



# Virtual Research Environment + Datenmanagement für kosmologische Simulationen und astronomische Surveys

H. Enke

A. Khalatyan, A. Partl, J. Klar, K. Riebe

Leibniz-Institut für Astrophysik Potsdam (AIP)

WGL Workshop Forschungsdaten, Berlin 2012

1

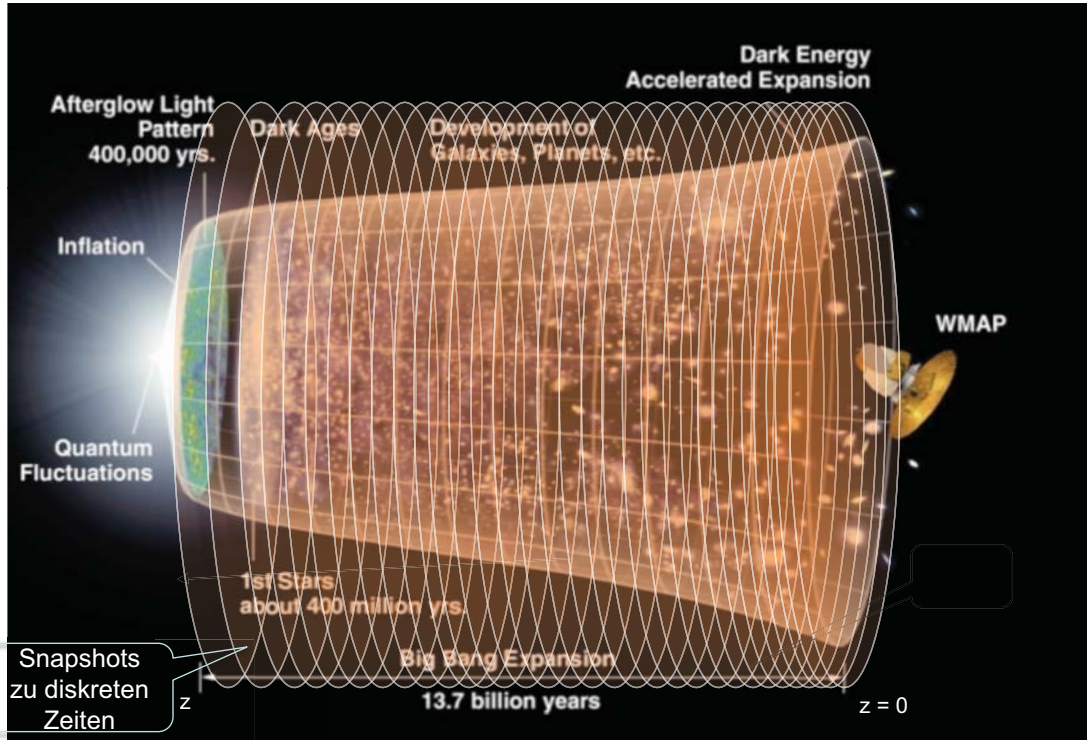


## Astrophysikalische Simulationen

Jahr	Simulationen	Daten	CPUh
1990	Zusammenstoss zweier Galaxien 32 <sup>3</sup> Partikel	~ 20 Megabyte	10 000
2009	Zusammenstoss von Galaxien 1024 <sup>3</sup> Partikel	~100 Terabyte	1 Mio.
2009	Grossräumige Strukturen im Universum Lokale Auflösung bis 4096 <sup>3</sup> Partikel	~1 Petabyte	20-50 Mio.
2009	Sterndynamos	~50 Gigabyte	100 000
	Black Hole Collisions	~5 Terabyte	1 Mio.

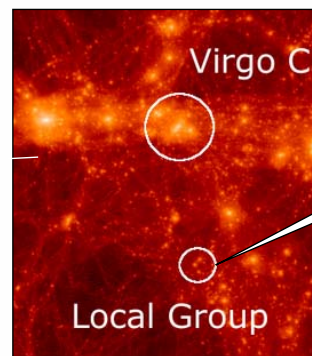
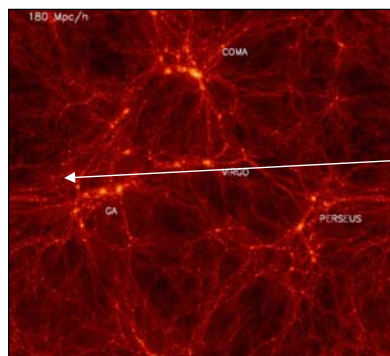
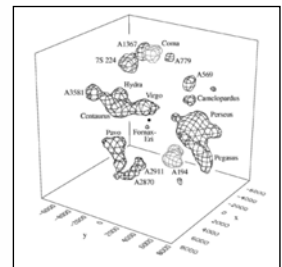
2

# Kosmologische Simulationen

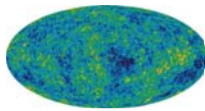


# CLUES Projekt

- **C**onstrained **L**ocal **U**niverse **E** Simulations
- mehr als 40 Wissenschaftler aus über 10 Ländern
- Reproduktion des "lokalen" Universum:
  - grossräumige kosmologische Simulationen ( $10^9$  Mpc/h)
  - Anfangsbedingungen für lokale Resimulationen mit hoher Auflösung
- Untersuchung der Bildung von Galaxien

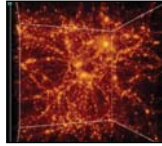


# CLUES Workflow



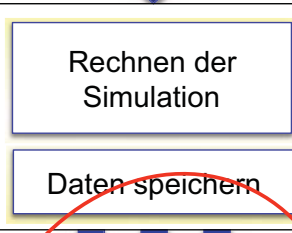
CMB+weitere Komponenten

Modell:  
Anfangsbedingungen



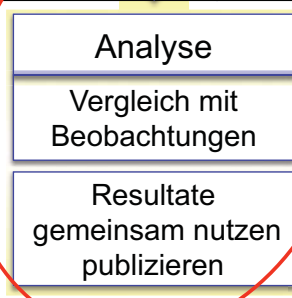
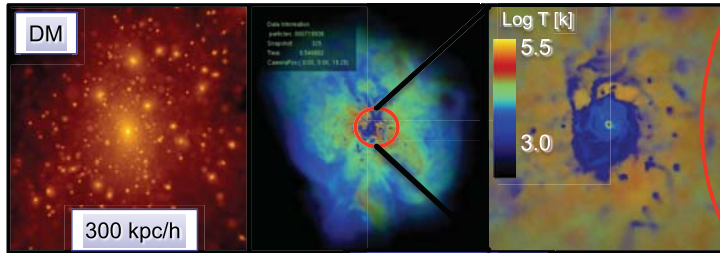
Simulation Snapshots

Computing Cluster



>10MCPUh

VRE

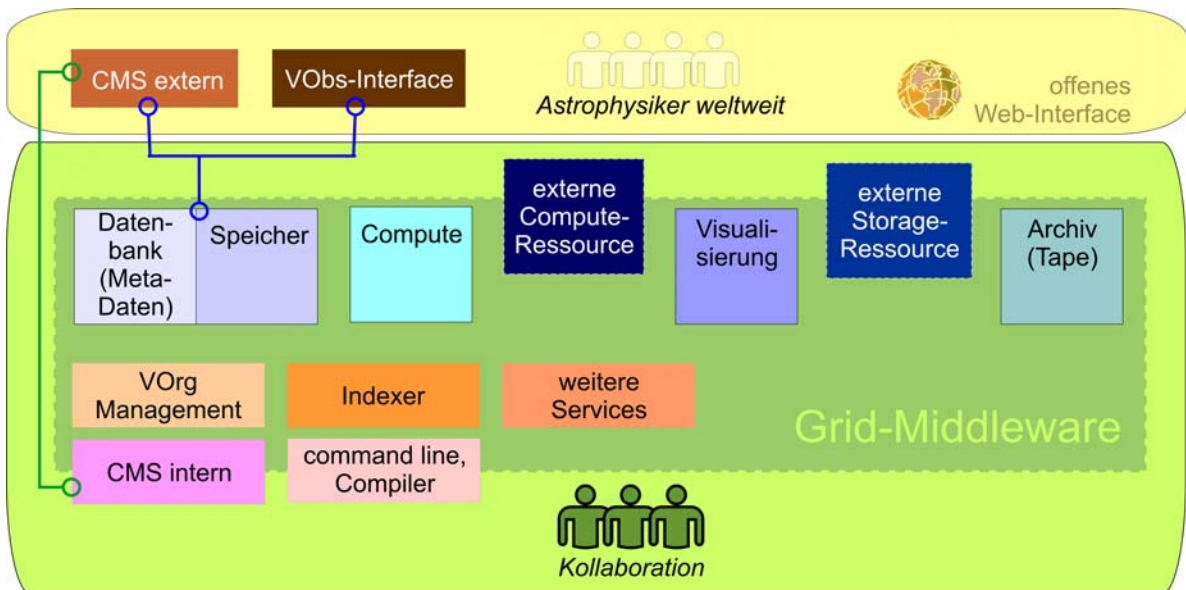


>100Tb

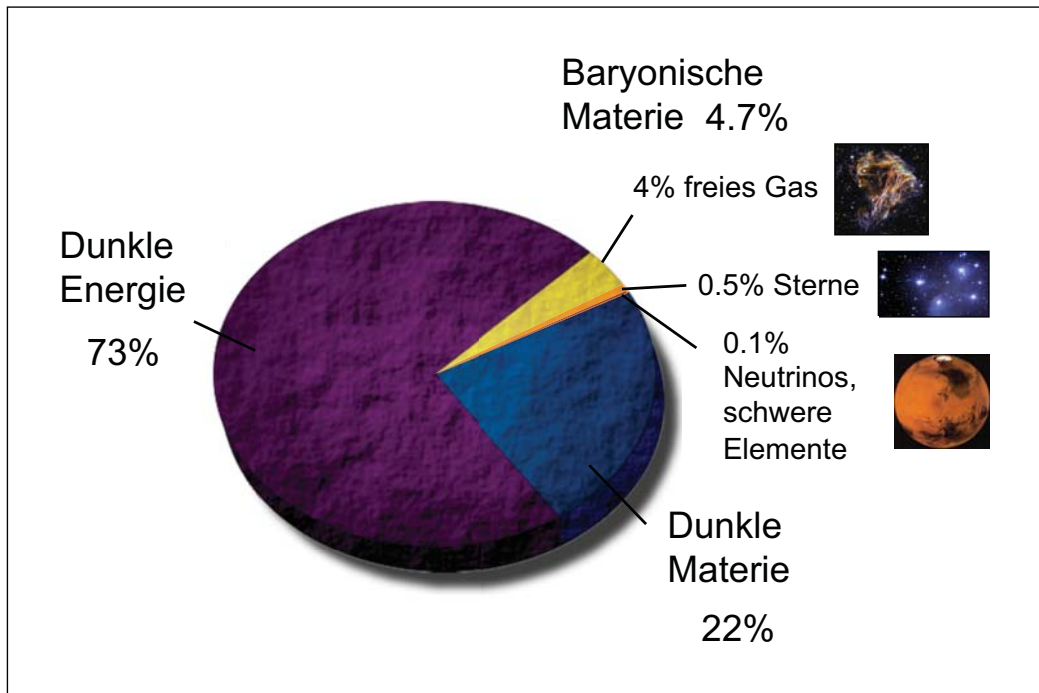


# Kollaborative VRE

Elemente einer Virtuellen Forschungsumgebung (VRE)

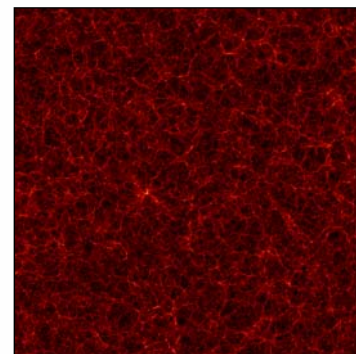




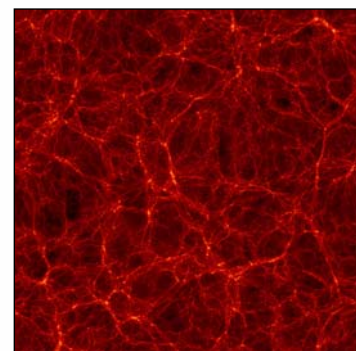


## WMAP5 Kosmologie

- MultiDark Run 1 (MDR1)
  - $(1 \text{ Gpc}/h)^3$  box (1 Gpc  $\sim 3.26 \cdot 10^9$  Lichtjahre)
  - $2048^3$  particles ( $\sim 8.6$  Milliarden)
  - Mass resolution  $8.7 \cdot 10^9 M_{\text{sun}}/h$
  
- Bolshoi Simulation (Klypin et al. 2010)
  - $250 (Mpc/h)^3$
  - $2048^3$  Partikel
  - Mass resolution  $1.35 \cdot 10^8 M_{\text{sun}}/h$
  - Bessere Auflösung als Millennium,

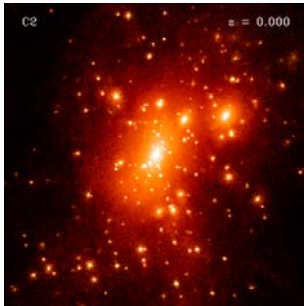
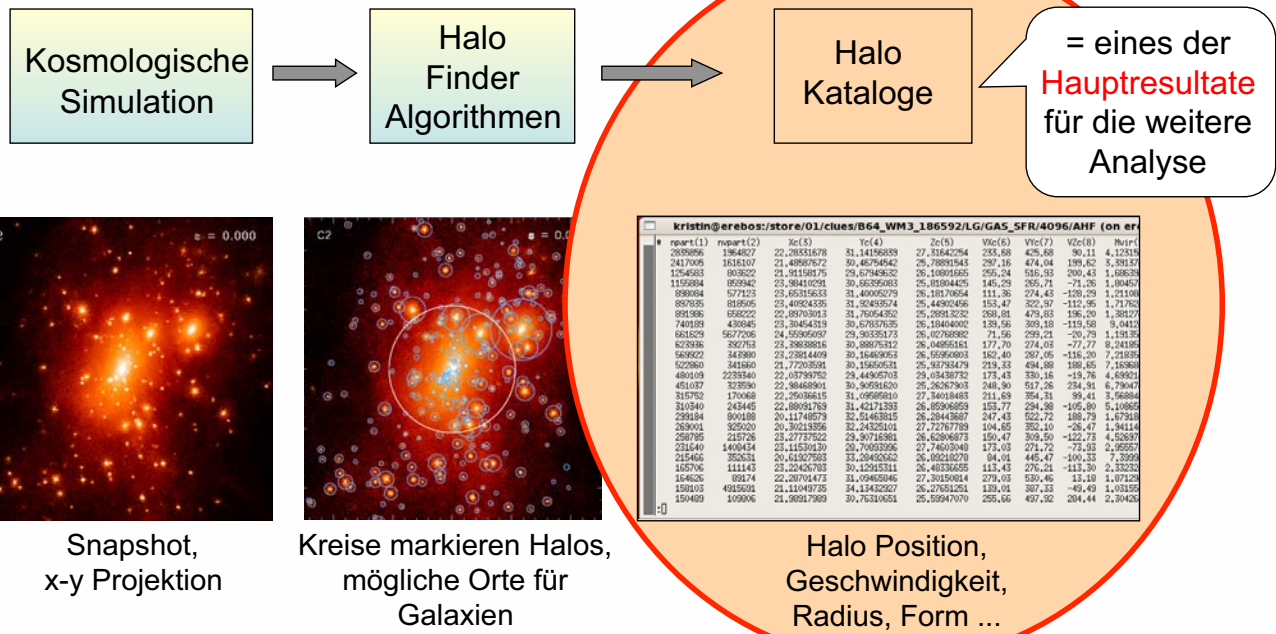


MDR1, 1 Gpc/h

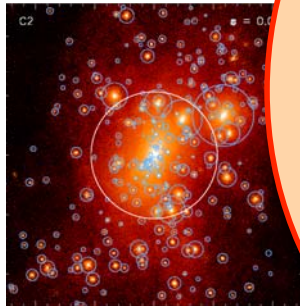


Bolshoi, 250 Mpc/h

# Halo Kataloge



Snapshot, x-y Projektion



Kreise markieren Halos, mögliche Orte für Galaxien

```

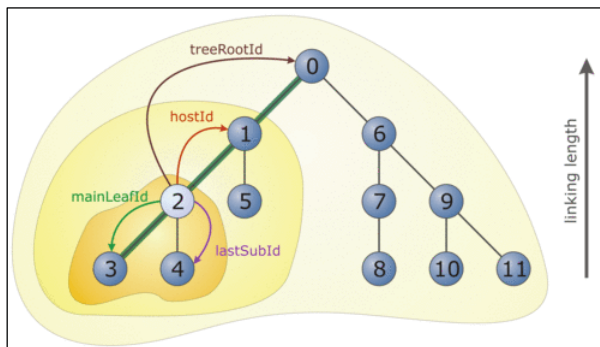
kristin@erebos:/store/01/clues/B64_WM3_1B6592/LG/GAS_SFR/4096/AHF (on er
rpart(1) rpart(2) Xc(5) Yc(4) Zc(5) Vxc(5) Vyc(7) Vzc(8) Hvir(1
285886 1864827 22,28351678 31,14156839 27,51842254 235,68 429,68 90,11 4,12319
2417005 1513107 21,48987872 30,48279442 25,78891545 257,16 474,04 159,62 5,39351
1254583 893222 21,91158175 29,5794632 26,10801585 295,24 516,33 200,43 1,68838
1158884 893942 23,58410291 30,66296083 25,81884425 145,29 255,71 -71,26 1,88467
988884 677123 23,63316633 31,4009579 25,18176564 111,36 274,45 -129,29 1,21188
897635 818505 23,40524335 31,50439374 25,44902456 153,47 322,97 -112,95 1,71782
891386 658222 22,85792013 31,79545252 25,28932252 258,01 478,88 196,20 1,88227
740189 436845 23,20464319 30,87637636 25,18484802 139,56 369,89 -119,58 5,0412
661629 5677206 24,55956957 29,39335173 26,02789882 71,56 259,21 -20,79 1,19135
623386 382753 23,28838816 30,58876312 25,04851161 177,70 274,03 -77,77 8,24185
568922 343880 23,23814409 30,16483063 26,59569803 162,40 267,05 -116,20 7,21838
522860 341680 21,77203591 30,15656531 26,51798479 219,33 494,88 189,85 7,16388
480169 2239340 22,07959752 29,44669703 29,03487732 173,43 330,16 -15,76 4,88821
451037 323990 22,98468901 30,30591620 25,26267903 248,90 517,26 234,91 6,79047
315752 170668 22,25036615 31,09989810 27,34018485 211,69 394,31 99,41 3,56884
310940 249445 22,88961769 31,42171533 26,85948959 153,77 294,88 -165,80 5,16865
289184 800188 20,11748579 32,51463815 26,28443887 247,43 522,72 188,79 1,67818
269001 925020 20,30218396 32,24325101 27,72767789 104,85 352,10 -26,47 1,94114
238785 218726 23,27737922 29,30748561 26,63898871 150,47 368,50 -122,73 4,92837
231440 1408434 23,11537430 28,74693306 27,74603048 173,03 271,72 -73,53 2,95557
215446 352531 20,61327953 33,20426352 26,18216270 94,01 445,47 -109,37 7,3988
182796 111143 23,24242703 30,12545311 26,48326655 113,43 276,21 -113,30 2,32323
164626 89174 22,28701473 31,09485846 27,30190814 279,03 520,46 13,10 1,07128
159193 4016501 21,11849735 34,13432827 26,27651251 139,91 387,53 -63,49 1,03155
150489 109006 21,98817989 30,76310651 25,55947070 285,66 497,92 284,44 2,30428

```

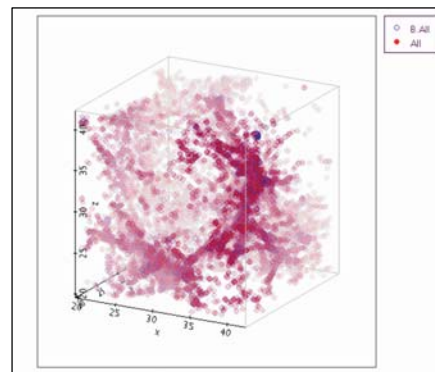
Halo Position, Geschwindigkeit, Radius, Form ...

# Resultate der Analyse

- Halo Kataloge: (halo = Materie-Zusammenballungen)
  - Halo-Finder identifizieren Dark Matter Halos
  - Zentren der Halos, in denen Galaxien zu finden sind



Indexing substructure trees of dark matter halos according to a depth first ordering. In the database we store for each halo the root of the tree (treeRootId), the direct host of the halo (hostId), the last substructure in its subtree (lastSubId) and the main leaf index (mainLeafId) for easy lookup.



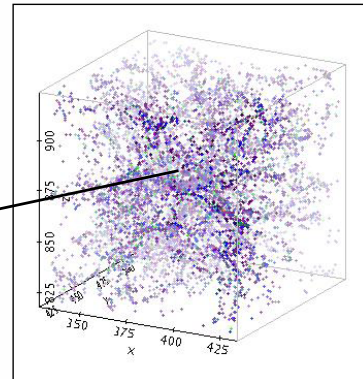
Comparing positions of dark matter halos for a 64 Mpc/h box, using two different halo catalogues: AHF (red) and Friends-of-Friends (blue)

# MultiDark Simulations Datenbank

- Halo + Galaxien Kataloge für verschiedene Halofinder (Position, Geschwindigkeit, Masse, Grösse, Form etc.)
- kinematische and Dichte - Profile der Halos
- Dunkle Materie Partikel-Verteilung einer Auswahl von Zeitschritten
- Feinstruktur und Merger-Trees



Massivster Cluster in einer Resimulation (dark matter density distribution)



Teilvolumen der MultiDark Simulation Box, jeder Punkt = ein Dark Matter Halo

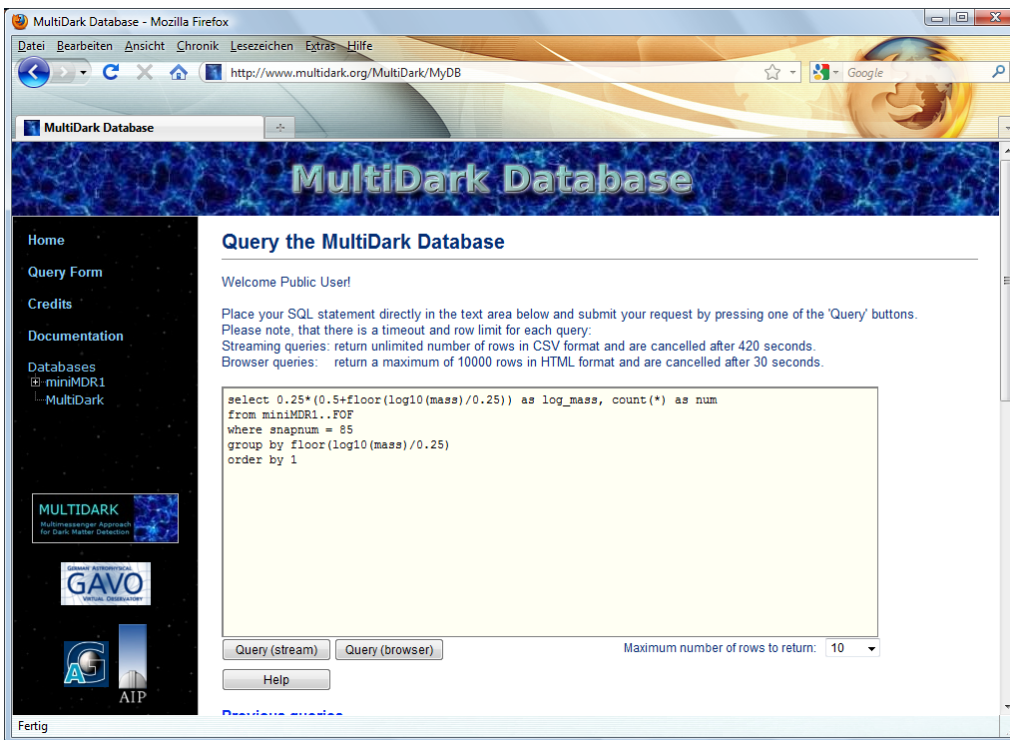
# Web Interface zur Datenbank

- Web interface: [www.multidark.org](http://www.multidark.org)
  - direkter Zugriff zu Daten via Standard SQL Queries
  - Weitere Zugriffe: webservice via wget, IDL, R, Topcat
  - History und Resultat der eigenen Queries in eigener Datenbank (registrierte Nutzer)
  - ausführliche Dokumentation
  - FAQ

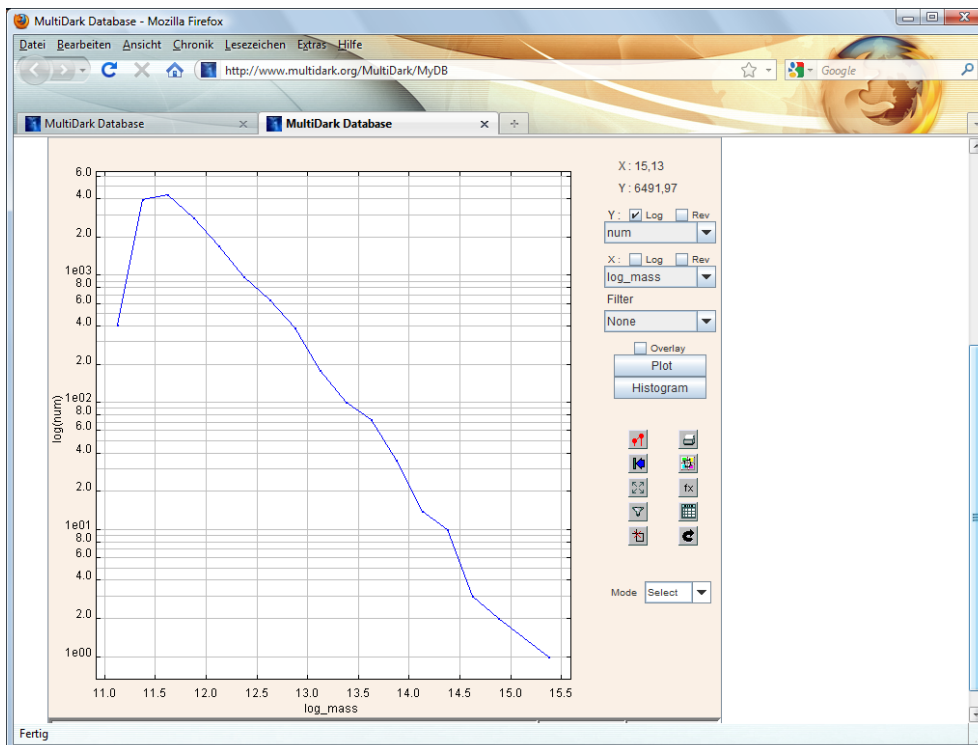




# SQL Abfrage der Datenbank



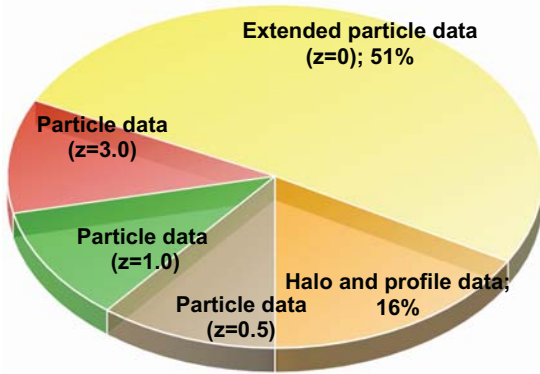
# Massenverteilung von Halos bei z=0





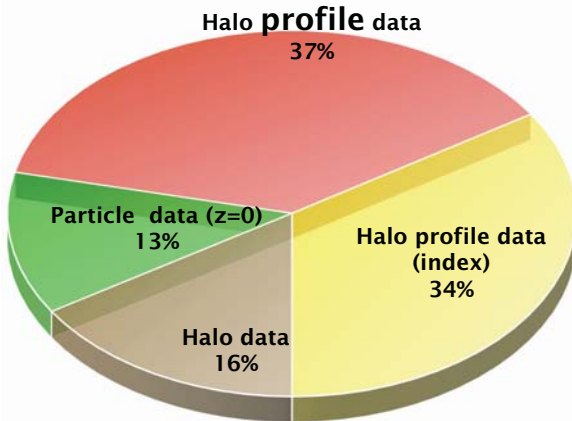
# Tabellendaten MDR1

Particle data (z = 0.5)	966 GB
Particle data (z = 1)	966 GB
Particle data (z = 3)	966 GB
Extended particle data (z = 0)	4575 GB
Halo and profile data	1503 GB
Total:	8976 GB



# Tabellendaten Bolshoi

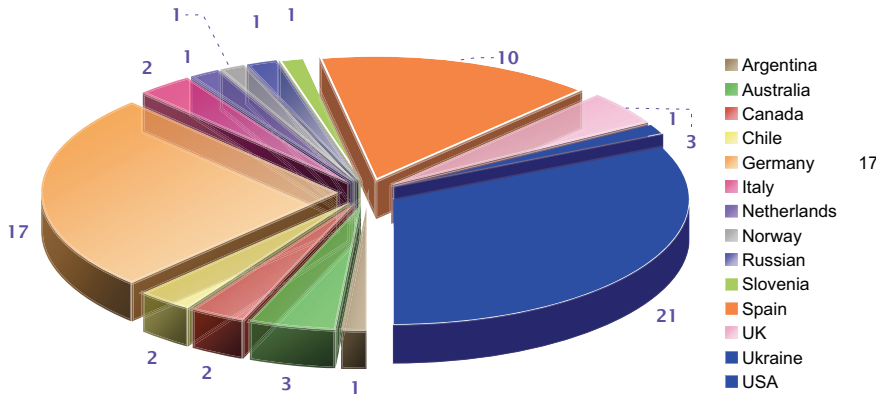
Halo data	1195 GB
Particle data z = 0	943 GB
Halo profile data (no index)	2785 GB
Halo profile data (index)	2550 GB
Total:	7473 GB







# Nutzung

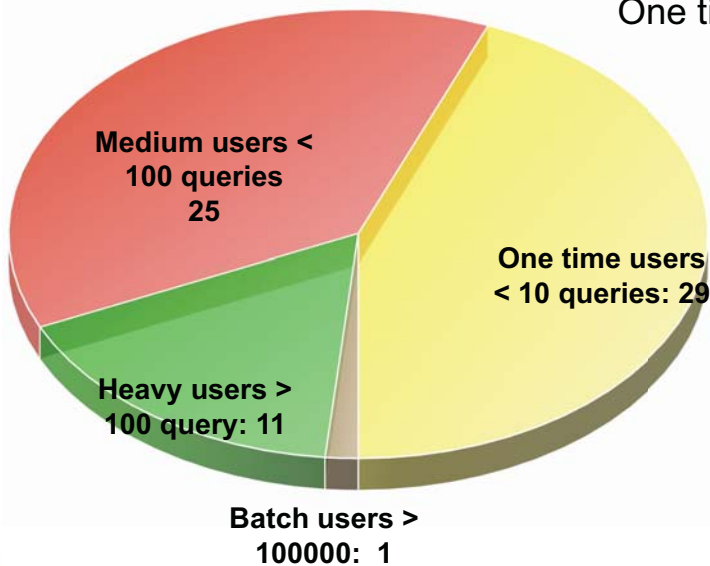


Argentina	1
Australia	3
Canada	2
Chile	2
Germany	17
Italy	2
Netherlands	1
Norway	1
Russian	1
Slovenia	1
Spain	10
UK	3
Ukraine	1
USA	21



# Nutzung

Batch users > 100000:	1
Heavy users > 100 query:	11
Medium users < 100 queries:	25
One time users < 10 queries:	29





## Herausforderung: Datenmanagement

- MultiDark DB Rohdaten (Simulation 'snapshots', Halo Kataloge...) sehr inhomogen:
  - von unterschiedlichen Gruppen erzeugt
  - an unterschiedlichen Supercomputern berechnet
  - mit unterschiedlichen Programmen
- Sammlung an einem Speicherort
- Zeitlich versetzte Verfügbarkeit der Daten
- Formatänderungen durch 'data quality assurance'
- Iterative Prozesse mit Feedback zwischen Simulatoren und DB Datenaufbereitern



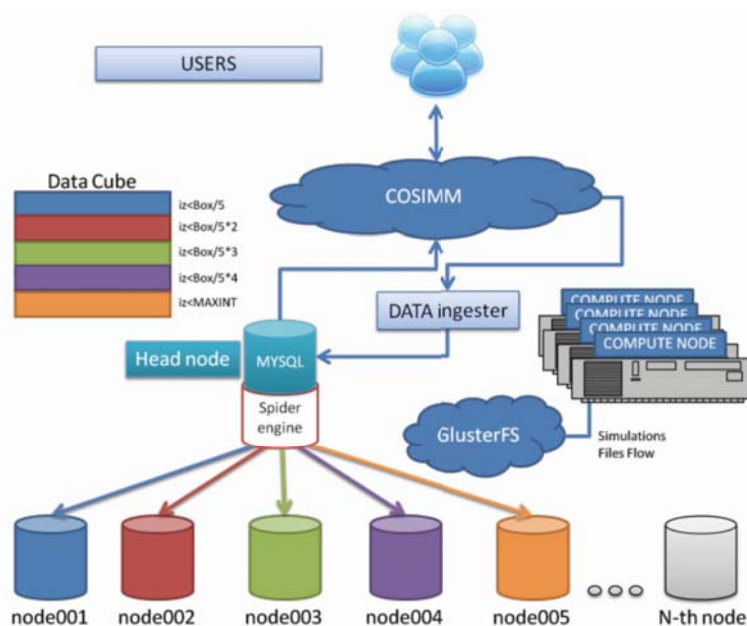
## data quality assurance / ingest

- Daten müssen verlässliche Qualität haben
- Sind Daten korrekt nach Transfer zum zentralen Server?
- Sind Daten komplett oder integer?
- Sind Daten vertrauenswürdig? (NaNs, Infs)?
- Sind Daten mit gleicher Version des Programms erzeugt (Bugfixes wg. quality assurance checks)?
- Sind Daten wissenschaftlich valide?

# Limitierungen und Herausforderungen

- Nutzerseitig:
  - Komplizierte Queries mit Joins dauern Stunden:
  - Spatiale Queries auf Partikel sind nur bei Einzelobjekten machbar
  - Limitierung bei statistischen Fragestellungen (e.g. suche alle Partikel in Halos oberhalb einer gegebenen Masse)
  - Zugriff auf Partikel Daten ziemlich langsam
  
- Serverseitig:
  - Lange Ingest-Zeiten
  - Last-Verteilung
  - Performanz derzeitig verfügbarer (kommerzieller) DBs an der Grenze

# Derzeitige Strategie





# RAVE Survey:

## RAVE (RADial Velocity Experiment)

Survey (Durchmusterung) des Südhimmels mit einem Multi-Fiber Spektroskop am 1.2-m UK Schmidt Teleskop des Australian Astronomical Observatory (AAO). Beobachtet werden Sterne der Milchstrasse. Die RAVE collaboration umfasst Wissenschaftler aus mehr als 20 Institutionen aus aller Welt. Das Projekt wird koordiniert vom Leibniz-Institut für Astrophysik Potsdam.

### IM AIP:

Datenreduktions-Pipeline für Spektren

Datenreduktions-Pipeline für chemische Elementhäufigkeiten

Rohdaten, reduzierte Daten

Webservices für alle Daten der Kollaboration

Erstellung, Kuratierung der Datenbanken/Kataloge für Analyse- Resultate für kollaborationsinterne und publizierte Daten

Website: [www.rave-survey.org](http://www.rave-survey.org)

RAVEDB Query Interface



# RAVE Survey:

## RAVE (RADial Velocity Experiment)

Survey (D  
Spektrosk  
Observato  
collaborat  
Welt. Das  
Potsdam.

### IM AIP:

Datenredu

Datenredu

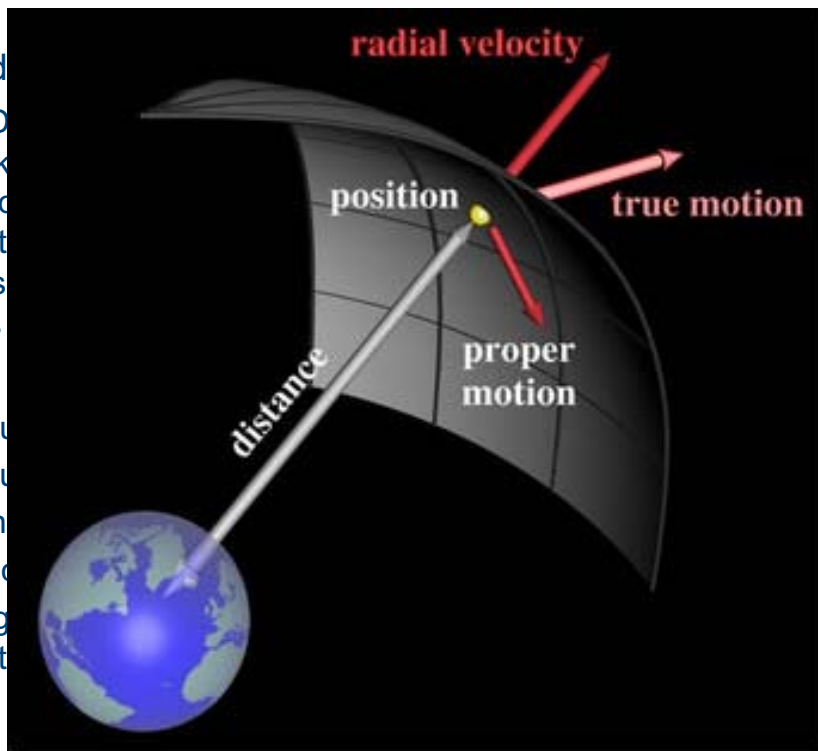
Rohdaten

Webservic

Erstellung

kollaborat

Website:



nomical  
Die RAVE  
n aus aller  
sik

esultate für





# RAVE Survey:

Rohdaten: ca. 1.5 TB  
 red. Daten I: ca. 2.0 TB  
 red. Daten II: ca. 4.0 TB

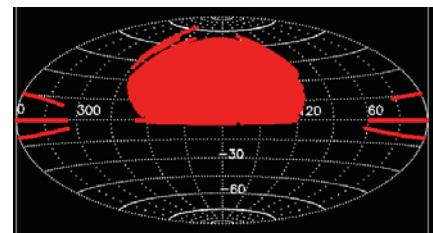
publizierte Kataloge: Radialgeschwindigkeiten für ca. 80000 Sterne  
 Distanzen für ca. 42000 Sterne  
 Chem. Häufigkeiten für ca. 38000 St.

Kataloge f. das Projekt: Radialgeschwindigkeiten, Spektren, ... für ca. 460000 Sterne  
 Distanzen für ca. 240000 Sterne  
 Chem. Häufigkeiten für ca. 205000 St.

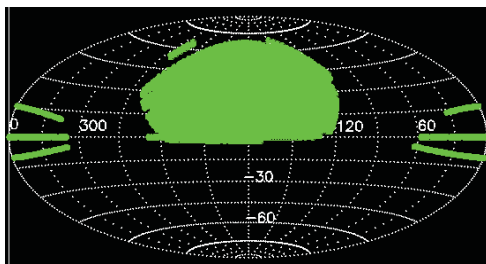


# SDSS Mirror Potsdam (DR7)

Daten: ca. 100 TB  
 Images: ca. 70 TB  
 Spektren: ca. 20 TB  
 Kataloge: ca. 12 TB



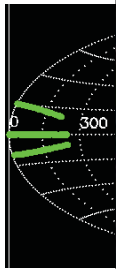
357 Millionen Objekte (SEGUE: 127 M, Legacy: 230 M)  
 Spektroskopischer Katalog



Class	N(total)	N(main)	N(SEGUE)
All	1,640,960	1,374,080	266,880
Galaxies	929,555	928,567	988
Quasars (z <2.3)	104,740	103,121	1,619
Quasars (z ≥2.3)	16,633	15,411	1,222
M stars and later	84,047	76,125	7,922
Other stars	380,214	150,748	229,466
Sky spectra	97,398	75,209	22,189
Unknown	28,383	24,767	3,616



Daten  
Ima  
Sp  
Ka  
35  
Spe



**Sloan Digital Sky Survey / SkyServer**  
SDSS Mirror site at AIP

Home Tools Schema Projects Astronomy SDSS Contact Us Download Site Search Help

Welcome to the DR7 Mirror site at AIP !!!

The SDSS mirror on ALMAGEST provides the Catalog server and (currently) part of the Data files.

This website presents data from the Sloan Digital Sky Survey, a project to make a map of a large part of the universe. We would like to show you the beauty of the universe, and share with you our excitement as we build the largest map in the history of the world.

**News**  
The site hosts data from **Data Release 7 (DR7)**. **What's new in DR7, what's new on this site, and known problems.** [More...](#)

**For Astronomers**  
A separate branch of this website for professional astronomers (English) [More...](#)

**SkyServer Tools**  
Famous places  
Get Images  
Visual Tools  
Explore  
Search  
Object Cross-ID  
CasJobs

**Science Projects**  
Basic  
Advanced  
Challenges  
For Kids  
Games and Contests  
Teachers  
Links to other projects

**Info Links**  
About Astronomy  
About the SDSS  
About the SkyServer  
SDSS Data Release 7  
SDSS Project Website  
Open SkyQuery  
Images of RC3 Galaxies

**Help**  
Getting Started  
FAQ  
How To  
Glossary  
Schema Browser  
Sample SQL Queries  
Details of SDSS Data

SDSS is supported by  
NSF  
NASA  
MEXT

Powered by **Microsoft**

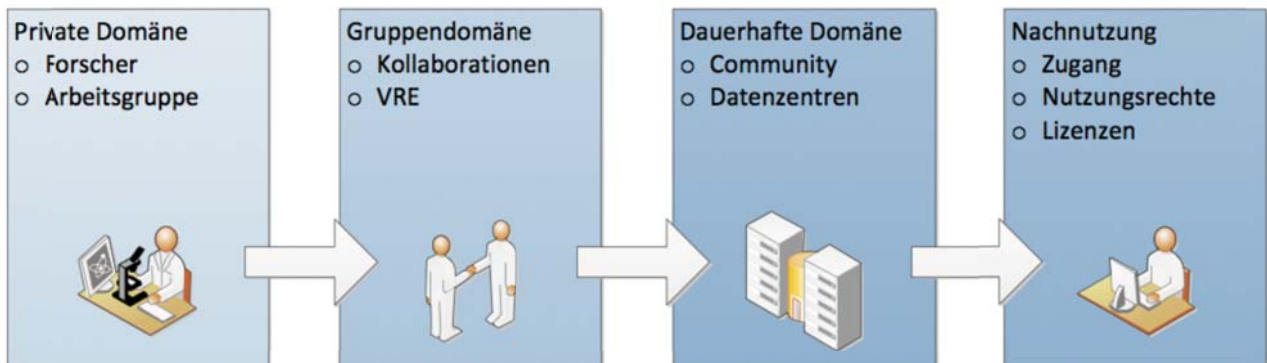
Site Traffic  
Privacy Policy

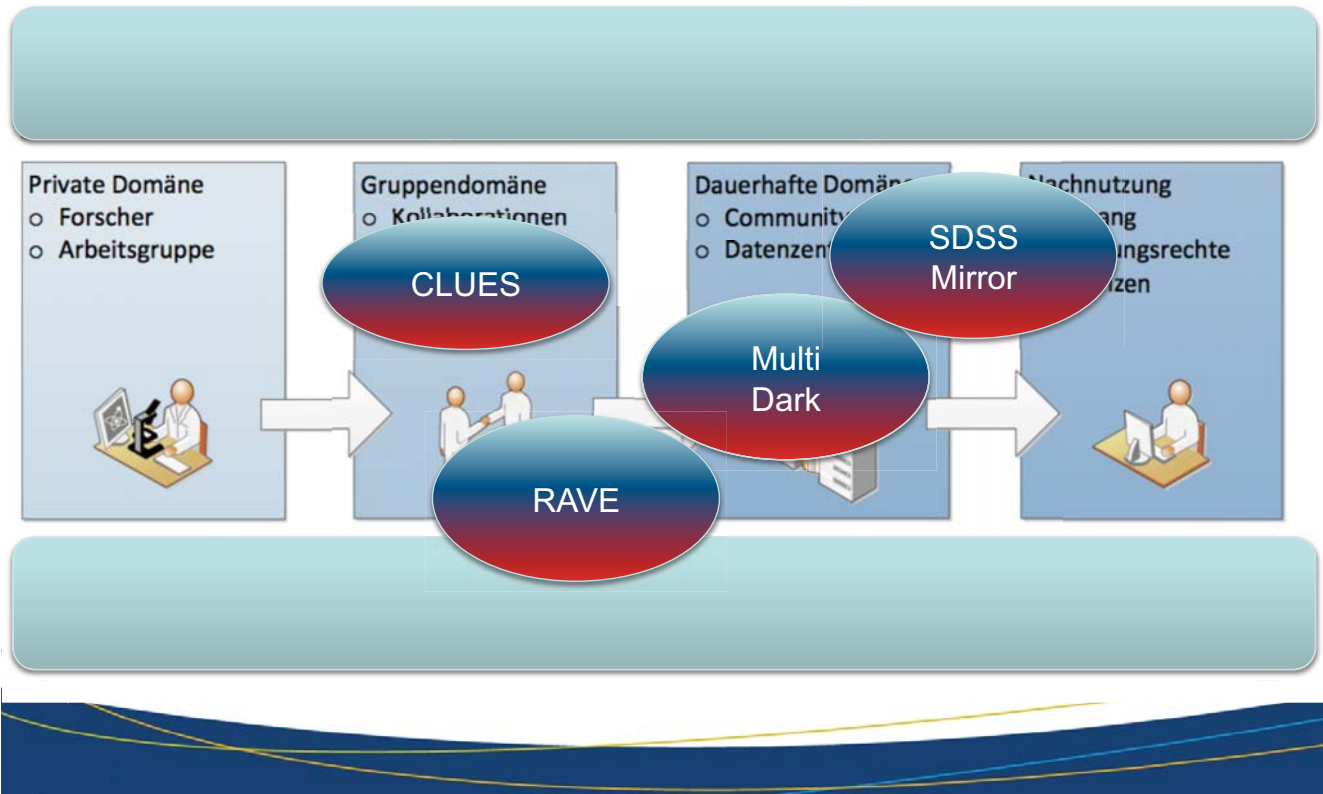
**Contact Us**



# Einordnung (Radieschen Schema)

Rahmenbedingungen disziplinübergreifender Forschungsdateninfrastruktur (DFG)





## eAstronomy : eScience+Astronomy

