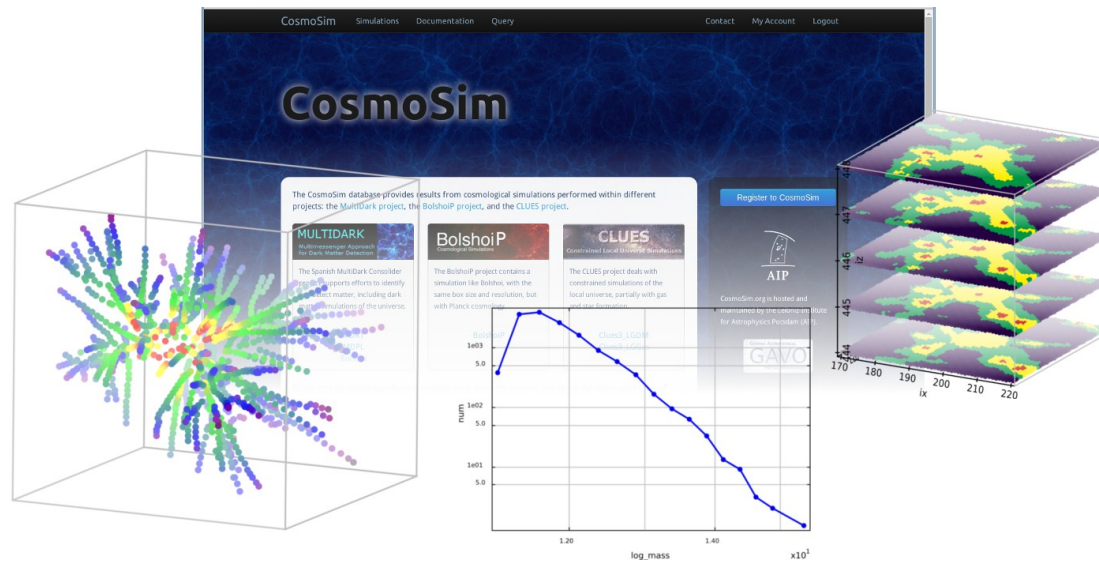


# CLUES data in the CosmoSim database



Kristin Riebe  
E-Science group @AIP

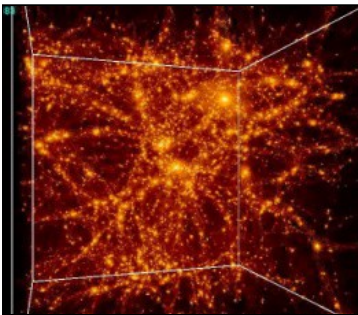


# Outline

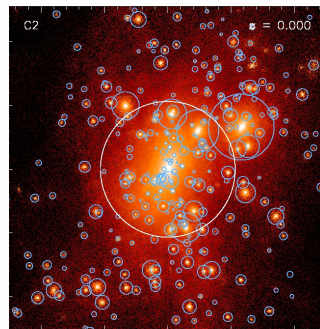
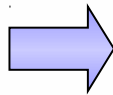
- General: Simulations and databases
- CosmoSim: database server, web interface
- Data products
- CLUES data
- Short demo
- Discussion

# Cosmological Simulations

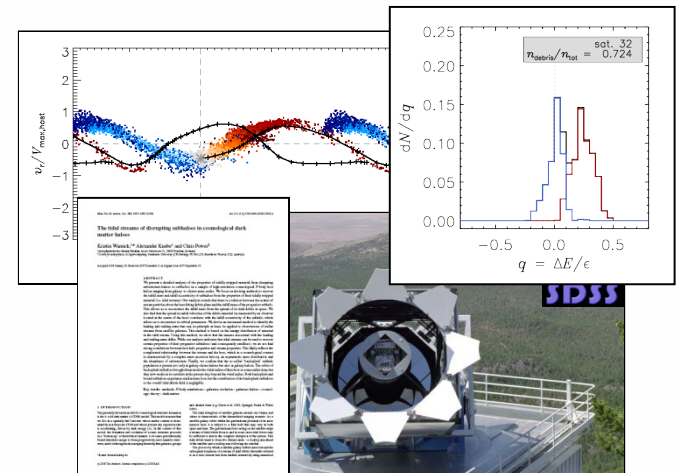
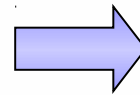
- Produce terabytes of data
- Post-processing:
  - Halo catalogues, merger trees, mock galaxy catalogues, ...
  - Variety of data formats
- How to share data within collaboration? How to publish?



Cosmological Simulation



Post-Processing  
(e.g. halo finding)



Scientific Output

# Databases

- Structured Query Language: SQL
- Uniform data formats
- Retrieve only subsets, results, **not** complete catalogues
- Examples: sort/filter halos, calculate mass functions, extract particles for halos, merger trees, substructures, ...

```
SELECT * FROM MDR1.FOF  
WHERE snapnum=85  
ORDER BY mass DESC LIMIT 10
```

10 most massive FOF  
groups at  $z=0$   
< 1 s

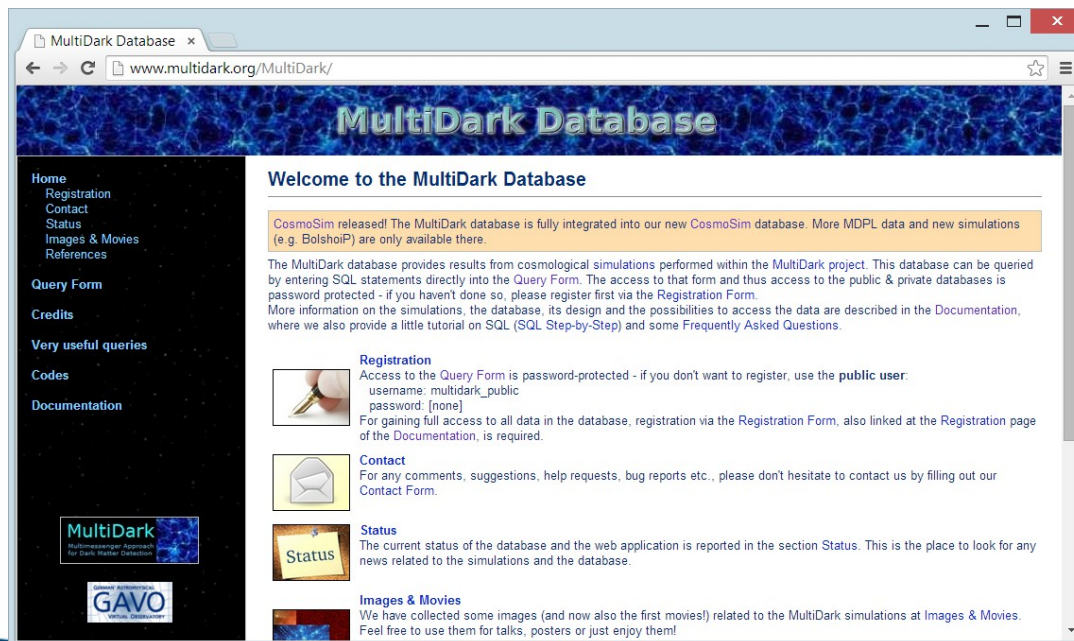
```
SELECT * FROM MDR1.FOF  
WHERE snapnum=85  
ORDER BY spin DESC LIMIT 50
```

50 FOF groups  
with biggest spin  
parameter  
~ 26 s

=> Just share the query, don't need to share the subset!

# Simulation Databases

- Millennium Run Database: very successful, hundreds of papers, still being used
- MultiDark Database: collaboration with Spanish MultiDark project, ~ 20 papers, also from people outside of collaboration



# Database server



- Previous setup: 1 Microsoft SQL server
  - expensive **license**, harder to share
  - **slow** retrieval times on **full table scans** (cannot have index for everything)
  - creating **index** on particle data ( $\sim 10^{10}$  particles) takes  $\sim 1$  week
- Solution: MyISAM+Spider with MariaDB
  - use MyISAM engine of MySQL/MariaDB
  - Spider engine (Kentoku Shiba) for distributed queries
    - => **data distributed** over 10 nodes, queries much faster!
  - **open source**
  - own developments in E-Science group:  
PaQu, QueryQueue, libhilbert, mysql\_sprng
    - => see <http://github.com/adrpar>





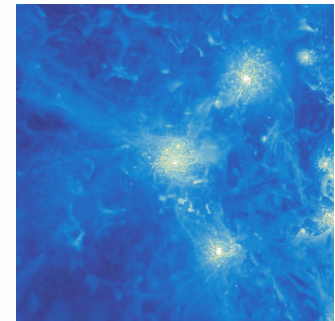
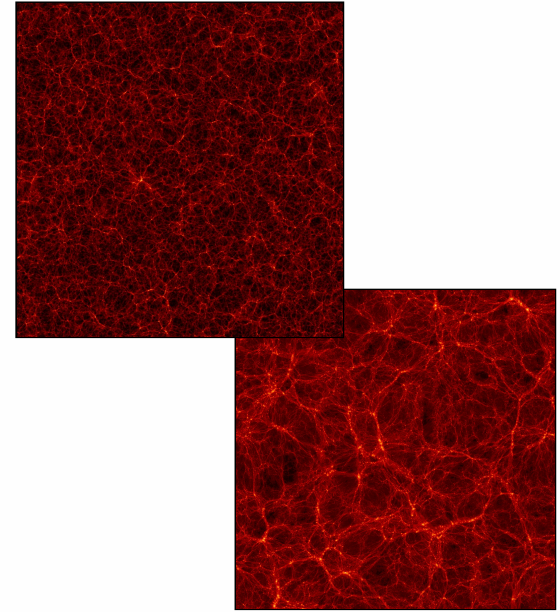
# CosmoSim

- new web interface:
  - Daiquiri web application, <http://escience.aip.de/daiquiri>
  - developed by J. Klar and A. Partl
  - modular, highly customizable
  - using PHP, Zend-framework
  - modern interface
  - authentication, query interface
  - wordpress integration
  - open source
  - also used for MUSIC, Jubilee and Curie databases in Madrid, project webpages, workshops (e.g. CLUES registration page!)



# Data at CosmoSim

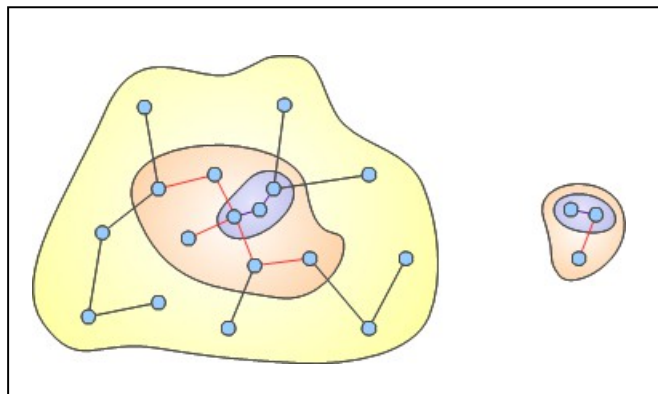
- 25 TB data, ~ 230 billion rows
- 6 simulations:
  - **MDR1**: WMAP5, 1000 Mpc/h,  $2048^3$  p.
  - **MDPL**: Planck1, 1000 Mpc/h,  $3840^3$  p.
  - **Bolshoi**: WMAP5, 250 Mpc/h,  $2048^3$  p.
  - **BolshoiP**: Planck1, 250 Mpc/h,  $2048^3$  p.
  - **2 CLUES** simulations:  
B64\_WM3\_186592, WMAP3:
    - Clues3\_LGDM: 2 Mpc in 64 Mpc/h box,  $4096^3$  p.
    - Clues3\_LGGas: same, but with gas+sfr
  - in preparation: SMDPL: Planck1, 400 Mpc/h,  $3840^3$  p.





# Data products

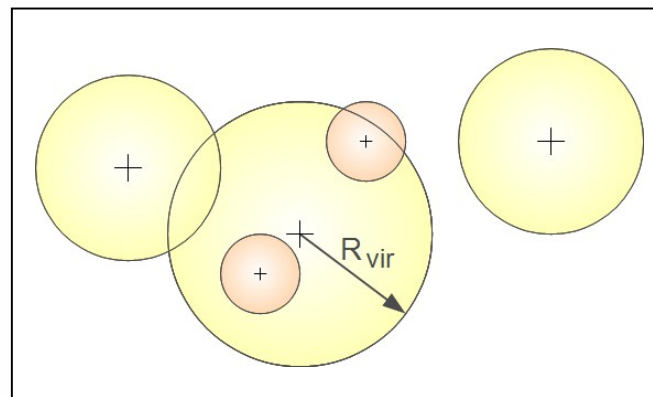
- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths



bdmId	snapnum	NInCat	hostFlag	x	y	z
8511186098	85	11186098	-1	548.8973	143.9528	895.7
8512166221	85	12166221	-1	384.92	468.0325	872.9
8512459068	85	12459068	-1	737.7123	482.5123	972.9
8505410295	85	5410295	-1	947.049	971.8178	267.4
8506742613	85	6742613	-1	276.4744	488.3719	548.7
8506941522	85	6941522	-1	665.9299	986.9252	815.9
8513458743	85	13458743	-1	665.9299	986.9252	815.9

unique identifier for  
each halo/group

1 halo/group per row



# Data products

- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths
- AHF/BDM halo profiles
  - inner structure of halos

bdmId	snapnum	NInCat	hostFlag	x	y	z
8511186098	85	11186098	-1	548.8973	143.9528	895.7
8512166221	85	12166221	-1	384.92	468.0325	872.9
8512459068	85	12459068	-1	737.7123	482.5123	972.9
8505410295	85	5410295	-1	947.049	971.8178	267.4
8506742613	85	6742613	-1	276.4744	488.3719	548.7
8506941522	85	6941522	-1	665.9299	986.9252	815.9
8513458743	85	13458743	-1	665.9299	986.9252	815.9

unique identifier for  
each halo/group

1 halo/group per row

bdmId	snapnum	NinCat	R_Rvir	Rbin	np	mass
8506742613	85	6742613	1.0629464	2828.978	10582	1.880109
8506742613	85	6742613	1.1953862	3174.1658	12201	1.985928
8506742613	85	6742613	1.3382614	3561.4724	11217	2.083212
8506742613	85	6742613	1.4941747	3996.0378	7967	2.152309
8506742613	85	6742613	1.6838918	4483.628	6245	2.206469
8506742613	85	6742613	1.8875381	5030.7134	7078	2.267854
8506742613	85	6742613	0.9454057	2521.329	10907	1.788326

distance from center

# Data products

- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths
- AHF/BDM halo profiles
  - inner structure of halos

bdmId	snapnum	NInCat	hostFlag	x	y	z
8511186098	85	11186098	-1	548.8973	143.9528	895.7
8512166221	85	12166221	-1	384.92	468.0325	872.9
8512459068	85	12459068	-1	737.7123	482.5123	972.9
8505410295	85	5410295	-1	947.049	971.8178	267.4
8506742613	85	6742613	-1	276.4744	488.3719	548.7
8506941522	85	6941522	-1	665.9299	986.9252	815.9
8513458743	85	13458743	-1	665.9299	986.9252	815.9

1 halo/group per row

unique identifier for each halo/group

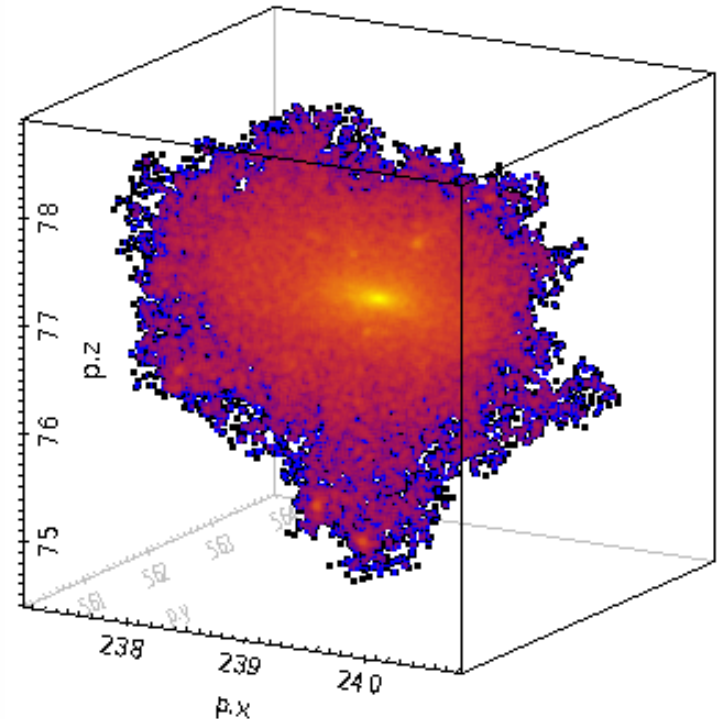
bdmId	snapnum	NInCat	R_Rvir	Rbin	np	mass
8506742613	85	6742613	1.0629464	2828.978	10582	1.880109
8506742613	85	6742613	1.1953862	3174.1658	12201	1.985928
8506742613	85	6742613	1.3382614	3561.4724	11217	2.083212
8506742613	85	6742613	1.4941747	3996.0378	7967	2.152309
8506742613	85	6742613	1.6838918	4483.628	6245	2.206469
8506742613	85	6742613	1.8875381	5030.7134	7078	2.267854
8506742613	85	6742613	0.9454057	2521.329	10907	1.788326

link to BDM halo

distance from center

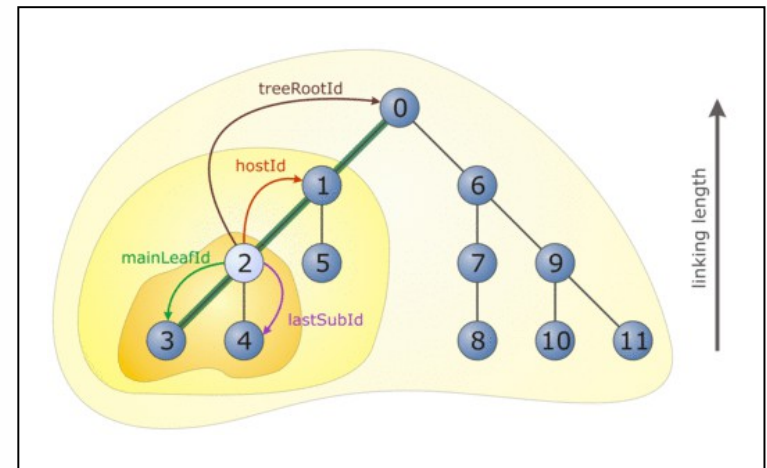
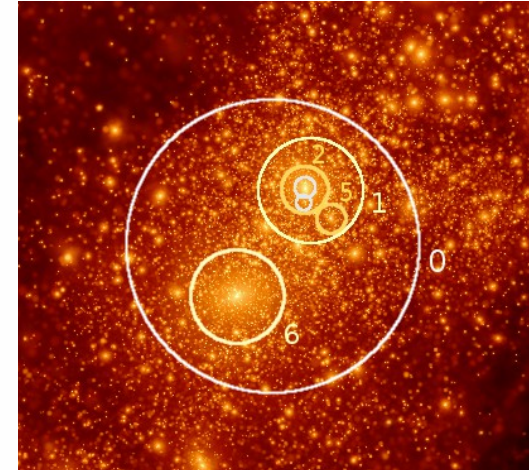
# Data products

- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths
- AHF/BDM halo profiles
  - inner structure of halos
- Particle snapshots
  - access to all particles
- FOF particles
  - link between halos and particles



# Data products

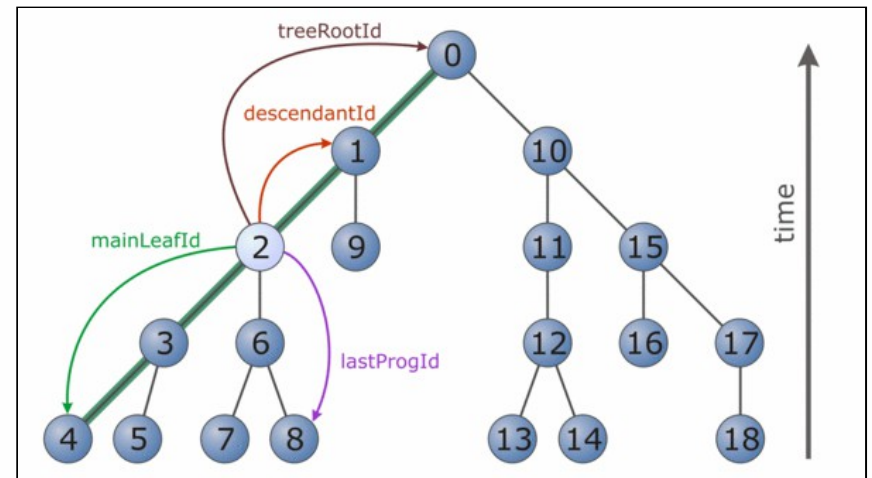
- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths
- AHF/BDM halo profiles
  - inner structure of halos
- Particle snapshots
  - access to all particles
- FOF particles
  - link between halos and particles
- FOF substructures, superclusters
  - hierarchy of structures





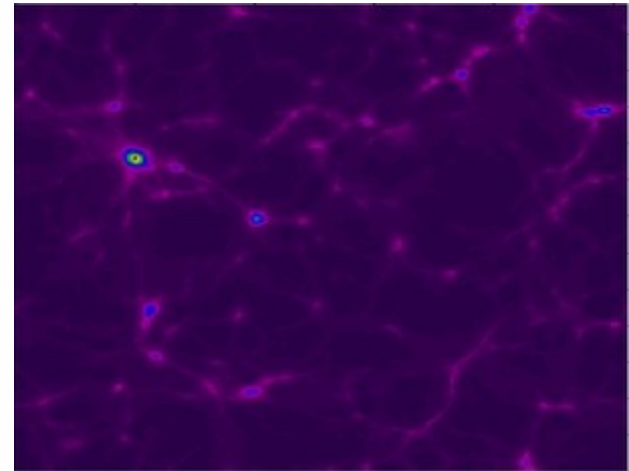
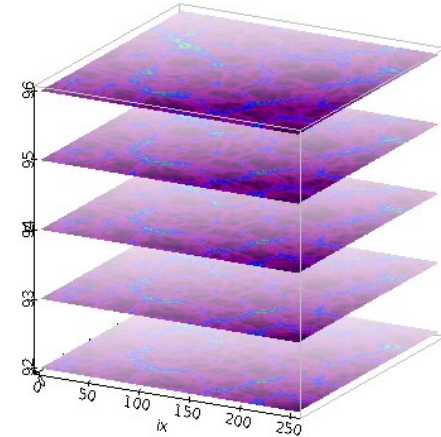
# Data products

- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths
- AHF/BDM halo profiles
  - inner structure of halos
- Particle snapshots
  - access to all particles
- FOF particles
  - link between halos and particles
- FOF substructures, superclusters
  - hierarchy of structures
- FOF merger trees
  - evolution of FOF groups

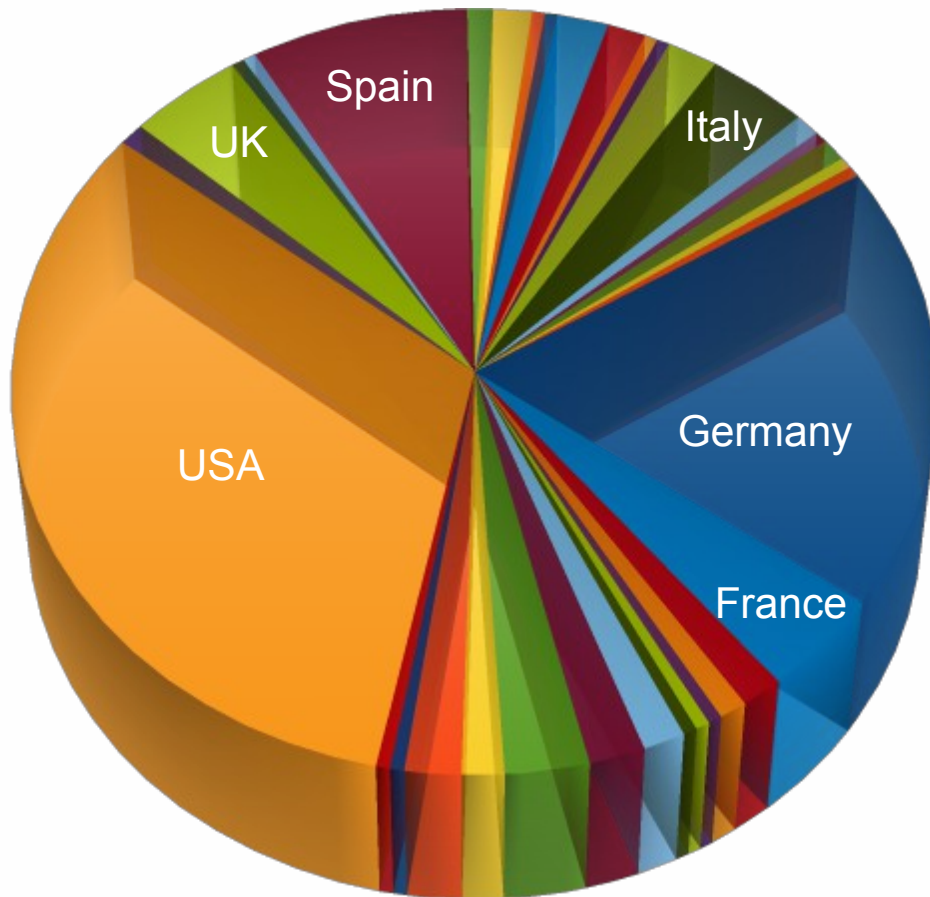


# Data products

- Halo catalogues
  - AHF, BDM, FOF
  - multiple density thresholds, linking lengths
- AHF/BDM halo profiles
  - inner structure of halos
- Particle snapshots
  - access to all particles
- FOF particles
  - link between halos and particles
- FOF substructures, superclusters
  - hierarchy of structures
- FOF merger trees
  - evolution of FOF groups
- density, cosmic web
  - large scale structure



# Database usage: users per country

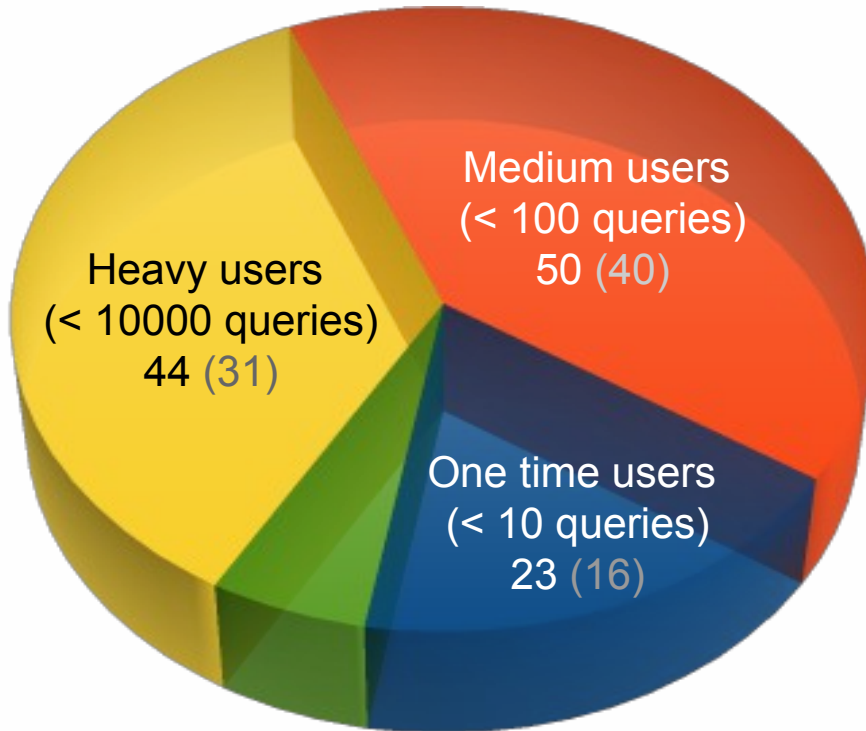


Users from 35 different countries

USA	69
Germany	40
Spain	17
France	9
UK	9
Italy	8

Argentina	Australia	Canada
Chile	China	Colombia
Czech Republic	Denmark	Ecuador
Egypt	Finland	France
Germany	Guatemala	Hong Kong
India	Iran	Israel
Italy	Japan	Korea
Malta	Mexico	Netherlands
Norway	Poland	Russia
Slovenia	Spain	Sweden
Taiwan	UK	Ukraine
USA	Venezuela	

# Queries per user (03/2014)



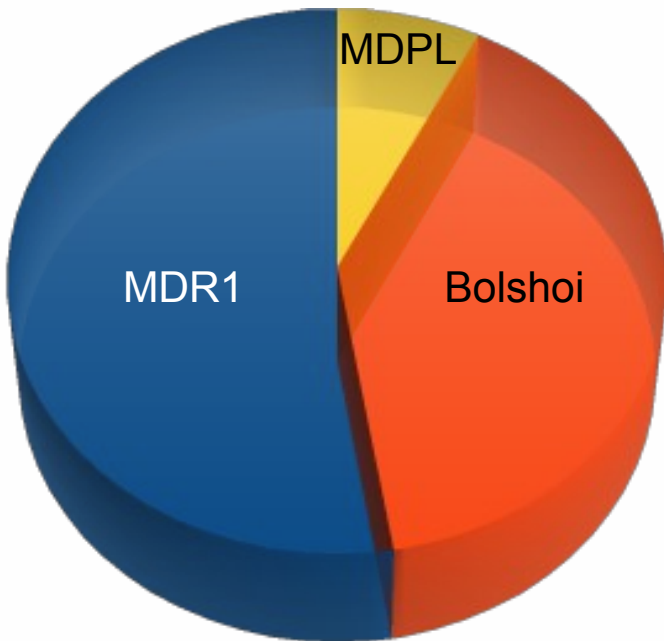
Batch users	> 10000 qry.	7
Heavy users	< 10000 qry.	44
Medium users	< 100 qry.	50
One time users	< 10 qry.	23

Total number of users: 124  
(+ 67 registered users with no queries)

Batch users  
(> 10000 queries)  
7 (4)

(Numbers in gray:  
from 04/2013)

# Most wanted simulations 08/2014



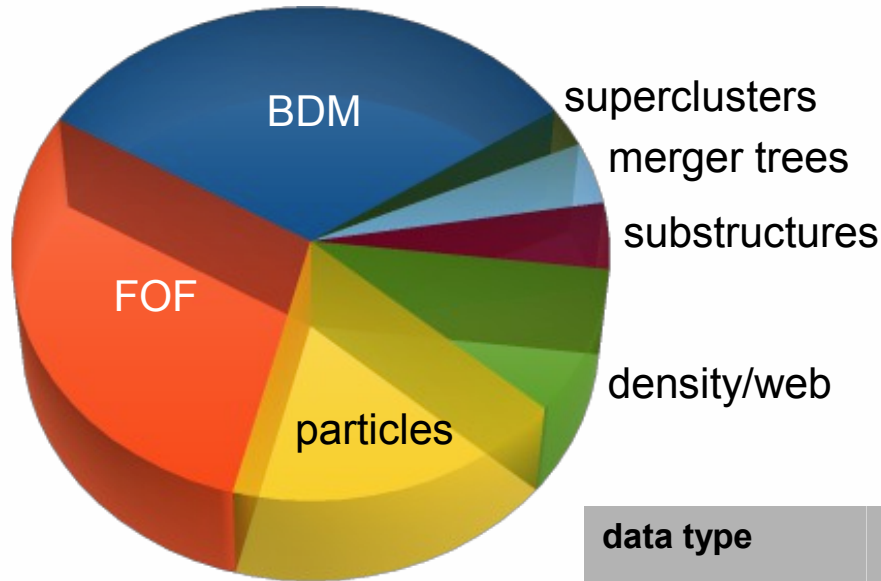
number of users per  
simulation

simulation	#users	#queries
MDR1	121	47658702
Bolshoi	92	16733420
MDPL	17	5175

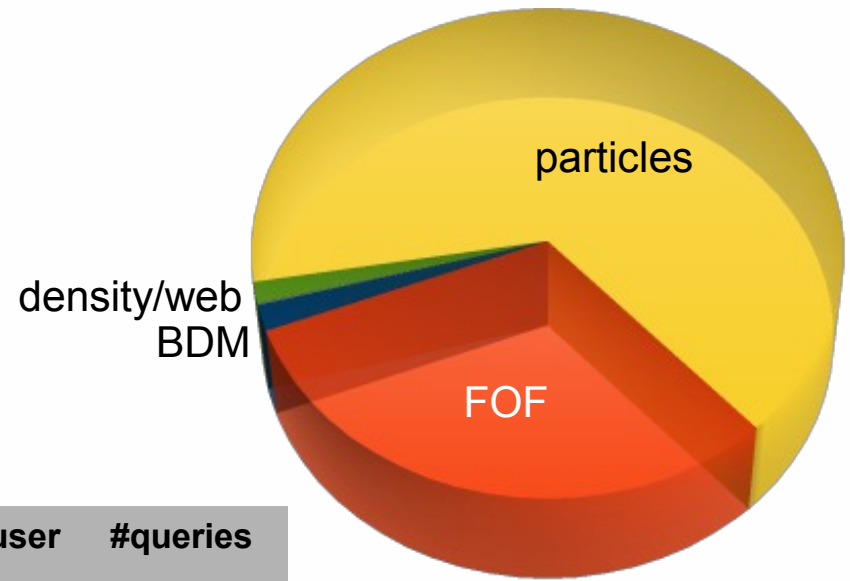


# Most wanted tables 08/2014

number of **users** per data type



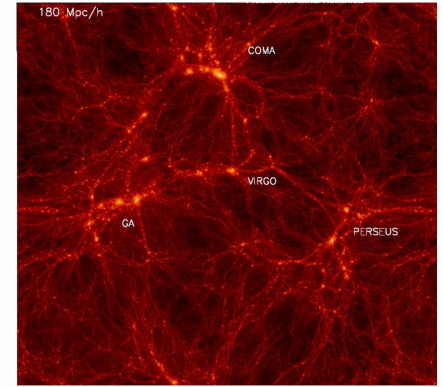
number of **queries** per data type



data type	#user s	#queries
BDM	112	45324
FOF	102	825134
particles	61	1751862
density/web	33	38524
substructures	15	151
mergertree	14	1317
superclusters	9	184

# CLUES data?

- first simulations now included in CosmoSim database
- AHF analysis rerun by Alexander Knebe, for consistent catalogues
- included in DB:
  - Clues3\_LGDM, Clues3\_LGGas
  - AHF, AHFGas, AHFStars: catalogues with halo properties
  - AHFProf: profiles of halos
  - AHFLum, AHFMag: luminosities, magnitudes for gas+sfr simulation
  - AHFMatch: links halos of gas and dm simulation



# Demo: querying data from CosmoSim



CosmoSim - Chromium

www.cosmosim.org

CosmoSim Simulations Documentation Query Contact My Account Logout

## CosmoSim

The CosmoSim database provides results from cosmological simulations performed within different projects: the [MultiDark project](#), the [BolshoiP project](#), and the [CLUES project](#).

**MULTIDARK**  
Multimessenger Approach for Dark Matter Detection

The Spanish MultiDark Consolider project supports efforts to identify and detect matter, including dark matter simulations of the universe.

MDR1  
MDPL  
Bolshoi

**BolshoiP**  
Cosmological Simulations

The BolshoiP project contains a simulation like Bolshoi, with the same box size and resolution, but with Planck cosmology.

BolshoiP

**CLUES**  
Constrained Local Universe Simulations

The CLUES project deals with constrained simulations of the local universe, partially with gas and star formation.

Clues3\_LGDM  
Clues3\_LGGas

Please visit the linked sites for more information about the projects and about the appreciated form of acknowledgment, if the data is used in a scientific publication or proposal. The MultiDark simulations MDR1 and MDPL as well as the Bolshoi simulation are also available via the [MultiDark database](#).

### Database access

The database can be queried by entering SQL statements directly into the [Query Form](#). If you haven't done so, please register first via the [Registration Form](#) to get your own private database where the results of your queries will be stored for you. You can also submit queries as a guest, but the result data can then be accessed and removed by any other guest as well.

**Register to CosmoSim**

**AIP**

CosmoSim.org is hosted and maintained by the Leibniz-Institute for Astrophysics Potsdam (AIP).

**GAVO**  
GERMAN ASTROPHYSICAL VIRTUAL OBSERVATORY

It is a contribution to the German Astrophysical Virtual Observatory.

The MultiDark and Bolshoi simulations were run on the NASA's Pleiades supercomputer at the NASA Ames Research Center.

**PRACE**



# Database access

- Web interface: <http://www.cosmosim.org>
- SAMP for sending results to VO clients
- UWS interface
  - Create, execute jobs, get results via command line
  - => allows scripting, thousands of queries at once
  - UWS-client in python (<https://github.com/adrpar/uws-client>)

# Discussion

- More data from CLUES simulations?
  - other halo catalogues? FOF?
  - AHF/FOF merger trees?
  - better format for substructures?
  - particle snapshots => how many? AHFParticles?
  - density fields?
- Which simulations next? (WM5?)
- What about merger trees for MD-simulations?
  - Rockstar-catalogues finished? For which simulations?
  - SAMs? see new Millennium-interface, TAO!
  - => create galaxy formation catalogues online!
- Feedback from users:
  - What else would people want to be able to do?
  - 
  - 
  -