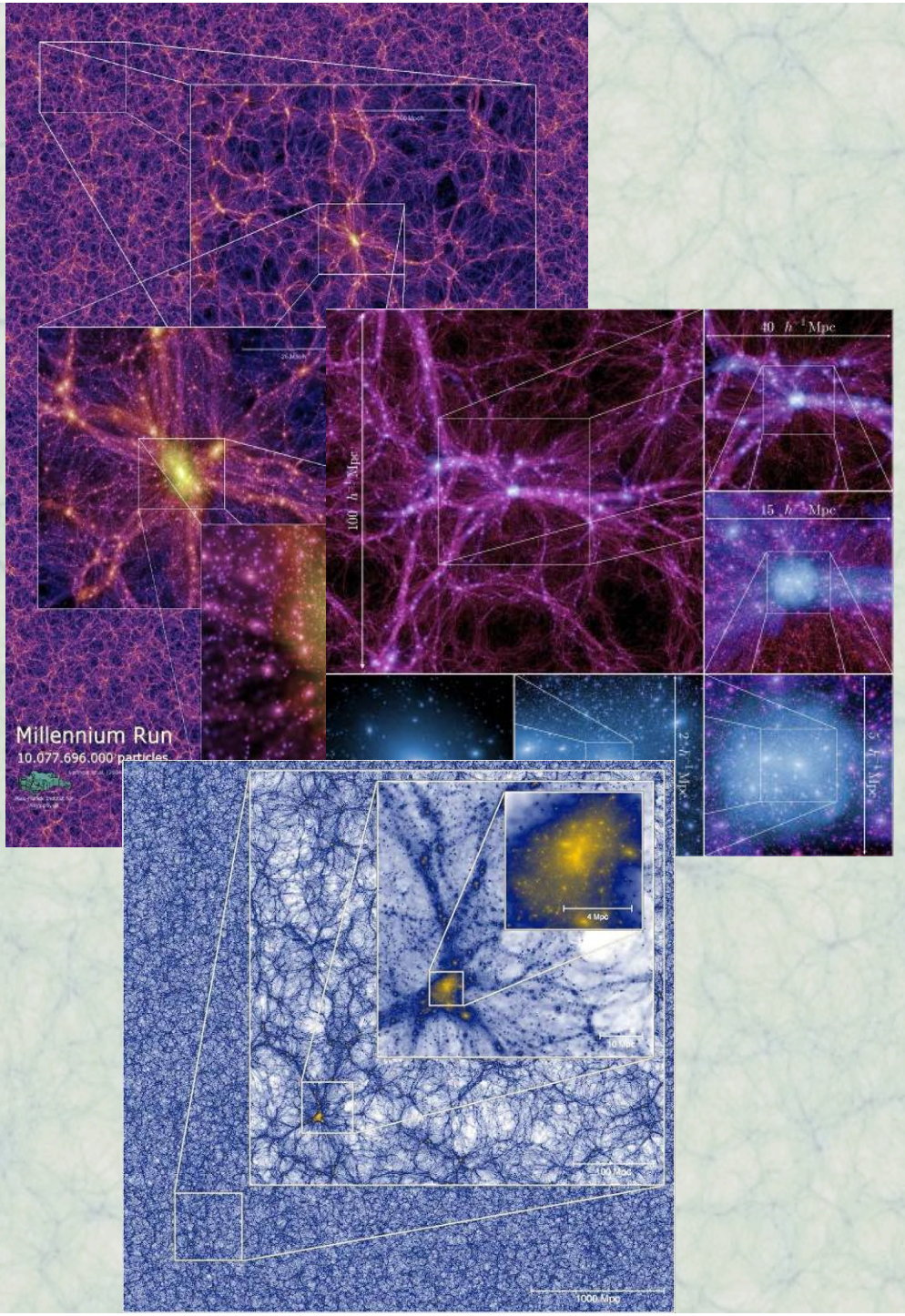
Mining and Disseminating Virtual Universes



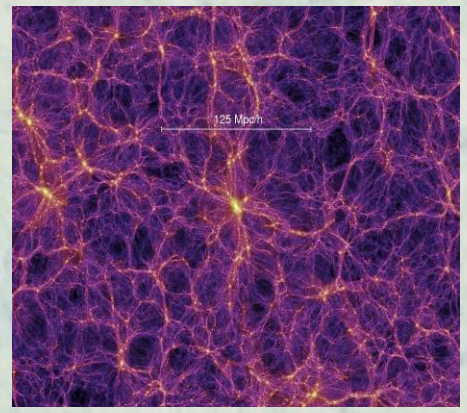
Gerard Lemson, MPA Garching

thanks to
Raul Angulo, Matthias Egger,
Bruno Henriques, Roderik Overzier,
Simon White (MPA), Guo Qi (Bejing),
Gabriel Dominique-Marleau (Canada),
Jeremy Blaizot (Lyon)

Overview

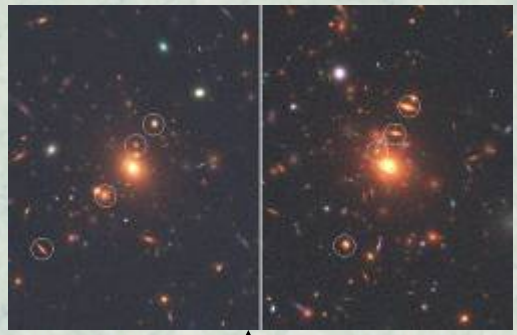
- Simulation data products: not just large, complex
  - Examples form Millennium suite
- Using relational database
  - Millennium Run DB
- Remote services
  - JobRunner
- Standards
  - IVOA
- Disseminating Simulations to Observers
  - MRObs

Raw data:
Particles

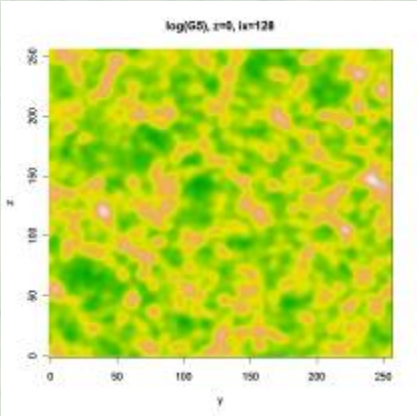


FOF groups and Subhalos

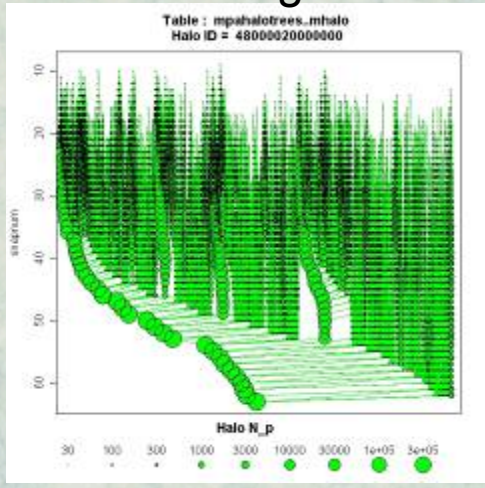
Mock images



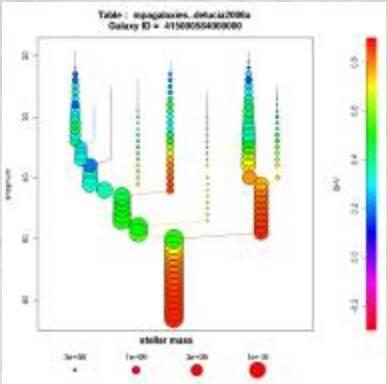
Density fields

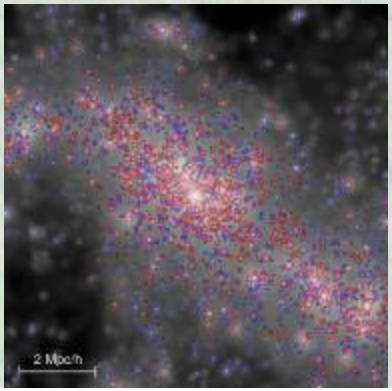


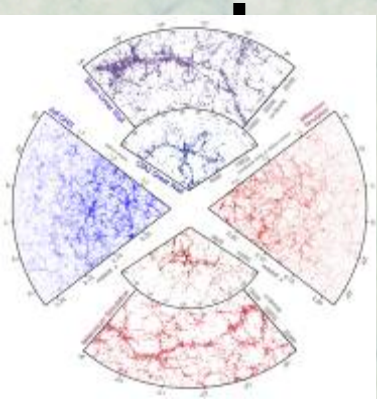
Subhalo merger trees



Synthetic galaxies (SAM)



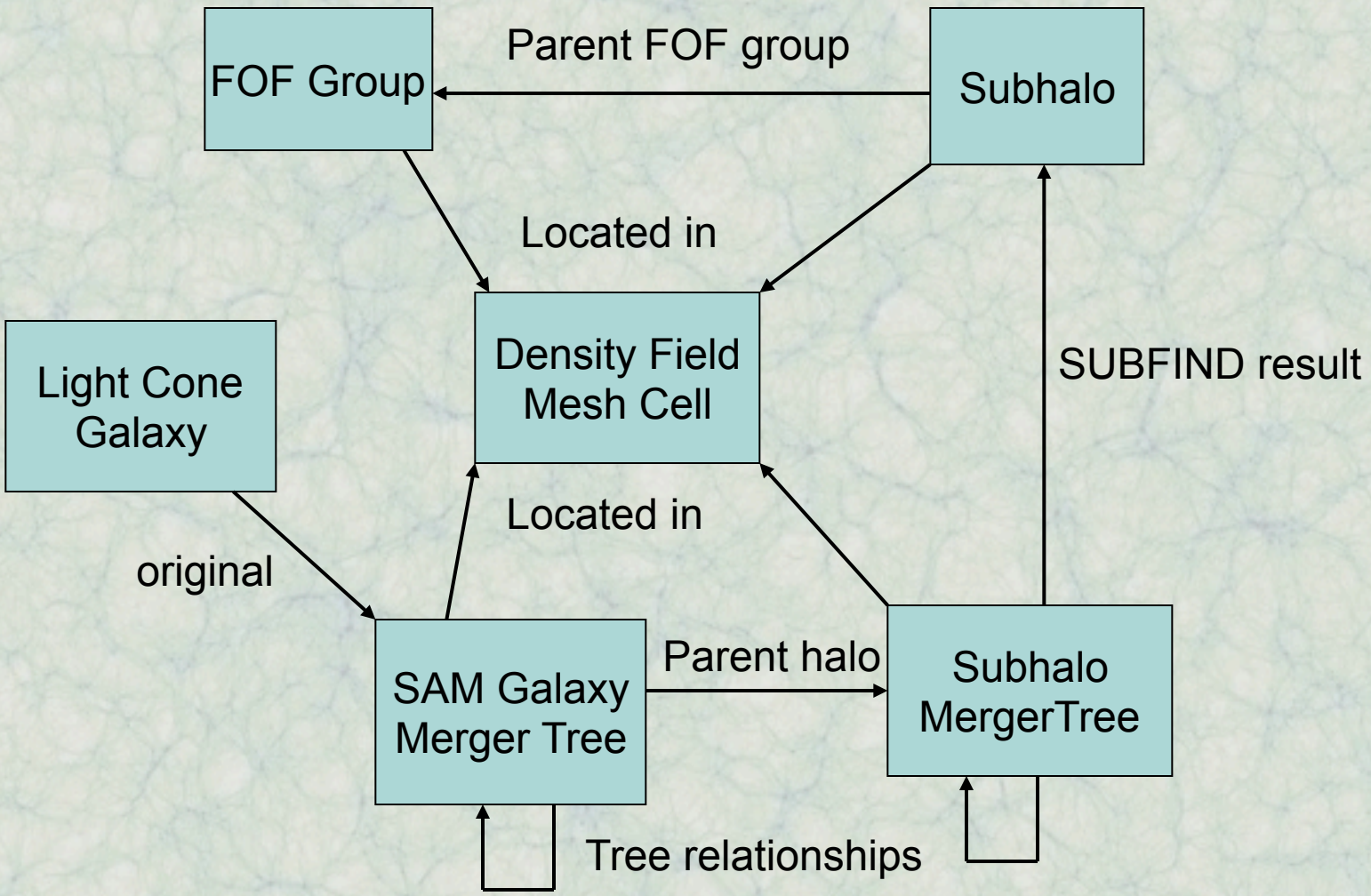




Mock catalogues

LDiA&C, 2011-08-08

Hierarchy of Data Products



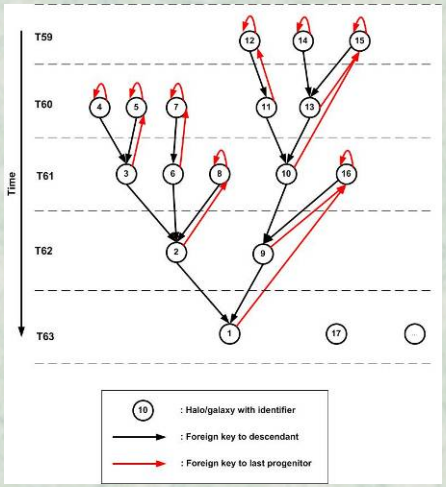
Why store simulations in relational database?

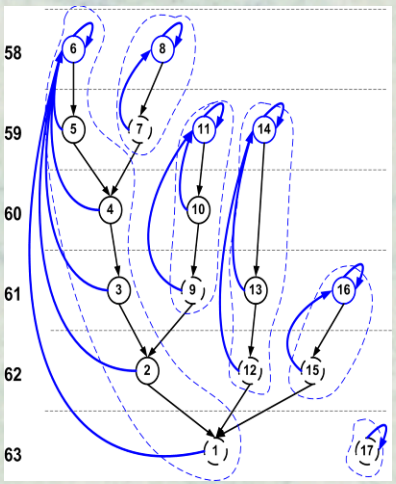
- encapsulation of data in terms of logical structure, *no need to know about internals of data storage*
- standard *query language* (SQL) for remote filtering
- advanced *query optimizers* (indexes, clustering)
- transparent parallelization
- authenticated remote access for multiple concurrent users
- (relatively) easy to add special functionality (spatial3d)

- **forces one to think carefully about data structure**
- **speeds up path from science question to answer**
- **facilitates communication (query code is cleaner)**
- **facilitates adaptation to IVOA standards (ADQL, TAP)**

Special features

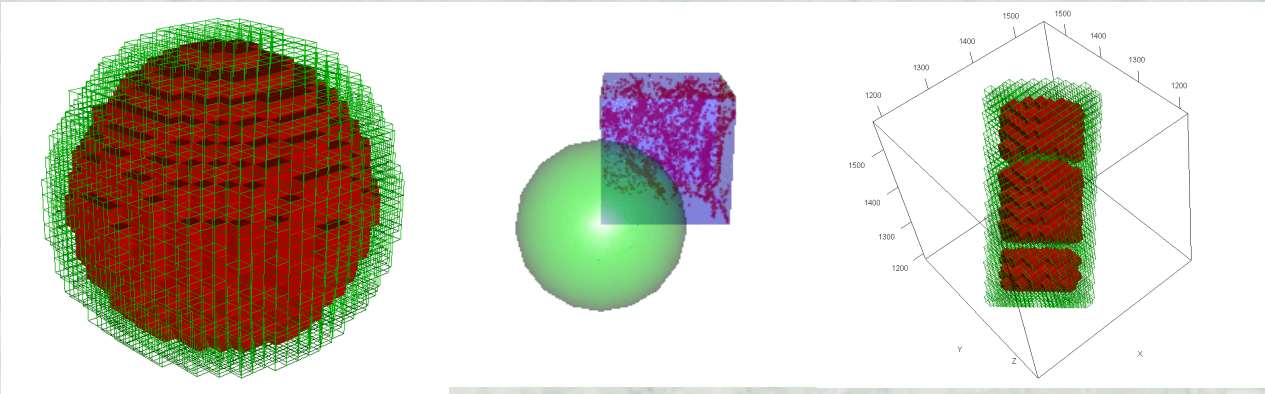
Encoding halo and galaxy merger trees





Lemson, Springel, ADASS 2005

Custom built library for spatial queries using Hilbert curve

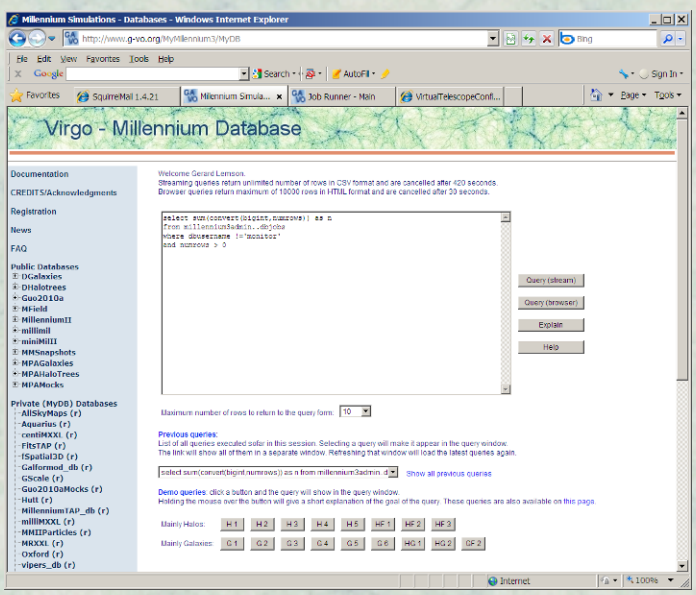


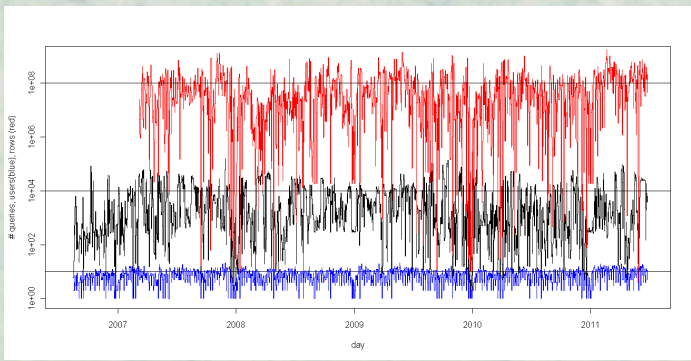
LDiA&C, 2011-08-08

Lemson, Budavari, Szalay SSDBM 2011

Virgo-Millennium Database

- 5 years
  - > 450 registered users
  - > 13 million queries
  - ~ 170 billion rows
- “mirror”@Durham
- copy@AIP: multidark.org

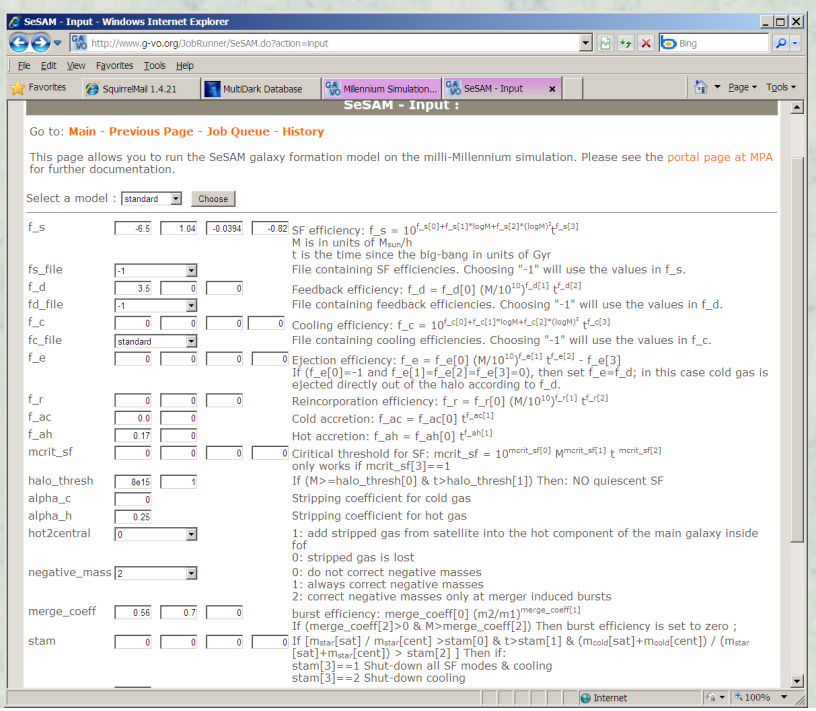


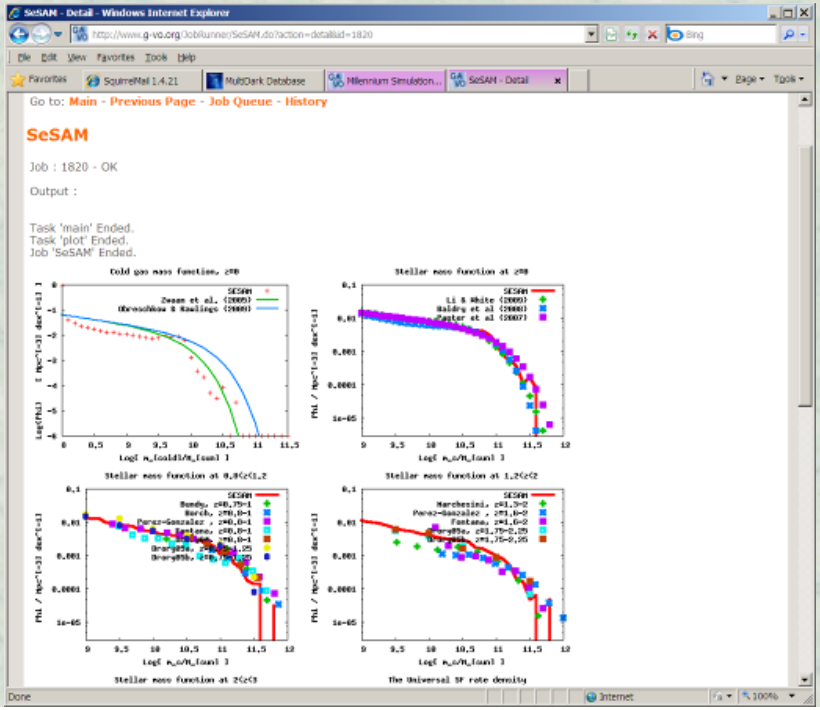


Lemson and the Virgo Consortium, astro-ph/060819

ad hoc services

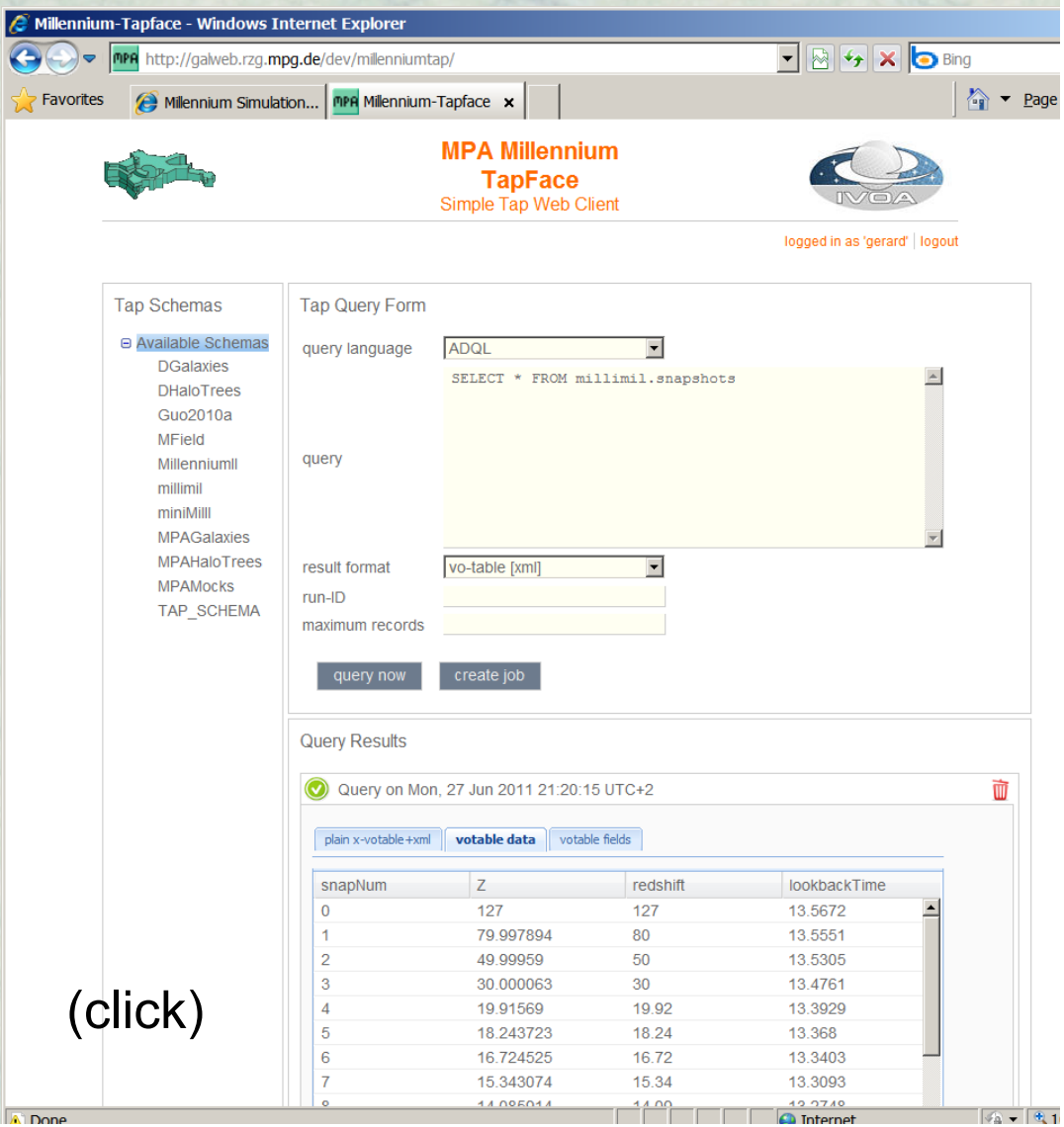
- SAM online
  - SeSAM, Neistein & Weinmann (2009), http://www.mpa-garching.mpg.de/galform/sesam/
  - soon
- Smac online
  - Visualisation of hydroclusters, Dolag etal, http://www.mpa-garching.mpg.de/HydroSims
- Soon more following UWS standard : L-Galaxies, light cones, mock images:





Implementing standards: TAP

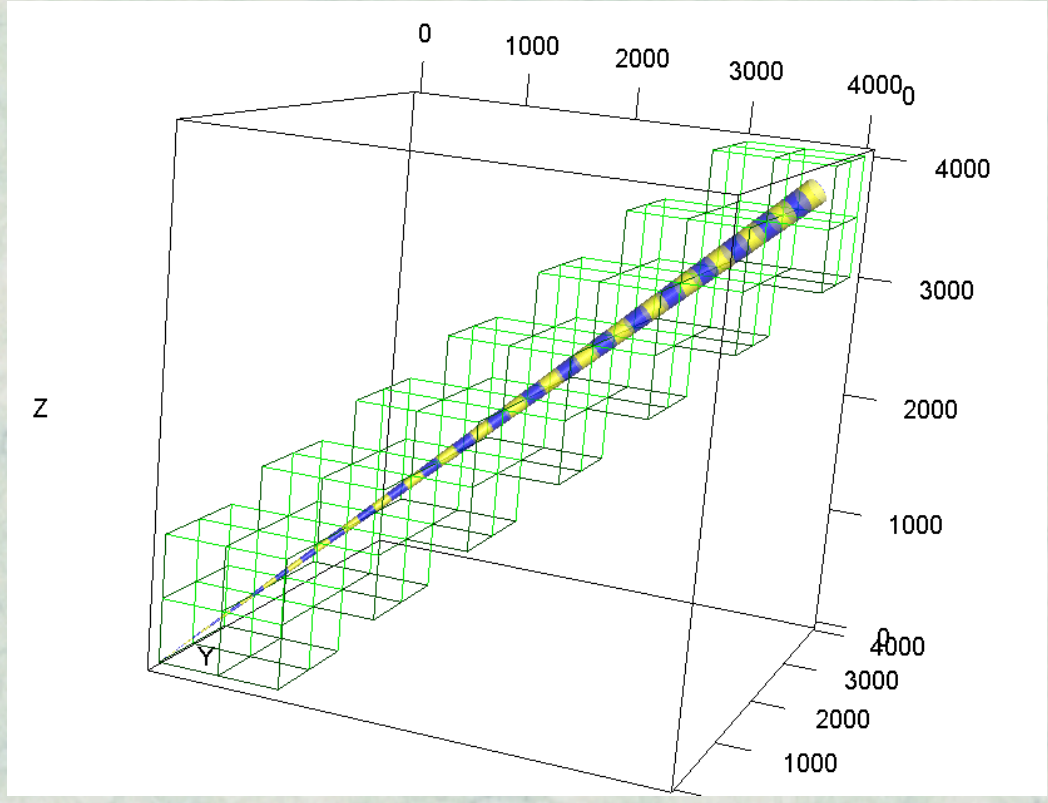
- IVOA standard for publishing relational databases
- async batch queries included

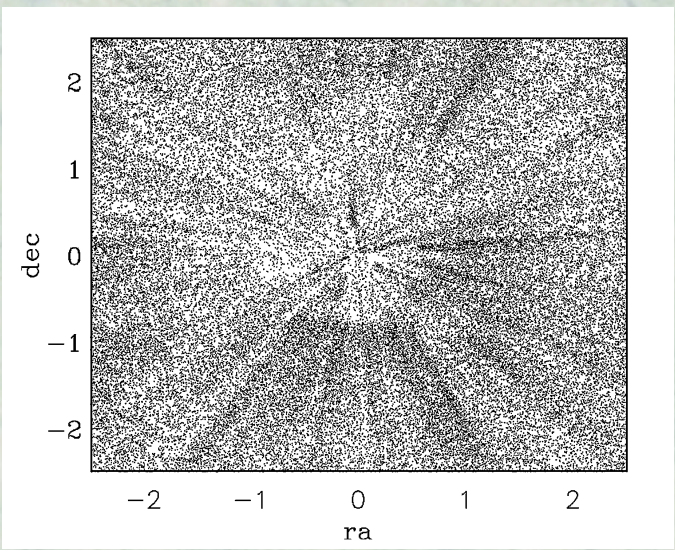


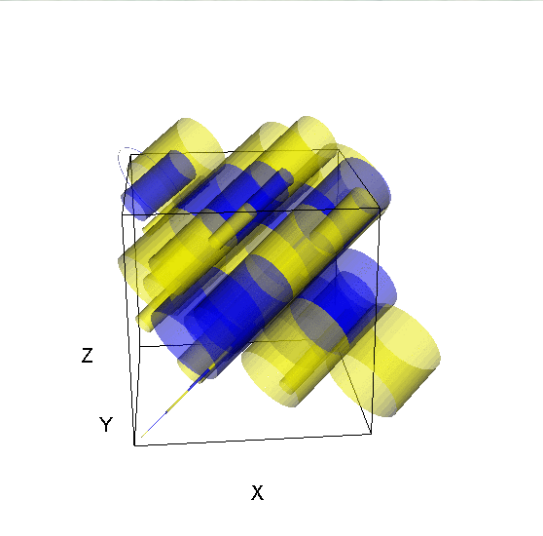
(click)

Reaching out to observers: Mock catalogues ...

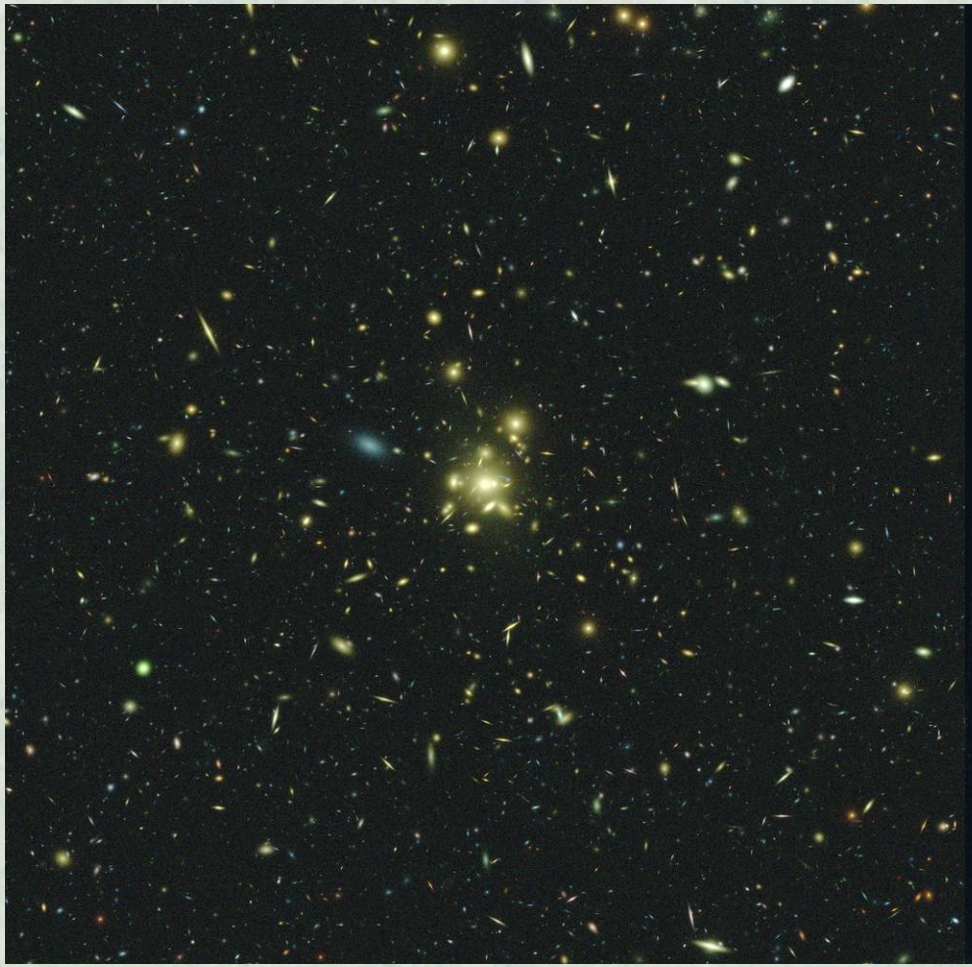
- Approximate observing on light cone
  - Blaizot etal (2005)
  - Kitzbichler & White (2006)







... and mock images
Clusters at z=0.4



MRObs, 7e14 Msun

HST/ACS, Harsono & De Propis, 2007

Message

- Large data can benefit from remote filtering
  - SQL, *ad hoc* services
- Complex data can benefit from relational model
  - *Federating cosmological databases*
- Unknown data can benefit from standardisation
  - data formats, (meta)data models
  - IVOA, VAO
- Unfamiliar data can benefit from special treatment
  - e.g. virtual telescopes

- Realistic use cases very important !

TODO

- Brain storming on use cases.
- Intro to IVOA standardisation process
  - what standards do we want/need?
- *Federation of Cosmological Databases*
  - why, how, what?
- Intro to RDBs?
- ...