



# Big Data und Smart Data in der Astronomie

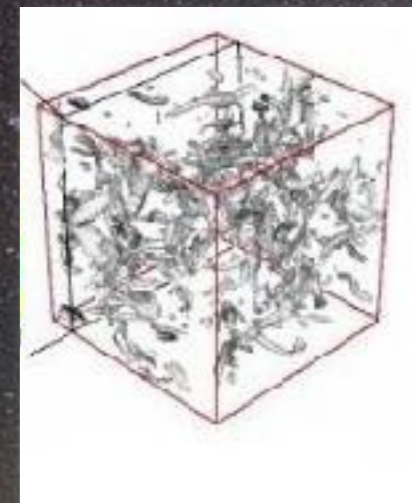
Matthias Steinmetz (AIP) @GalacticRAVE

# Vier Arten, Forschung zu betreiben

- Seit einigen tausend Jahren –  
Experimentelle Forschung
  - Beschreibung der verschiedenen Phänomene in der Natur
- Seit einigen hundert Jahren –  
Theoretische Forschung
  - Newton's Gesetze, Maxwellsche Gleichungen
- Seit einigen zehn Jahren –  
Computergestützte Forschung
  - Simulation komplexer Systeme
- Seit einigen Jahren –  
Datengetriebene Forschung
  - Wissenschaftler sind überwältigt von Daten aus verschiedenen Quellen
    - Instrumente
    - Simulationen
    - Sensor-Netzwerke



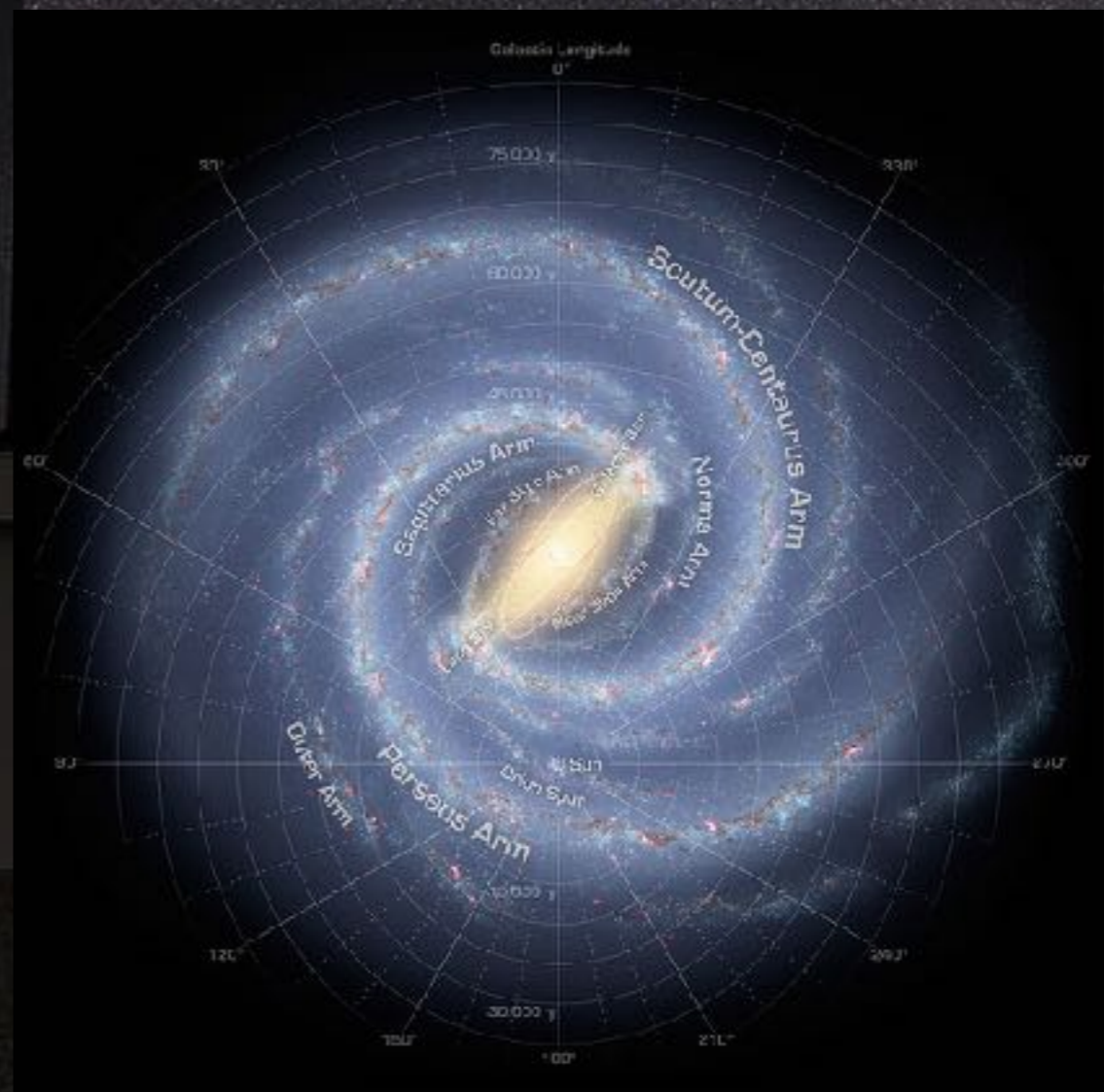
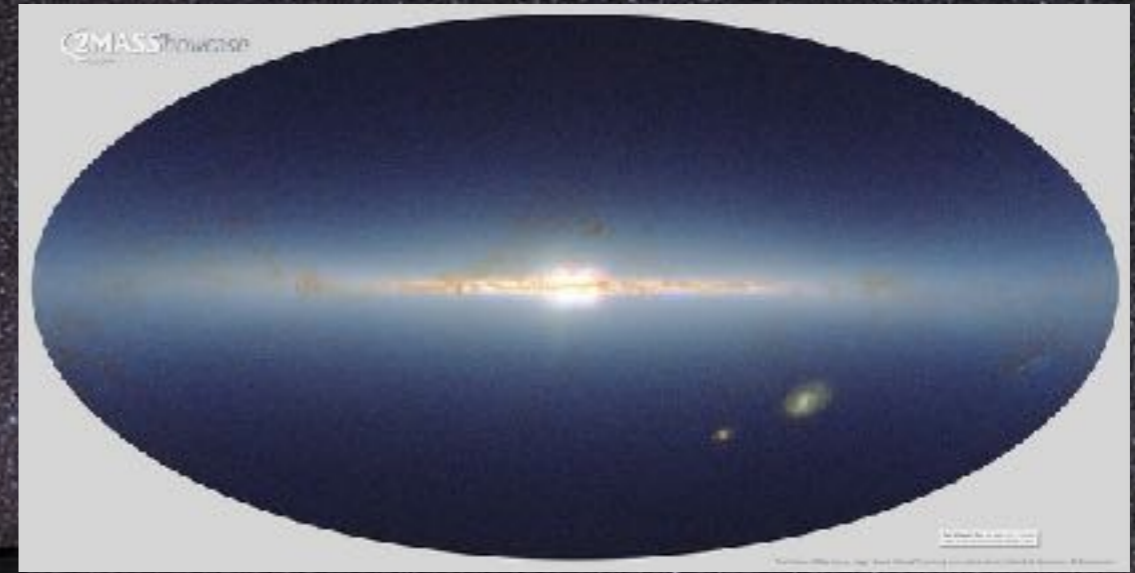
$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$



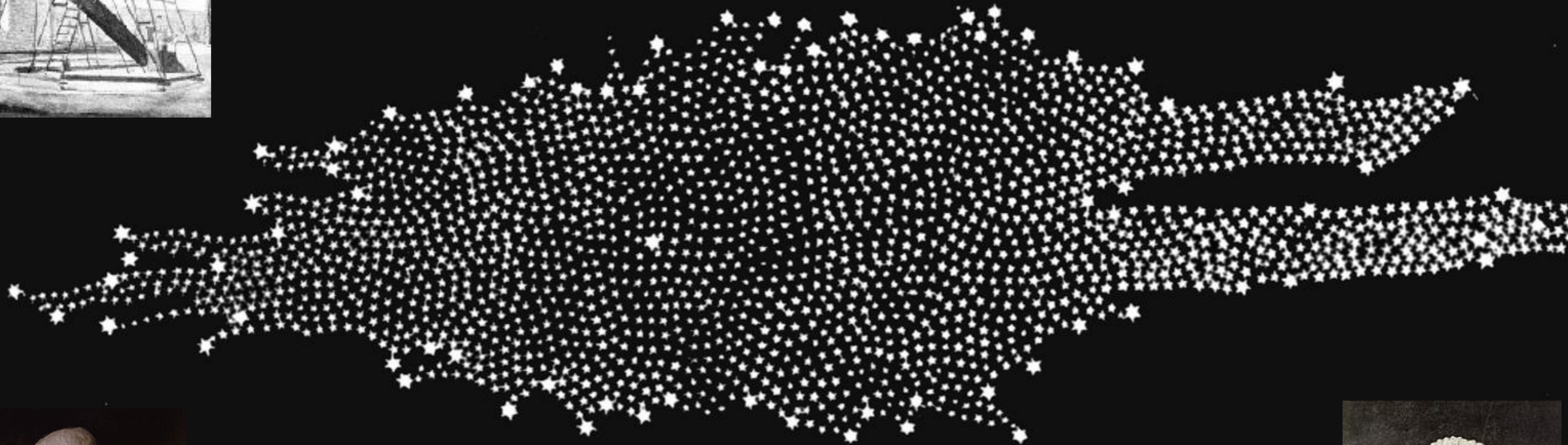
# Astronomie aus der historischen Perspektive

- statischer Himmel (Fixsterne),  
etwa 6000 Sterne sind mit bloßem Auge sichtbar  
⇒ Navigation
- einige wenige sich bewegende Objekte  
(Sonne, Mond, 5 Planeten „Wandersterne“)  
⇒ Zeitmessung, Kalenderwesen
- gelegentlich unerwartete Ereignisse  
(Kometen, Novae, Supernovae)

# Beispiel Milchstraße



# Spätes 18. Jahrhundert : erste Karte der Milchstraße

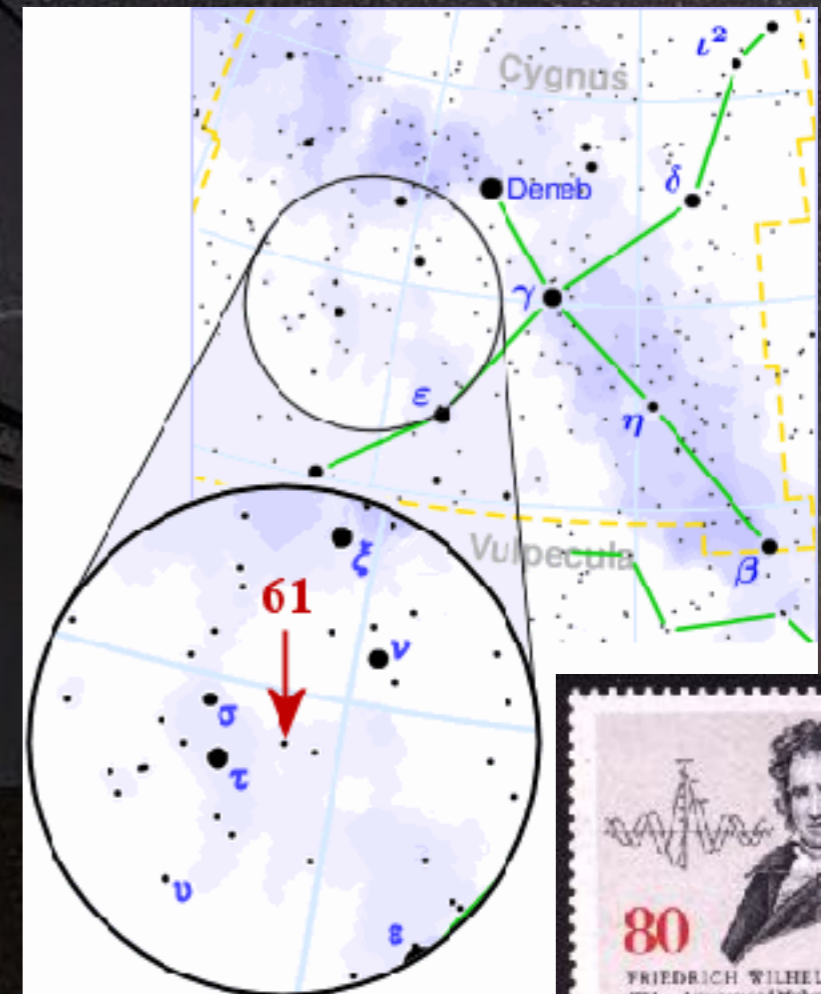
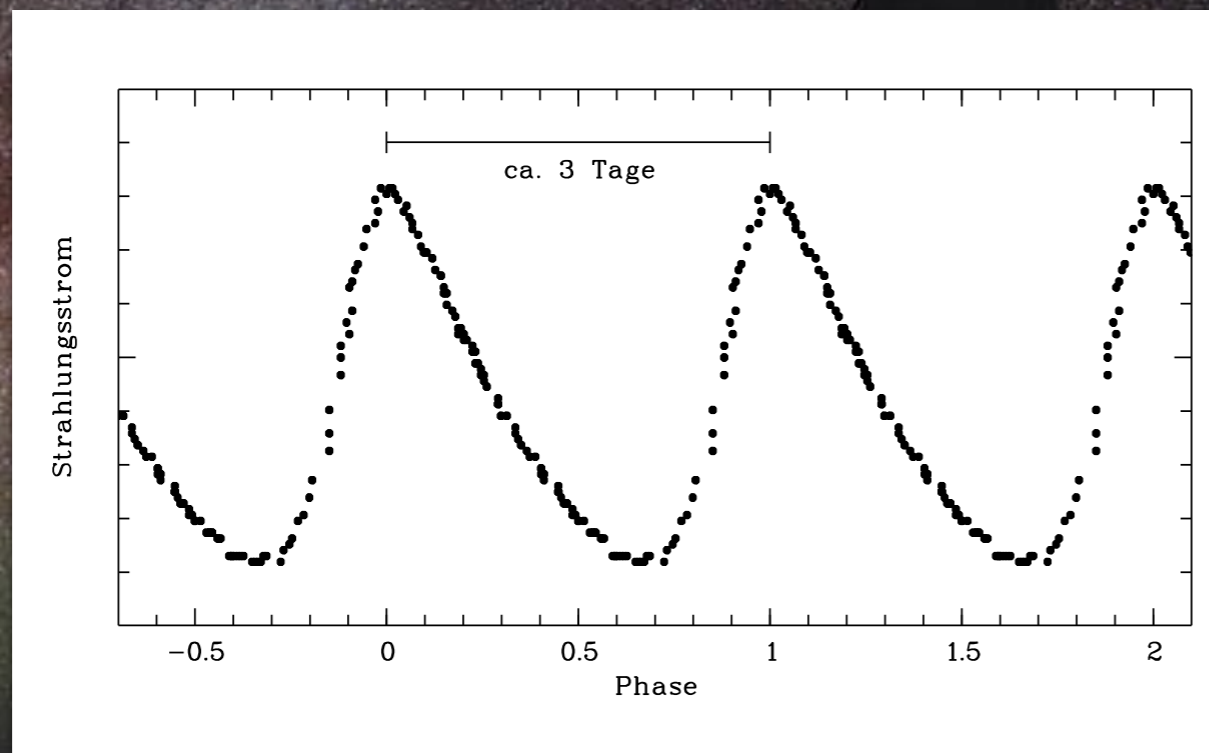


Einzelobjekt  $\Rightarrow$  Durchmusterung (Herschel, Bonner Durchmusterung)

# Fixsterne sind nicht fixiert!

- Sterne bewegen sich
- Veränderliche Sterne
- Novae (Supernovae)
- Parallaxen

Entfernungen,  
⇒ Astronomie  
wird 3D

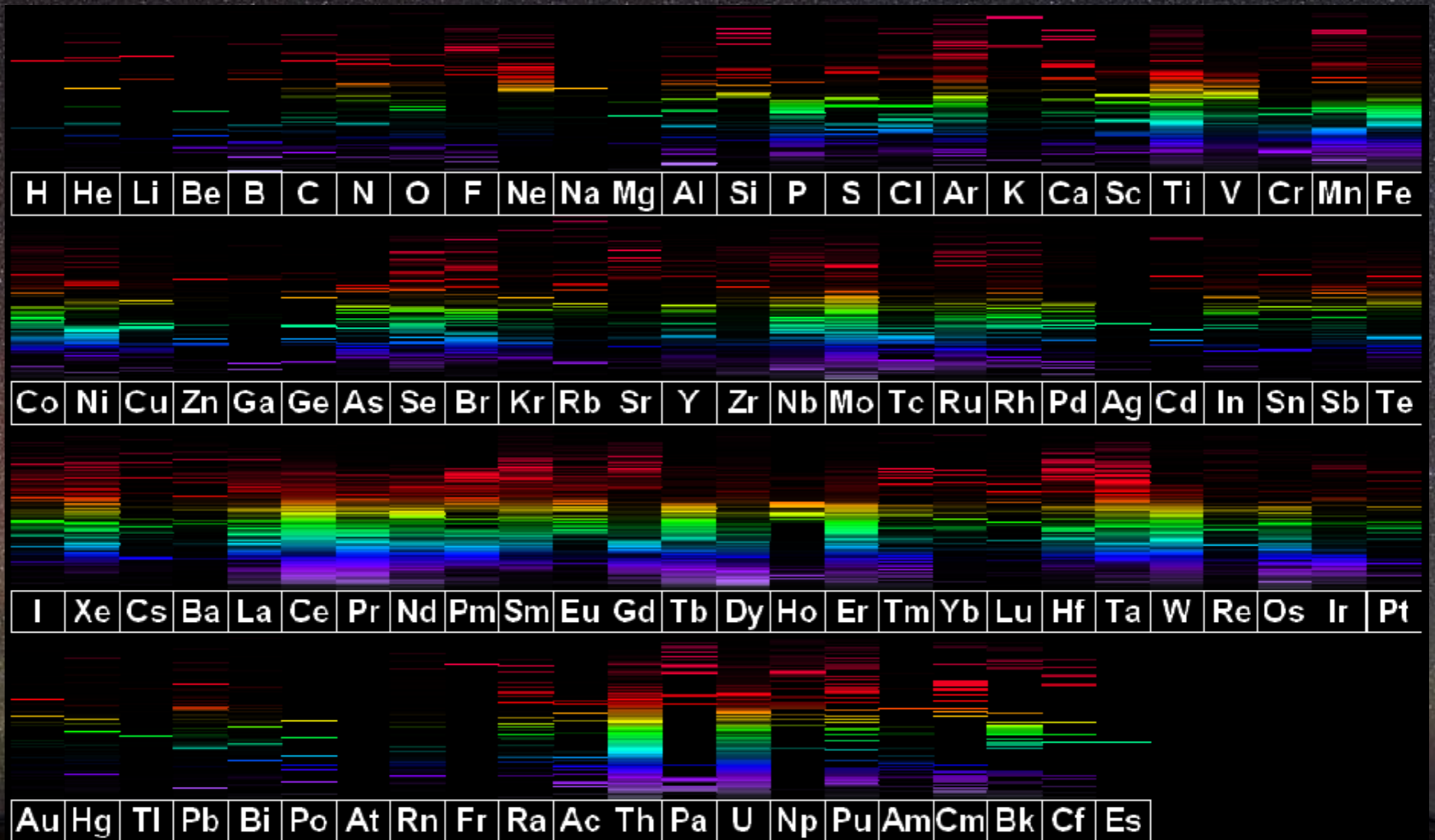


⇒ „time domain“

# Wesentliche Erfindungen im 19.&20. Jahrhundert

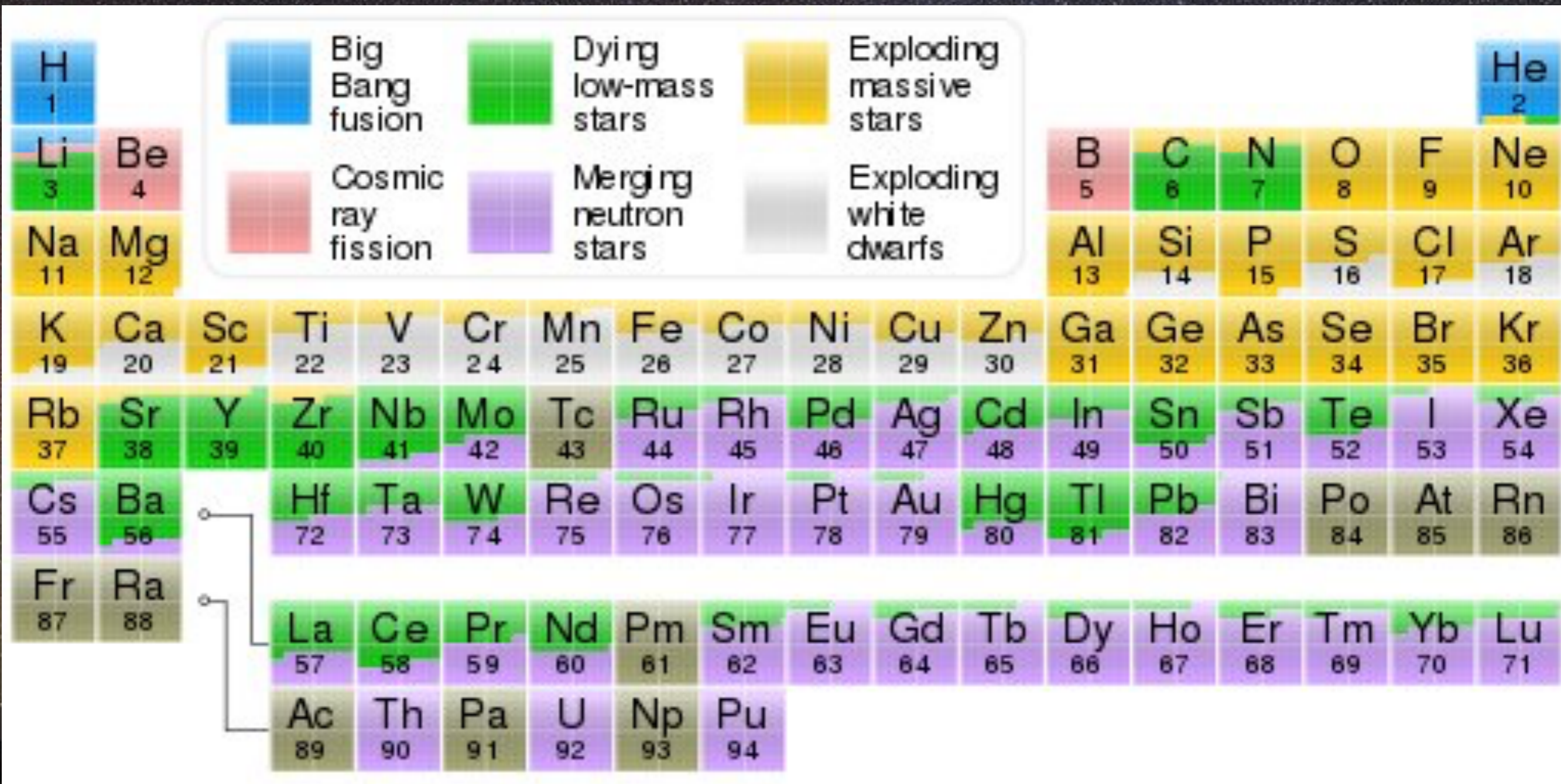
- Spektroskopie (1859) - Astrophysik
  - herausfinden, woraus die Objekte gemacht sind,
  - Bewegungen entlang der Sichtlinie, Astronomie wird 6D
  - Entdeckung des Mediums zwischen Sternen and Galaxien
- Photographie (1850s)
  - Aufnahme von Beobachtungen
  - Analyse nach der Beobachtung (Arbeiten mit Archivdaten)
  - erhöhte Sensitivität
- CCD (1980s)
  - digital Verarbeitung
  - gewaltige Steigerung in der Effizienz
  - lineare Detektoren (Beobachtung von Objekten deutlich lichtschwächer als der Himmel)
- Neue Wellenlängenbereiche (Radio, IR, UV, Röntgen, Gammastrahlung)
- Neue Zugänge (kosmische Strahlung, Neutrinos, Gravitationswellen)

# Beispiel: Ursprung der Elemente



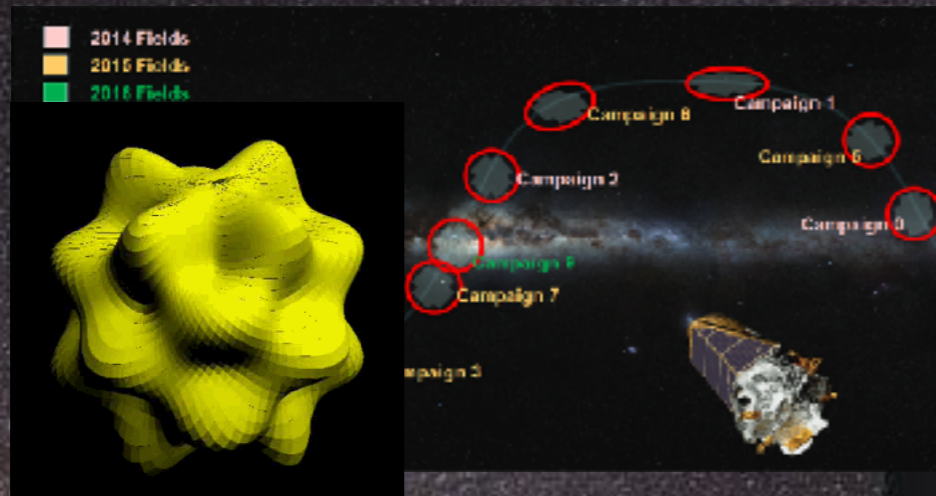


# Beispiel: Ursprung der Elemente



# Präzisionsphotometrie mit Zeitserien

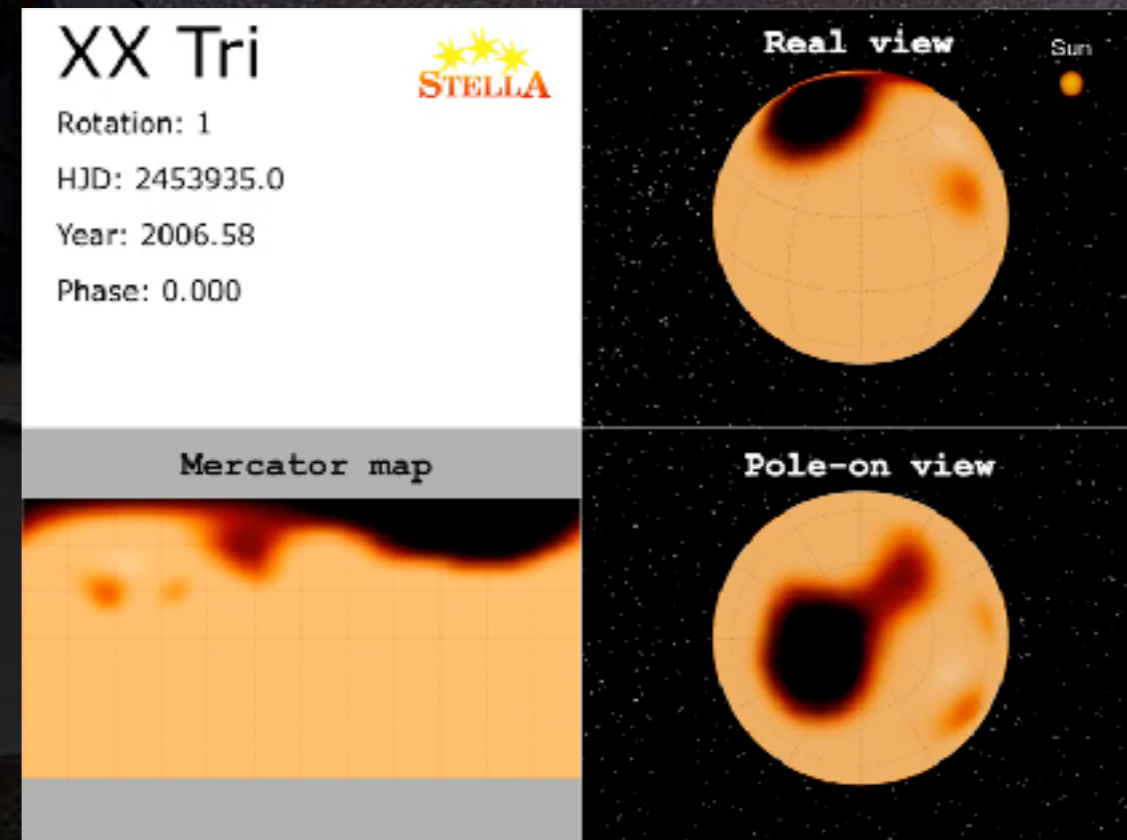
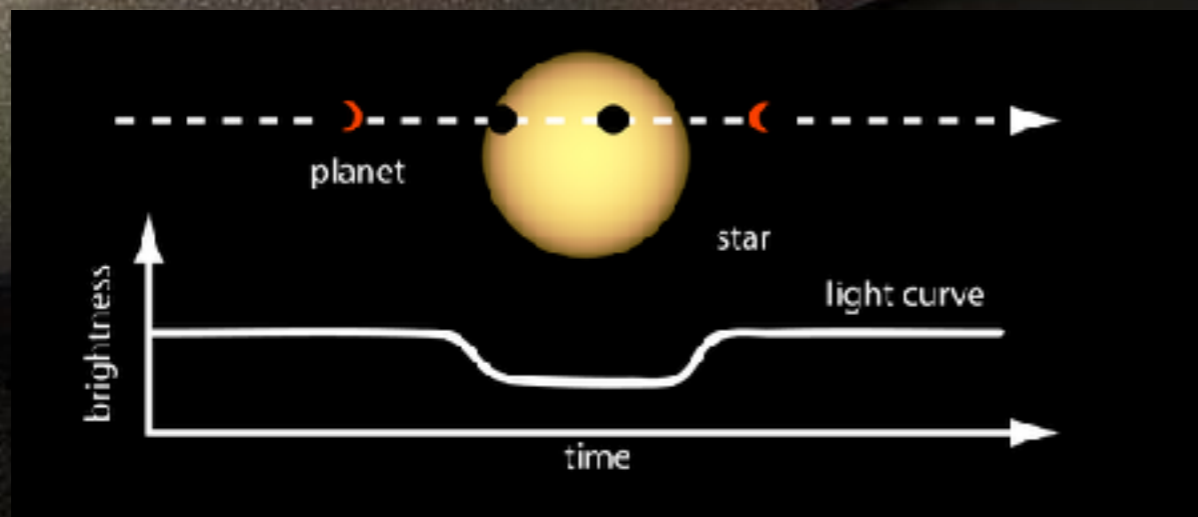
Helioseismologie  $\Rightarrow$  Asteroseismologie



Sternaufbau & Altersbestimmung

## Sternoberflächen

## Entdeckung von Exoplaneten



# Astronomie in den 2010ern - die großen Fragen

- Wann und wie begann das Universum und woraus besteht es?
- Wie haben sich die Objekte des Kosmos (Planeten, Sterne, Galaxien) gebildet und wie entwickeln sie sich?
- Ist unsere Sonne/ unser Planetensystem etwas besonderes? Sind wir allein?
- Inwieweit gelten die Gesetze der Physik anderswo im Kosmos (z.B. schwarze Löcher, Urknall)? Können wir etwas über den Mikrokosmos lernen (Neutrinos, dunkle Materie, dunkle Energie) indem wir den Makrokosmos studieren?



# Denkschrift 2017

## Perspektiven der Astrophysik in Deutschland 2017-2030

Von den Anfängen des Kosmos bis zu Lebensspuren  
auf extrasolaren Planeten

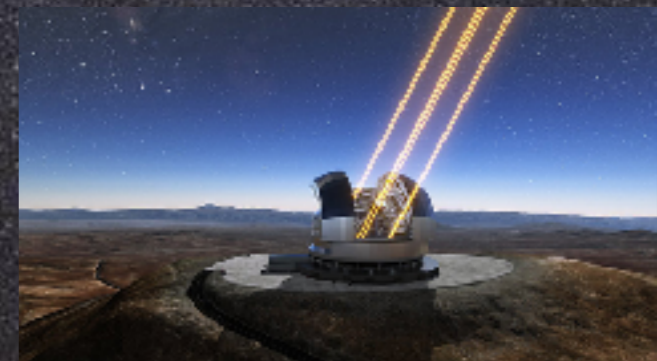
---

Matthias Steinmetz, Marcus Brüggen, Andreas Burkert, Eva Schinnerer, Jürgen Stutzki,  
Linda Taccini, Joachim Wambsgänß, Jörn Wilms (Redaktionskomitee des Rates deutscher Sternwarten)

# Geräte der 2010er: 8m-Teleskope



Klassische Observatorien  
„ausrichten und starren“

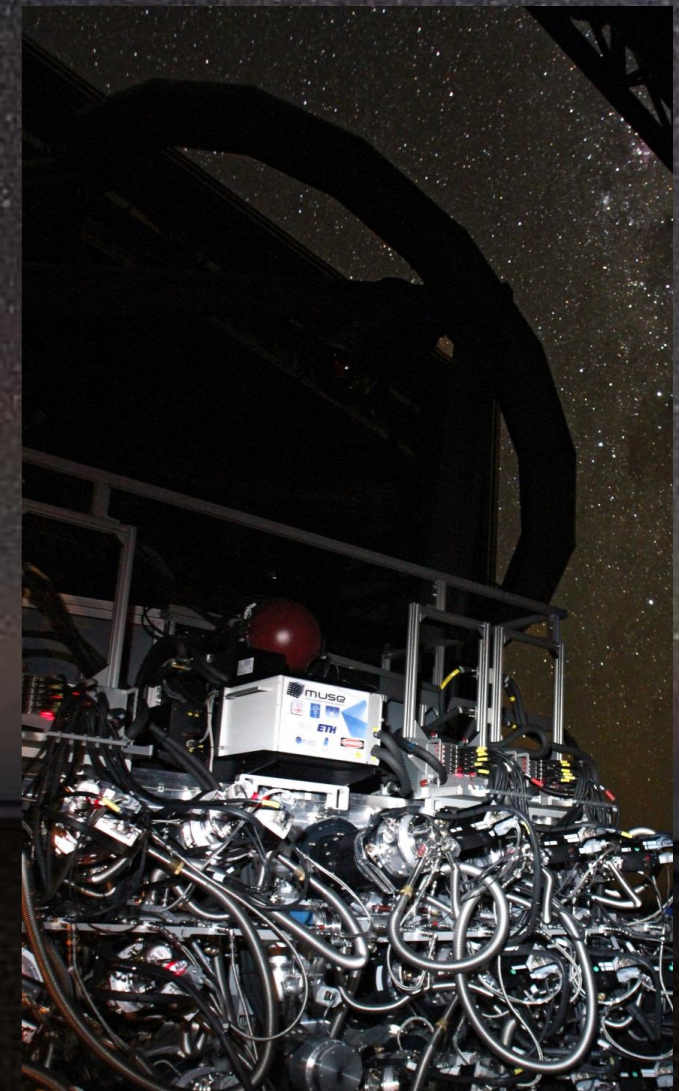
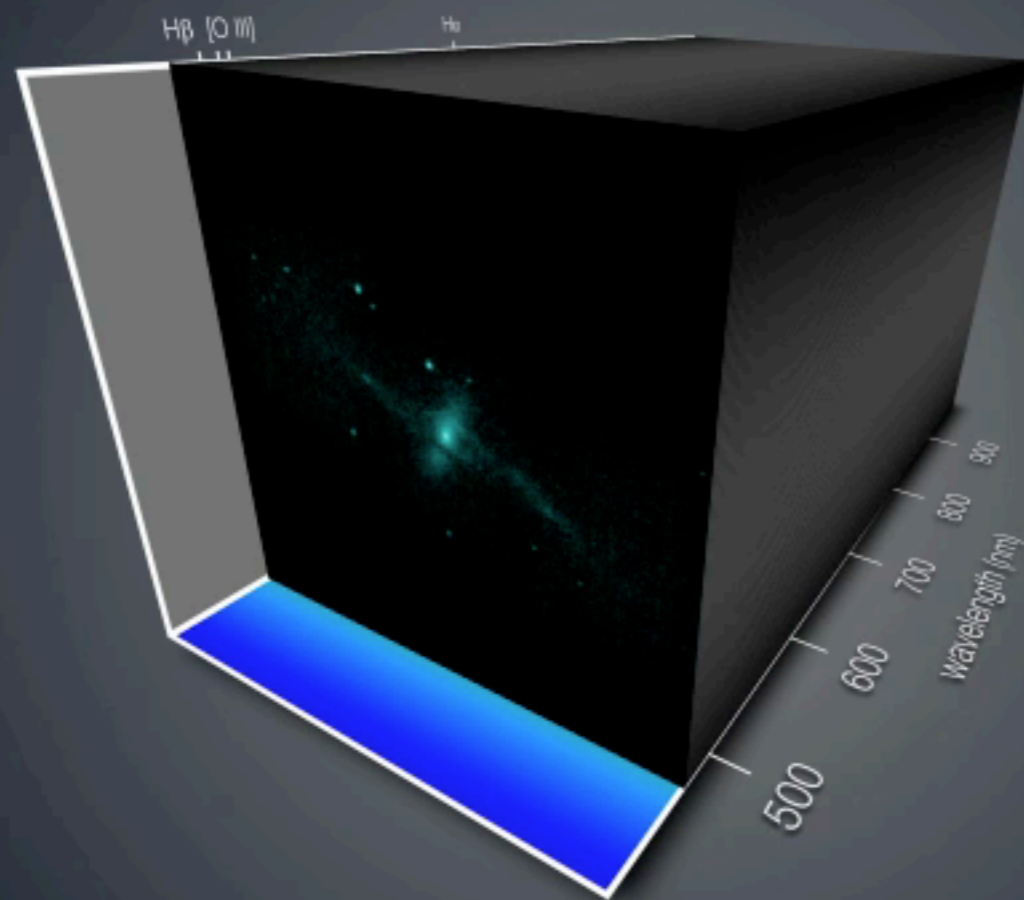
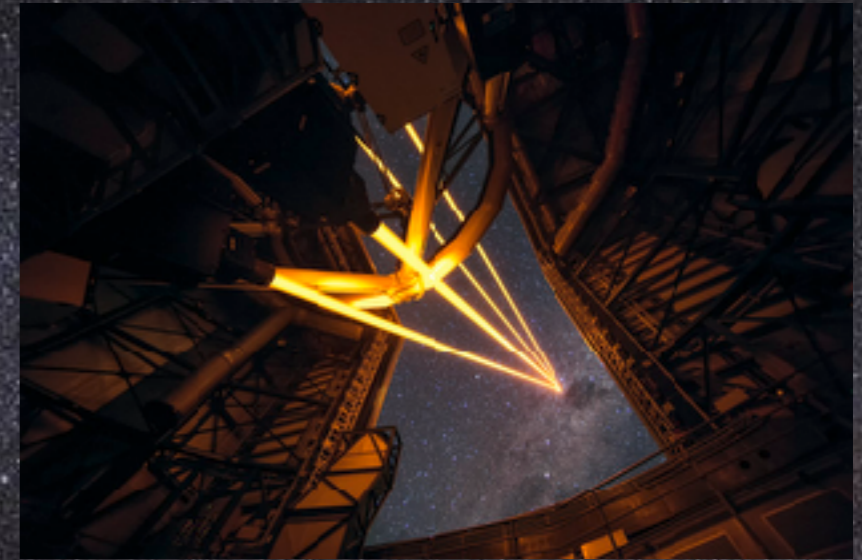


im Bau: 39m ELT



# Geräte der 2010er: MUSE@VLT

Integrale Feld-Spektroskopie:  
verbindet Abbilden und  
Spektroskopie



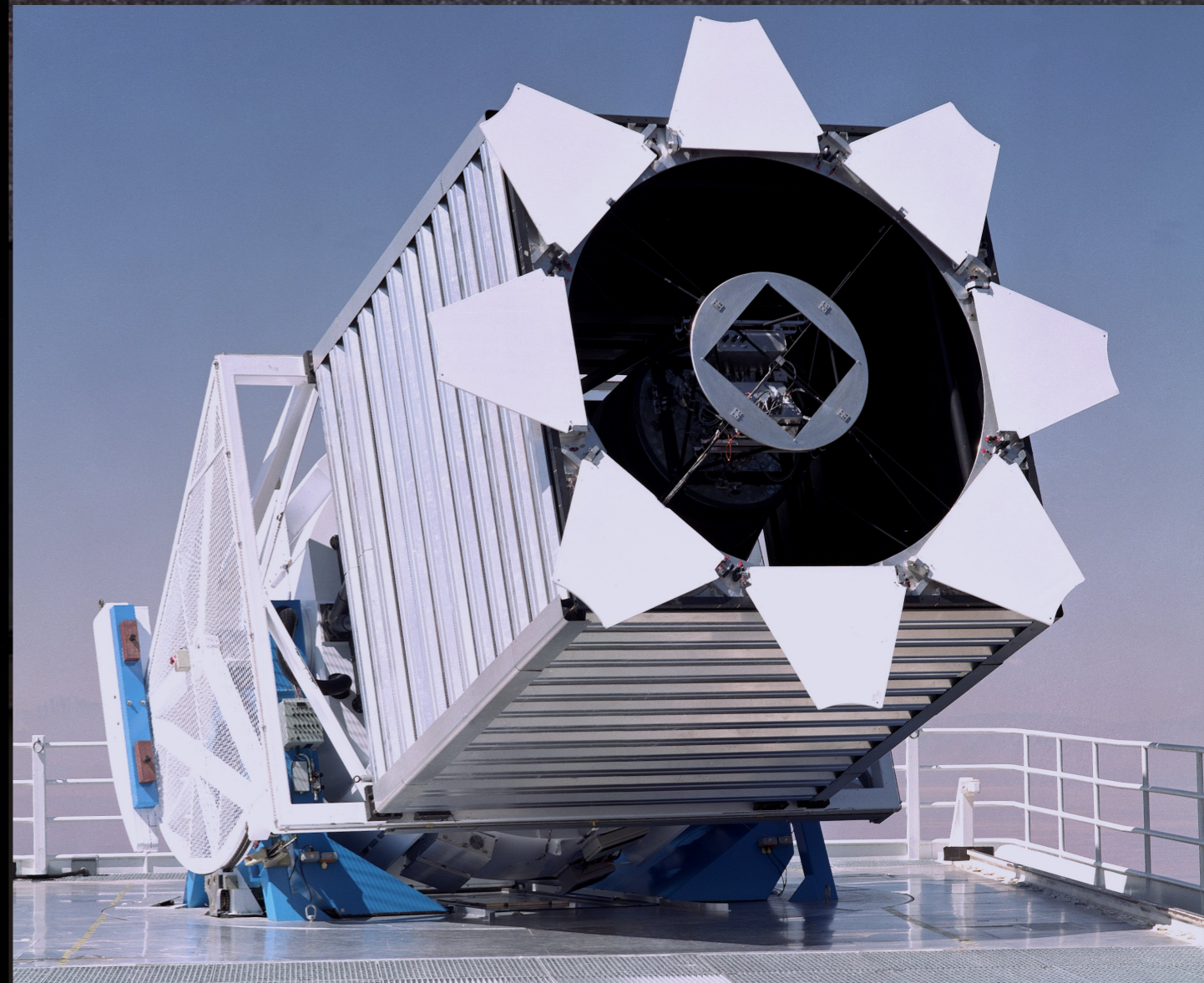
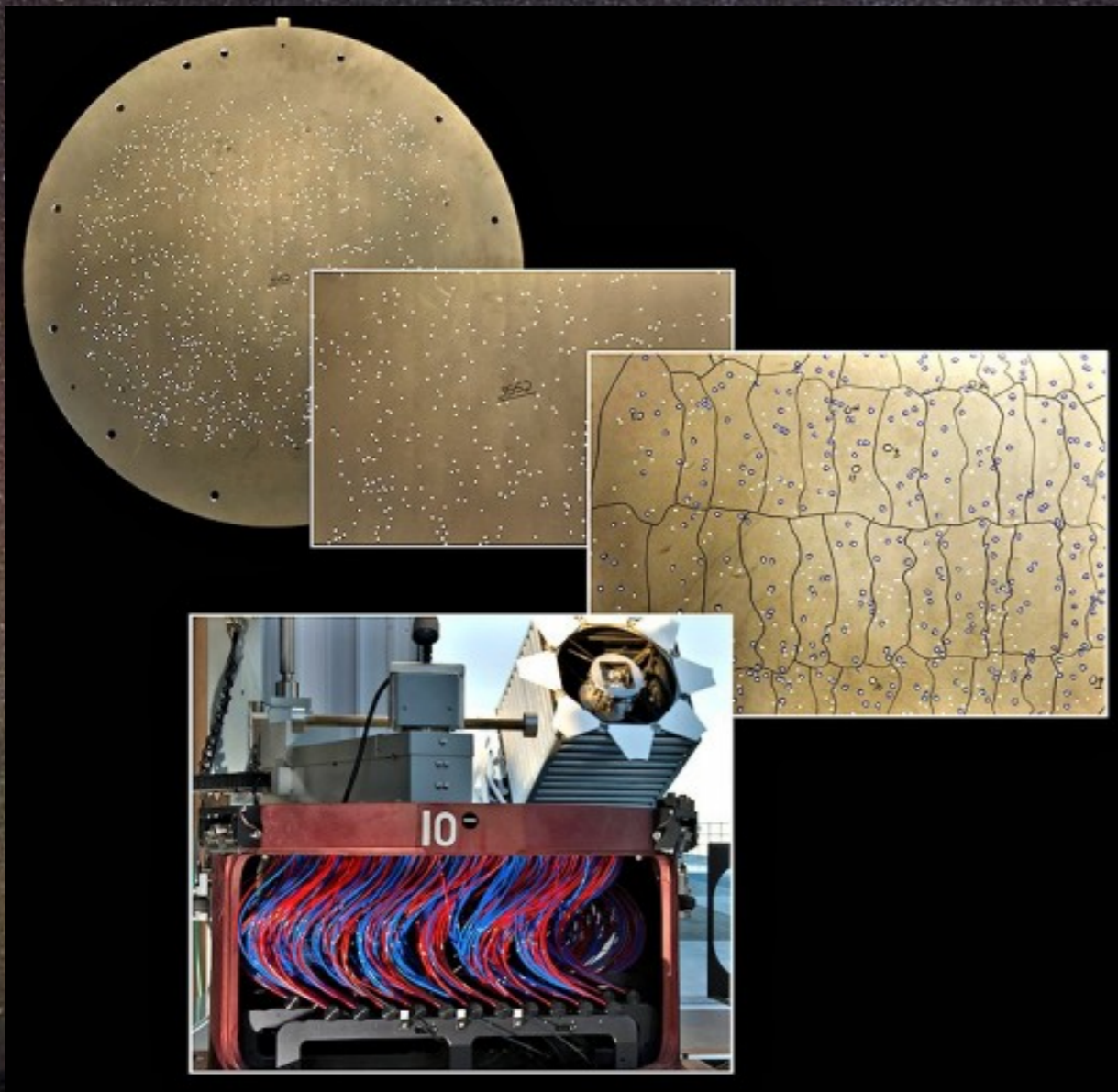
# Geräte der 2010er: „Time domain“

Überwachung: automatische and robotische Teleskope



# Geräte der 2010er: Durchmusterungen

Sloan Digital Sky Survey: Digitale Revolution der Astronomie



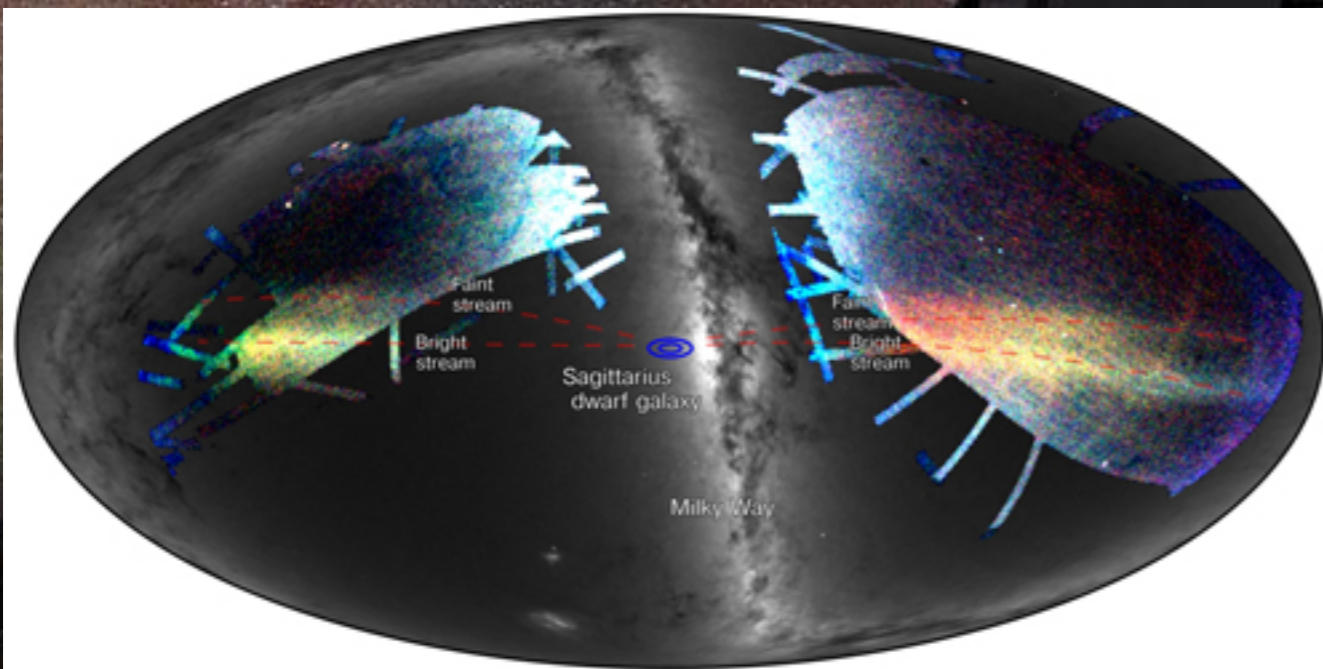


# 18 Jahre Sloan Digital Sky Survey

- 2000-2005 SDSS-I
- 2005-2008 SDSS-II
- 2008-2014 SDSS-III
- 2014-2020 SDSS-IV  
(einschl. Südhimmel)
- 2020-2025 SDSS-V (tbc)

TABLE 1  
HIGH-IMPACT OBSERVATORIES

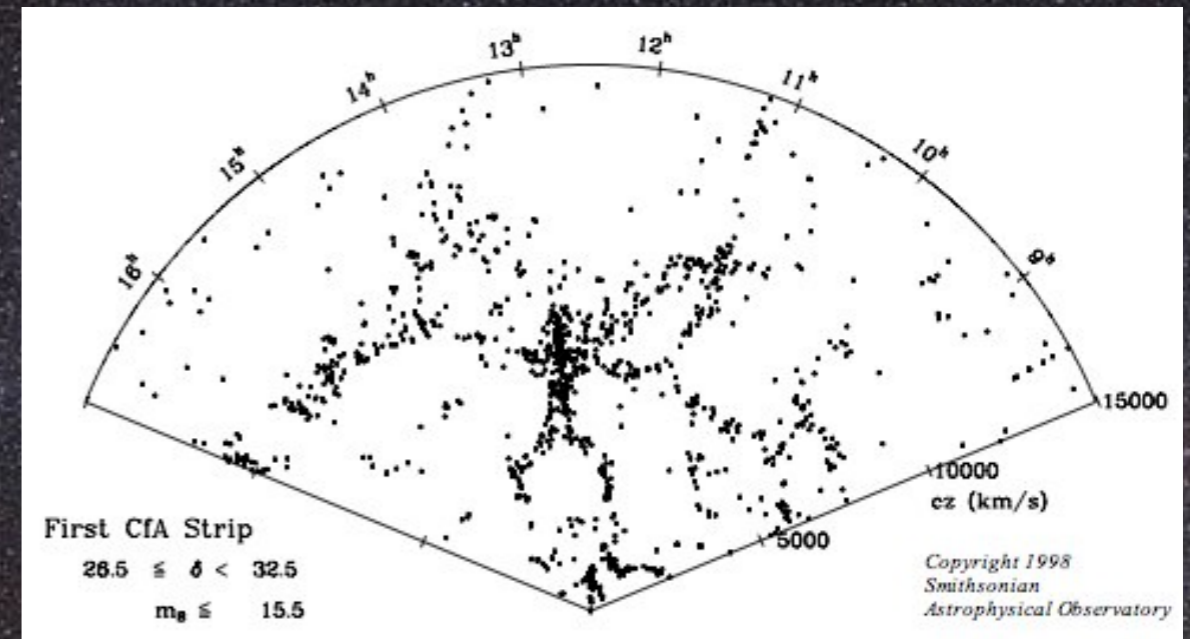
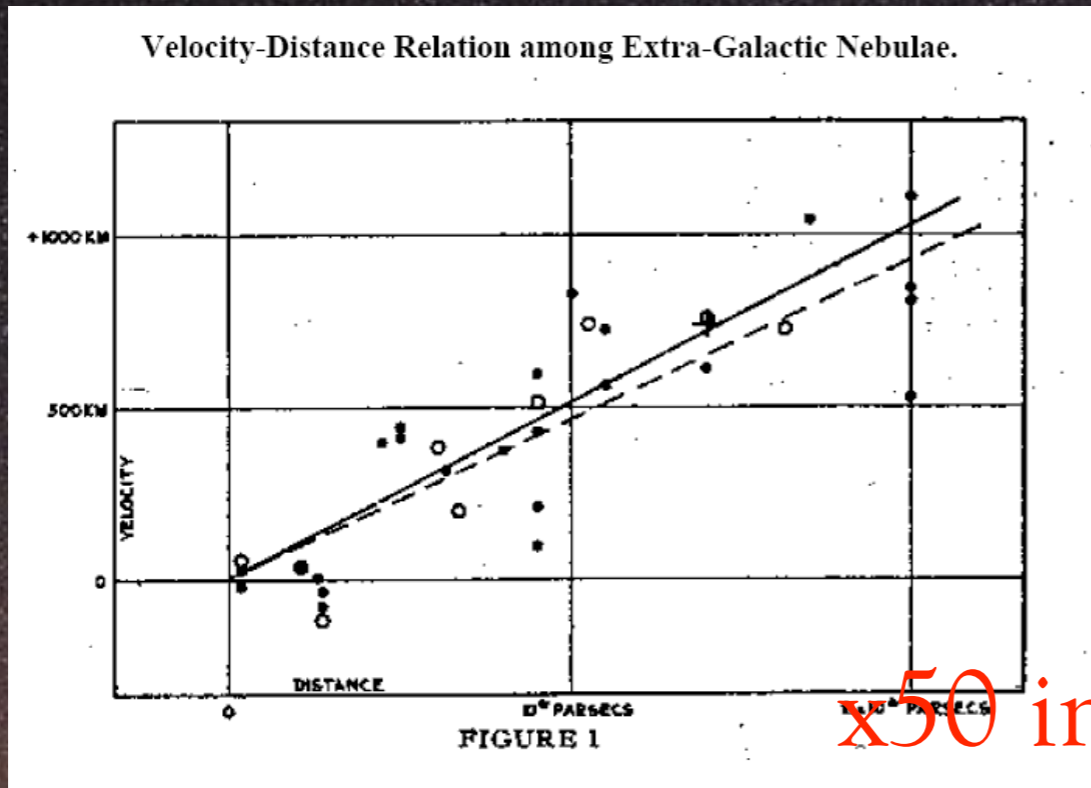
Rank	Facility	Citations	Participation
1	SDSS	1892	14.3%
2	Swift	1523	11.5%
3	HST	1078	8.2%
4	ESO	813	6.1%
5	Keck	572	4.3%
6	CFHT	521	3.9%
7	Spitzer	469	3.5%
8	Chandra	381	2.9%
9	Boomerang	376	2.8%
10	HESS	297	2.2%



Unter den einflussreichsten astronomischen Observatorien:

- 13 Datenveröffentlichungen
- 5000+ wissenschaftliche Veröffentlichungen
- zusammen mehr als 200000 mal zitiert
- h-index: 180

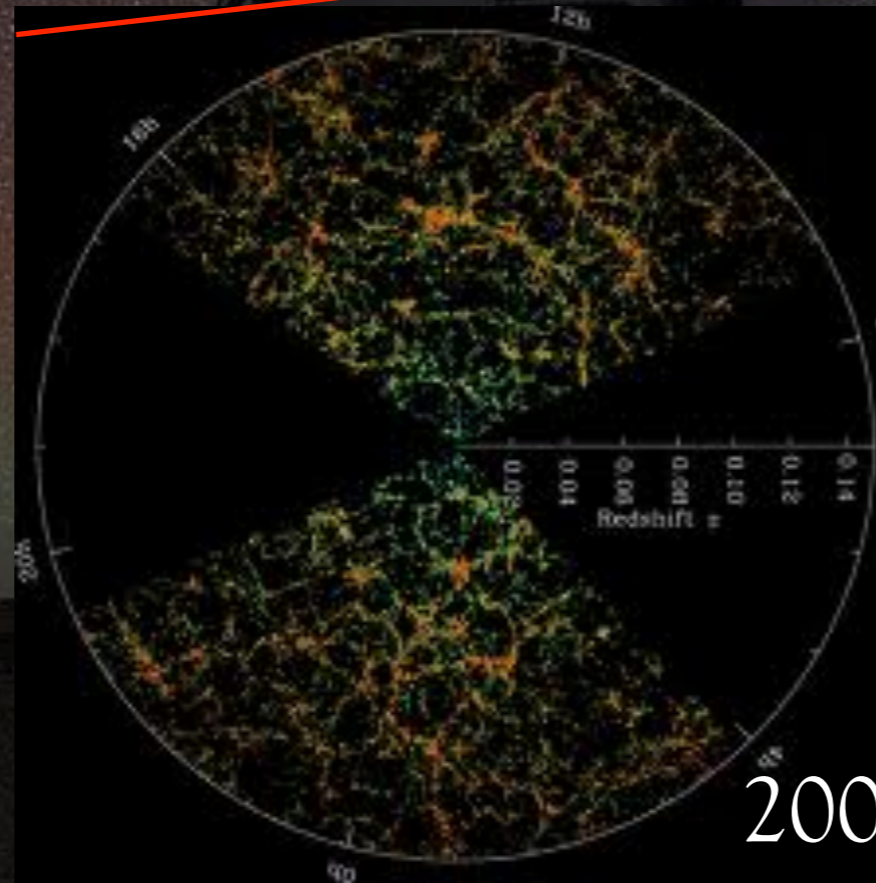
# Zahl der Galaxien, für die wir Rotverschiebungen haben



1929: 24 Galaxien

x50 in 60 Jahren

1986: 1100 Galaxien



x 1000  
in 25 Jahren

2008: 1 Millionen Galaxien

# SDSS: fortgeschrittene Datenprodukte (Kataloge)

- **Katalog:**

- eine Liste der gemessenen Objekte (Sterne, Galaxien)
- gemessene Parameter (Größe, Farbe, Helligkeit)

- **Der Nutzen von Sternkarten:**

- **Entdeckung:** Ist das ein neuer Asteroid oder ist er bereits erfasst?
- **Klassifikation:** Welche Arten von Galaxien gibt es?
- **Populationen:** Ändern Quasare ihre Eigenschaften mit der Zeit?
- **Seltene Objekte:** Ist dies ein merkwürdiges/besonderes Objekt?
- **Kosmologie:** Wie schnell expandiert das Universum?

- **Erfolgsrezept:**

Science ready database - Messungen können analysiert werden ohne komplexe Bildanalysen selbst durchzuführen

★ ASTRONOMERS ★  
★ AT WORK ★

Let's dispel  
some common  
beliefs.

I NEVER USE  
A TELESCOPE

I NEVER  
GO NEAR A  
TELESCOPE

I NEVER  
EVEN  
LOOK UP.

J. WATTS



# Big data means open data!

© Zeljko Ivezic

1. Frühe Datenveröffentlichung verbessern das endgültige Datenprodukt
2. Frühe Datenveröffentlichung ermöglichen gleichzeitig Wissenschaft mit Datenaufnahme
3. Mehr Wissenschaft mit den vorhandenen Daten
4. Ermöglicht, wissenschaftliche Ergebnisse zu reproduzieren
5. Synergien zwischen verschiedenen Datensätzen
6. Fachübergreifende Wissenschaft
7. Mehr Zitierungen und Prestige für das Team
8. Ausbildung und Öffentlichkeitsarbeit

# Warum ist Astronomie „special“ für Big Data

© Jim Gray/Alex Szalay

- große Wirkung auf die Öffentlichkeit
- kein kommerzieller Wert
  - kaum Probleme mit Datenrechten und Datenschutz (jedoch Sicherheit: Satellitenpositionen und Bahnen)
  - ideal, um mit Algorithmen zu experimentieren
- „echte“ Daten mit all den Problemen
  - multi-dimensional
  - verteilt im Raum und Zeit
- verschiedenartig und verteilt
  - viele Observatorien
  - an verschiedenen Orten
  - zu verschiedenen Zeiten
- es gibt sehr viele Astro-Daten (bald 100te von Petabytes)

© Karen Masters

## How to Take Part - Tutorial

Your job is very simple! All you need to do is look out for the features that mark out spiral and elliptical galaxies. In fact, as you're a human and not a computer, most galaxies should be easy to classify since they're obviously spirals or obviously ellipticals. On this page, you will practice classifying galaxies. On the next page, you will take a short trial to test your skills. If you don't pass the trial, you can try again. Once you pass the trial, you can start contributing to Galaxy Zoo science!

### Part 1A ... Spiral or Elliptical Galaxies?

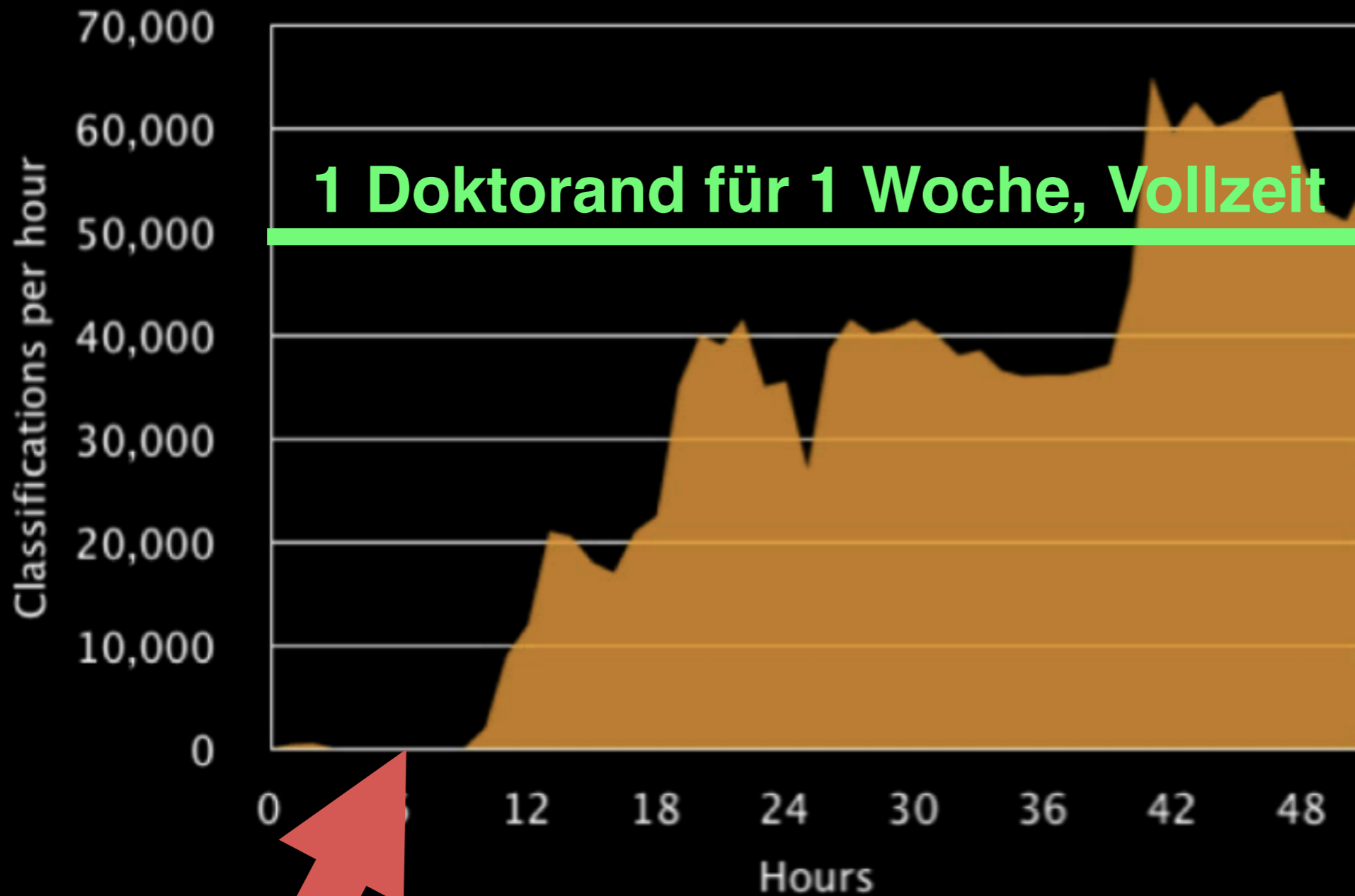


This is a face-on **Spiral Galaxy**. You can clearly see the spiral arms and a central bulge.



This **Elliptical Galaxy** is composed entirely of a bulge of stars. There is no disk or spiral arms.

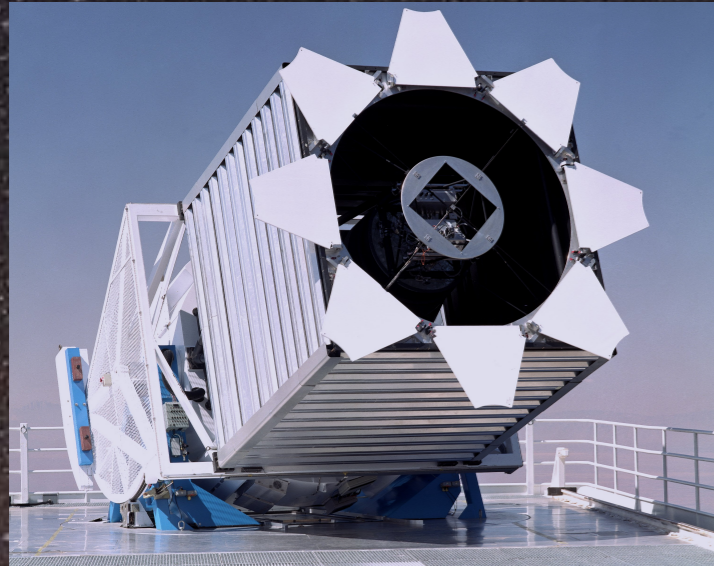
# Die ersten paar Tage



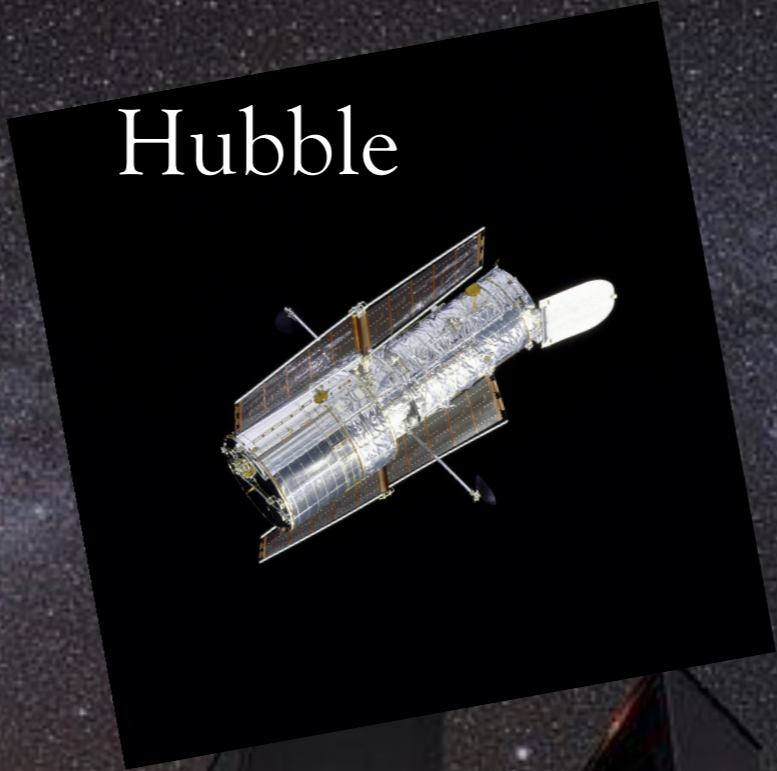
**Server überlastet!**



# GALAXY ZOO



SDSS



Hubble



UKIDSS



Dark Energy Camera



Illustris  
Simulation



# ... harte, publizierte Wissenschaft !

## SPACE

- Galaxy Zoo (44)
- Moon Zoo
- Solar Stormwatch (2)
- Galaxy Zoo - Mergers
- Galaxy Zoo - Supernova (5)
- Planet Hunters (6)
- Milky Way Project (2)
- Ice Hunters (1)
- Andromeda Project
- Planet Four
- Space Warps
- Radio Galaxy Zoo
- Stardate M83
- Disk Detective
- Sunspotter

## CLIMATE

- Old Weather (3)
- Cyclone Center

## HUMANITIES

- Ancient Lives (1)
- Operation War Diary

## NATURE

- Whale FM (2)
- Seafloor Explorer



**Galaxy Zoo: The Ultraviolet Attenuation Law in Backlit Spiral Galaxies, Keel+ 2014. Available here**



**Galaxy Zoo: an independent look at the evolution of the bar fraction over the last eight billion years from HST-COSMOS, Melvin+ 2014. Available here**



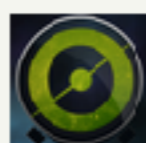
**The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies, Schawinski+ 2014. Available here**



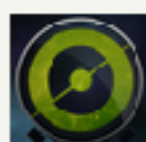
**Galaxy Zoo: quantifying morphological indicators of galaxy interaction, Casteels+ 2013. Available here**



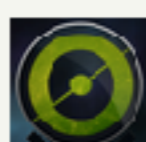
**Galaxy Zoo: Observing Secular Evolution Through Bars, Cheung+ 2013. Available here**



**Galaxy Zoo: A Catalog of Overlapping Galaxy Pairs for Dust Studies, Keel+ 2013. Available here**



**Galaxy Zoo: Motivations of Citizen Scientists, Raddick+ 2013. Available here**



**The different star-formation histories of blue and red spiral and elliptical galaxies, Tojeiro+ 2013. Available here**



# Die Suche nach Gravitationswellen

**EINSTEIN@HOME**  
EINSTEIN@HOME  
Catch a Wave From Space

About | Get Involved | Gravitational Waves | Ask a Physicist | In the News | Resources | Contact | Home

## Welcome to Einstein@Home

[Join Einstein@Home](#)

According to Albert Einstein, we live in a universe full of gravitational waves. He suggested that the movements of heavy objects, such as black holes and dense stars, create waves that change space and time. We have a chance to detect these waves, but we need your help to do it!

Einstein@Home uses computer time donated by computer owners all over the world to process data from gravitational wave detectors. Participants in Einstein@Home download software to their computers, which process gravitational wave data when not being used for other computer applications, like word processors or games. Einstein@Home doesn't affect the performance of computers and greatly speeds up this exciting research.

[Learn more about the project.](#)

Einstein@Home screensaver

APS  
physics

# Space

Sort by



## How do galaxies form?

NASA's Hubble Space Telescope archive provides hundreds of thousands of galaxy images.

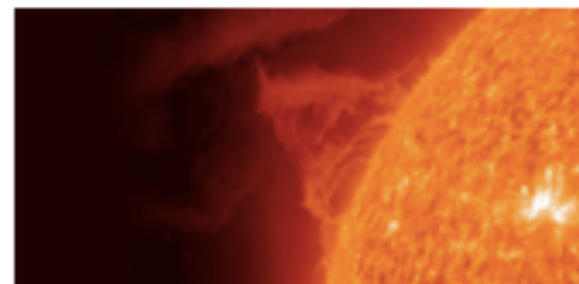
GALAXY ZOO



## Explore the surface of the Moon

We hope to study the lunar surface in unprecedented detail.

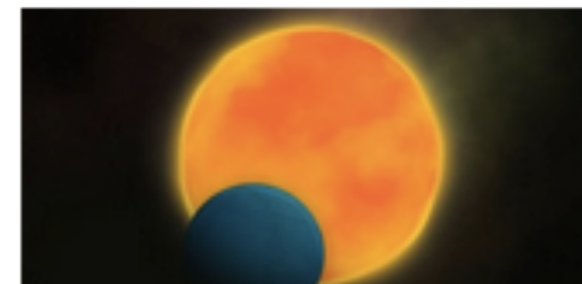
MOON ZOO



## Study explosions on the Sun

Explore interactive diagrams to learn about the Sun and the spacecraft monitoring it.

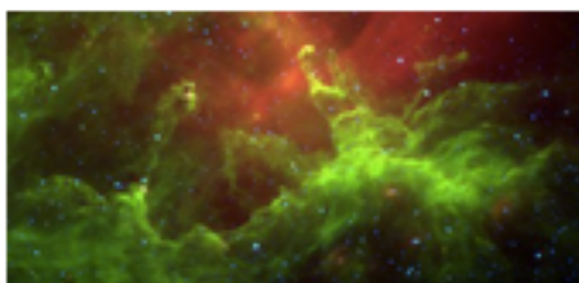
SOLAR STORMWATCH



## Find planets around stars

Lightcurve changes from the Kepler spacecraft can indicate transiting planets.

planethunters.org



## How do stars form?

We're asking you to help us find and draw circles on infrared image data from the Spitzer Space Telescope.

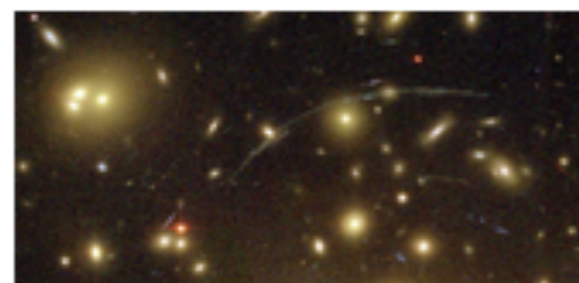
THE MILKYWAY PROJECT



## Explore the Red Planet

Planetary scientists need your help to discover what the weather is like on Mars.

PLANET FOUR



## Help us find gravitational lenses

Imagine a galaxy, behind another galaxy. Think you won't see it? Think again.

SPACEWARPS



@kevinschawinski

## Climate



### Model Earth's climate using historic ship logs

Help scientists recover Arctic and worldwide weather observations made by US Navy and Coast Guard ships.

Weather



### Classify over 30 years of tropical cyclone data.

Scientists at NOAA's National Climatic Data Center need your help.

CycloneCenter

## Humanities

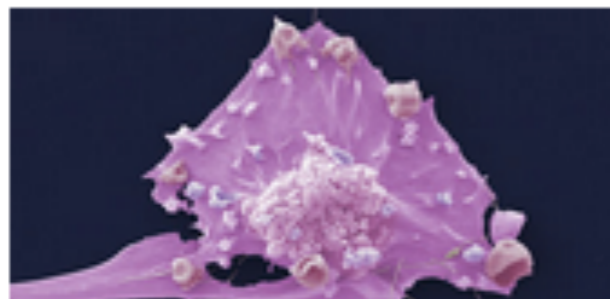


### Study the lives of ancient Greeks

The data gathered by Ancient Lives helps scholars study the Oxyrhynchus collection.

ANCIENT LIVES

## Biology



### Analyse real life cancer data.

You can help scientists from the world's largest cancer research institution find cures for cancer.

Cell Slider



### Track genetic mysteries

We can better understand how our genes work by spotting the worms laying eggs.

WORM WATCH LAB

# Citizen Science

© Kevin Schawinski

200 billion hours  
a year spent watching TV by US adults

source: BBC

Gesamtaufwand bis heute,  
Wikipedia zu befüllen

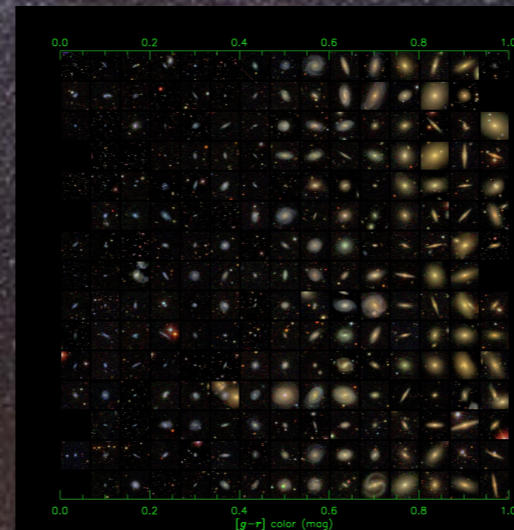
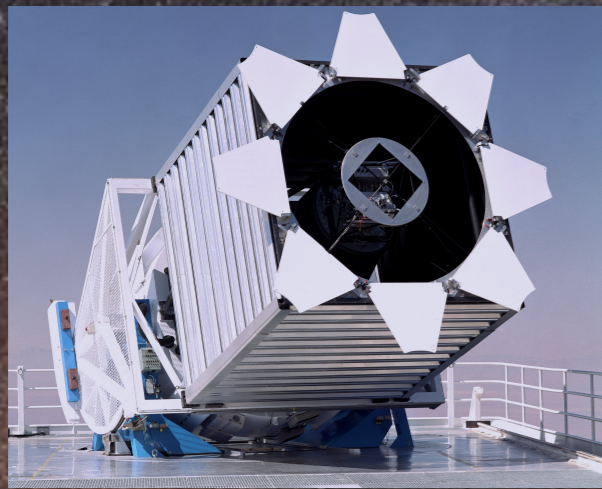


# Nächstes Jahrzehnt: LSST

(Large Synoptic Survey Telescope):

Verbindung von Durchmusterung und time domain

SDSS machte ein Farbbild des Himmels



LSST wird ein Farbfilm des Himmels machen



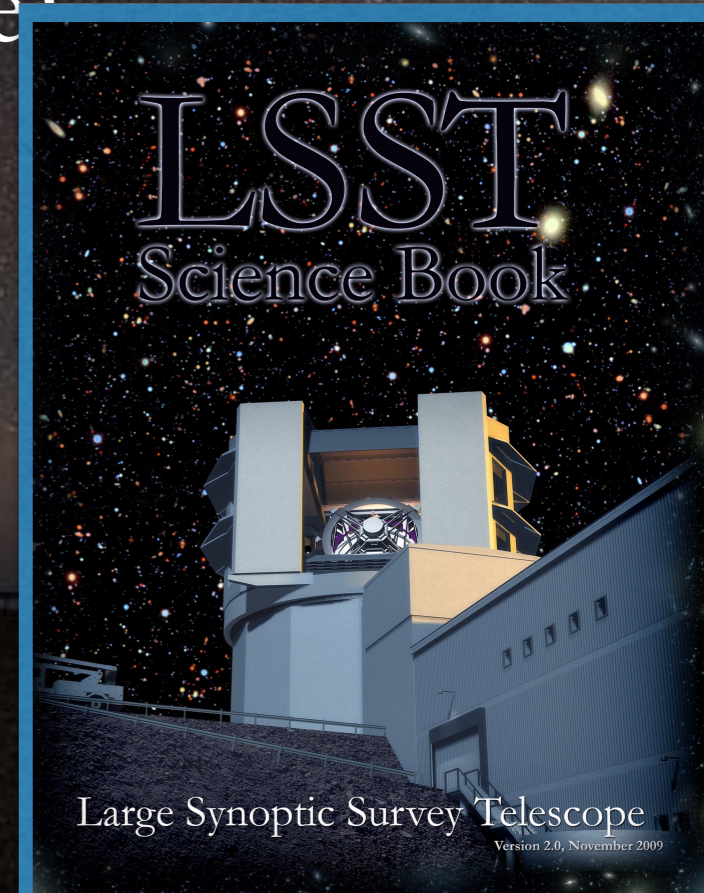
Eine optische/nah-IR Durchmusterung es halben Himmels in 6 Farben bis zu  $r \sim 27.5$  basierend auf  $\sim 1000$  Wiederholungen über eine 10-Jahres-Periode

Ein Katalog mit 20 Milliarden Sternen und 10 Milliarden Galaxien und exquisiter Photometrie, Astrometrie and Bildqualität!



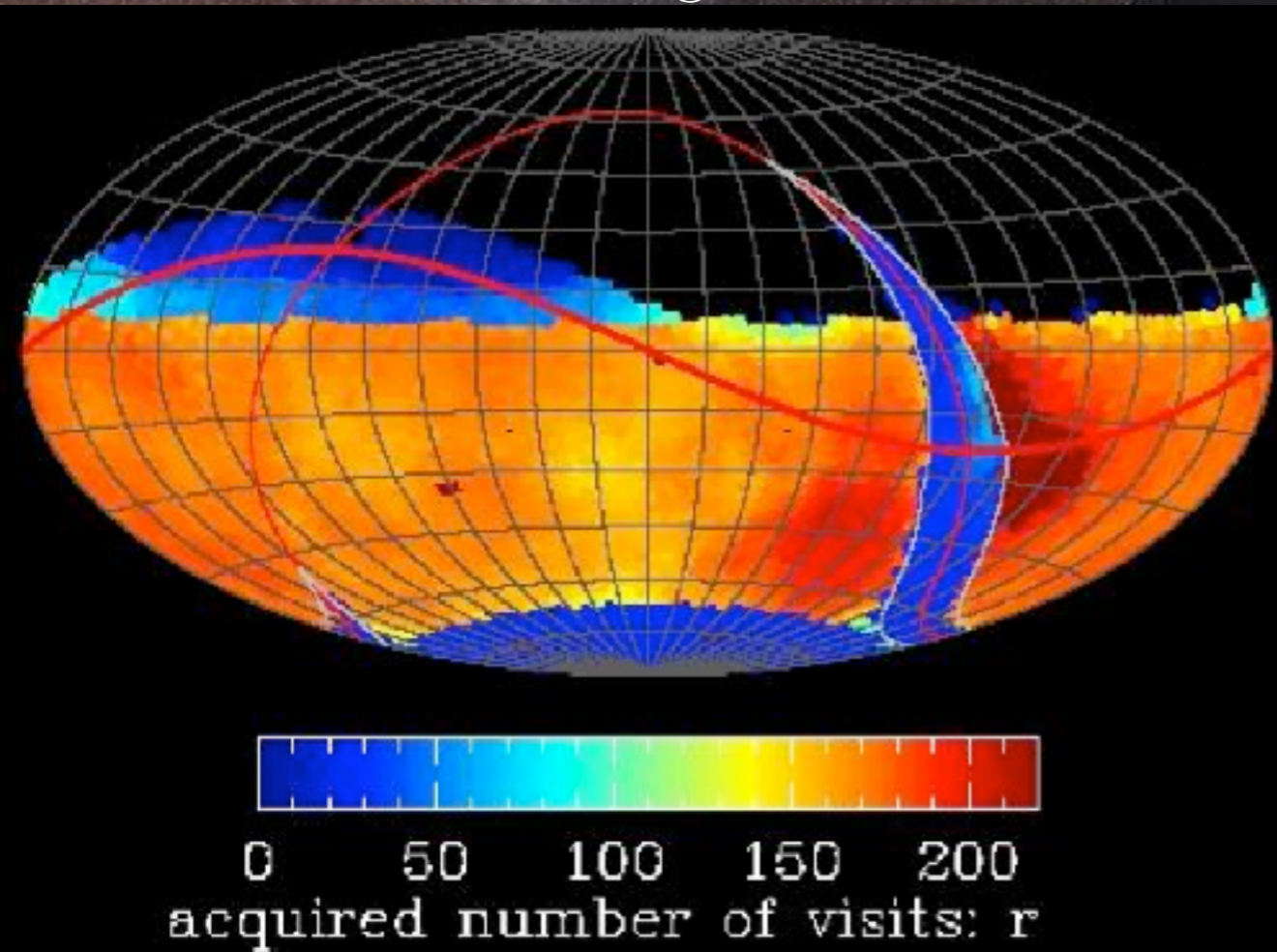
# LSST Wissenschaftsthemen

- Dunkle Materie, dunkle Energie, Kosmologie (räumliche Verteilung der Galaxien, Gravitationslinsen, Supernovae, Quasare)
- Time domain (kosmische Explosionen, Veränderliche Sterne)
- Aufbau des Sonnensystems (Asteroide)
- Aufbau der Milchstraße (Sterne)



# Idee hinter LSST: einheitliche Durchmusterung

- 90% der Beobachtungszeit gehen an einen einheitliche Durchmusterung: alle 3-4 Nächte wird der gesamte Himmel 2 mal pro Nacht abgetastet
- Nach 10 Jahren wird der halbe Himmel etwas 1000 mal abgebildet (in 6 Farben): ein digitaler Farbfilm des Himmels!
- ~100 PB an Daten: etwa 1 Milliarde 16 Mpix Bilder, erlaubt die Vermessung von 40 Milliarden Objekten!



**Links:** Simulation einer 10-jährigen LSST-Durchmusterung: Farbe entspricht der Zahl der Wiederholungen im r-Band (Aitoff Projektion, äquatoriale Koordinaten)

# Vergleich der Gesichtsfelder: Gemini vs. LSST

Primary Mirror Diameter

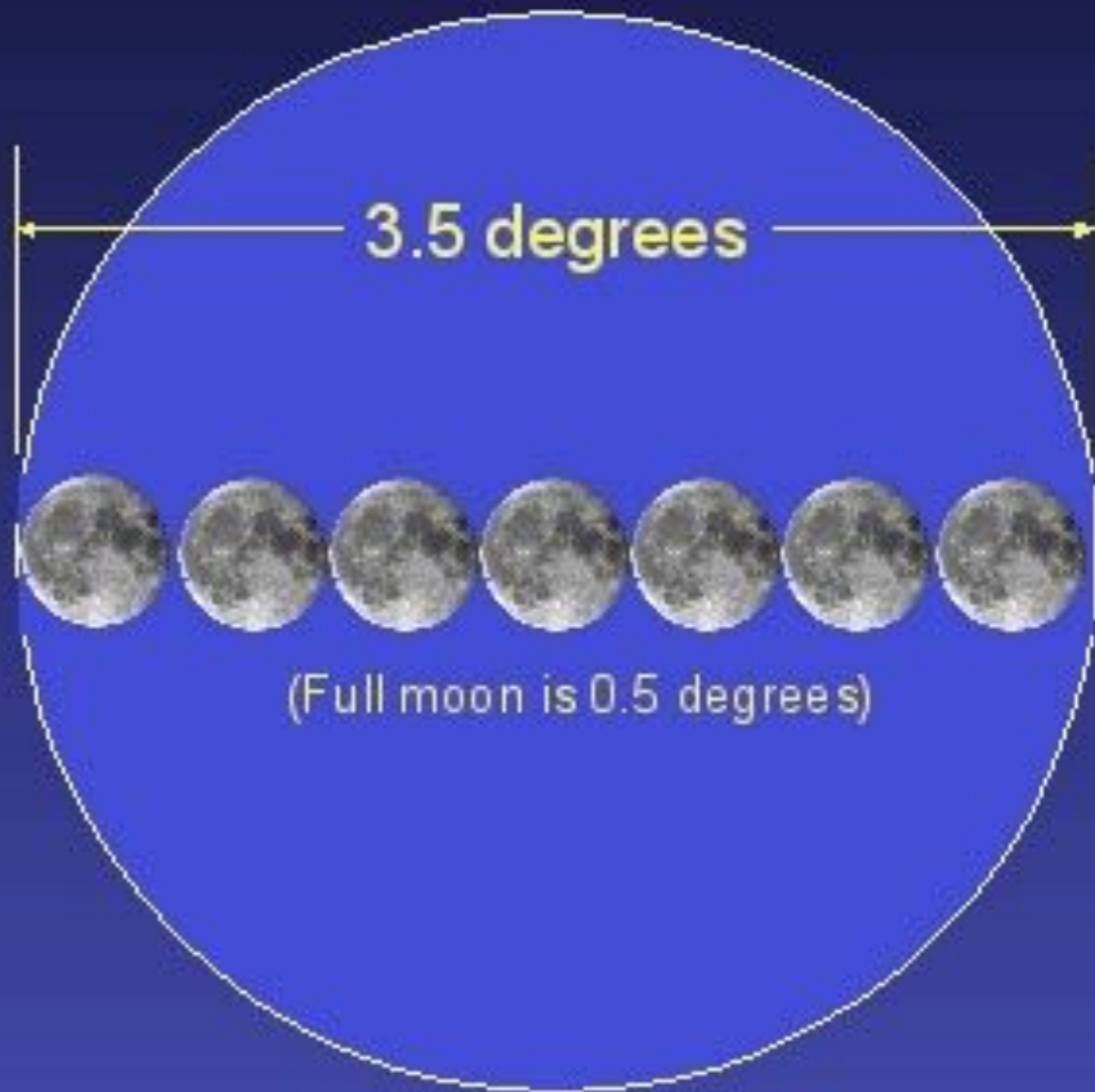
Field of View



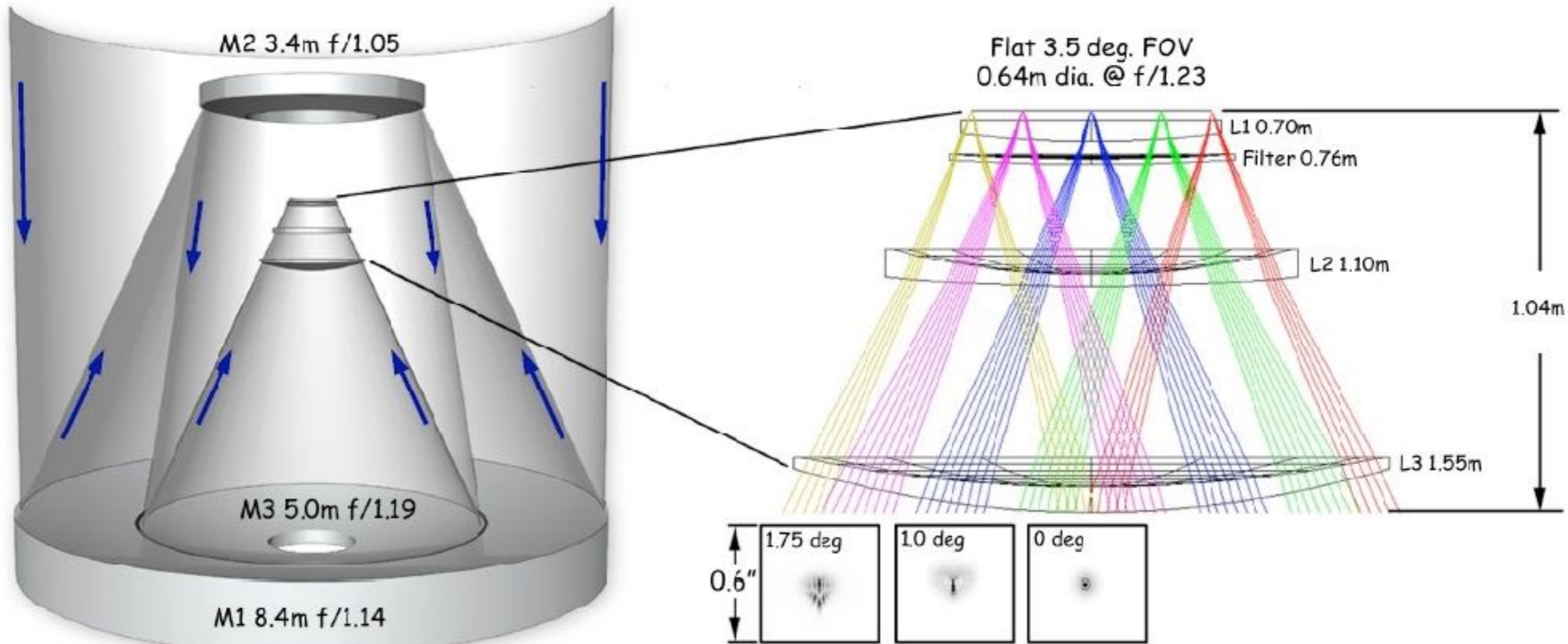
Gemini South Telescope



LSST



# Optisches Design des LSST



## Drei-Spiegel-System nach Paul-Baker

- großes Gesichtsfeld mit exzellenter Bildqualität:
- Bildqualität dominiert von atmosphärischem Seeing



# Large Synoptic Survey Telescope

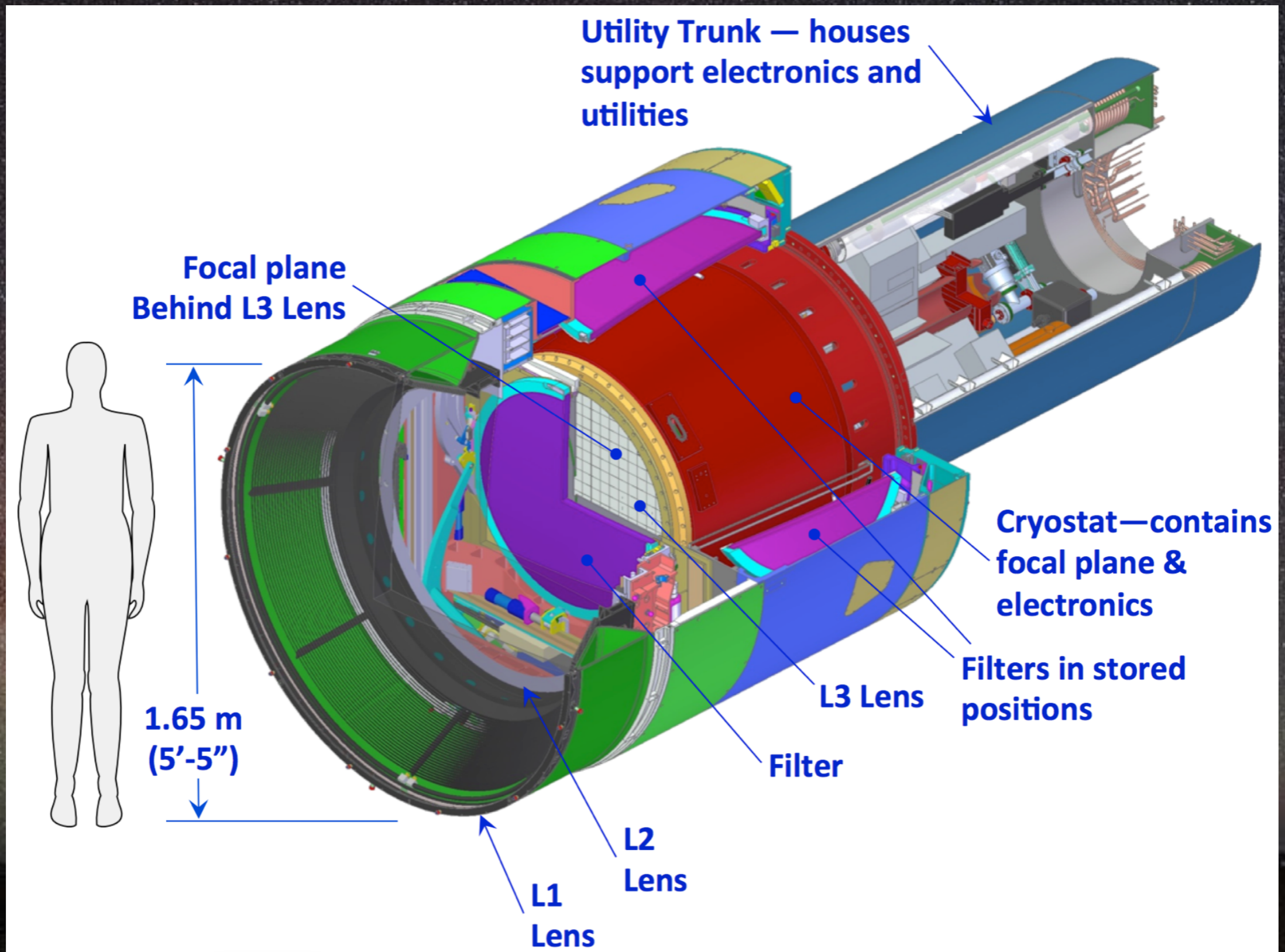


# LSST Telescope

Erstes Licht: 2019

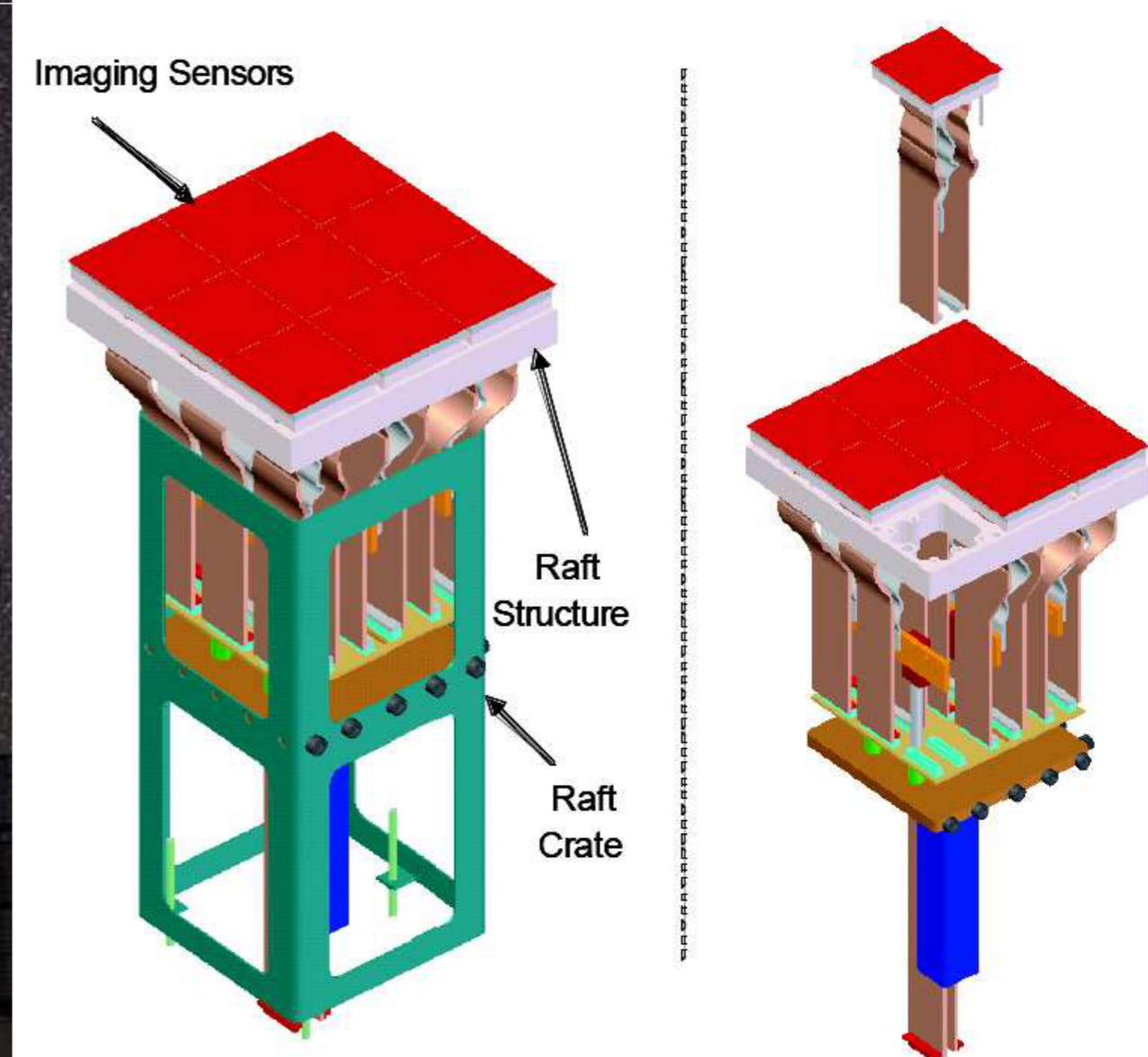
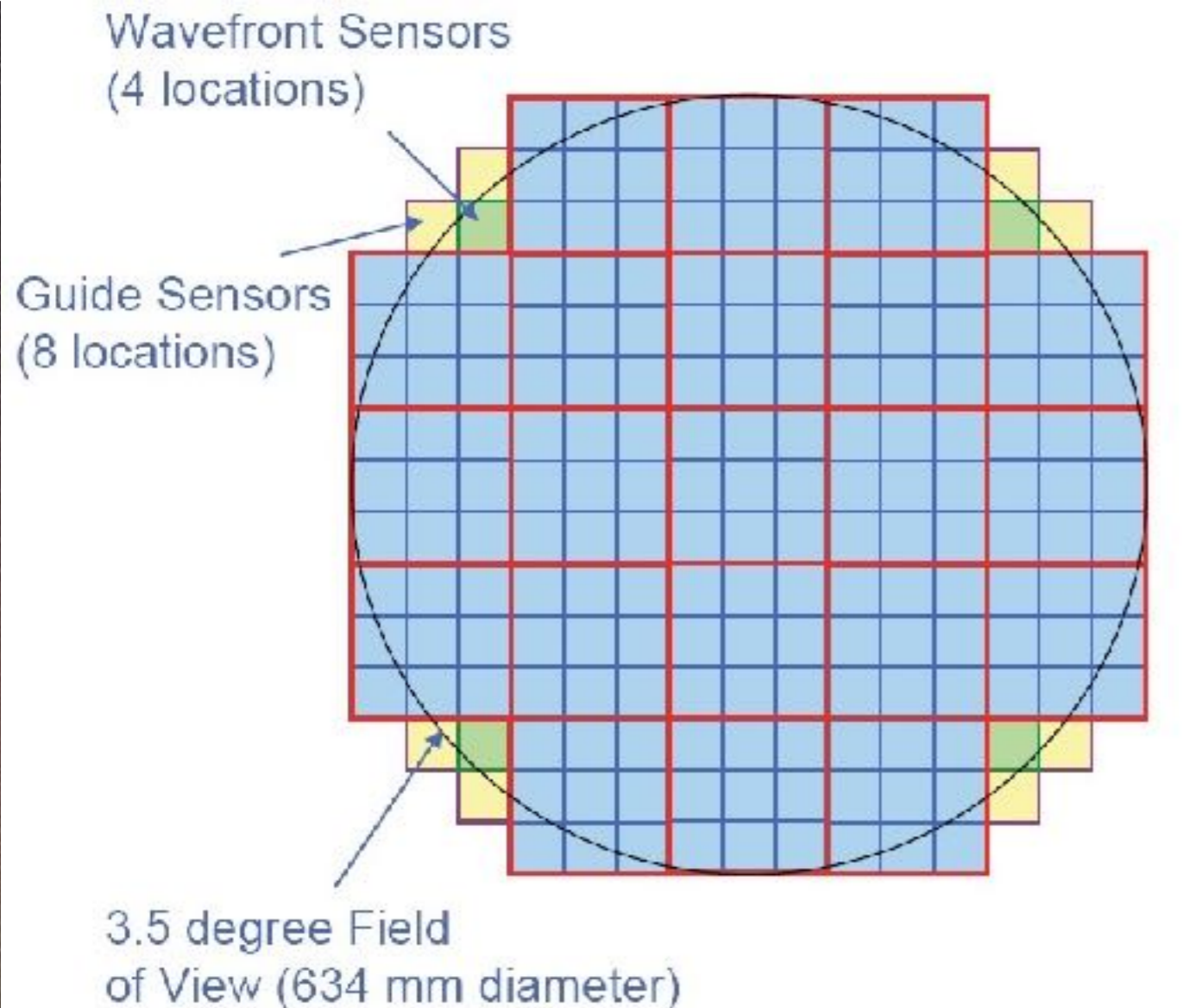


# LSST camera



Größte astronomische Kamera: 2800 kg, 3.2 Gpix

# LSST Kamera



- Modulares Design: 3200 Megapix = 189 x16 Megapix CCD
- 9 CCDs gemeinsame Elektronik: raft (=camera)
- Problematische rafts können recht einfach ausgetauscht werden



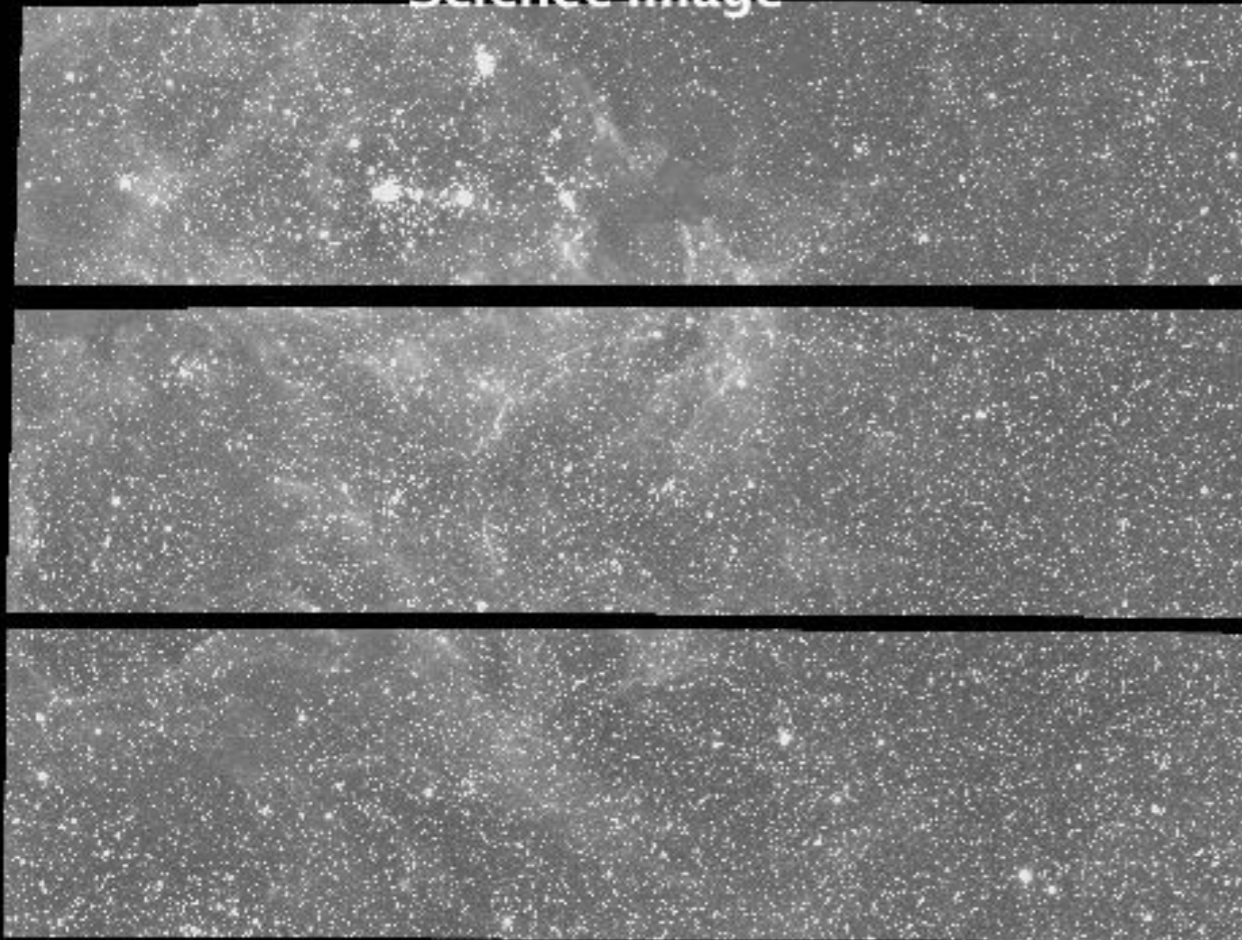
# Time Domain: Objekte ändern sich!

**Positionen:** Asteroide und Eigenbewegung der Sterne

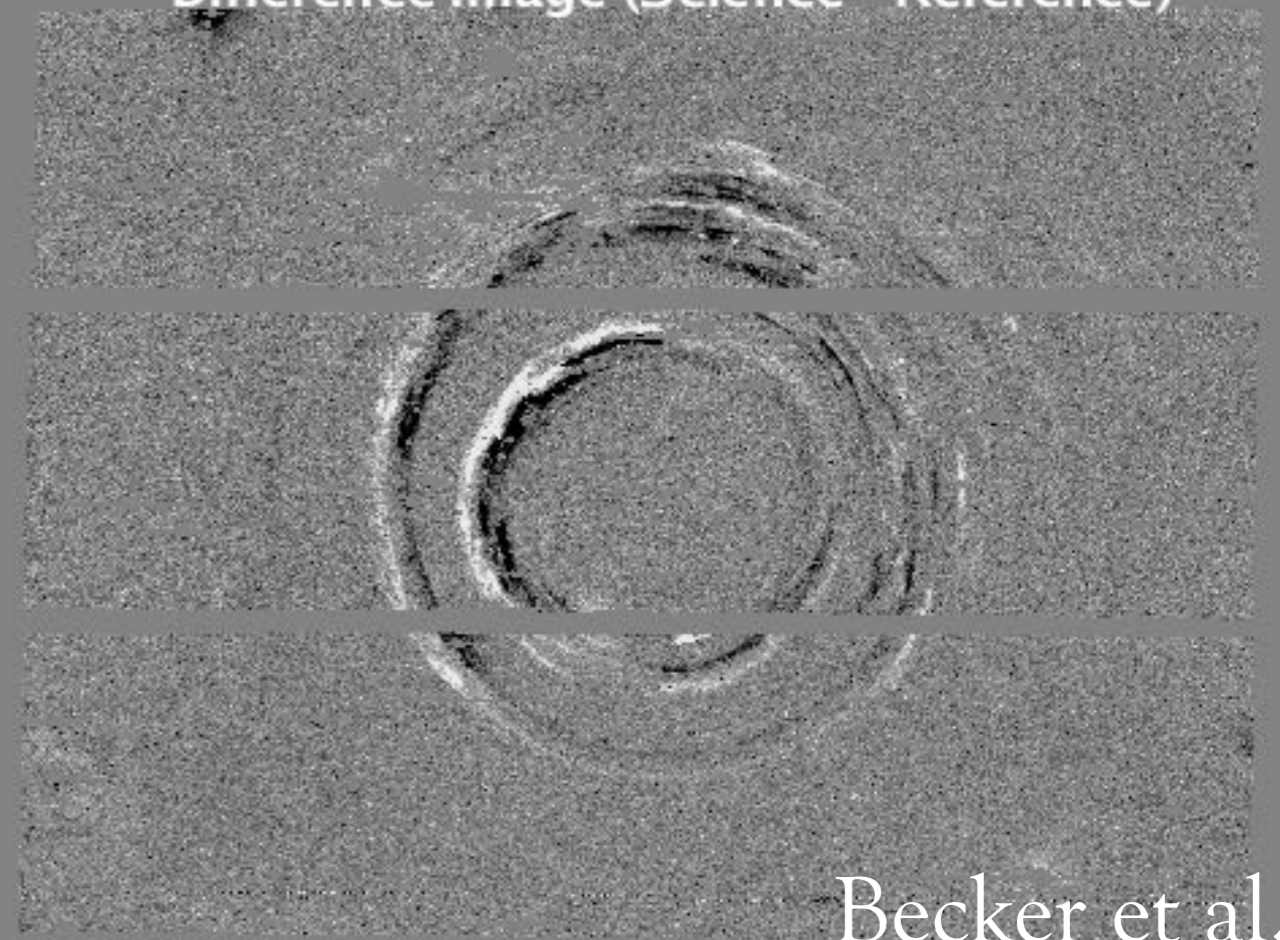
**Helligkeit:** Kosmische Explosionen und Veränderliche

Nicht nur Punktquellen - Lichtecho einer Supernova

Science Image



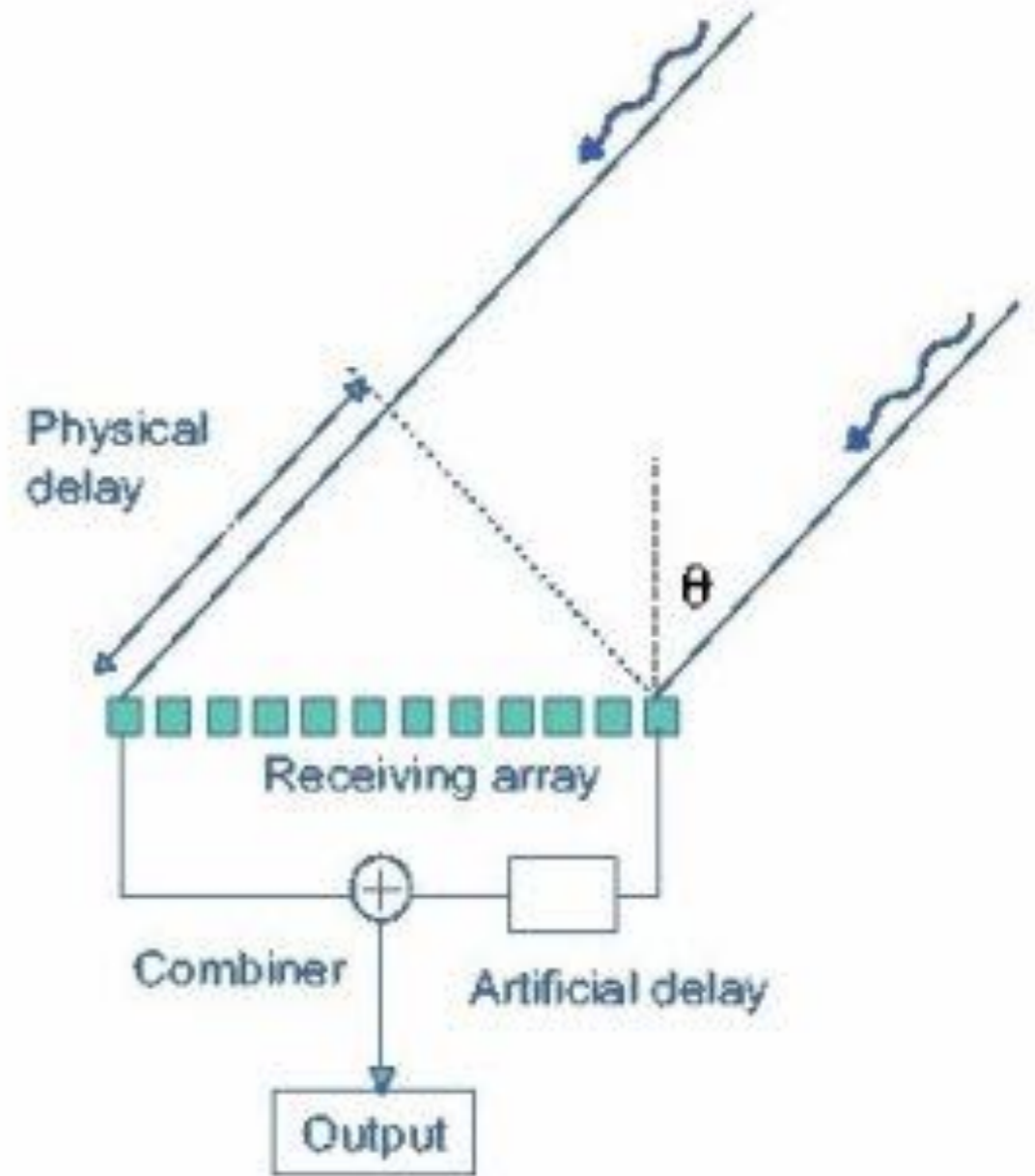
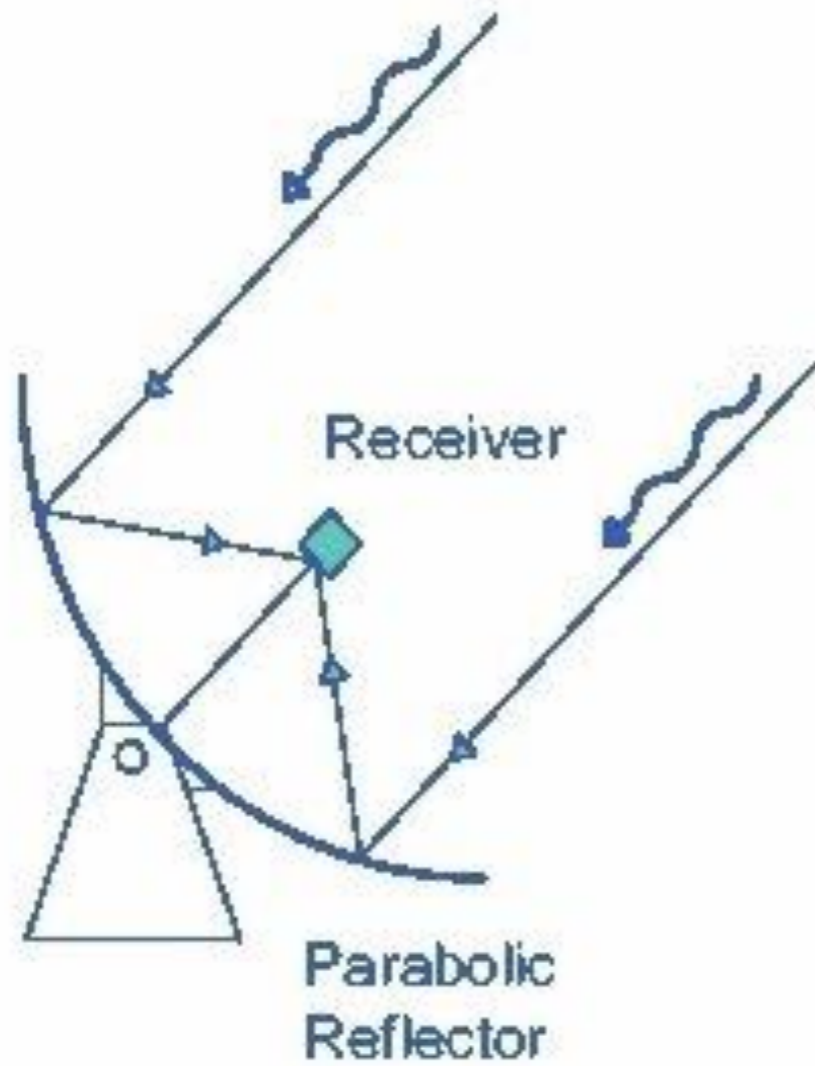
Difference Image (Science - Reference)



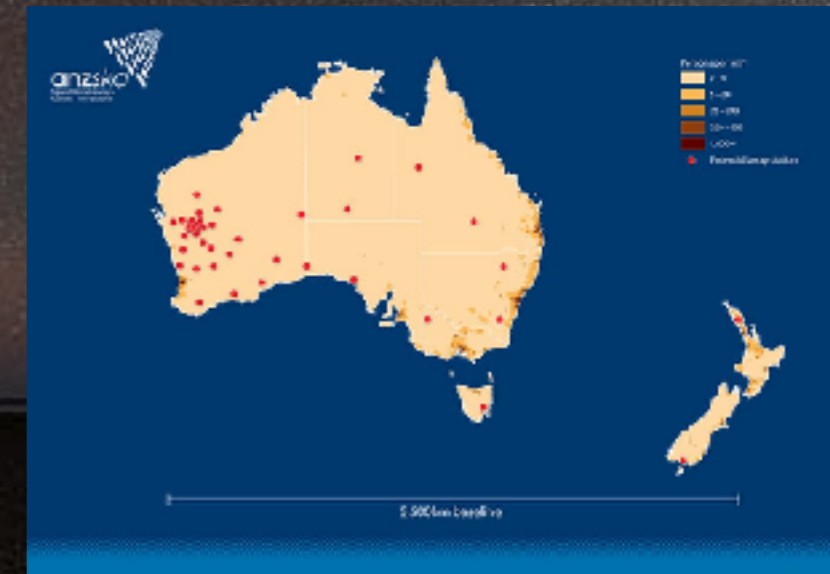
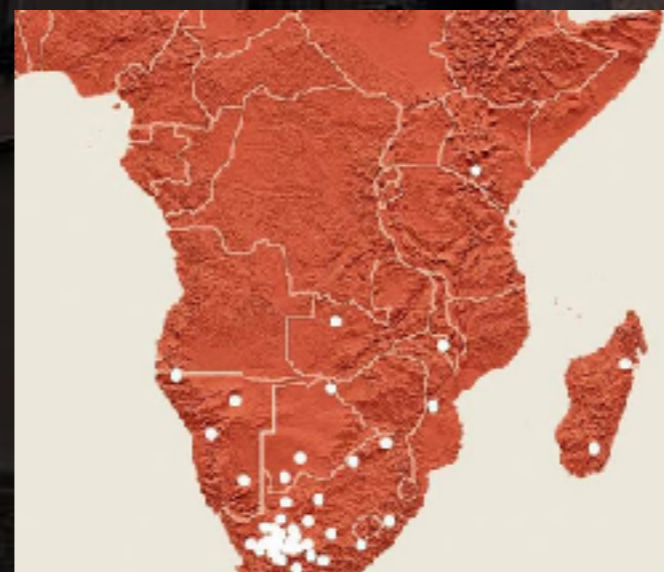
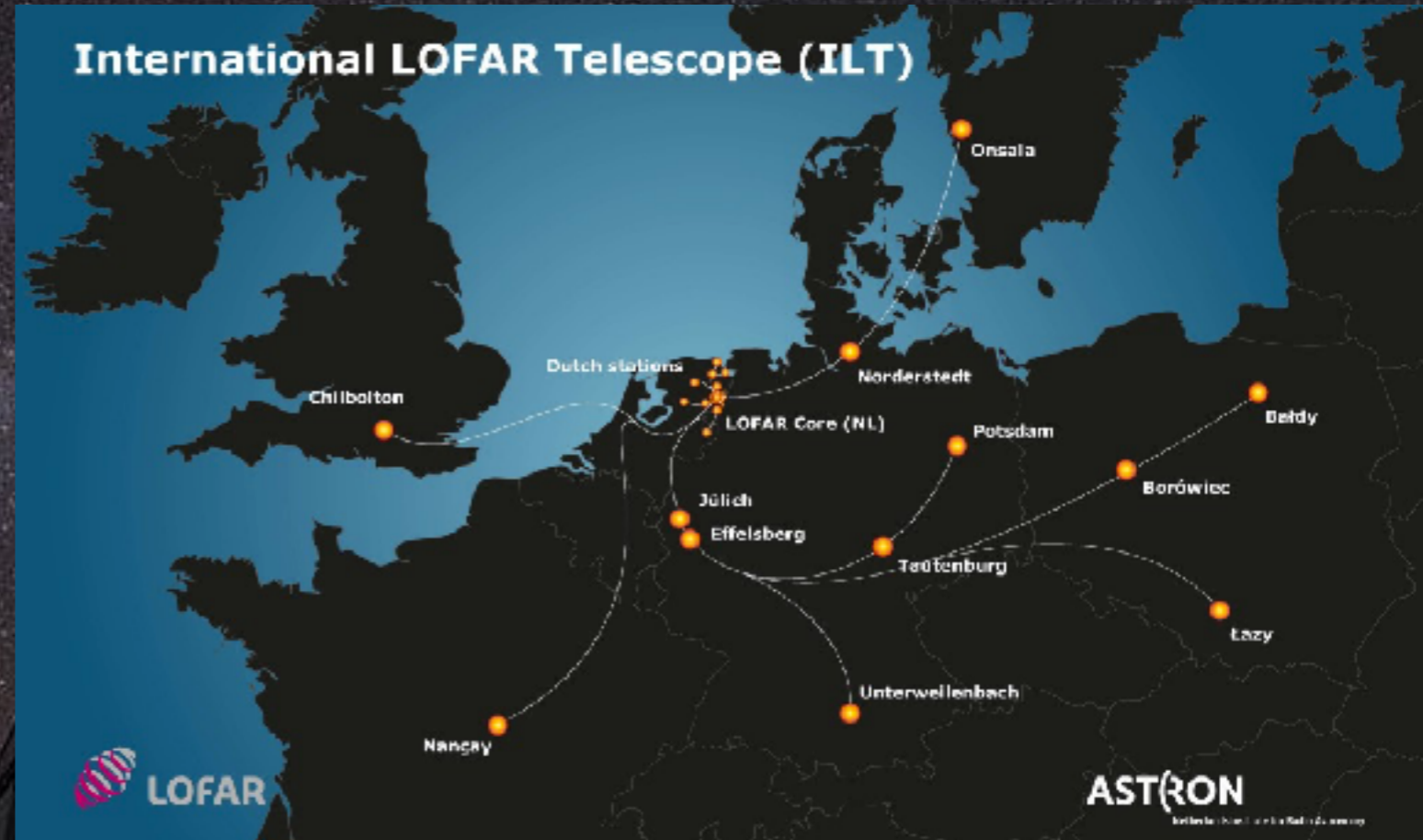
Becker et al.

LSST liefert mehr veränderliche Sterne wie normale Sterne im SDSS Webstream mit Daten für Transients innerhalb 60s. Echtzeit-Alarm!

# Software-Teleskope in der Radioastronomie



# LOFAR & SKA



# Und Spektroskopie?

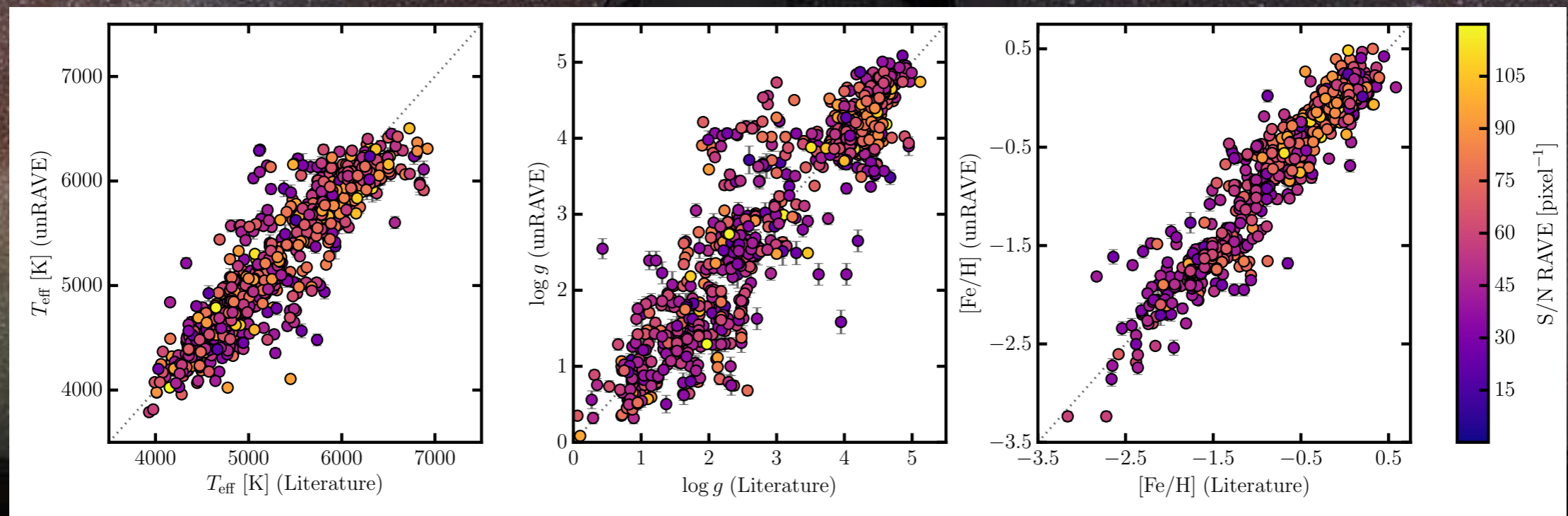
- Einige spektroskopische Durchmusterungen in Vorbereitung (aber „nur“ 4m-Teleskope):
  - WEAVE & DESI (Nordhalbkugel),
  - 4MOST (Südhalbkugel)
- Einige zehn-Millionen Spektren mittlerer Auflösung ( $R=5,000$ ), einige Millionen mit höherer ( $R=20,000$ )
- Datenvolumen signifikant aber beherrschbar. Hauptthema ist Datenkomplexität (Physik)
- etwa Hunderttausend HiRes ( $R=100,000$ )
- Asteroseismologie nur für etwa dreißigtausend Sterne
- Längerfristig (Ende oder 2020er): dediziertes spektroskopisches 12m-Teleskop der ESO

# Pyramide der Datenqualität

- 40 Milliarden: Photometrie bei mehreren Epochen
- 100 Millionen: niedrig aufgelöste Spektroskopie
- 10 Millionen: mittel-aufgelöste Spektroskopie
- 100,000: hoch aufgelöste Spektroskopie
- 30,000: Asteroseismologie
- analog in anderen Gebieten (Einzelaufnahmen mit hoher räumliche Auflösung vs seeing begrenzte Durchmusterungen)
- ditto für verschiedenen Wellenlängen und andere Messenger

# Datengetriebene Methoden

- Unmengen an Information ist verborgen in niederaufgelösten Daten, aber schwierig jeweils für sich zu extrahieren (Entartungen)
- Datengetriebene Verfahren erlauben es, statistisch das Problem anzugehen. Beispiel: individuelle chemische Häufigkeiten mit niederaufgelösten Spektren („The Cannon“)



Casey et al, 2017

# Big Data in der Astronomie

- Astronomie sucht nach neuen Entdeckungen
- Um das volle Potenzial auszuschöpfen: Big data heisst open data!
- Einige faszinierende big-data-Projekte in der Umsetzung
- Ideale „Spielwiese“ (keine kommerziellen Interessen, (fast) keine Datenschutzbedenken)
- Datengetriebene Algorithmen sind der Schlüssel um die Wissenschaft in den Daten auszuschöpfen (Pyramide der Daten-Qualität)