Calling these operations creates a new experiment.



# Relax

- Using the data model doesn't require managing a relational database.
- Using the data model doesn't mean a different XML format.
- But, using the concepts from the data model means that services can be related.
- E.g., standard SimDAP operation can use data from a SimDB service.

# One Data Access Service

- My suggestion: Merge S3 and SimDAP.
- Can be view as S3 plus some standard operation,
- or, as SimDAP, plus service-defined custom operations.
- Either way, service return access to theory data.
- Unified interface and response means clients can access more services.

# Questions & Decisions

- 1 Is this reasonable, does it meet our needs?
- 2 Can we use the simulation data model?
- 3 Pick a name (again, I don't care what it is).
- 4 Agree on required and optional service operations.
- 5 What are the required and suggested columns of the VOTable response?
- 6 Implementations, VODataService extension schema, lots of talking. . .

# Discussion Time

La Vache Qui Rit image removed to avoid copyright infringement.