



Leibniz-Institut für  
Astrophysik Potsdam

# Erfahrungen des AIP bei der Datenkuratierung

Jochen Klar

Leibniz-Institut für Astrophysik Potsdam (AIP)

[jklar@aip.de](mailto:jklar@aip.de) / [@jochenklar](https://twitter.com/jochenklar)

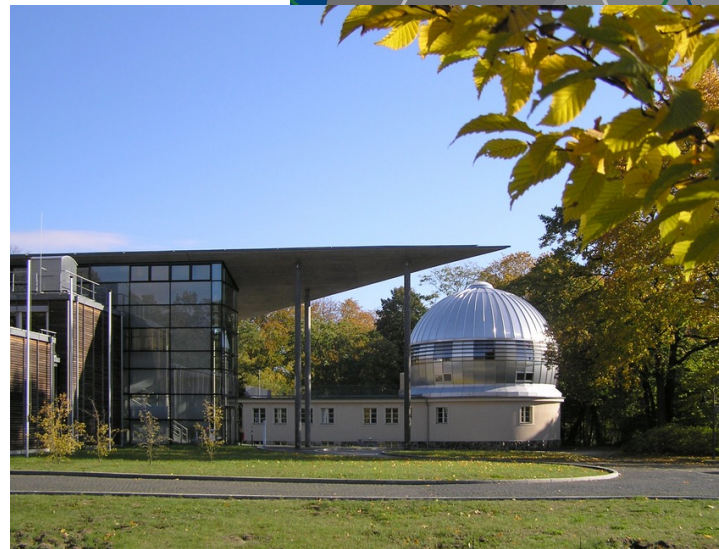
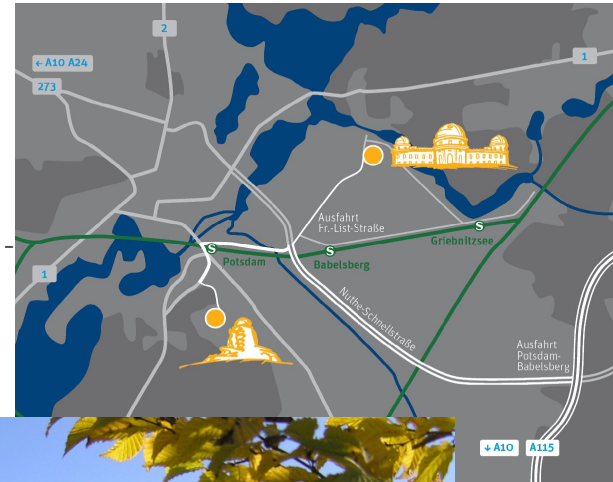
# Leibniz-Institut für Astrophysik Potsdam (AIP)

## Forschung

- Kosmische Magnetfelder
- Extragalaktische Astrophysik
- Entwicklung von Forschungsinfrastruktur und

## E-Science

- Infrastruktur
- Services
- Softwareentwicklung
- Support
- Community



# E-Science und Forschungsdatenprojekte

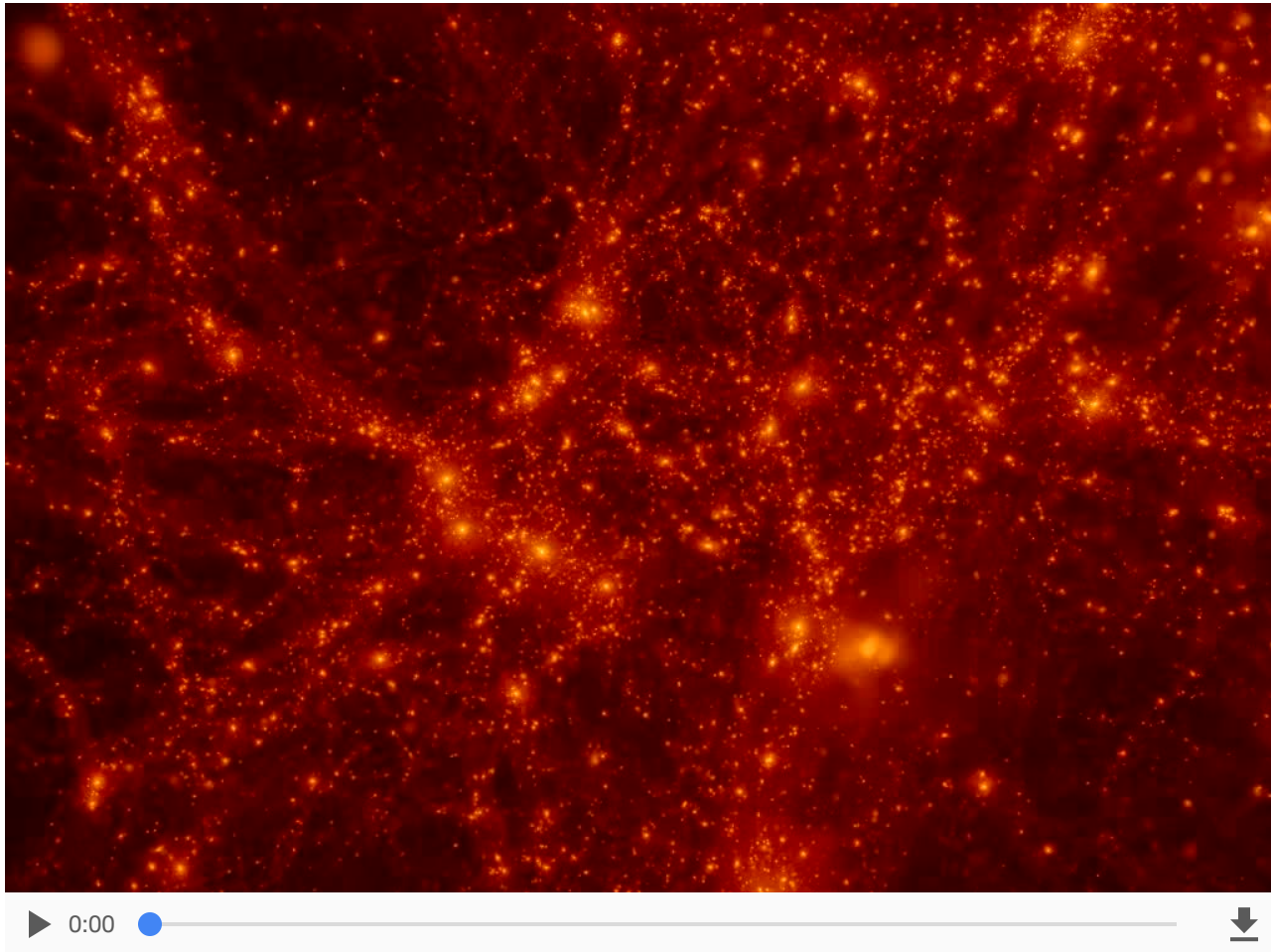
## Astronomie und Astrophysik

- seit 2003 [German Astrophysical Virtual Observatory \(GAVO\)](#)
- 2005 - 2009 [AstroGrid-D](#)
- seit 2017 [e-inf-astro: Interoperable e-Science-Infrastruktur](#)

## interdisziplinär

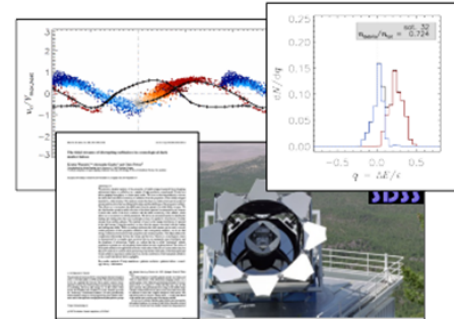
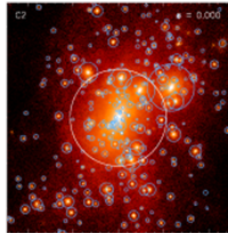
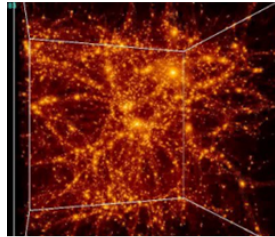
- 2009-2012 [WissGrid](#)
- 2011-2013 [Rahmenbedingungen einer disziplin-übergreifenden Forschungsdaten-Infrastruktur \(Radieschen\)](#)
- 2013-2014 [Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen](#)
- seit 2015 [Research Data Management Organiser \(RDMO\)](#)

# Kosmologische Simulationen



K. Riebe and the CLUES team, [www.clues-project.org](http://www.clues-project.org)

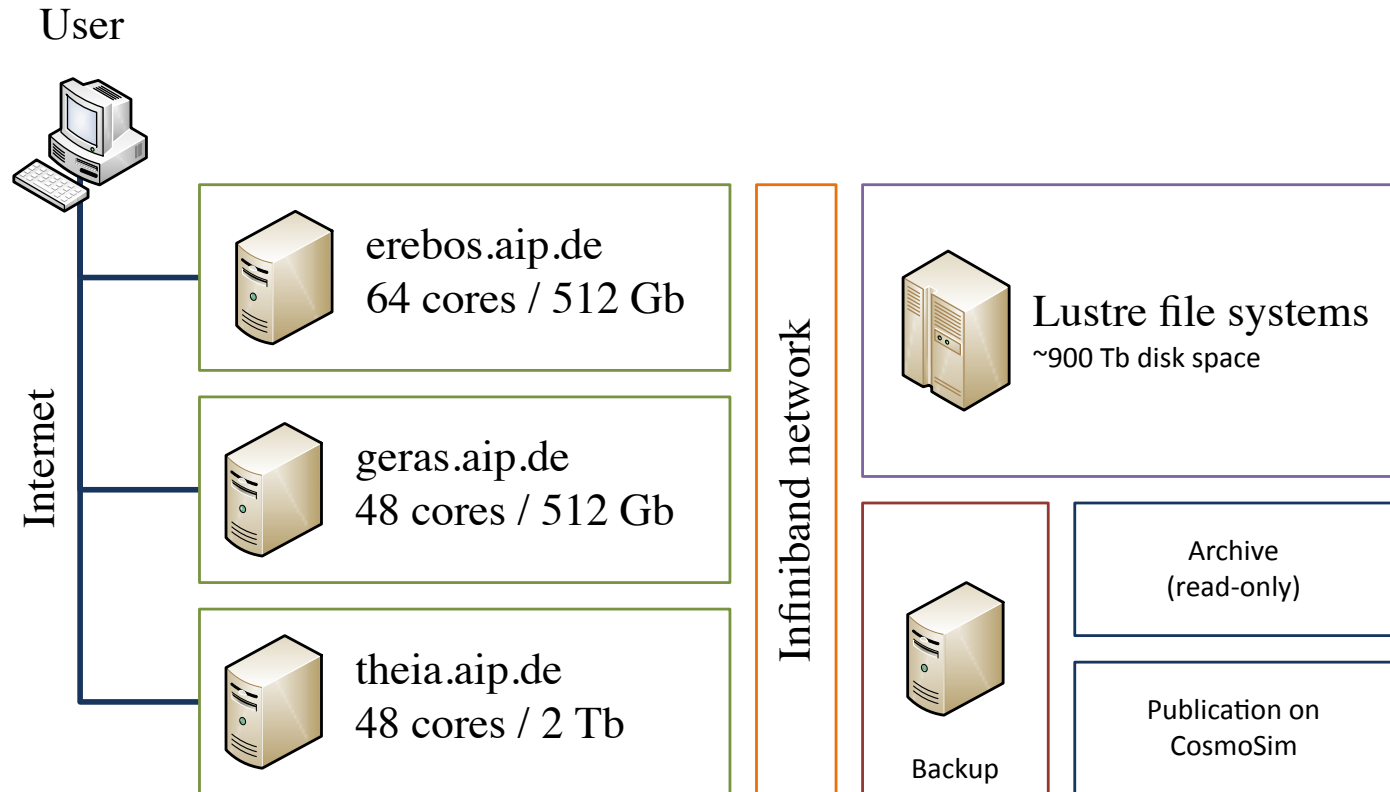
# EREBOS CRE



## CLUES: Constrained Local Universe Simulations

- ~ 70 Forschende aus Potsdam, Jerusalem, Madrid, Las Cruces, Santa Cruz, Lyon, Hawaii, Bogota, ...
- große Datenmengen aus Simulationen (~ 1 Pb)
- Rohdaten aus internationalen Rechenzentren (LRZ, NASA Ames, MareNostrum)
- *Post processing* zu weiteren Datenprodukten
- Wissenschaftliche Programme, Compiler, Bibliotheken
- Login per Secure Shell (SSH)

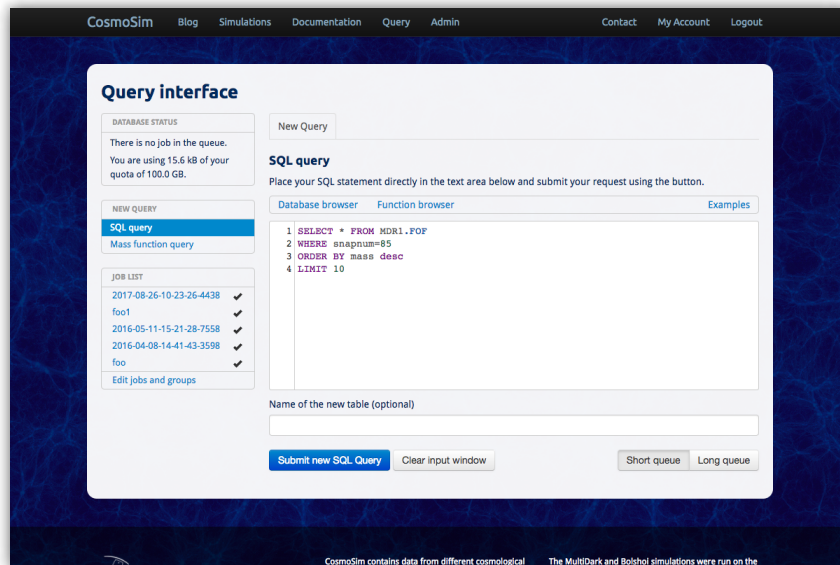
# EREBOS CRE



# CosmoSim

- Online unter <https://www.cosmosim.org>
- Abfrage von (Teil-)Datensätzen mit SQL, *share the query not the data*
- ~ 100 TB Daten, CCO

```
SELECT * FROM MDR1.FOF WHERE snapnum=8 ORDER BY mass DESC LIMIT 10
```

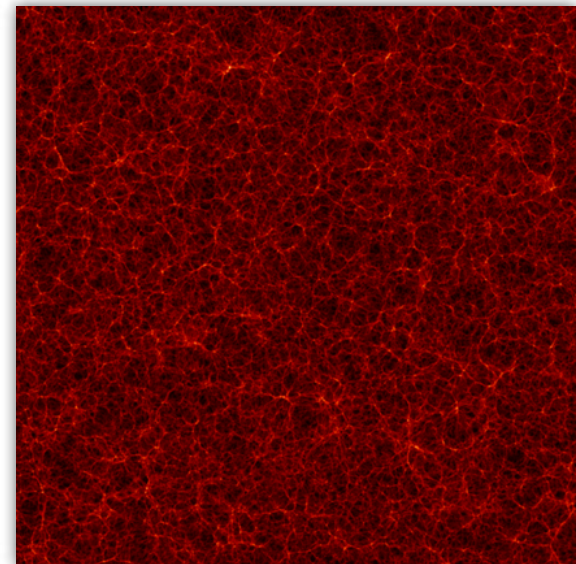


The screenshot shows the CosmoSim web interface. At the top, there is a navigation bar with links for CosmoSim, Blog, Simulations, Documentation, Query, Admin, Contact, My Account, and Logout. The main content area is titled "Query interface" and contains several sections:

- DATABASE STATUS:** A box indicating "There is no job in the queue. You are using 15.6 kB of your quota of 100.0 GB."
- NEW QUERY:** A section with a "New Query" button and a "SQL query" input field. Below the input field, there are buttons for "Database browser", "Function browser", and "Examples".
- SQL query:** A text area containing the SQL query: 

```
1 SELECT * FROM MDR1.FOF
2 WHERE snapnum=85
3 ORDER BY mass desc
4 LIMIT 10
```
- JOB LIST:** A table showing a list of jobs with their IDs, names, and status (checked or unchecked).
- Submit new SQL Query:** A button to execute the query, along with "Clear input window", "Short queue", and "Long queue" options.

At the bottom of the interface, there is a footer that reads: "CosmoSim contains data from different cosmological The MultiDark and Bolshoi simulations were run on the".



# MUSE



## Multi Unit Spectroscopic Explorer

- *First light* am 31 Januar 2014
- Besteht aus 24 Integral-field units (IFU)
- Produziert Datenkuben: 2 räumliche Dimensionen + Wellenlänge



Iztok Boncina/ESO

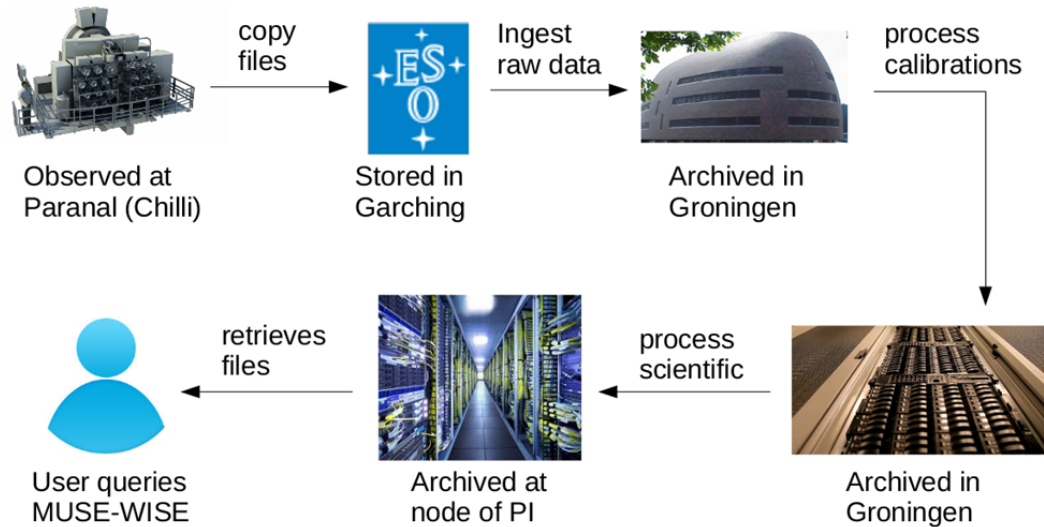


Eric Le Roux / Service Communication / UCBL / MUSE

Website: <http://muse-vlt.eu/science/>

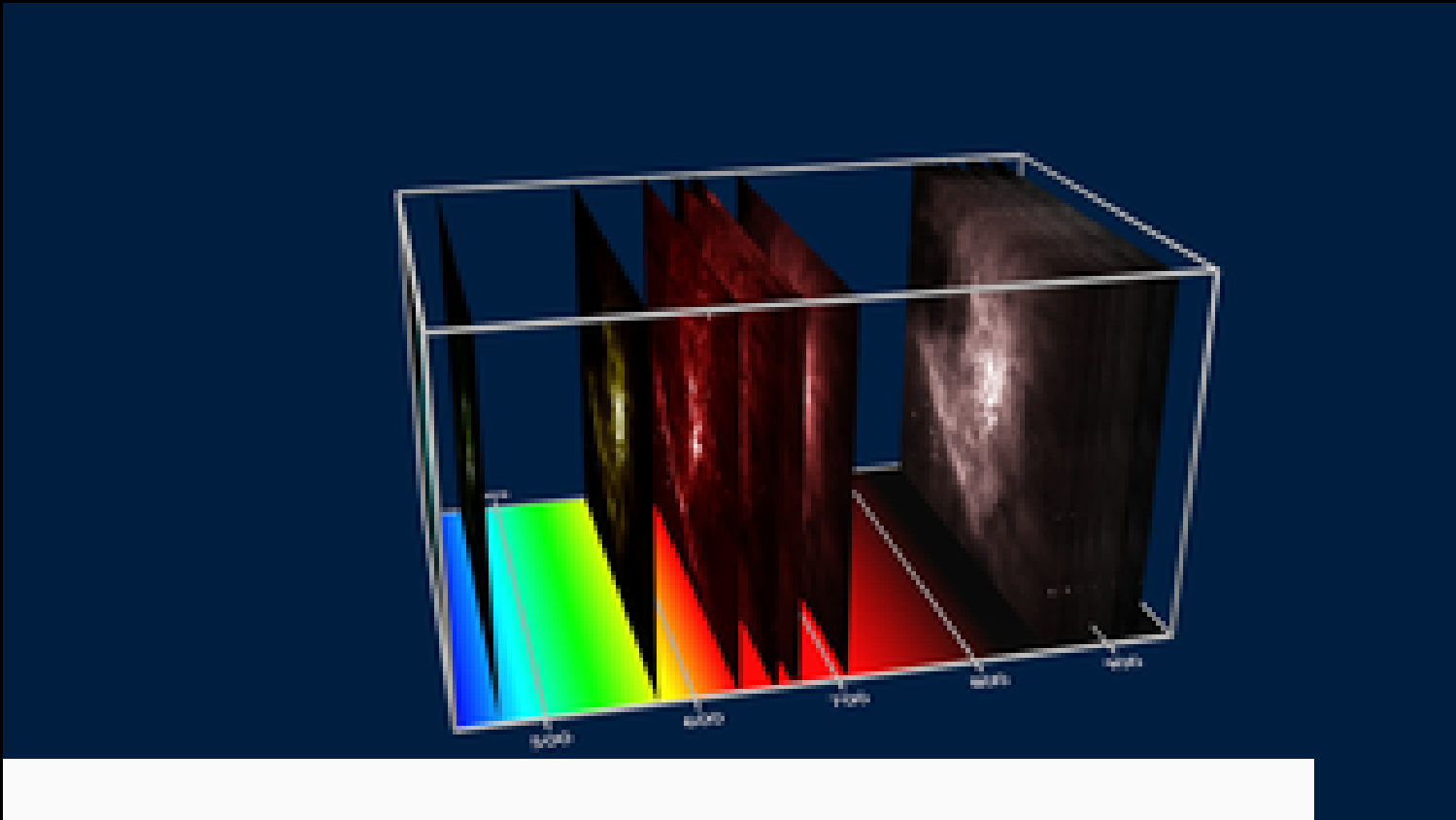


# MUSE CRE



## MUSE-WISE environment

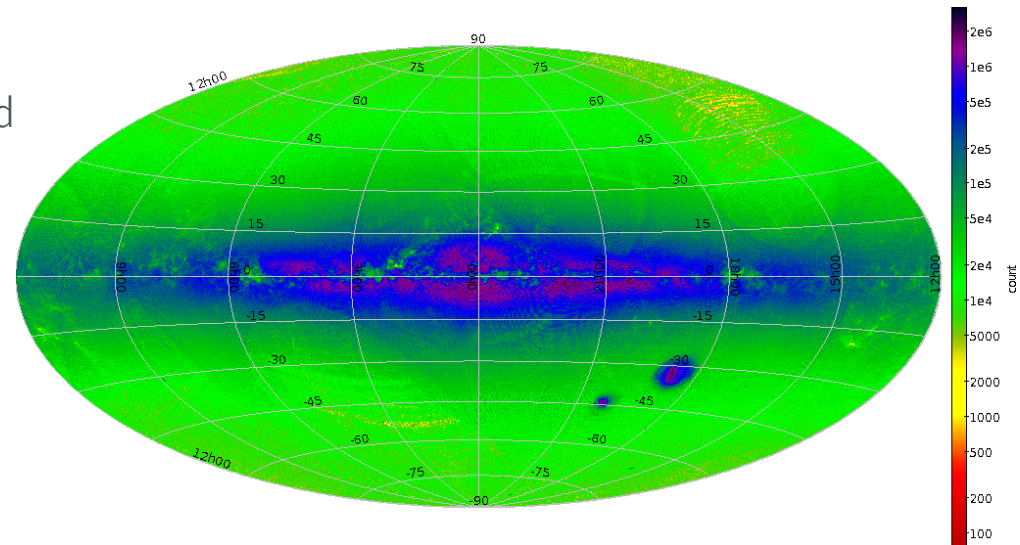
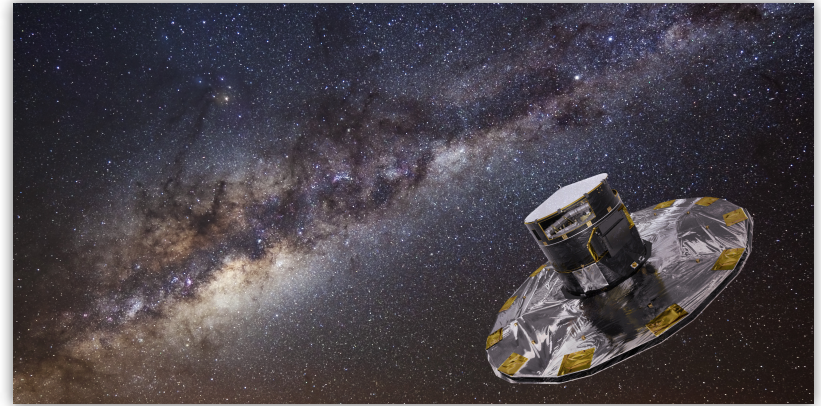
- 2 Datenbankserver in Potsdam and Groningen (Metadata, Provenience)
- 7 Nodes mit Datenserver (Images, graphs, previews, logs) and Distributed Processing Units (DPU)
- Einheitlicher Software stack, Einheitliches Datenmodell



Credit: Kristin Riebe, Peter Weillbacher, MUSE collaboration

# Gaia Survey

- Satellitenmission der ESA
  - Start: 19. 12. 2013
  - Betrieb: 2014-2018
- Hochpräzise Vermessung
  - Parallaxen
  - Radialgeschwindigkeiten
  - Sterneigenschaften
- Auswertung und Publikation durch das Data Processing and Analysis Consortium (DPAC)
  - Steuerung (CU 1)
  - Auswertung (CU 2-8)
  - Archiv und Katalog (CU 9)



# Gaia Archive

- Data Release 1 (GDR1) am 14. 9. 2016
- [Zentrales Archiv der ESA](#), Mirrors in [Potsdam](#), [Heidelberg](#), ...

The screenshot shows the Gaia@AIP Query interface. At the top is a navigation bar with links: Gaia@AIP, Query, Documentation, Database tables, Blog, FAQ, Contact, My Account, Admin, Logout.

### Query interface

**DATABASE STATUS**  
There is no job in the queue.  
You are using 1.5 GB of your quota of 100.0 GB.

**NEW QUERY**  
**SQL query**

**JOB LIST**

GDR1	📁
TGAS	✓
GUMS10	📁
GOG11	📁
UNASSIGNED	
2017-08-26-09-37-53-0297	✓
2017-06-23-10-30-09-6691	✓
100	✓
2017-03-01-12-03-05-4298	✓
2016-11-18-10-54-32-8797	✓
111	⊘
TGAS3	✓
TGAS2	⊘
<a href="#">Edit jobs and groups</a>	

**New Query**

### SQL query

Place your SQL statement directly in the text area below and submit your request using the button.

Database browser   Function browser   Simbad object search   [Examples](#)

```
1 SELECT gmag * 0.1 AS gmag_bin, COUNT(gmag) AS number
2 FROM
3 (
4   SELECT FLOOR(`phot_g_mean_mag` * 10) AS gmag
5   FROM `GDR1`.`gaia_source`
6 ) AS gmag_tab
7 GROUP BY gmag;
```

Name of the new table (optional)

**Submit new SQL Query**   Clear input window   Short queue   Long queue

# Digitalisierung von Fotoplatten



## **DFG-Projekt: Digitalisierung astronomischer Fotoplatten und ihre Integration in das internationale Virtual Observatory**

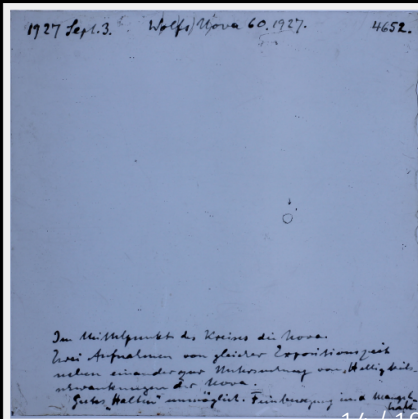
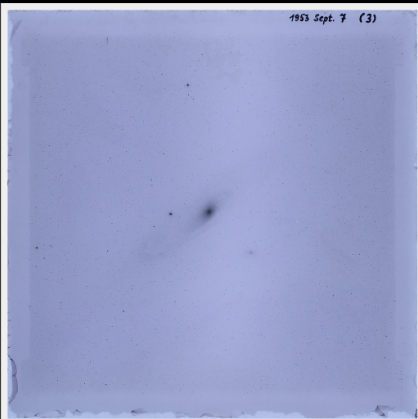
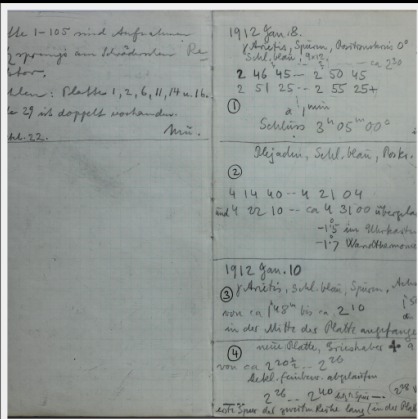
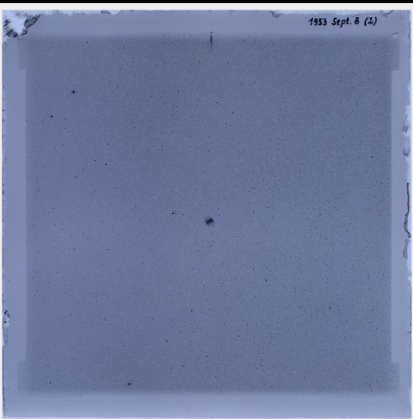
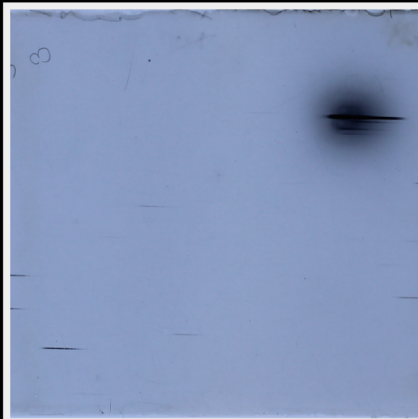
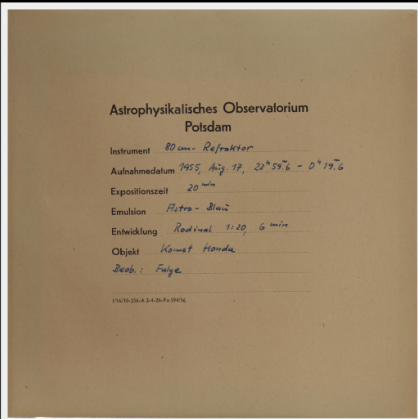
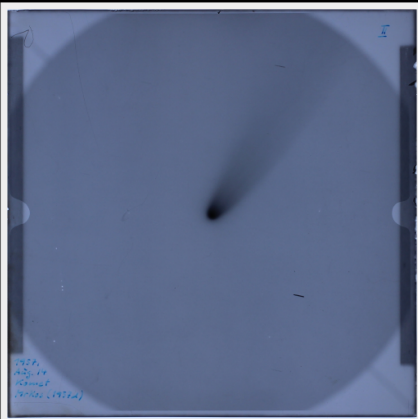
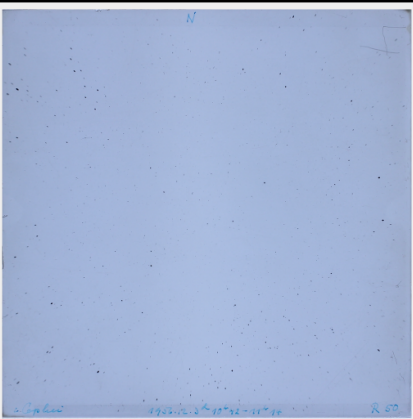
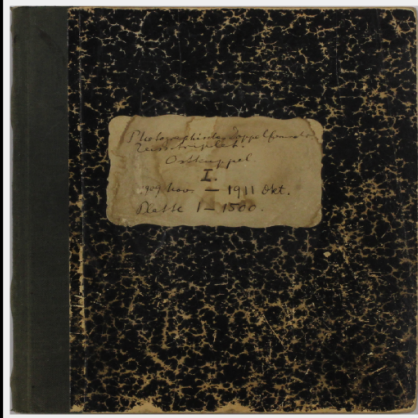
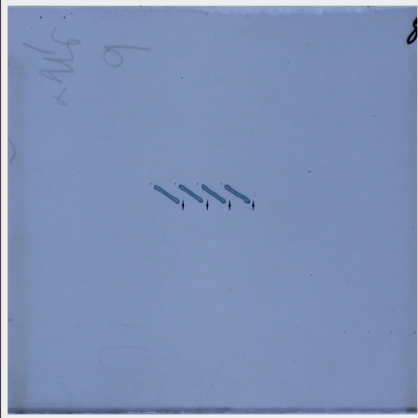
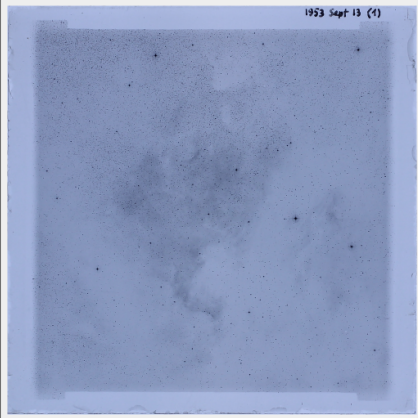
Fotoplatten aus Bamberg, Hamburg und Potsdam aus den Jahren 1909-1976

Astrophysikalisches Observatorium  
Potsdam

Instrument R 60  
 Aufnahmezeit 1958 Juli 26-28  
 Expositionszeit 11<sup>h</sup> 47 - 11<sup>h</sup> 49  
 Emulsion via R 1  
 Entwicklung final, 10 min.  
 Objekt  $\epsilon$  Cephei

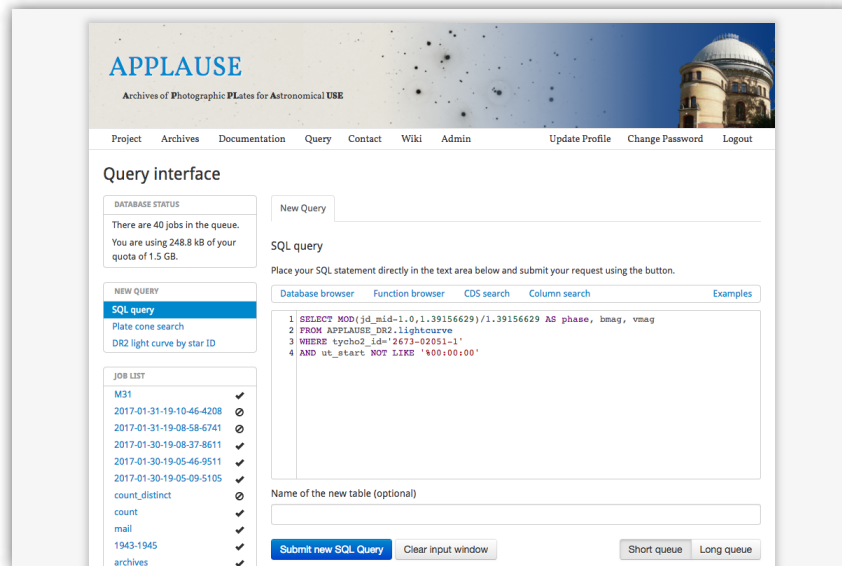
1958-200-4-3-24-2-1958

1958 Sept 13 (4)



# APPLAUSE Datenbank

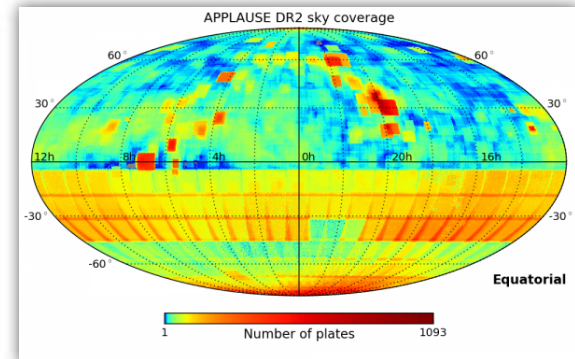
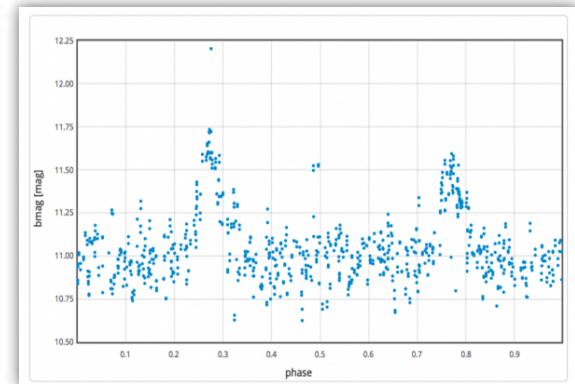
- Online unter <https://www.plate-archive.org>
- 51.517 Scans von 42.789 Fotoplatten
- 58.115 digitalisierte Plattenhüllen
- Logbuch-Einträge aus 149 Logbüchern
- ~ 30 TB an Daten, CCO, DOI, Europeana



The screenshot shows the APPLAUSE database query interface. At the top, there is a header with the APPLAUSE logo and the text "Archives of Photographic Plates for Astronomical USE". Below the header, there is a navigation menu with links for Project, Archives, Documentation, Query, Contact, Wiki, Admin, Update Profile, Change Password, and Logout. The main content area is titled "Query interface" and contains several sections:

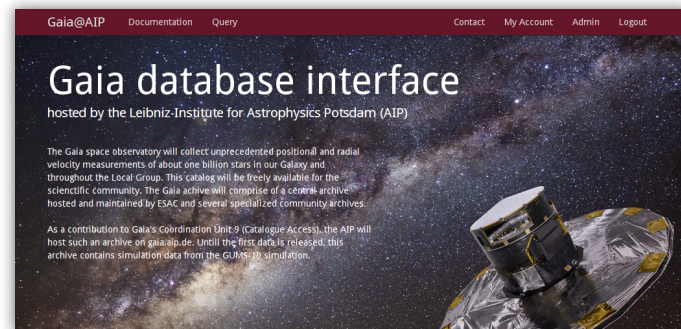
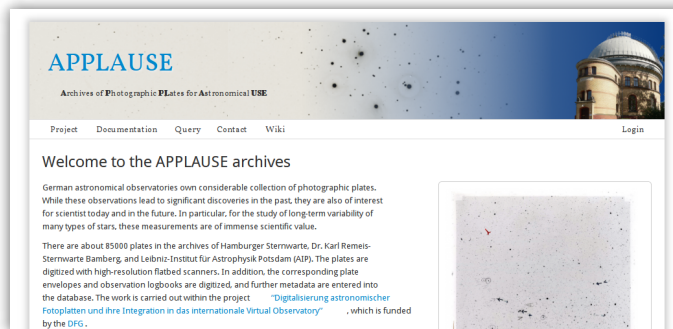
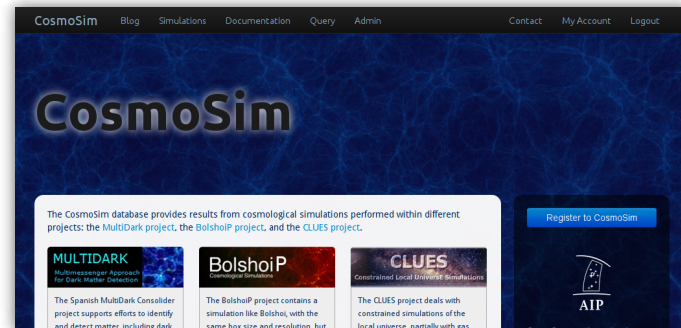
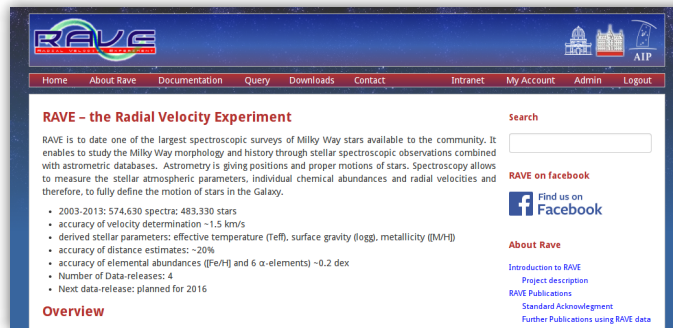
- DATABASE STATUS:** "There are 40 jobs in the queue. You are using 248.8 kB of your quota of 1.5 GB."
- NEW QUERY:** A section for creating a new query, including a "New Query" button and a "SQL query" input field. Below the input field, there is a "Submit new SQL Query" button and a "Clear input window" button.
- SQL query:** A section for entering a SQL query. It includes a "Database browser" button and a "Function browser" button. Below the input field, there is a "Name of the new table (optional)" input field and a "Submit new SQL Query" button.
- SQL query example:** A pre-filled SQL query: 

```
1 SELECT MOD(jd_mid-1,0,1,39156629)/1,39156629 AS phase, bmag, vmag
2 FROM APPLAUSE_DR2.lightcurve
3 WHERE tycho2_id='2673-02051-1'
4 AND ut_start NOT LIKE '%00:00:00'
```
- JOB LIST:** A table showing the status of various jobs, including job ID, date, and status.



# Datenpublikationen am AIP

- RAVE database query interface for DR3 (Siebert et al., 2011)
- MultiDark Database (Riebe et al., 2011)
- RAVE (2013), CosmoSim (2014), APPLAUSE (2015), Gaia@AIP (2016)





# Virtual Observatory



- Standards (Formate, Metadaten, Protokolle)
- Interoperable Tools und Anwendungen
- Verteilte Datenzentren, *VO registry*

# Datenkuratierung am AIP

## Datenmanagement

- Virtuelle Forschungsumgebungen zum kollaborativen Arbeiten
  - EREBOS CRE
  - MUSE CRE
  - GREGOR CRE
- Unterstützung von Surveys (z.B. RAVE) oder eigenen Observatorien (z.B. Stella)
- Nachträgliche Digitalisierung von Fotoplatten (Digital Heritage)

## Datenpublikation

- Zugriff über Webseiten auf SQL Datenbanken
- Nutzung von DOI für alle neueren Datenpublikationen
- Standardisierung und Formate durch das Virtual Observatory



Leibniz-Institut für  
Astrophysik Potsdam

**Danke für Ihre Aufmerksamkeit!**

Jochen Klar

[jklar@aip.de](mailto:jklar@aip.de) / [@jochenklar](https://www.instagram.com/jochenklar)

[science.aip.de](https://science.aip.de)

[science.aip.de/vis](https://science.aip.de/vis)

[www.aip.de](https://www.aip.de)