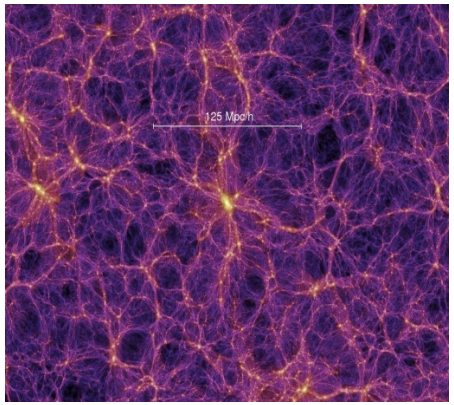
PUBLISHING SIMULATIONS IN THE VO AND ELSEWHERE

Gerard Lemson
MPA Garching, Germany

Max-Planck-Institut für Astrophysik

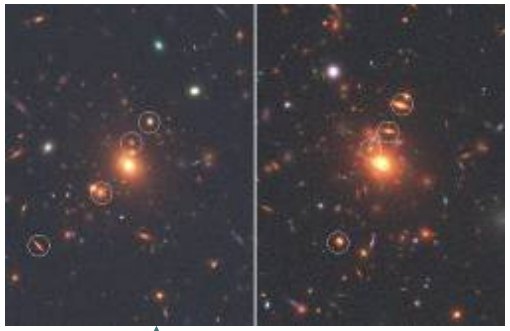
UC-HiPACC

Raw data: Particles

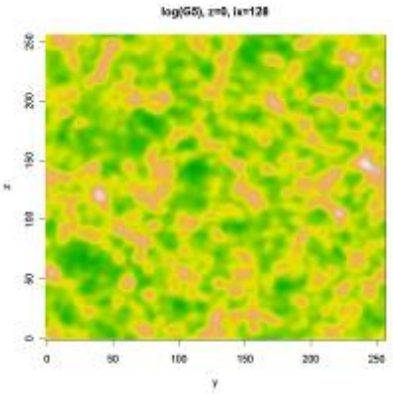


FOF groups and Subhalos

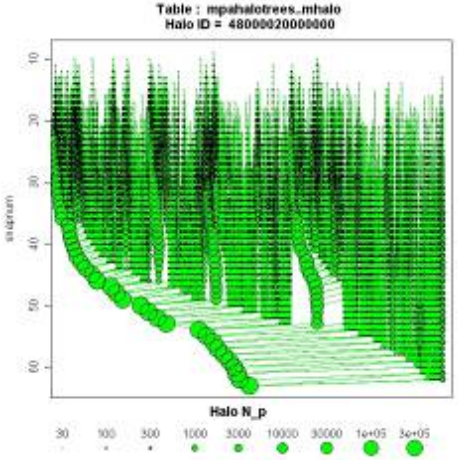
Mock images

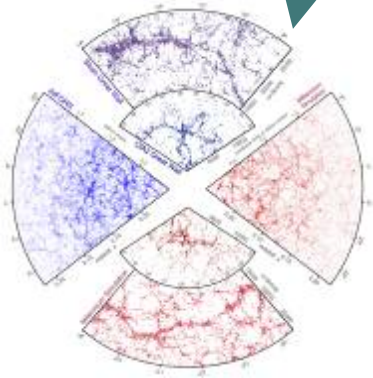


Density fields

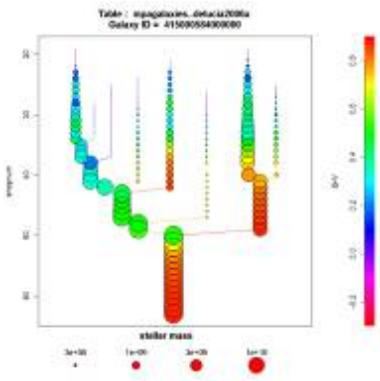


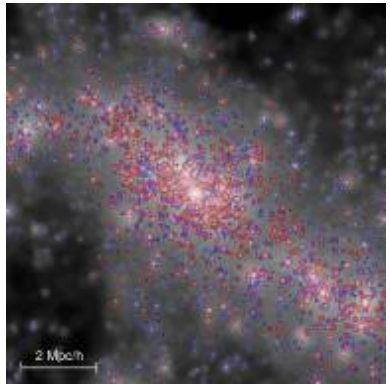
Subhalo merger trees





Synthetic galaxies (SAM)





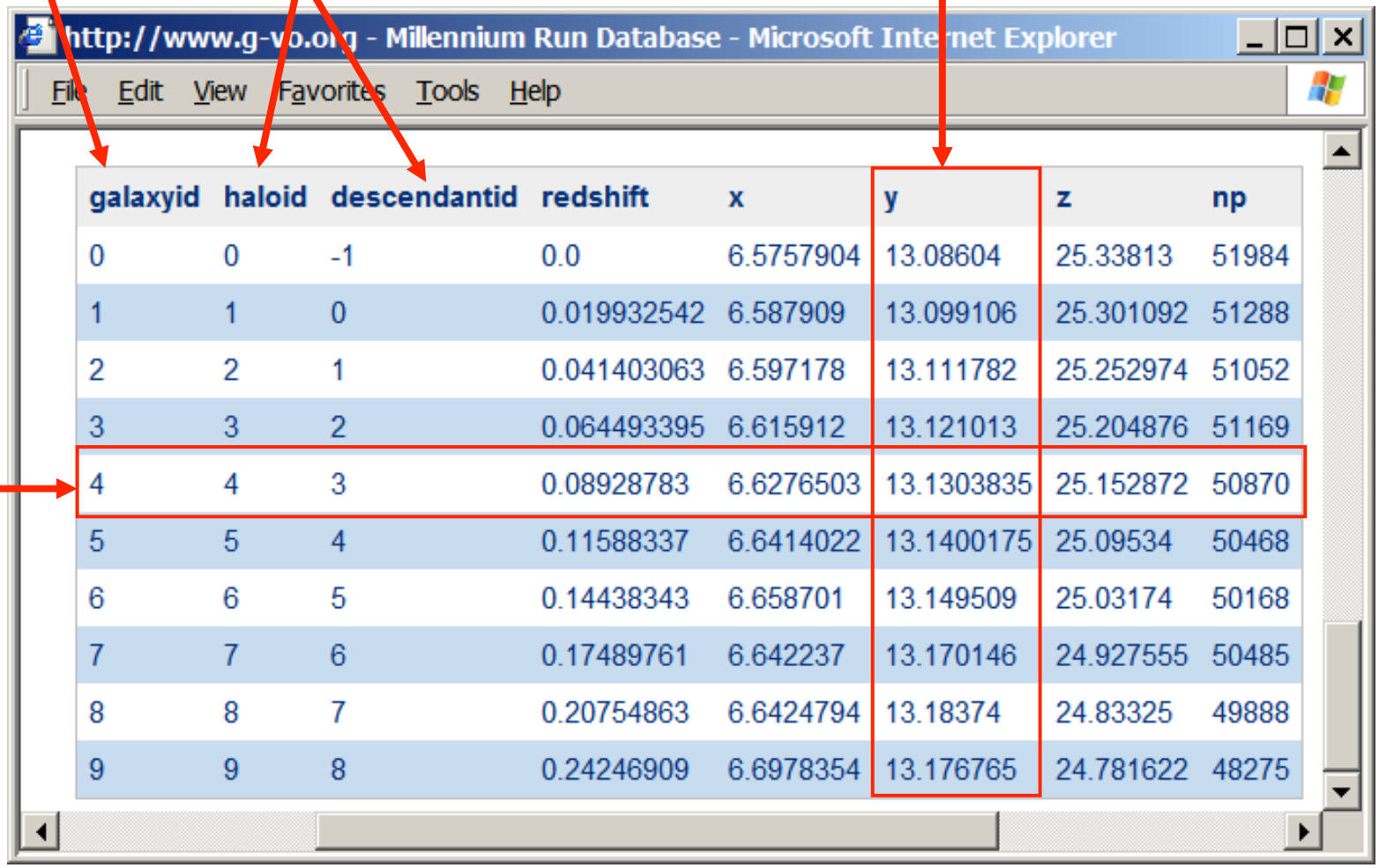
Mock catalogues

Primary Key Column

Foreign Key Columns

Column

Row



galaxyid	haloid	descendantid	redshift	x	y	z	np
0	0	-1	0.0	6.5757904	13.08604	25.33813	51984
1	1	0	0.019932542	6.587909	13.099106	25.301092	51288
2	2	1	0.041403063	6.597178	13.111782	25.252974	51052
3	3	2	0.064493395	6.615912	13.121013	25.204876	51169
4	4	3	0.08928783	6.6276503	13.1303835	25.152872	50870
5	5	4	0.11588337	6.6414022	13.1400175	25.09534	50468
6	6	5	0.14438343	6.658701	13.149509	25.03174	50168
7	7	6	0.17489761	6.642237	13.170146	24.927555	50485
8	8	7	0.20754863	6.6424794	13.18374	24.83325	49888
9	9	8	0.24246909	6.6978354	13.176765	24.781622	48275

Normalization

FOF

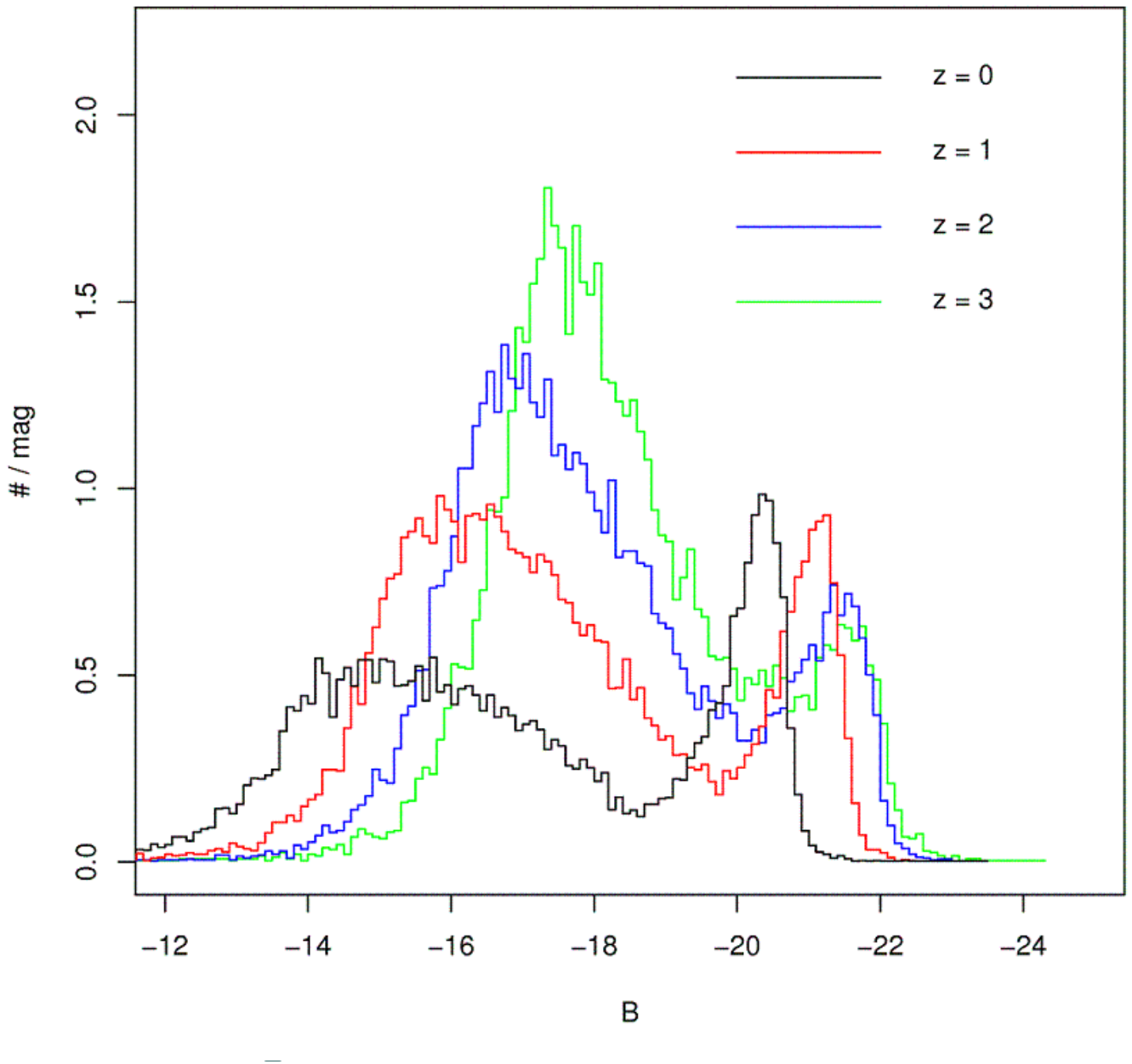
fofId	nSub	m200	x	…
123	2	445.77	7.6	…
456	2	101.32	35.1	…
789	1	70.0	67.0	…
…	…	…	…	…

Galaxy

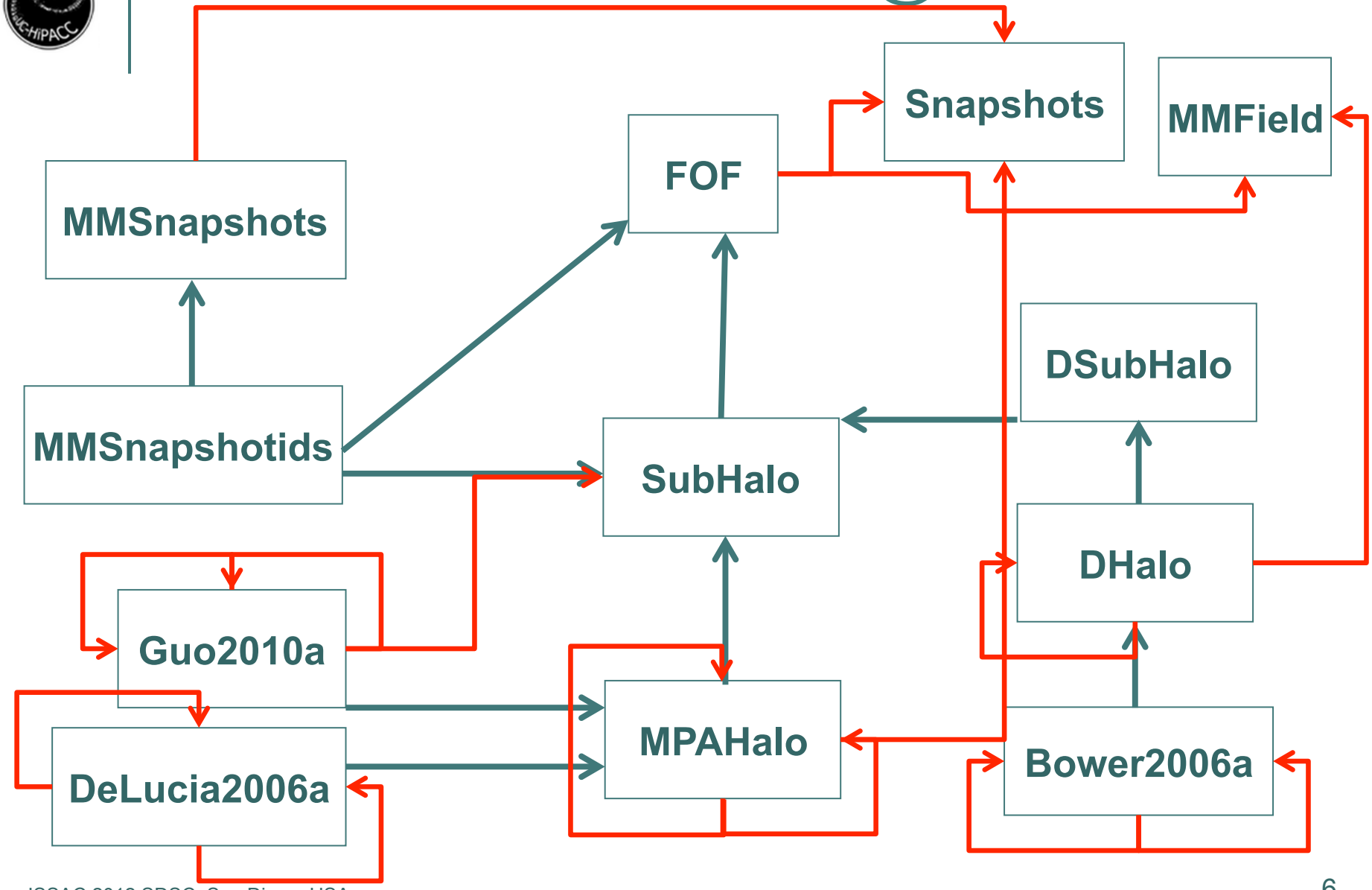
galId	haloId	mStar	magB	X	…
112	6625	0.215	-17.9	7.6	…
113	6625	0.038	-15.6	7.4	…
154	6626	0.173	-17.1	7.65	…
221	7883	1.20	-20.7	35.1	…
223	7883	0.225	-19.7	35.0	…
225	7883	0.04	-17.5	34.9	…
278	7884	1.54	-19.4	35.2	…
…	…	…	…	…	…

SubHalo

haloId	fofId	Np	X	vMax	…
6625	123	100	7.6	165	…
6626	123	65	7.9	130	…
7883	456	452	35.1	200	…
7884	456	255	35.2	190	…
9885	789	30	67.0	110	…
…	…	…	…	…	…


z = 0
z = 1
z = 2
z = 3
# / mag
0.0
0.5
1.0
1.5
2.0
−12
−14
−16
−18
−20
−22
−24
B

millimil database/schema @ISSACTAP



Motivation for data model

1.Return the (B-band luminosity function of) galaxies residing in halos of mass between 10^13 and 10^14 solar masses.
2.Return the galaxy content at z=3 of the progenitors of a halo identified at z=0
3.Return all the galaxies within a sphere of radius 3Mpc around a particular halo
4.Return the complete halo merger tree for a halo identified at z=0
5.Find positions and velocities for all galaxies at redshift zero with B-luminosity, colour and bulge-to-disk ratio within given intervals.
6.Find properties of all galaxies in haloes of mass 10**14 at redshift 1 which have had a major merger (mass-ratio < 4:1) since redshift 1.5.
7.Find all the z=3 progenitors of z=0 red ellipticals (i.e. B-V>0.8 B/T > 0.5)
8.Find the descendants at z=1 of all LBG's (i.e. galaxies with SFR>10 Msun/yr) at z=3
9.Make a list of all haloes at z=3 which contain a galaxy of mass >10**9 Msun which is a progenitor of BCG's in z=0 cluster of mass >10**14.5
10.Find all z=3 galaxies which have NO z=0 descendant.
11.Return the complete galaxy merging history for a given z=0 galaxy.
12.Find all the z=2 galaxies which were within 1Mpc of a LBG (i.e. SFR>10Msun/yr) at some previous redshift.
13.Find the multiplicity function of halos depending on their environment (overdensity of density field smoothed on certain scale)
14.Find the dependency of halo formation times on environment (“halo assembly bias”)

Max-Planck-Institut für Astrophysik

Some special design features in the Millennium Databases

Identifiers
Environment
Trees
Spatial queries (Tamas L1)

Identifiers

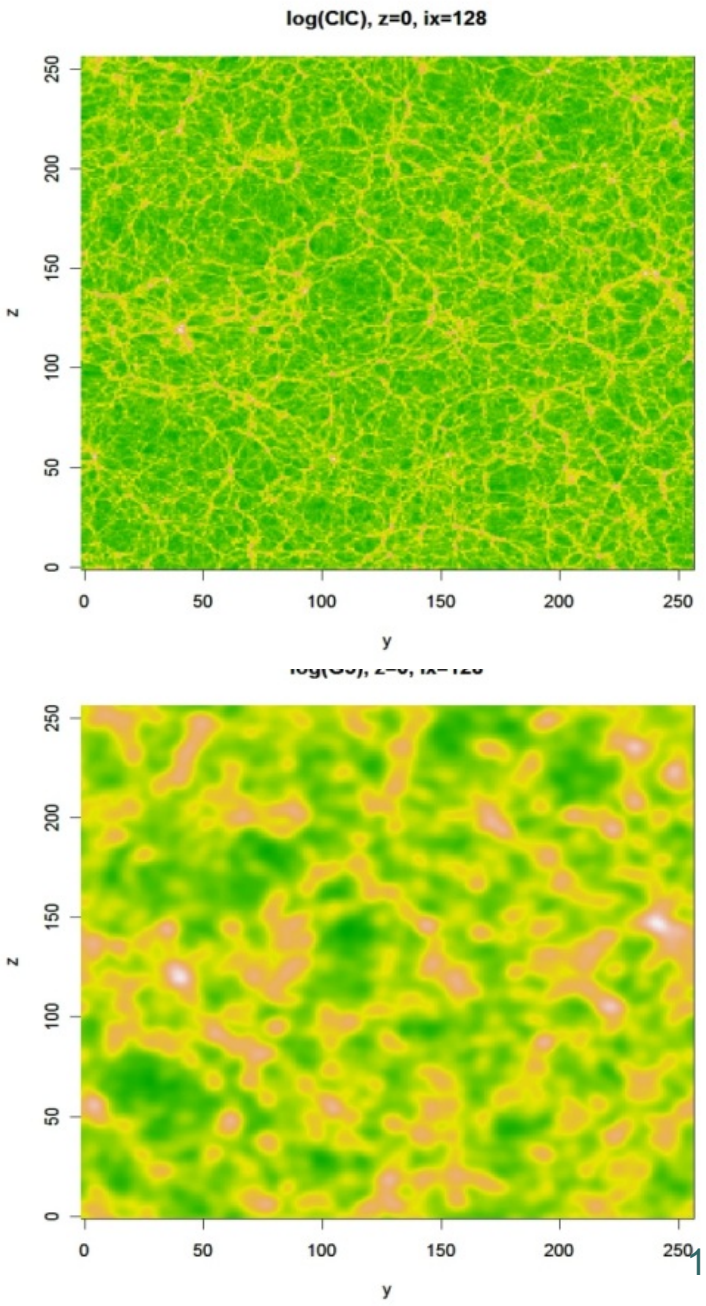
- Uniquely identify an object in a table
- May have extra structure for convenience
- E.g.
  - haloid = fileNr x 1e12+treeId x 1e6 + rank-in-tree
- Allows querying “in chunks”:
  - ```
    select ...
      from halos
     where haloid between
         :f1*1e12 and (:f1+:stride)*1e12-1
    ```
  - ```
    :f1 [0,511]
    :stride =1,10,50
    ```

Identifiers (cntd)

- Parent-child relations reflected in identifiers avoid need for associative tables
  - FOFs in snapnums
    - fofId=snapnum*10^{10}+filenr*10^{6}+rank-in-file
  - Subhalos in FOFs
    - subhaloId = fofId*10^{6}+rank-in-fof
  - Particles in FOFs (mini-Mil-II)
    - particleId = fofId*10^{6}+rank-in-fof
    - global id for tracking of orbits

Representing Environment

- “find void galaxies”
- Environment as density field on 256^3 grid
- Smoothed at various scales
  - CIC
  - G_5, G10
- Objects know their grid cell, identified by **`phKey`**

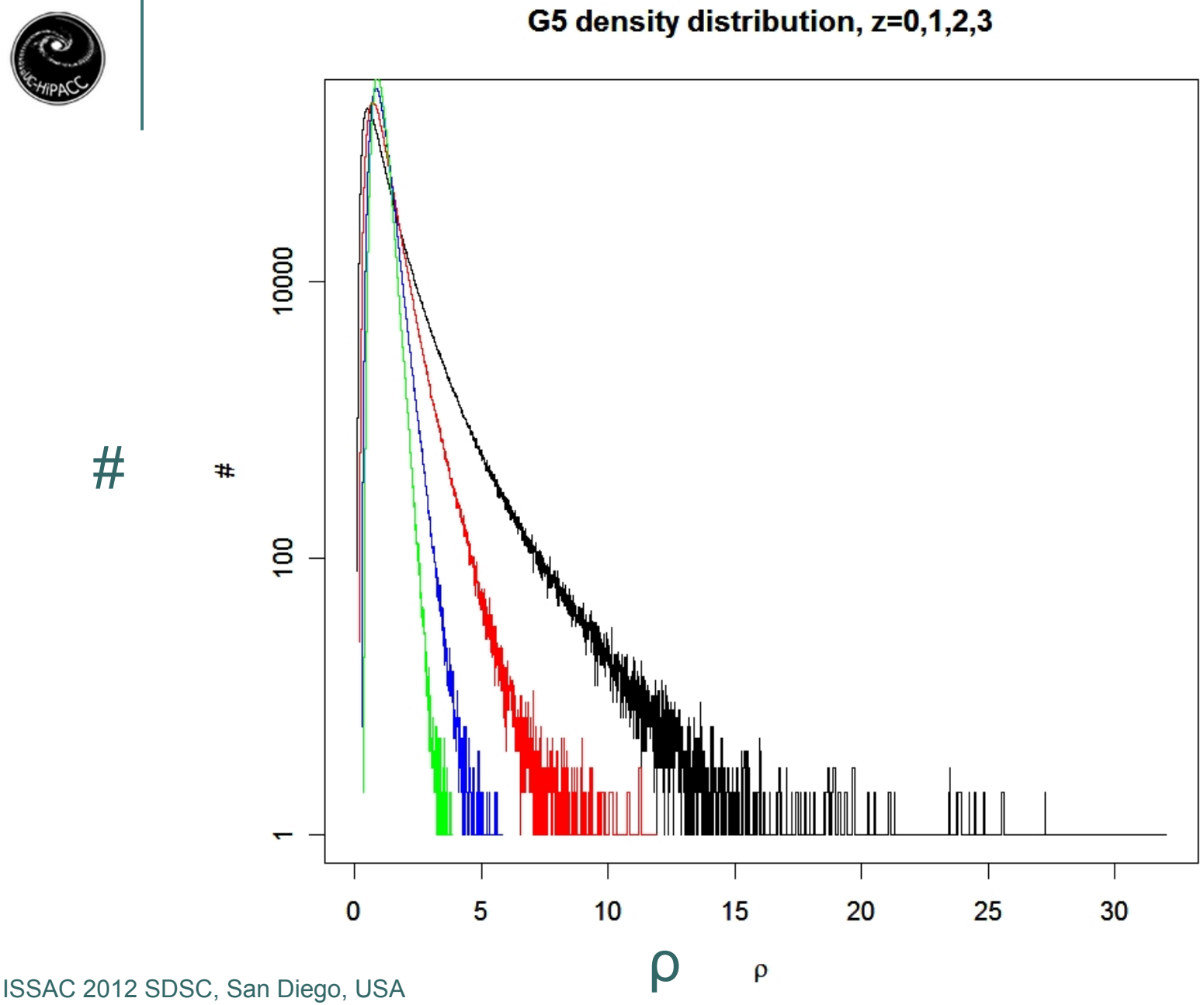


Millimil.mmfield
(no ix, iy,iz)

ix	iy	iz	snapnum	phkey	cic	g1_25	g2_5	g5	g10
7	3	5	63	167	0.34960523	0.2927515	0.39591226	1.6946505	1.9613136
7	3	6	63	168	0.55104446	0.44405165	0.43340233	1.5855794	1.9591872
7	3	7	63	169	0.61097676	0.51611525	0.5319513	1.5662786	1.945689
6	3	7	63	170	0.25970677	0.30828816	0.552665	2.0592175	2.0777974
6	3	6	63	171	0.2796842	0.35897163	0.5428122	2.204627	2.1025789
6	2	6	63	172	0.9006497	0.5354726	0.5335305	1.8476957	2.0909338
6	2	7	63	173	0.19145049	0.3474926	0.5078265	1.7384003	2.0719836
7	2	7	63	174	0.8240695	0.6702552	0.5201308	1.3223413	1.9361987
7	2	6	63	175	0.7857794	0.71089053	0.5162294	1.3444736	1.9434941
7	1	6	63	176	2.265775	1.5351428	0.68073857	1.2364181	1.9360421

Histogram of density field at redshifts 0,1,2,3; Gaussian smoothing 5 Mpc/h (full millennium density field)

```
select snapnum
,      .01*floor(f.g5/.01) as g5
,      count(*) as num
  from mfield..mfield f
 where f.snapnum in (63,41,32,27)
 group by snapnum
,      .01*floor(f.g5/.01)
 order by 1,2
```

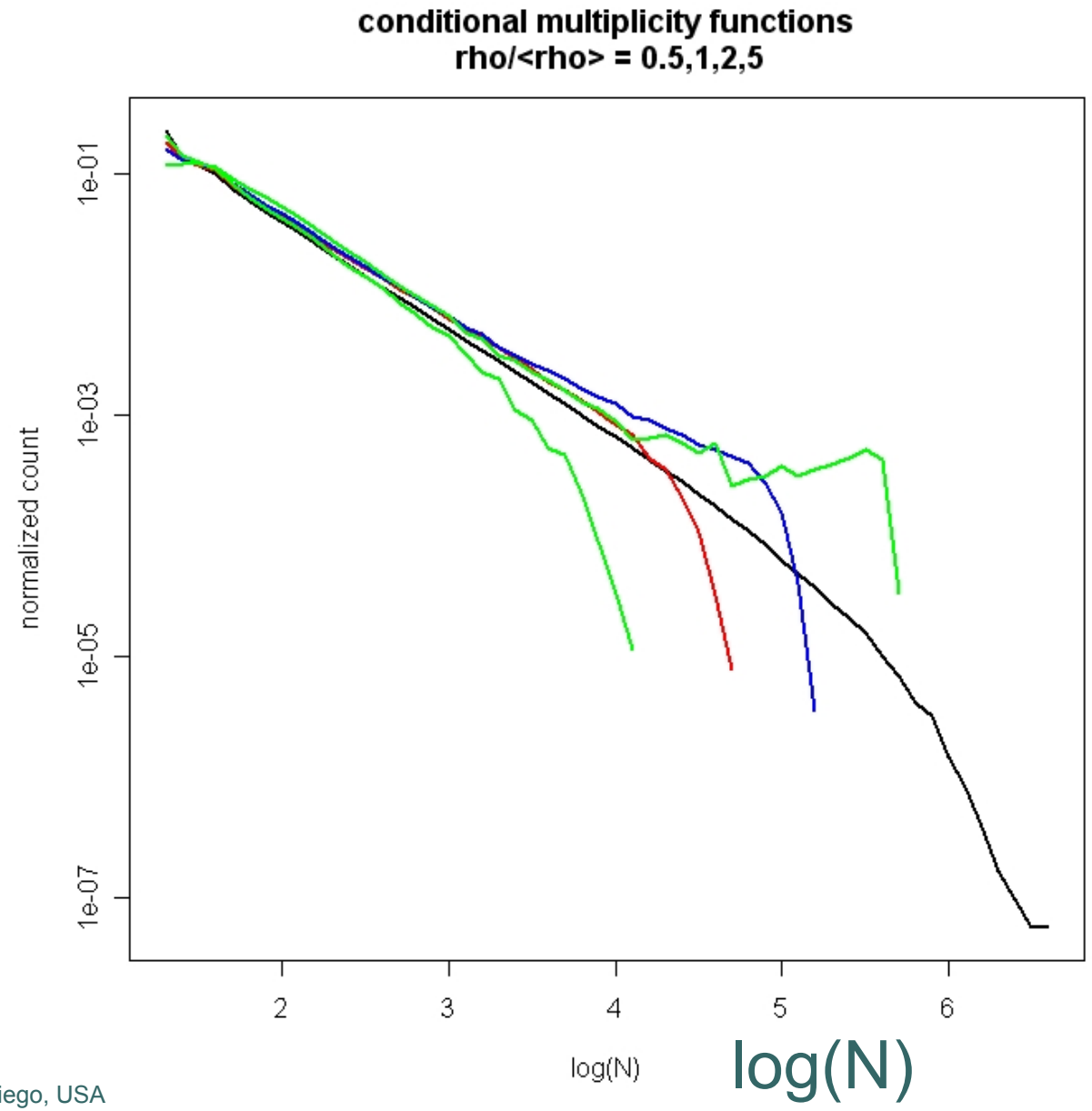

G5 density distribution, z=0,1,2,3
#
10000
100
1
0
5
10
15
20
25
30
ρ
ρ

FOF mass multiplicity function, conditioned on density in environment

```
select .1*floor(log10(fof.np)/.1) as lognp
,       count(*) as num
  from mfield..mfield f
  ,     mfield..fof fof
 where fof.snapnum=f.snapnum
   and fof.phkey = f.phkey
   and f.snapnum = 63
   and f.g5 between 1 and 1.1
group by .1*floor(log10(fof.np)/.1)
order by 1
```

(and similar for g5 = 0.5,2,5)

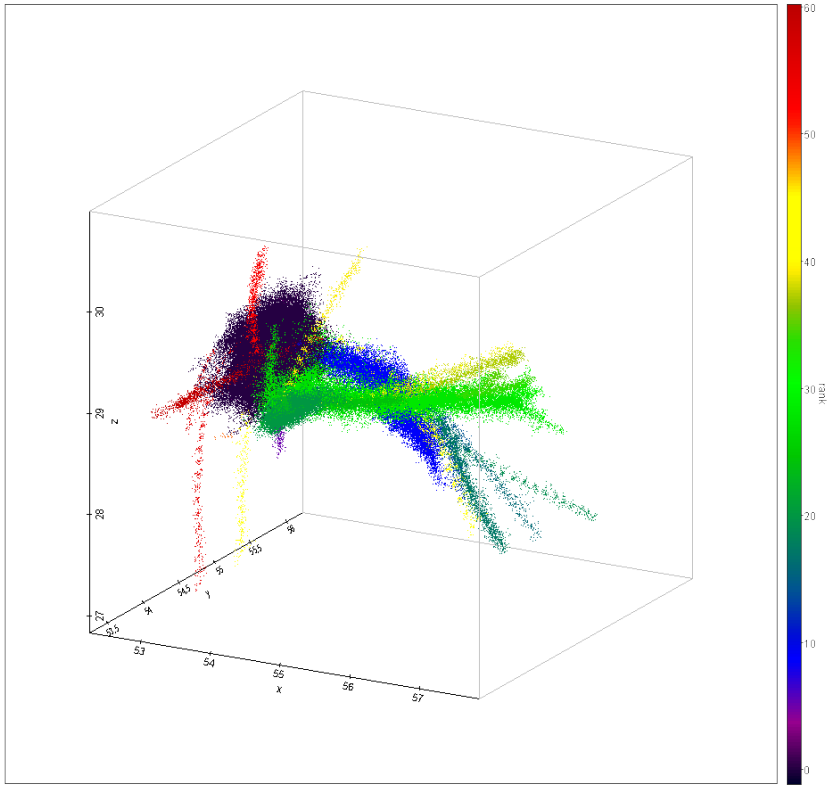
#


conditional multiplicity functions
rho/<rho> = 0.5,1,2,5
normalized count
1e-01
1e-03
1e-05
1e-07
2
3
4
5
6
log(N)

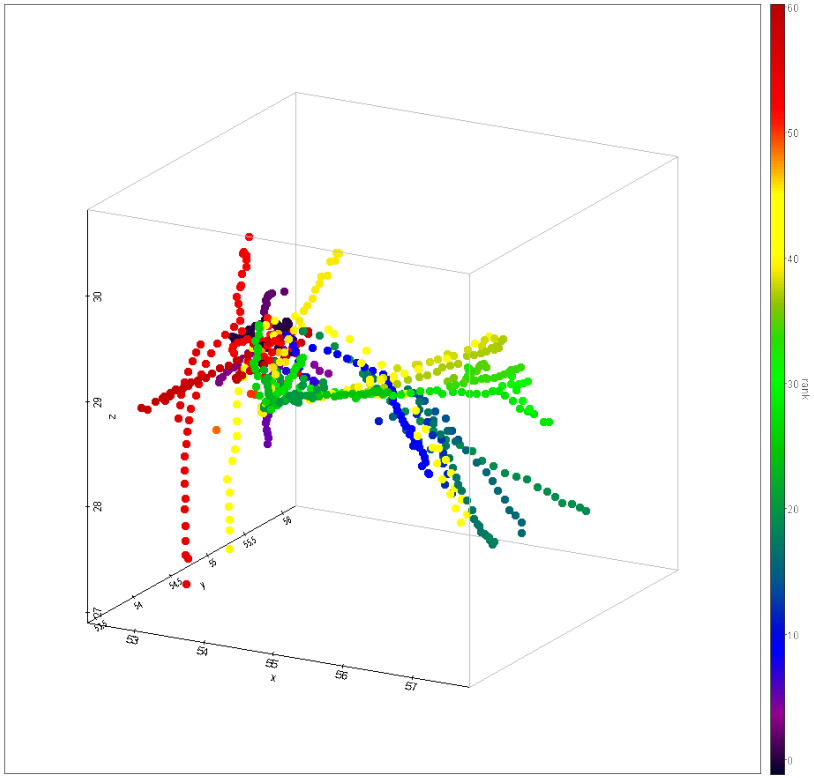
log(N)

Max-Planck-Institut für Astrophysik

Time evolution on merger trees



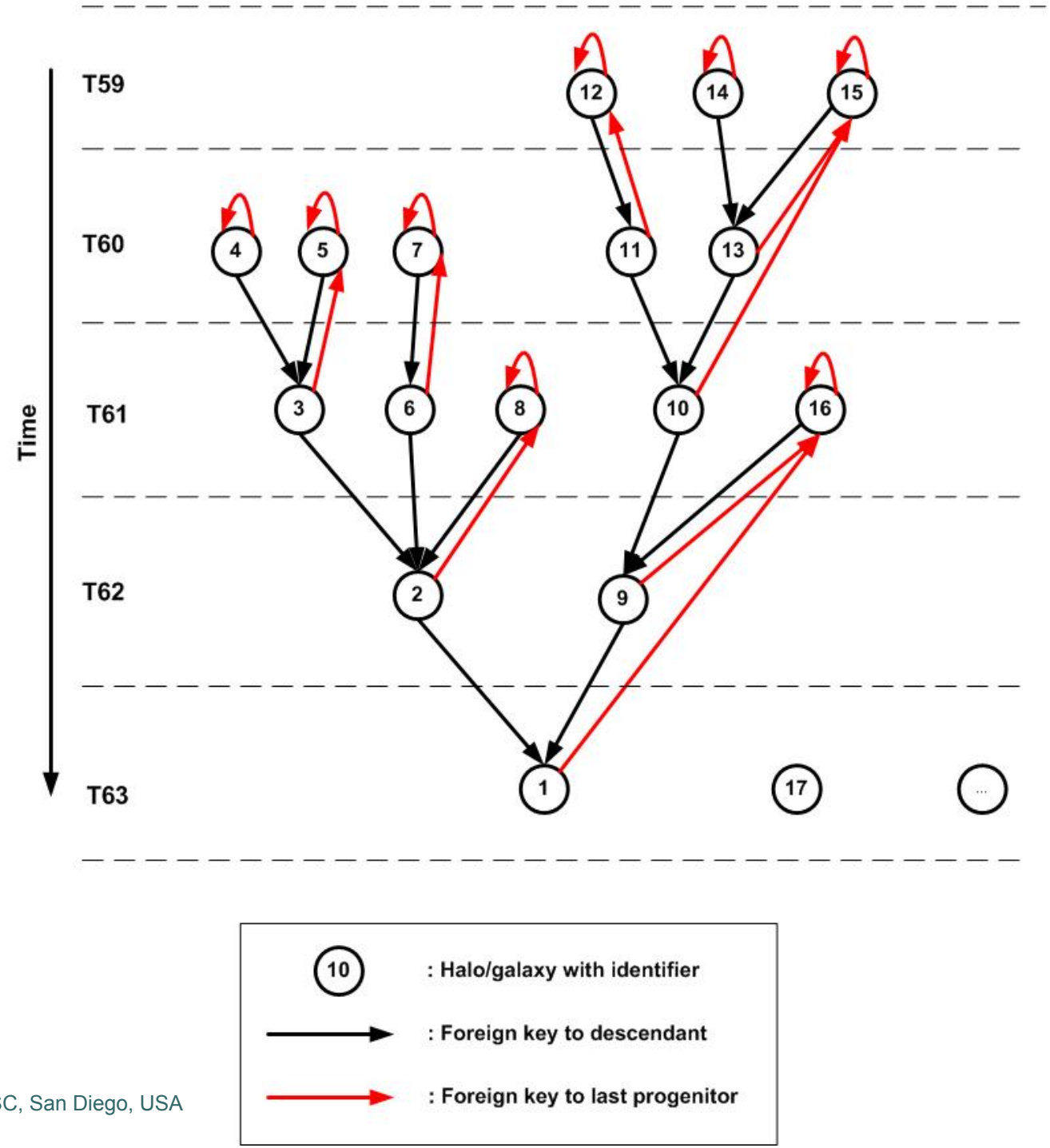
particles



halos

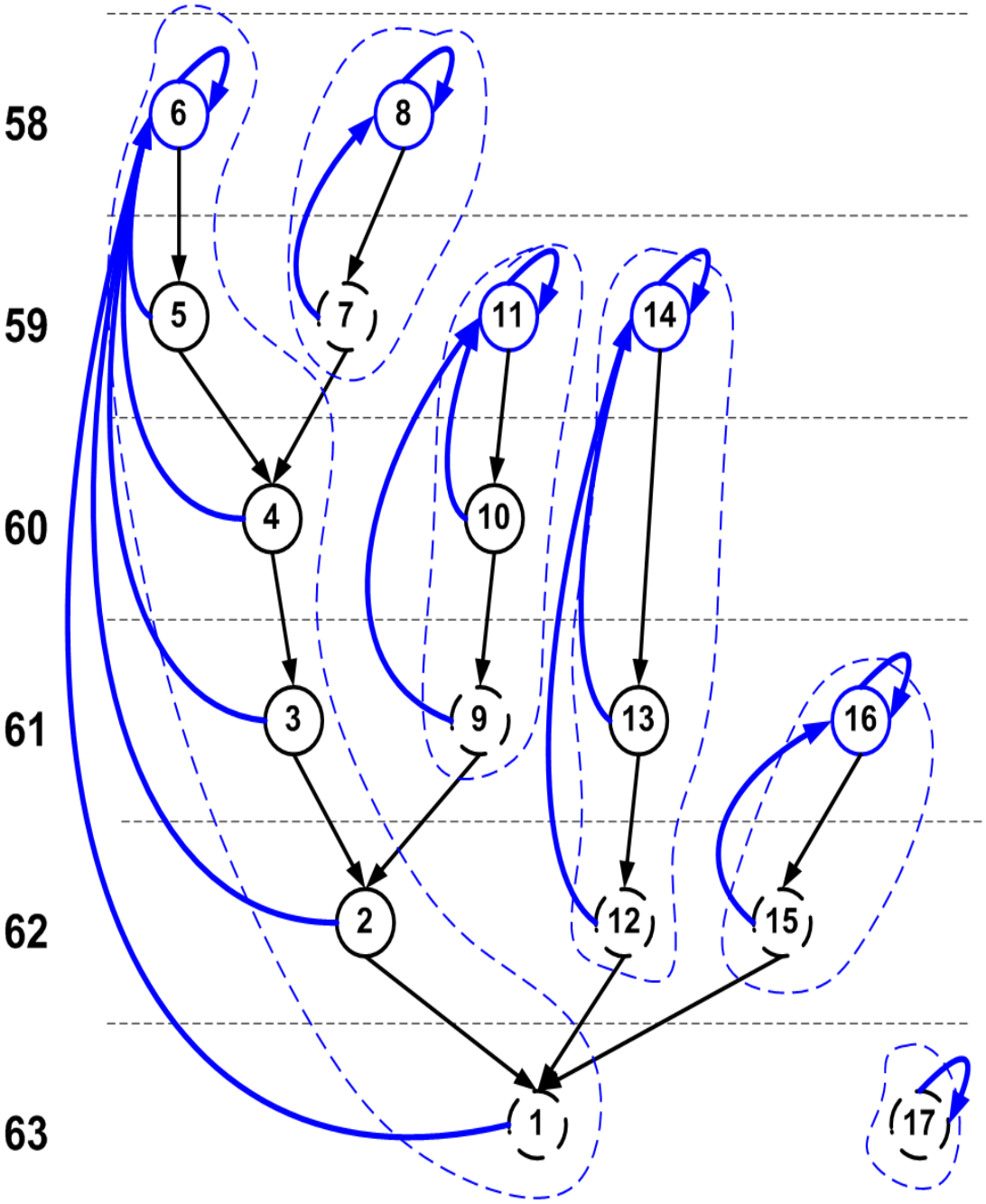
Trees in a database

- Recursion only partially supported
  - And not efficient
- Special solution
  - Indexing based on depth-first-order of progenitors
- Pointers to
  - descendant
  - last progenitor (finding all progenitors)
  - main leaf (finding main progenitors)
    - trees are getting very large (10^8)
    - branches ~100
  - tree root
    - finding descendants. indexing on intervals?


T59
T60
T61
T62
T63
Time
12
14
15
4
5
7
11
13
3
6
8
10
16
2
9
1
17
...
10 : Halo/galaxy with identifier
: Foreign key to descendant
: Foreign key to last progenitor

Main branches

- Track the object
- Pointer to *main leaf*



Merger trees (halos):

```
select prog.*
  from millimil.mpahalo des
  ,    millimil.mpahalo prog
 where des.haloId = 0
   and prog.haloId between
       des.haloId and des.lastProgenitorId
```

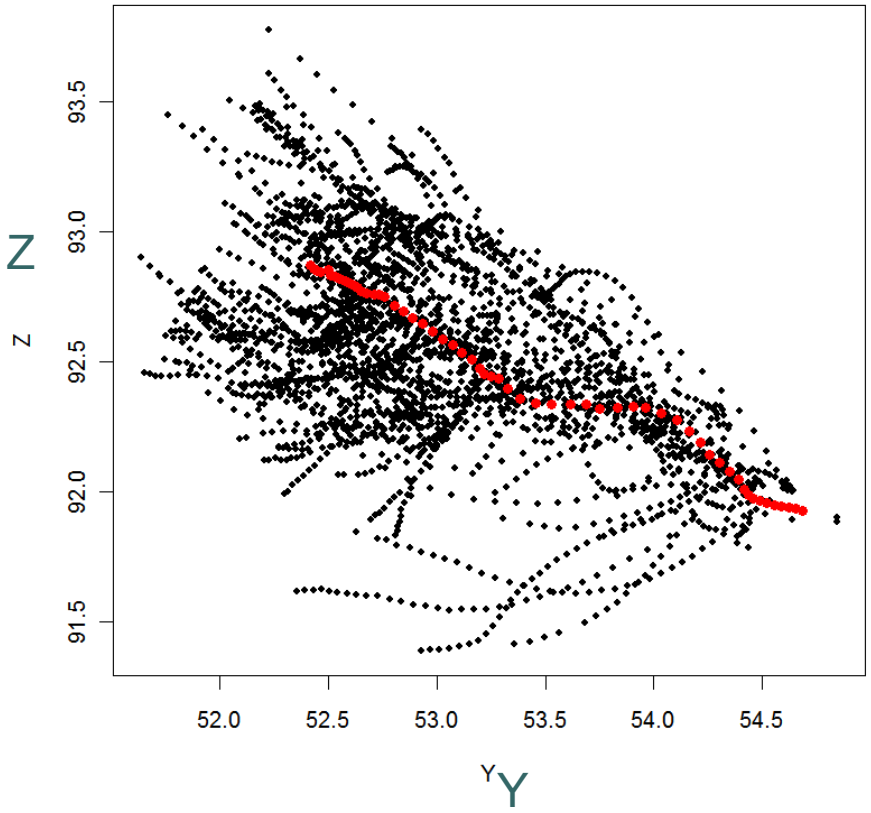
Main progenitors (galaxies):

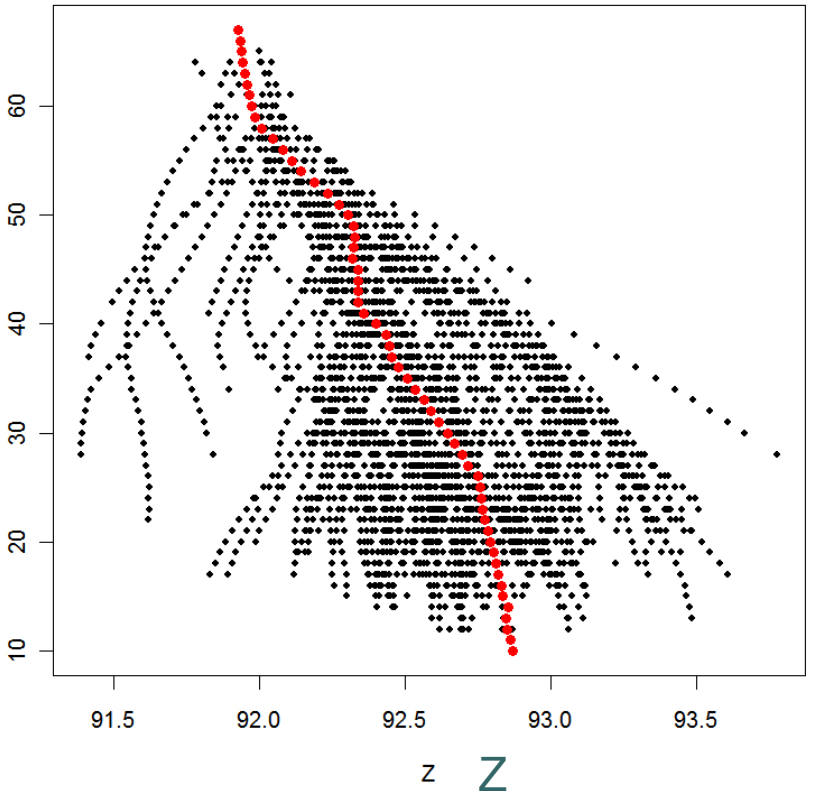
```
select prog.*
  from millimil.guo2010a des
  ,    millimil.guo2010a prog
 where des.galaxyId = 0
   and prog.galaxyId between
       des.galaxyId and des.mainLeafId
```

Descendants :

Hands on session

Merger tree rooted in particular halo (in Millennium-II database)

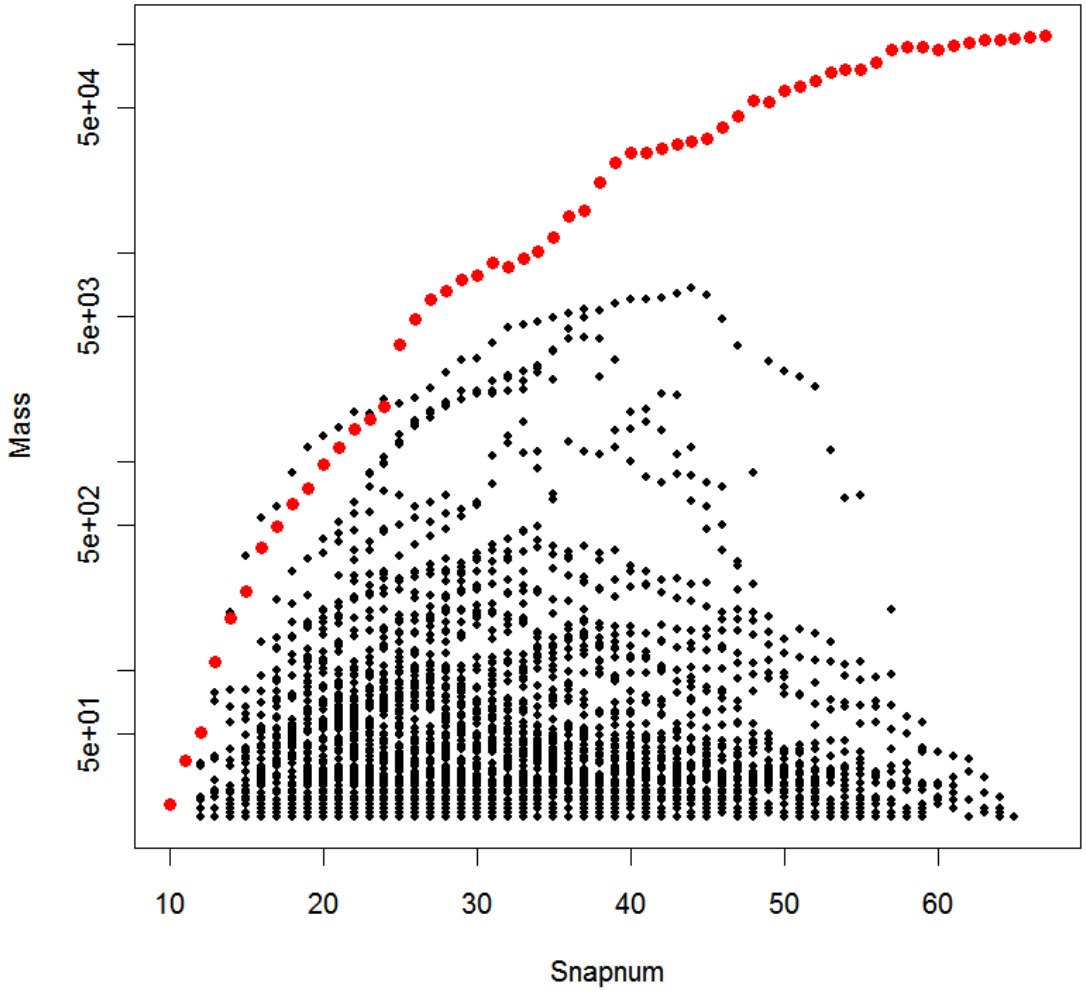




Time

Evolution of mass

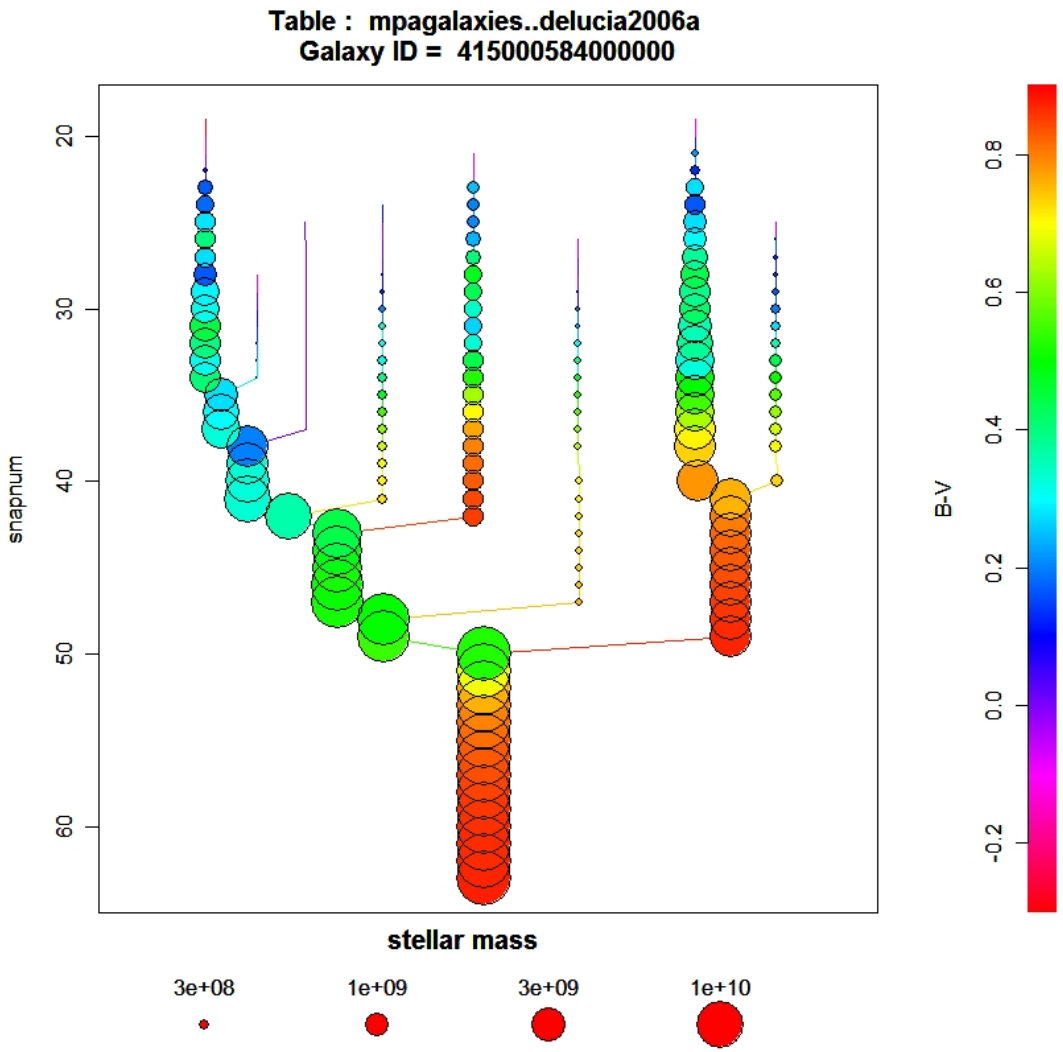
Mass

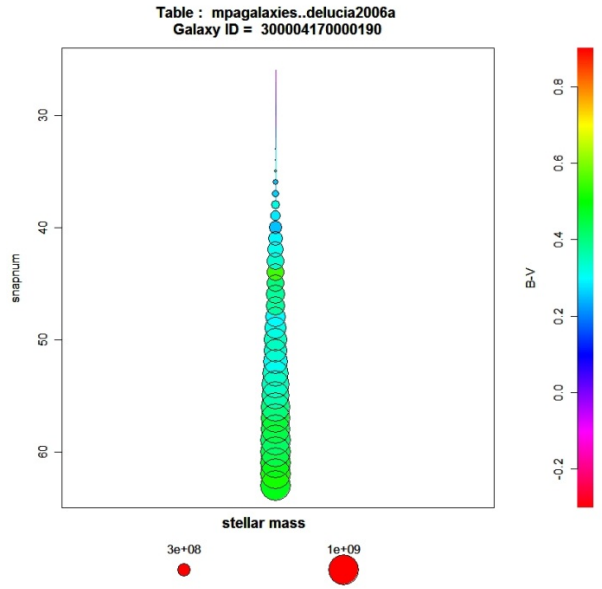


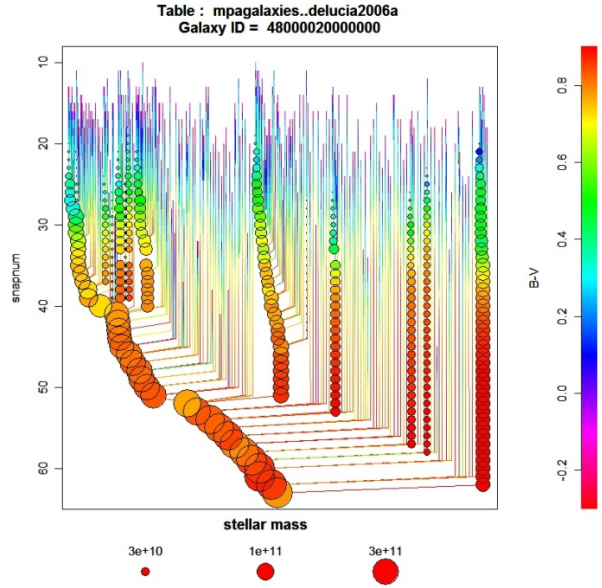
Time

Max-Planck-Institut für Astrophysik

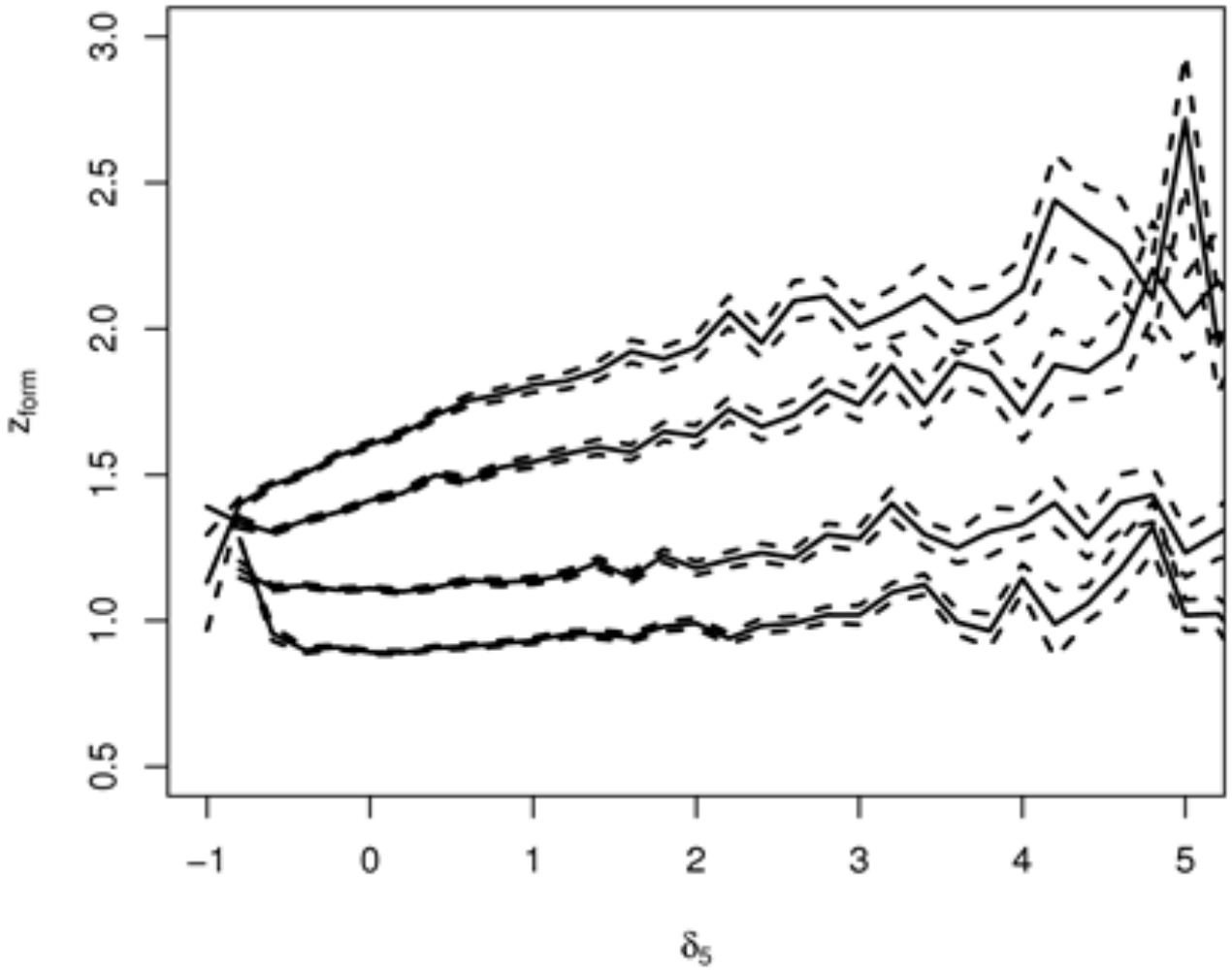
Galaxies







HO-1: reproduce halo assembly bias, question 14

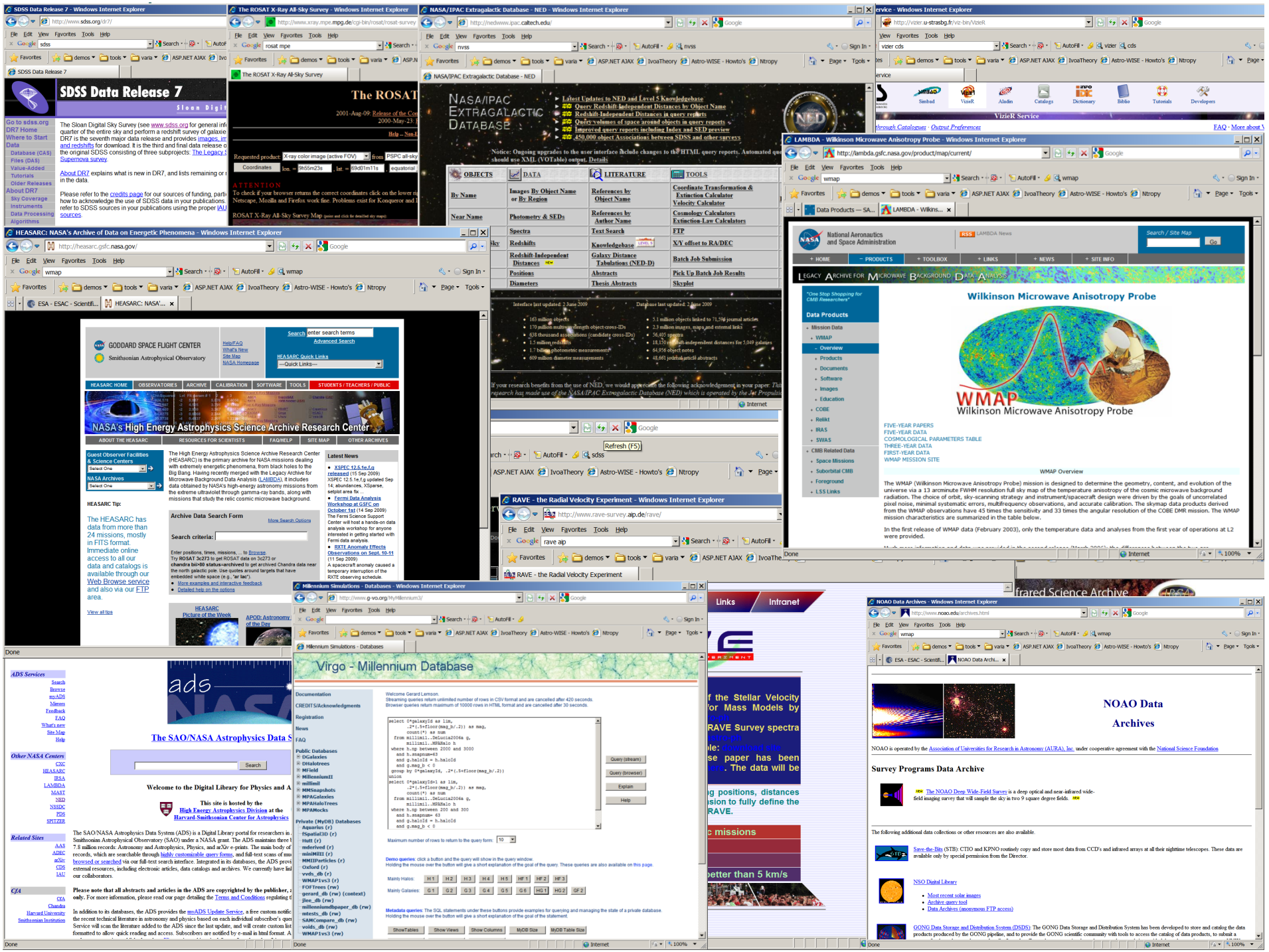


Back to Matt's categorization of questions.

- What are the hard questions in our approach?
  - SQL does not support them though data does.
  - Solution: download lots of our data, write your own code.
  - Ask DB managers to add more functions to your DB. E.g. Spatial3D, many @JHU
- What are impossible questions?
  - Not supported by our data.
  - Solution:
    - 1. create your own data (L-Galaxies online, light-cones online etc.)
    - 2. **Find it elsewhere!**

Max-Planck-Institut für Astrophysik

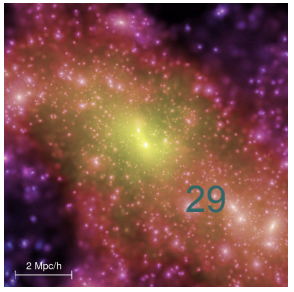
The Virtual Observatory (VO, VObs): motivation, approach, results


SDSS Data Release 7
The ROSAT X-Ray All-Sky Survey
NASA/IPAC Extragalactic Database
VizieR Service
LAMBDA - Wilkinson Microwave Anisotropy Probe
HEASARC: NASA's Archive of Data on Energetic Phenomena
RAVE - the Radial Velocity Experiment
Millennium Simulations - Databases
Virgo - Millennium Database
NOAO Data Archives
ads

Internet as telescope

- It has data on every part of the sky
- In every measured spectral band: optical, x-ray, radio..
- As deep as the best instruments (2 years ago).
- It is up when you are up
- It's a smart telescope: links objects and data to literature on them
- It even contains truly virtual data

(courtesy Alex Szalay)

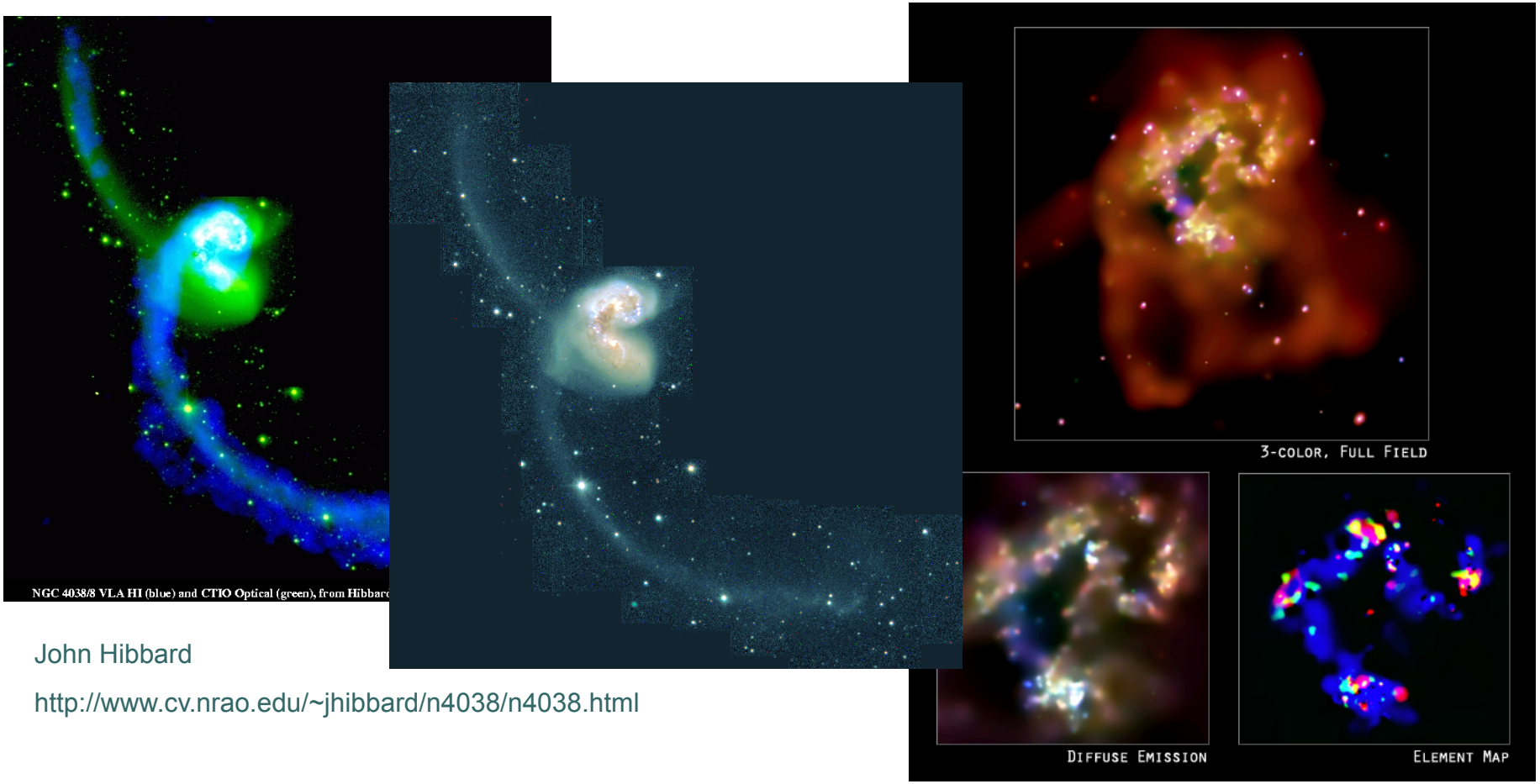


A multi-wavelength telescope

Radio

Optical

X-Ray



John Hibbard

http://www.cv.nrao.edu/~jhibbard/n4038/n4038.html

NASA/CXC/SAO/G. Fabbiano et al.

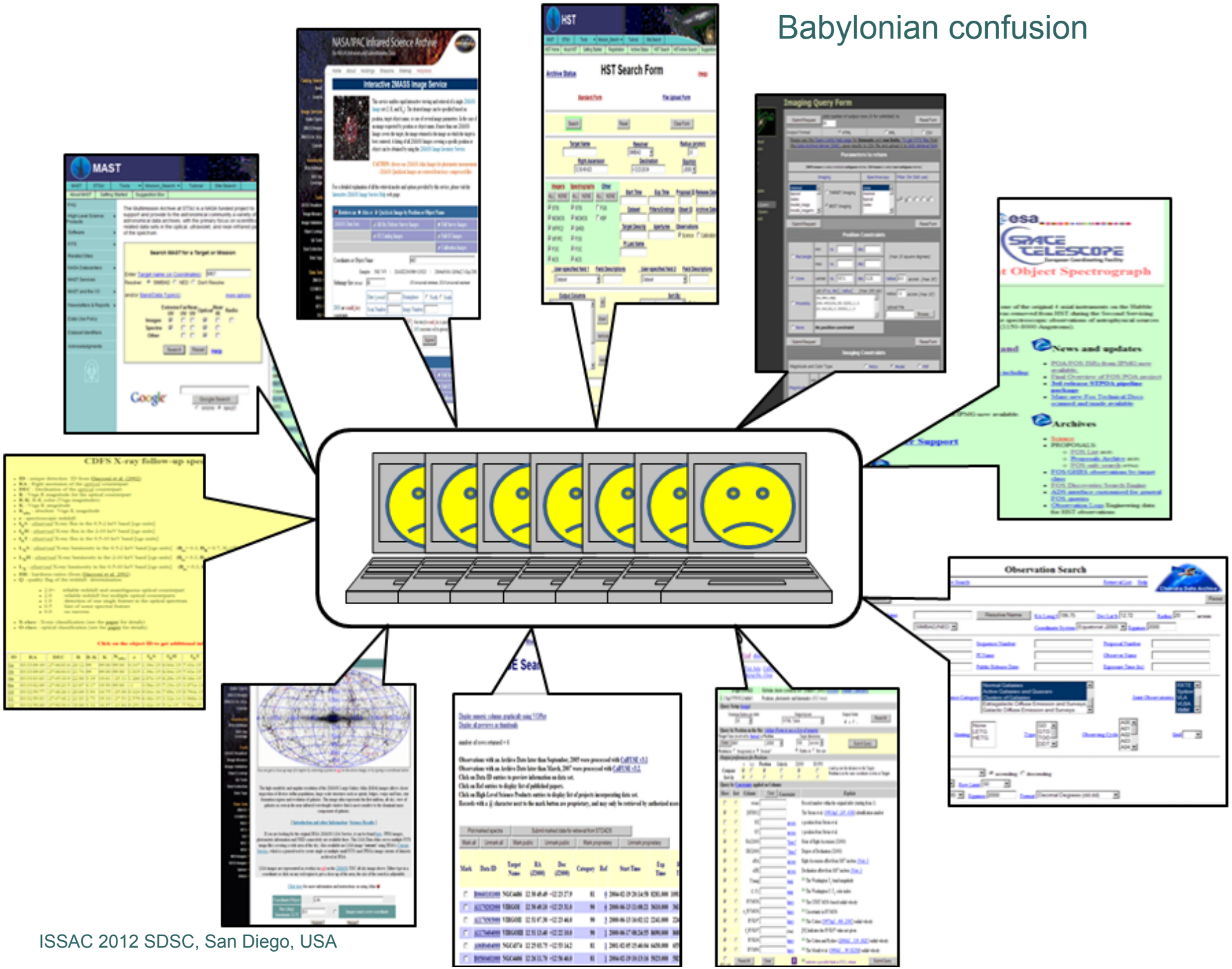
Virtual Observatory

Aims to facilitate access to online astronomical resources by *standardizing*:

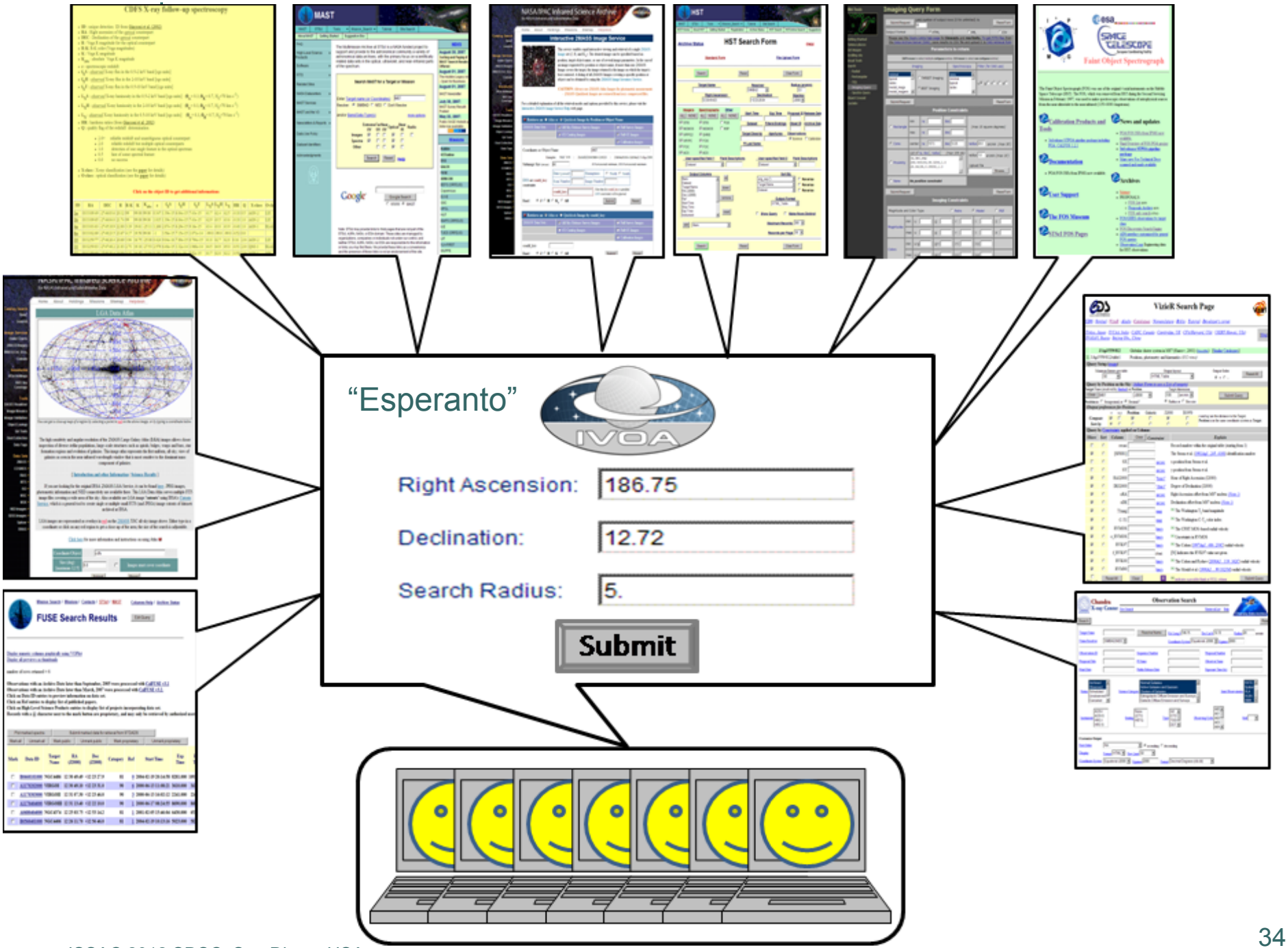
- Publication and Discovery
- Description/meta-data
- Selection/Retrieval
- Data formats
- Usage/value-added-services

Why standardization?

Babylonian confusion



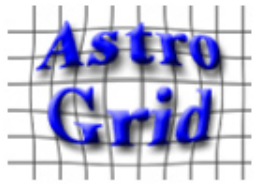
Max-Planck-Institut für Astrophysik

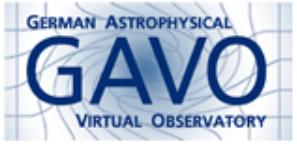

"Esperanto"
IVOA
Right Ascension:
186.75
Declination:
12.72
Search Radius:
5.
Submit
MAST
HST Search Form
Faint Object Spectrograph
FUSE Search Results
VizieR Search Page
Observation Search

The International Virtual Observatory Alliance (IVOA)

Facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory.

Current IVOA members





Working and Interest Groups

- WGs
  - **Standards and Pro**cess: how the IVOA works
  - **VOTable**: standard format for tabular data sets
  - **Semantics**: how to understand one another
  - **Data Access Layer**: very simple data access services
  - **Resource registry**: where to register and discover resources
  - **Applications**: stand alone, and together
  - **Data Modeling**: how to describe data sets
  - **VO Query Language**: more sophisticated data access
  - **Grid and web services**: programmatic accessibility
  - **VOEvent**: astronomical telegrams in XML
- IGs
  - **Theory**: virtual observations for virtual universes
  - **Data Curation and Preservation**: how not to loose your data
  - **Knowledge Discovery in Databases**: data mining algorithms

Warning up front

- VO can not (and does not aim to) be everything to everyone
- Users will have to be able to visit the underlying data in all gory detail: *provenance*
- Even then standardisation helps
- Agreement is hard to come by: politics (see FITS)
- Problems are hard !
- VO is a research project.

Data Access Protocols

- Simple protocols for discovering and retrieving data sets
  - Source catalogues
  - Images
  - spectra
- Query on
  - position on sky
  - observation time
  - wavelength range
- Return Formats
  - VOTable
  - FITS
- Recent:
  - Table Access Protocol (below more)
  - ObsTAP

VO's esperanto

SSA

POS: 186.75,12.72

SIZE: 5.

BAND: /

TIME: /

SCS

Right Ascension: 186.75

Declination: 12.72

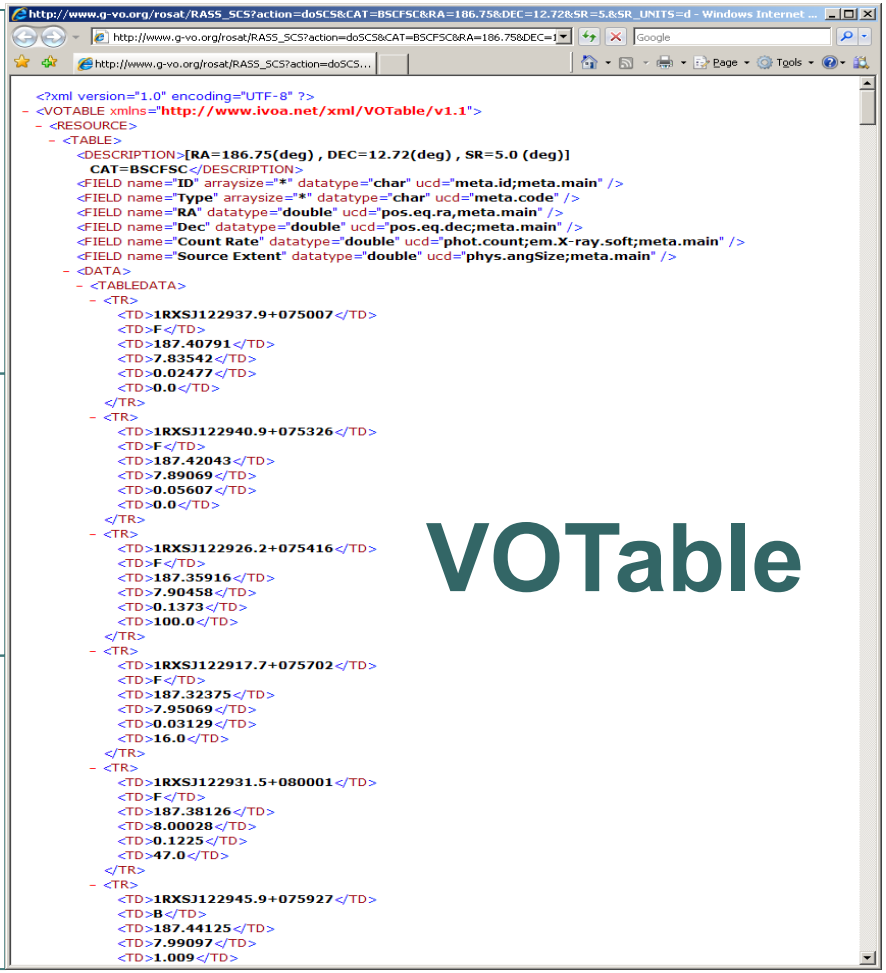
Search Radius: 5.

SIA

POS : 186.75,12.72

SIZE : 1.,1.

INTERSECT : OVERLAPS



VOTable


http://www.g-vo.org/rosat/RASS_SCS?action=doSCS&CAT=BSCFSC&RA=186.75&DEC=1
Google
http://www.g-vo.org/rosat/RASS_SCS?action=doSCS...
Page
Tools
<?xml version="1.0" encoding="UTF-8" ?>
<VOTABLE xmlns="http://www.ivoa.net/xml/VOTable/v1.1">
<RESOURCE>
<TABLE>
<DESCRIPTION>[RA=186.75(deg) , DEC=12.72(deg) , SR=5.0 (deg)]
CAT=BSCFSC</DESCRIPTION>
<FIELD name="ID" arraysize="*" datatype="char" ucd="meta.id;meta.main" />
<FIELD name="Type" arraysize="*" datatype="char" ucd="meta.code" />
<FIELD name="RA" datatype="double" ucd="pos.eq.ra,meta.main" />
<FIELD name="Dec" datatype="double" ucd="pos.eq.dec;meta.main" />
<FIELD name="Count Rate" datatype="double" ucd="phot.count;em.X-ray.soft;meta.main" />
<FIELD name="Source Extent" datatype="double" ucd="phys.angSize;meta.main" />
<DATA>
<TABLEDATA>
<TR>
<TD>1RXSJ122937.9+075007</TD>
<TD>F</TD>
<TD>187.40791</TD>
<TD>7.83542</TD>
<TD>0.02477</TD>
<TD>0.0</TD>
</TR>
<TR>
<TD>1RXSJ122940.9+075326</TD>
<TD>F</TD>
<TD>187.42043</TD>
<TD>7.89069</TD>
<TD>0.05607</TD>
<TD>0.0</TD>
</TR>
<TR>
<TD>1RXSJ122926.2+075416</TD>
<TD>F</TD>
<TD>187.35916</TD>
<TD>7.90458</TD>
<TD>0.1373</TD>
<TD>100.0</TD>
</TR>
<TR>
<TD>1RXSJ122917.7+075702</TD>
<TD>F</TD>
<TD>187.32375</TD>
<TD>7.95069</TD>
<TD>0.03129</TD>
<TD>16.0</TD>
</TR>
<TR>
VOTable

Messaging standard: VOTable

- http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOTable
- XML format for tabular data:

```
<VOTABLE>
 <RESOURCE>
  <TABLE>
    <FIELD name="ra" datatype="float"
           ucd="pos.eq.ra"/>
    <FIELD name="dec" datatype="float"
           ucd="pos.eq.dec"/>
    <DATA>
      <TABLEDATA>
        <TR><TD>123</TD>
            <TD>-45</TD>
        </TR>
         .....
```

Discovery: Resource Registry

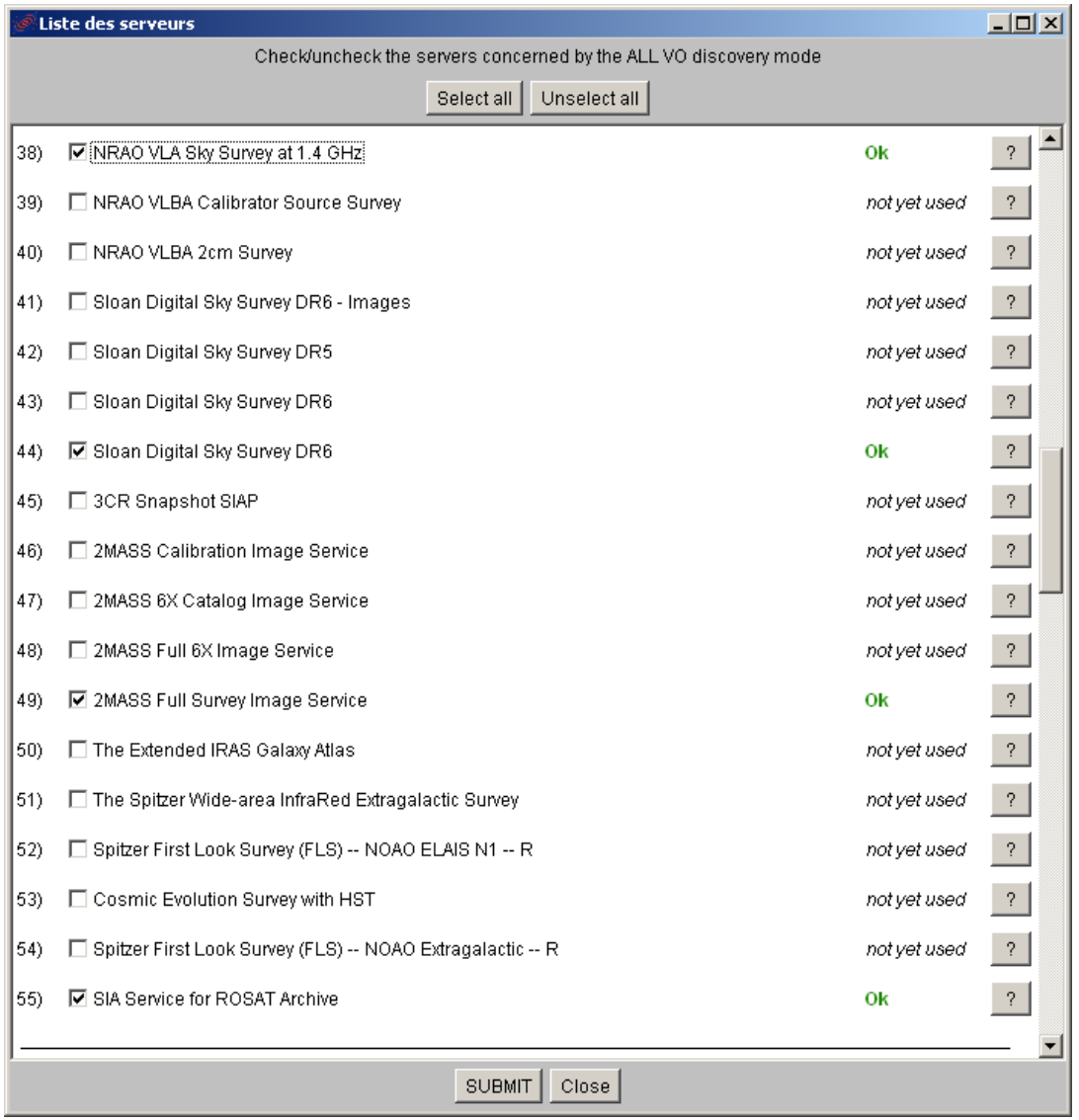
- Database containing descriptions of online *Resources*
  - data sets
  - protocol implementations
  - web applications
  - *anything that can be identified*
- XML schema for describing these
- Implementations:
  - VAO Searchable Registry at STScI
  - AstroGrid
  - GAVO
- Registry aware client tools:
  - VOExplorer (registry browser)
  - Splat, SpecView, Aladin, TOPCAT,…

Standardization facilitates interoperability

- VO aware tools:
  - Images: Aladin
  - Source lists and tables: TOPCAT, VOPlot
  - Spectra: Splat, SpecView, VOSpec
  - 3D, simulations: VisIVO
- Application interoperability: SAMP
  - Messaging standard
  - Tying TOPCAT to Aladin to Splat to …
  - Uses VOTable to send data from one app to another
  - All on your desktop
  - Even from browser (HO-1) !

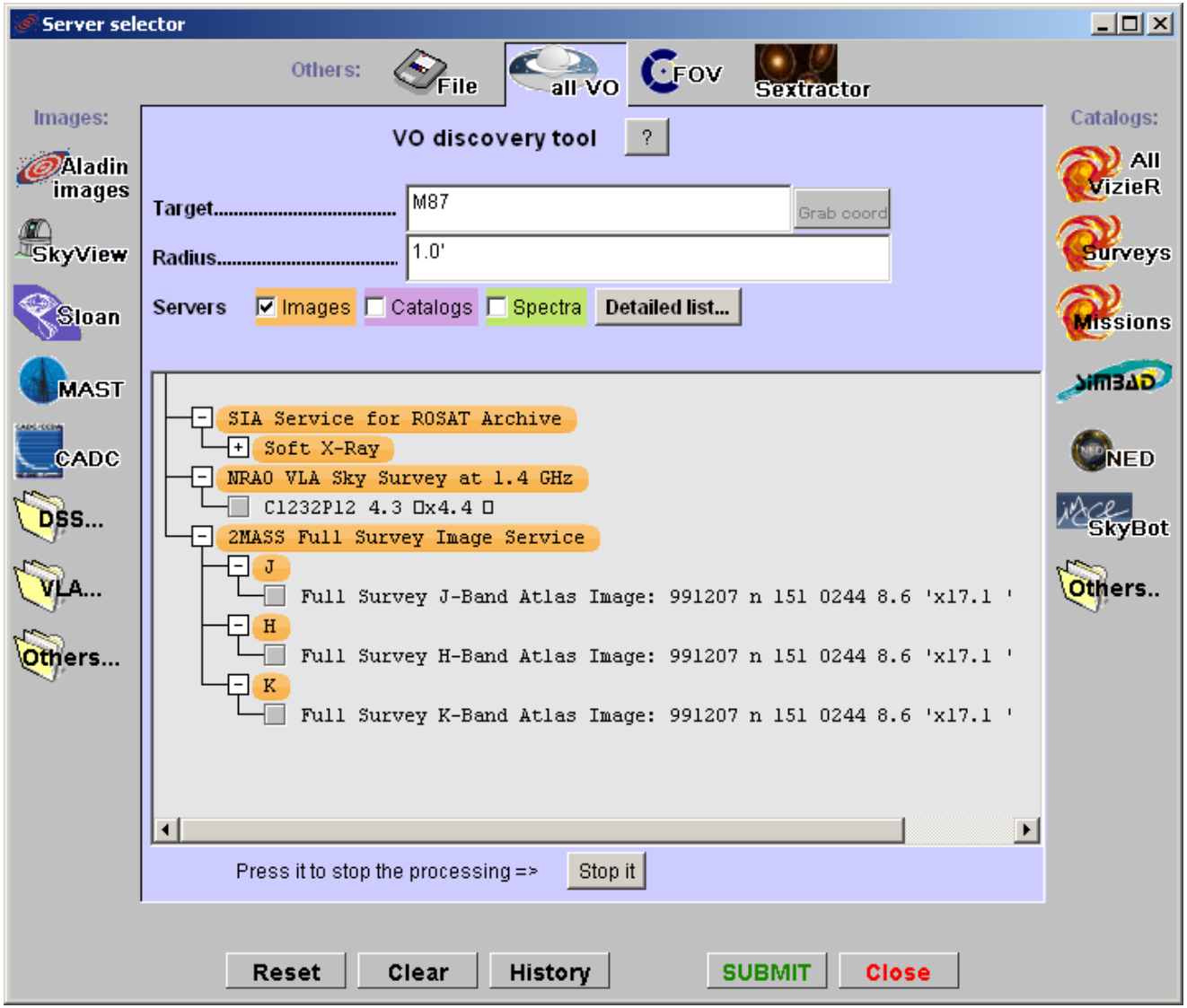
Registry + DAL protocols: Interoperability

Standard services, once registered, can be found by client tools …



Interoperability

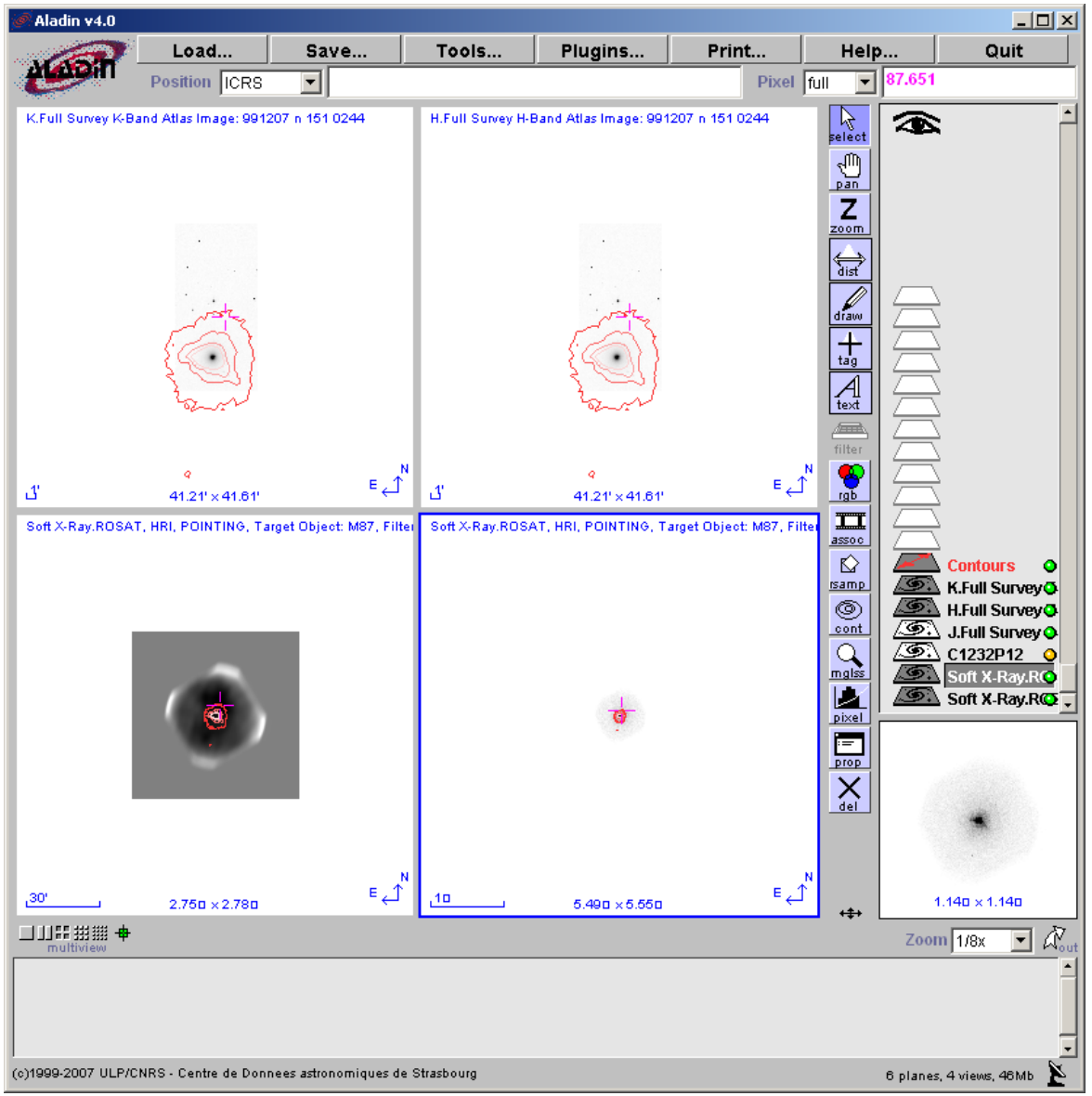
….and executed together (too many ROSAT results to show all here!)



Max-Planck-Institut für Astrophysik

Interoperability

… and shown together



Special: Theory in the VO

Observations in the VO

- Most VO efforts concentrate on observational data sets
  - simple observables: photons detected at a certain time from a certain area on the sky
  - long history of archiving
  - pre-existing standards (FITS)
  - valuable over long time (digitising 80 yr old plates)
- Standards observationally biased
  - common sky: cone search, SIAP, region
  - common objects: XMatch
  - data models: characterisation of sky/time/energy(/no polarisation yet)

Theory in the VO: issues

- Simulations not so simple
  - complex observables
  - no standardisation (not even HDF5)
  - archiving ad hoc, for local use
- Current IVOA standards somewhat irrelevant
  - no common sky
  - no common objects
  - requires data models for content, physics, code
- Moore's law makes useful lifetime relatively short: few years later can do better

History of simulations

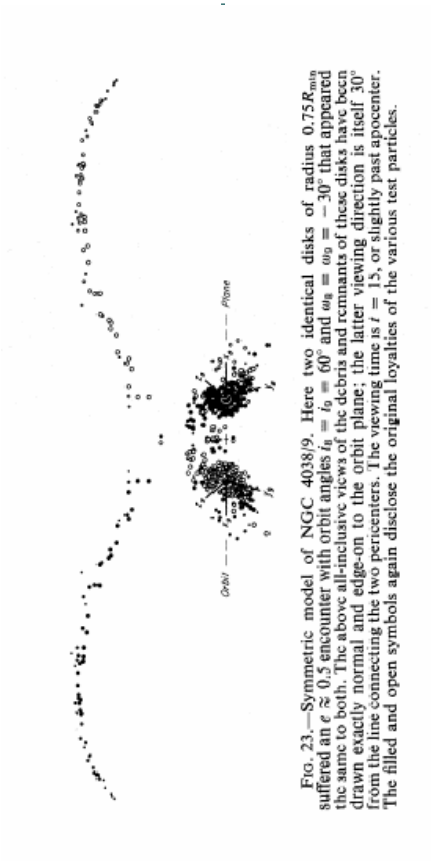
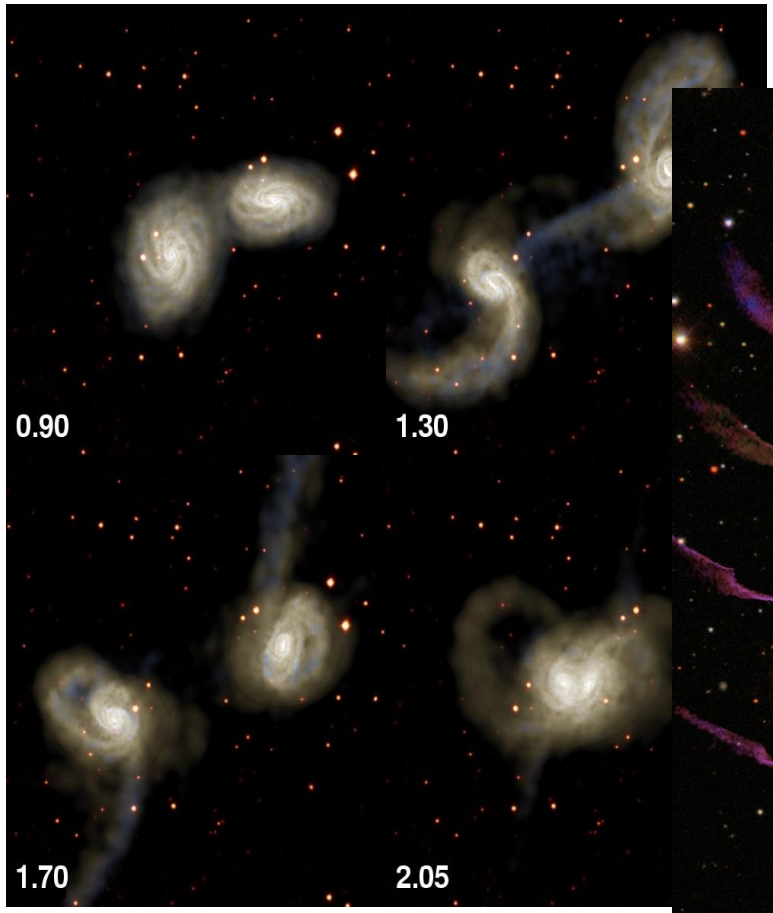
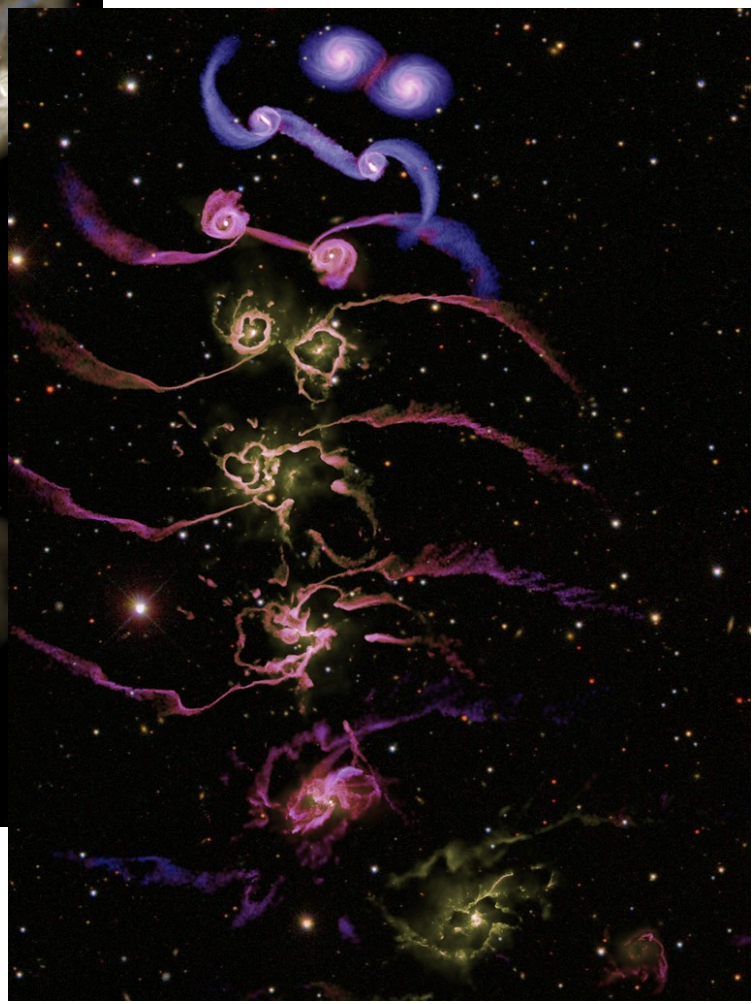


FIG. 23.—Symmetric model of NGC 4038/9. Here two identical disks of radius $0.75R_{min}$ suffered an $e \approx 0.5$ encounter with orbit angles $i_0 = i_0 = 60°$ and $\omega_0 = \omega_0 = -30°$ that appeared the same to both. The above all-inclusive views of the debris and remnants of these disks have been drawn exactly normal and edge-on to the orbit plane; the latter viewing direction is itself 30° from the line connecting the two pericenters. The viewing time is $t = 15$, or slightly past apocenter. The filled and open symbols again disclose the original loyalties of the various test particles.

Toomre & Toomre, 1972



Courtesy Volker Springel

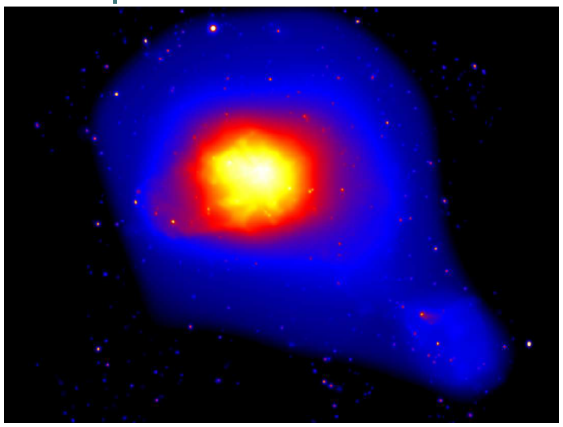


Di Matteo, Springel
and Hernquist, 2005

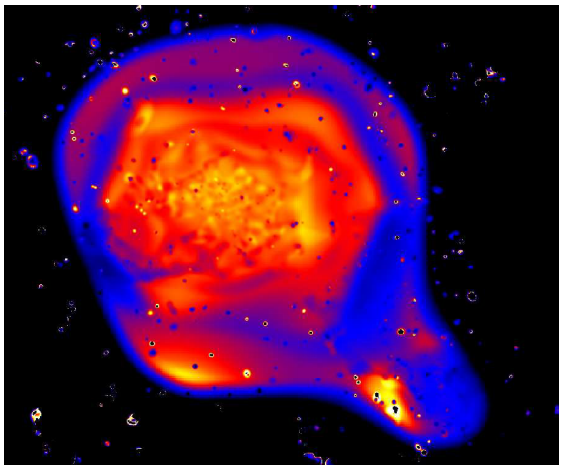
So why bother publishing simulations?

- Simulations are interesting:
  - For many cases only way to see processes in action
  - **Complex observations require sophisticated models for interpretation**
- Bridging gap in specializations: not everyone has required expertise to *create* simulations, though they *can analyze* them.
  - Persistent reference data sets
- Many use cases do *not* require the latest/greatest
  - Exposure time calculator
  - Survey design

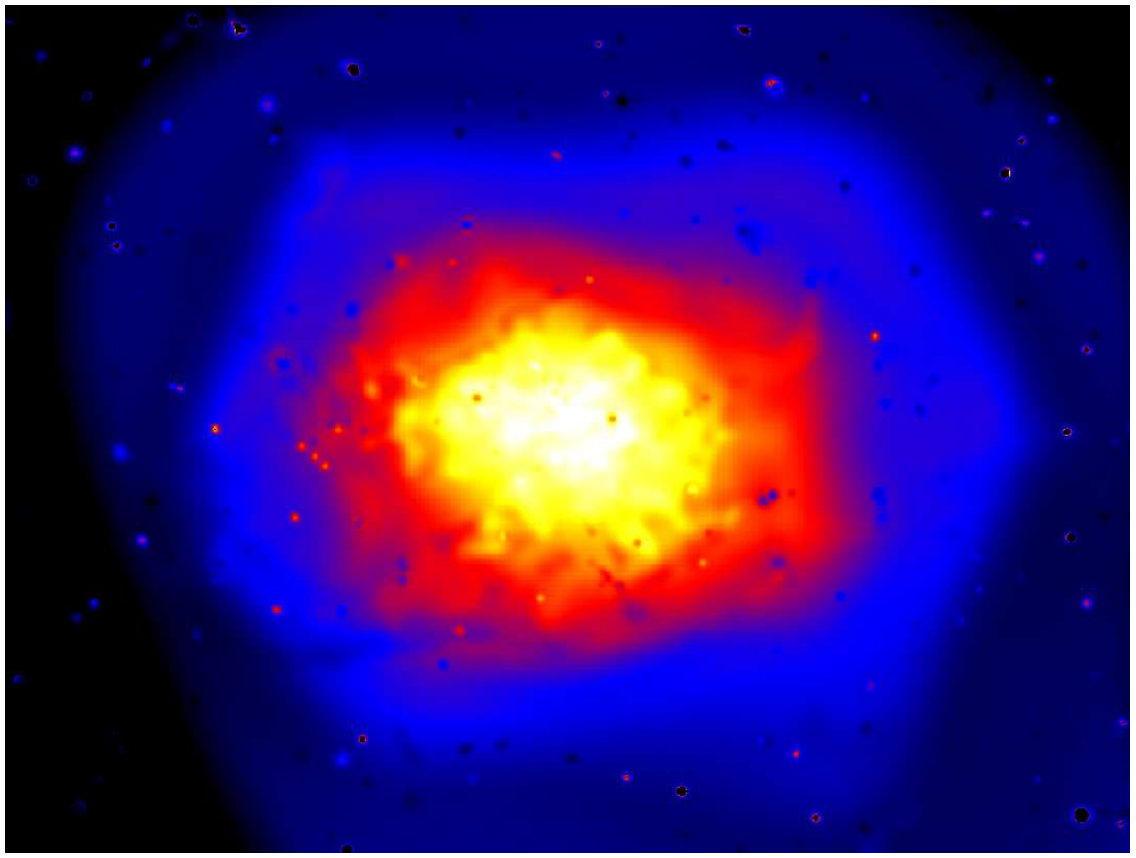
Detailed observations



electron density



gas temperature

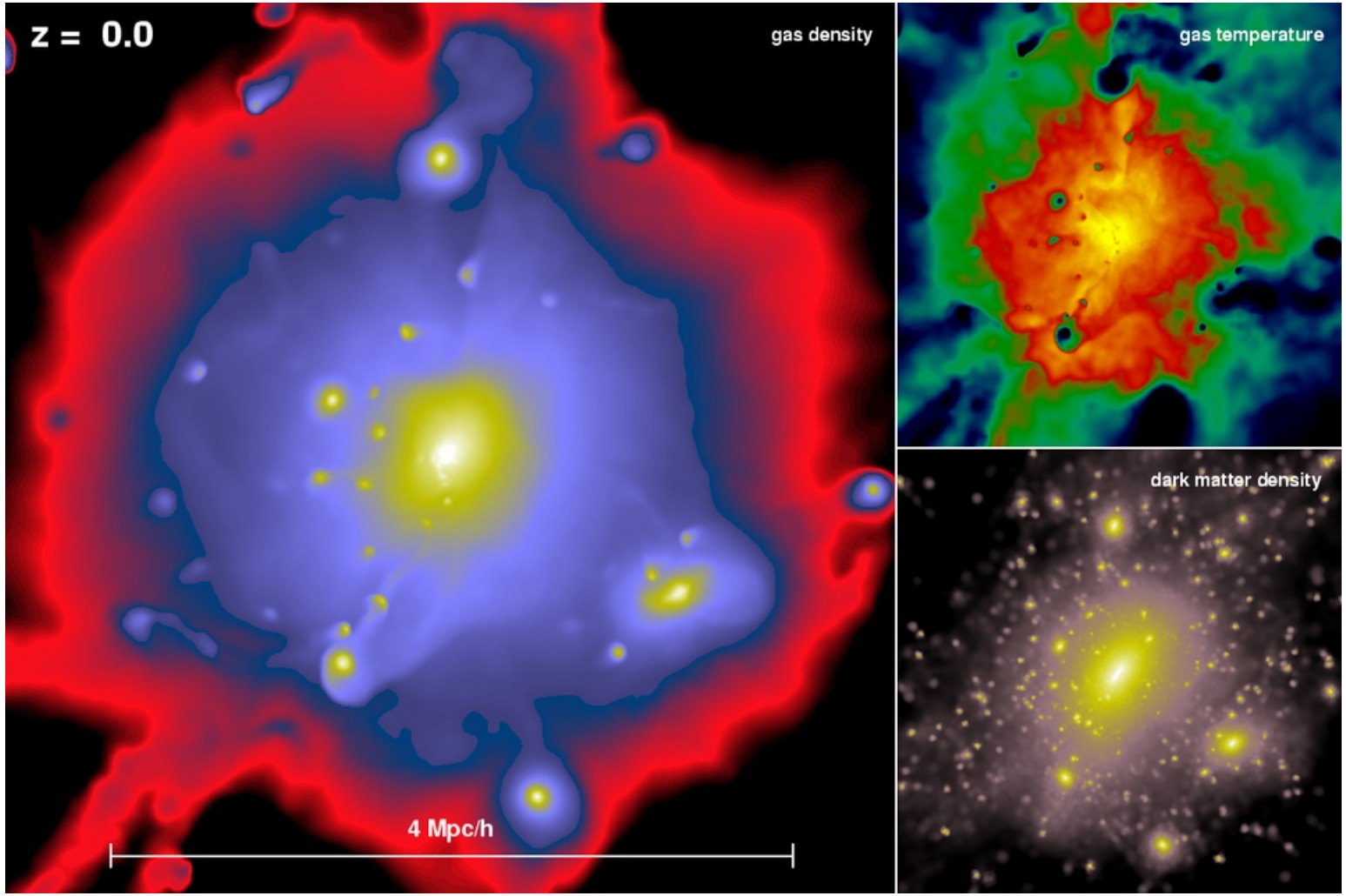


gas pressure

Courtesy Alexis Finoguenov, Ulrich Briel, Peter Schuecker, (MPE)

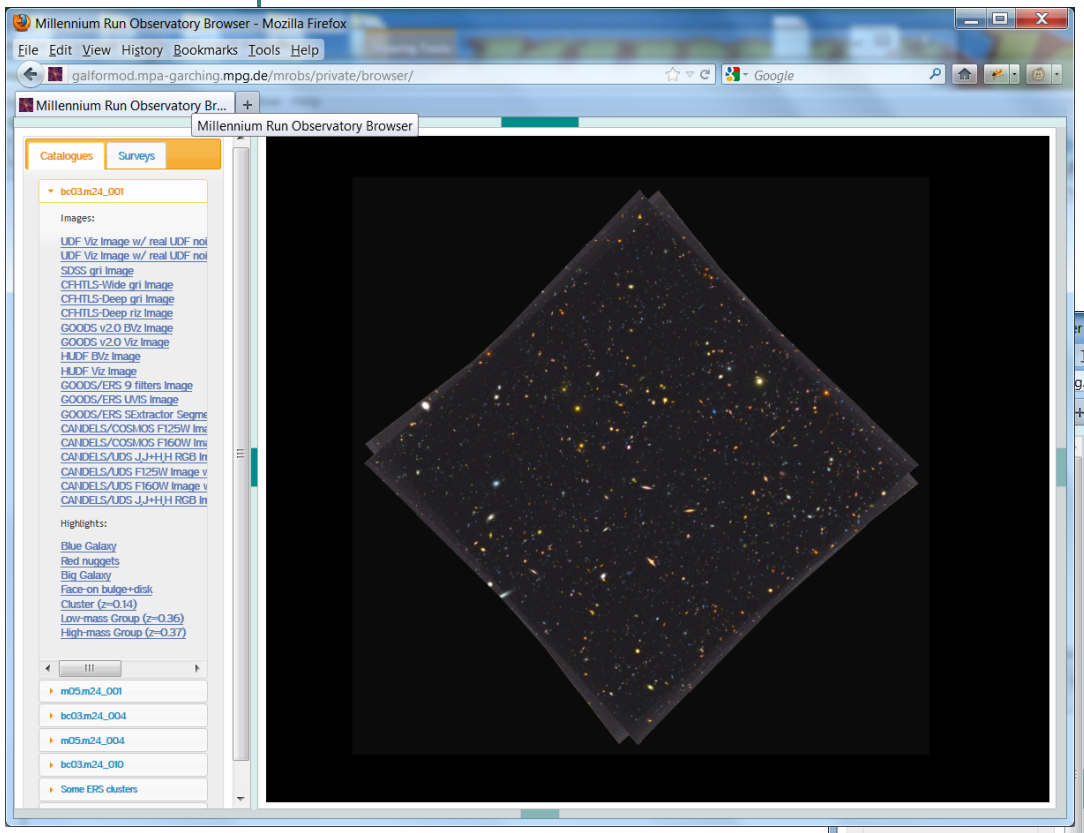
Max-Planck-Institut für Astrophysik

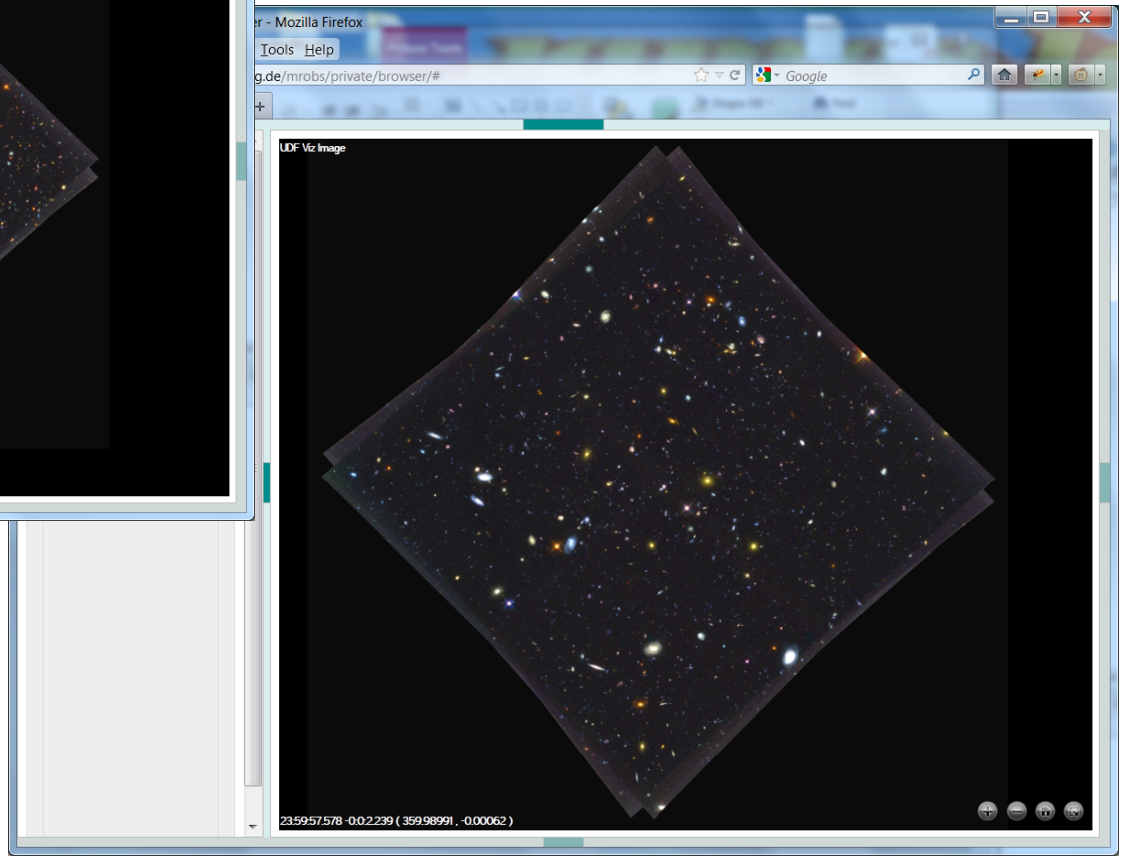
Detailed models



Courtesy Volker Springel

MRObs example: UDF





So why bother publishing simulations?

- Simulations are interesting:
  - For many cases only way to see processes in action
  - Complex observations require sophisticated models for interpretation
- Bridging gap in specializations: not everyone has required expertise to *create* simulations, but they *can analyze* them.
  - Persistent reference data sets
- Many use cases do *not* require the latest/greatest
  - Exposure time calculator
  - Survey design

Theory in the VO

- Theory interest group
- Simulation Data Model
  - Registry of simulations (under construction) http://galformod.mpa-garching.mpg.de/dev/SimDM-browser/
  - Maybe used in HO-2
- Simulation Data Access Layer
  - In progress
  - Role for yt?
- Ad hoc services always welcome
  - Millennium Run Database
  - Planck simulator
- Useful standards: TAP, UWS
  - MillenniumTAP
  - L-Galaxies online (under construction, maybe HO-2)

Table Access Protocol: TAP

- How to publish data in a relational database
- Defines protocol for
  - Retrieving metadata about database
    - TAP_SCHEMA
      - schemas
      - tables
      - columns
      - foreign keys
  - Sending queries to the database
    - Query language (ADQL-2.0)
    - **sync** and **async**
    - Uploading data (TAP_UPLOAD)
    - Execution parameters
  - Retrieving results
    - Formats

Example: ISSACTAP

- http://ion-21-11.sdsc.edu/issactap
- Metadata
  - http://ion-21-11.sdsc.edu/issactap/tables
- QUERYING
  - http://ion-21-11.sdsc.edu/issactap/sync?
    REQUEST=doQuery&
    LANG=SQL&
    QUERY=SELECT * FROM millimil.MPAHalo WHERE snapnum=63 AND
    np BETWEEN 100 AND 1000 AND x BETWEEN 10 AND 12&
    FORMAT=votable
- TOPCAT as TAP client tool (demo)
- More in hands-on sessions
  - this afternoon 4PM
  - Thu. 4PM

Hands-on session

- HO-1: getting familiar with database access tools and SQL
- HO-2: publishing data
- Usernames/passwords will be mailed to you

THANKS TO THE ORGANIZERS AND THANK YOU.

Acknowledgment:

Thanks to Matthias Egger for building the TAP interface. GL and Matthias Egger are supported by Advanced Grant 246797 GALFORMOD from the European Research Council.