Astroinformatics I (Semester 1 2024)

Introduction & Course Logistics

Nina Hernitschek
Centro de Astronomía CITEVA
Universidad de Antofagasta

April 22, 2024

What you will learn in this class

Astronomers spend about 100 % of their time working with a computer.

This course will **prepare you** for:

- How to use specialized computer programs to do research
- How to use a computer efficient to do research
- How to write programs for typical research tasks

What you will learn in this class

Astronomers spend about 100 % of their time working with a computer.

This course will **prepare you** for:

- How to use specialized computer programs to do research
- How to use a computer efficient to do research
- How to write programs for typical research tasks

What this course does **not** cover:

- advanced programming skills
- machine learning
- usage of LaTeX

(You might find some of that in other courses offered.)

Course Logistics

course content:

- lecture: Monday 11 - 13h
- tutorial: Wednesday 12 - 14h
- graded project practice
- project presentation

Course Logistics

course content:

- lecture: Monday 11 - 13h
- tutorial: Wednesday 12 - 14h
- graded project practice
- project presentation

grading:

- 4 graded project practices of 2 hours each (60% of total grade)
  content: application of what has been learned in class, report in English to produce in one week
  evaluation: capacity to solve a problem using astronomical software tools, written expression
- participation (10 % of total grade)
- project presentation (30 % of total grade)

Course Logistics

course content:

- lecture: Monday 11 - 13h
- tutorial: Wednesday 12 - 14h
- graded project practice
- project presentation

grading:

- 4 graded project practices of 2 hours each (60% of total grade)
  content: application of what has been learned in class, report in English to produce in one week
  evaluation: capacity to solve a problem using astronomical software tools, written expression
- participation (10 % of total grade)
- project presentation (30 % of total grade)

contact and course material:

- e-mail: `nina.hernitschek@uantof.cl`
- github: `https://github.com/ninahernitschek/astroinformatica_I_2024_1`

Course Logistics

April 22 Lecture 1: Introduction, **April 24** Tutorial 1

May 7 (!) Lecture 2: Linux, **May 8** Tutorial 2

May 13 Lecture 3: Astronomical Survey Data, **May 15** Tutorial 3

May 20 Graded Project Practice 1

May 27 Lecture 4: Programming Introduction, **May 29** Tutorial 4

June 3 Lecture 5: Python Introduction, **June 5** Tutorial 5

June 10 Graded Project Practice 2

June 17 Lecture 6: Data Exploration & Visualization, **June 19** Tutorial 6

June 24 Lecture 7: Astronomical Packages **June 26** Tutorial 7

July 1 Graded Project Practice 3

July 8 Lecture 8: Software Projects, **July 10** Tutorial 8

July 15 Graded Project Practice 4

July 22 Project Presentation

Rules for Coding, Presentations, Report

coding:

- use meaningful variable names
- use comments and documentation
- If you have a question when something doesn't work, summarize what you tried - often this will even lead to the solution.

Rules for Coding, Presentations, Report

coding:

- use meaningful variable names
- use comments and documentation
- If you have a question when something doesn't work, summarize what you tried - often this will even lead to the solution.

project practice reports and project presentation:

- use LaTeX
- figures: all own figures should be in vectorized pdf format
- for all data: data description (incl. citation)
- own work: properly cite what is not your own work; discuss how the previous work is similar to or different from your own work
- implementation: medium-level implementation description with libraries/ software frameworks (incl. citation)
- discussion: reflect your approach (strengths, weaknesses, limitations), lessions learned
- bibliography

Textbooks

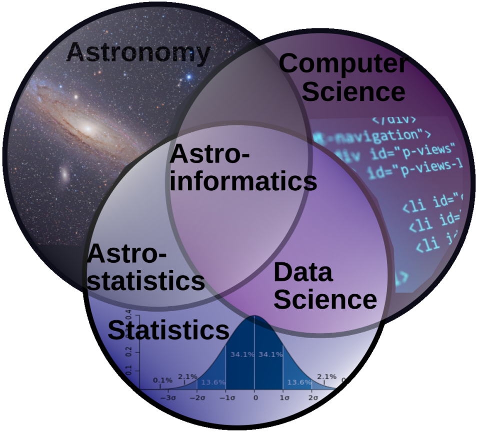
I can recommmend the following textbooks and online resources:

Python cookbook; Beazley, D. & Jones, B.; O'Reilly Media.

A Beginner's Guide to Working with Astronomical Data. M. Pössel, The Open Journal of Astrophysics, vol. 3, issue 1, id. 2, 2020.

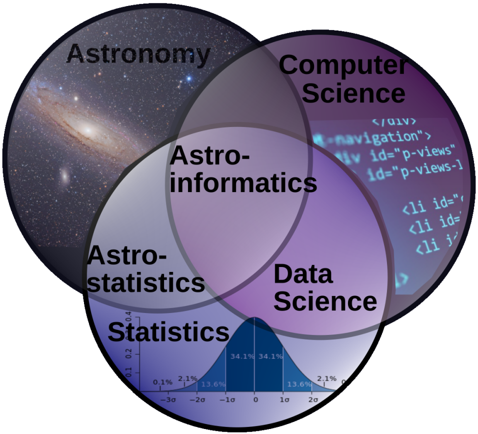
What is Astroinformatics?

Astroinformatics is an **interdisciplinary** field:



What is Astroinformatics?

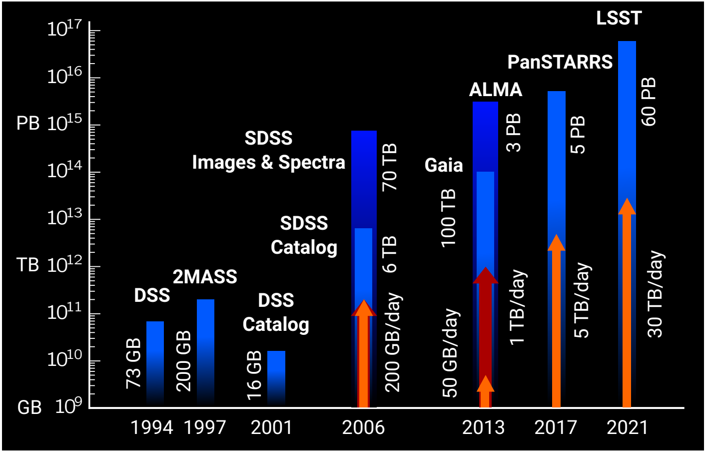
Astroinformatics is an **interdisciplinary** field:



Astroinformatics is primarily focused on developing the tools, methods, and applications for research in data-heavy astronomy.

Why we need Astroinformatics?

Modern All-Sky Surveys come with an increasing data volume:



Why we need Astroinformatics?

upcoming large all-sky surveys like (but not only) LSST

Challenge:

- enormous data volume, $<$ 60 PB total, 15 - 30 TB/ night
- follow-up opportunities should be identified immediately

Chance:

large data volume enables for

population studies

finding rare 'one-in-a-million', 'one-in-a-billion' events, often called *anomalies*

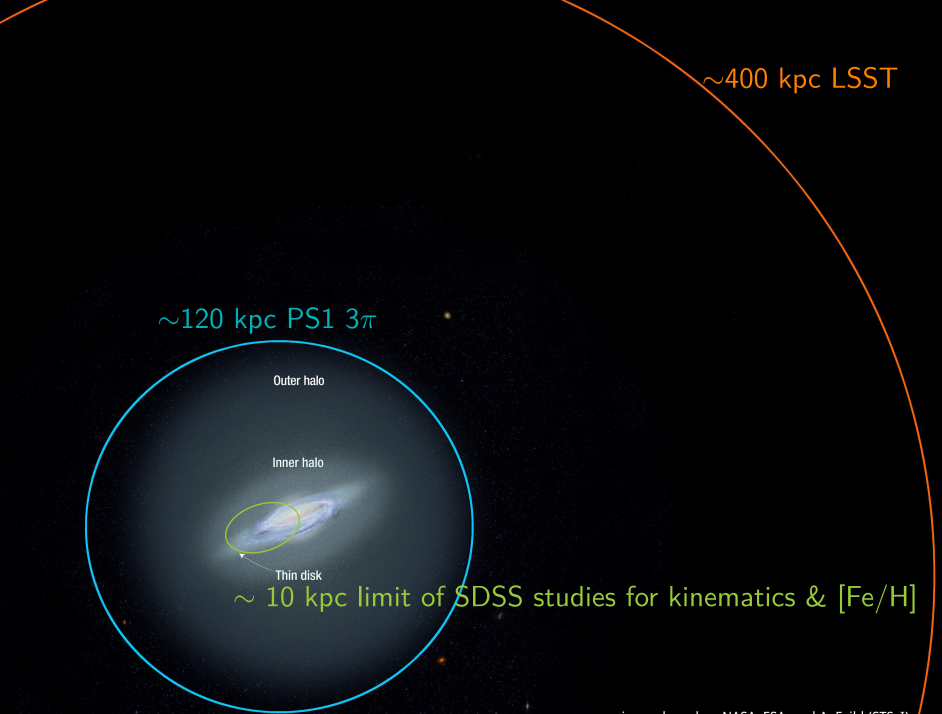


image based on NASA, ESA, and A. Feild (STScI)

Why we need Astroinformatics?

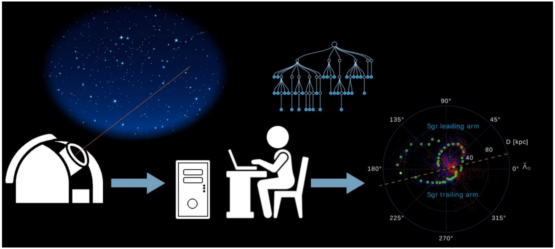
Astronomers need to solve equations, sometimes symbolic, sometimes numerically. Those are for example:

- Equations of movement of celestial bodies. (Compare to: The discovery of Neptune was made by solving equations by hand.)
- Computation of the stellar interior.
- Hydrodynamical simulations of which some now contain billions of particles.
- Machine learning for data from large all-sky surveys, for such as classification.
- Survey strategy planning, e.g. LSST survey strategy uses machine learning.
- Astroengineering problems, e.g. adaptive optics.

Example Use Cases

how I use computers/ computer programs in research

- planning observations



Motivation

Course Logistics

What is Astroinformatics?

Astronomical Data

Tools

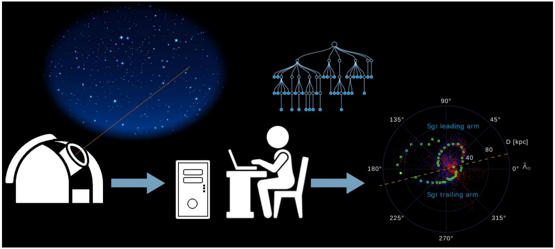
Survey

Outlook

Example Use Cases

how I use computers/ computer programs in research

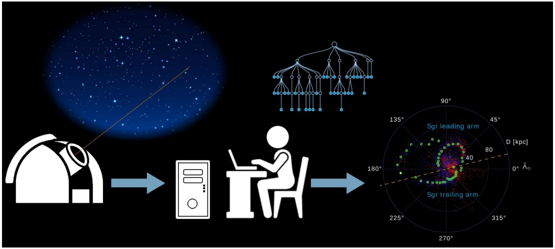
- planning observations
- downloading, plotting, preprocessing, merging data



Example Use Cases

how I use computers/ computer programs in research

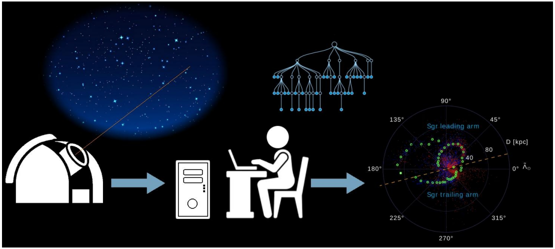
- planning observations
- downloading, plotting, preprocessing, merging data
- classifying data (machine learning)



Example Use Cases

how I use computers/ computer programs in research

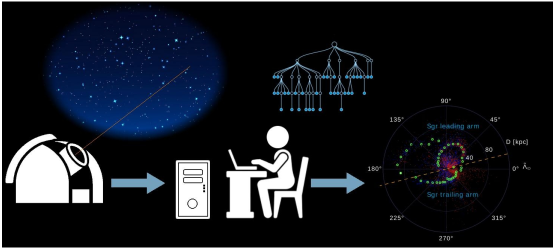
- planning observations
- downloading, plotting, preprocessing, merging data
- classifying data (machine learning)
- creating complex visualizations



Example Use Cases

how I use computers/ computer programs in research

- planning observations
- downloading, plotting, preprocessing, merging data
- classifying data (machine learning)
- creating complex visualizations
- developing a LSST broker software



Simulations

first astronomical simulation by Holmberg (1941):

1941ApJ....94..385H

THE ASTROPHYSICAL JOURNAL

AN INTERNATIONAL REVIEW OF SPECTROSCOPY AND ASTRONOMICAL PHYSICS

VOLUME 94 NOVEMBER 1941 NUMBER 3

ON THE CLUSTERING TENDENCIES AMONG THE NEBULAE

II. A STUDY OF ENCOUNTERS BETWEEN LABORATORY MODELS OF STELLAR SYSTEMS BY A NEW INTEGRATION PROCEDURE

ERIK HOLMBERG

ABSTRACT

In a previous paper[1] the writer discussed the possibility of explaining the observed clustering effects among extragalactic nebulae as a result of captures. The present investigation deals with the important problem of whether the loss of energy resulting from the tidal disturbances at a close encounter between two nebulae is large enough to effect a capture. The tidal deformations of two models of stellar systems, passing each other at a small distance, are studied by reconstructing, piece by piece, the orbits described by the individual mass elements. The difficulty of integrating the total gravitational force acting upon a certain element at a certain point of time is solved by replacing gravitation by light. The mass elements are represented by light-bulbs, the candle power being proportional to mass, and the total light is measured by a photocell (Fig. 1). The nebulae are assumed to have a flattened shape, and each is represented by 37 light-bulbs. It is found that the tidal deformations cause an increase in the attraction between the two objects, the increase reaching its maximum value when the nebulae are separating, i.e., after the passage. The resulting loss of energy (Fig. 6) is comparatively large and may, in favorable cases, effect a capture. The spiral arms developing during the encounter (Figs. 4) represent an interesting by-product of the investigation. The direction of the arms depends on the direction of rotation of the nebulae with respect to the direction of their space motions.

Simulations

first astronomical simulation by Holmberg (1941):

The paper is a study of the tidal disturbances appearing in stellar systems which pass one another at small distances.

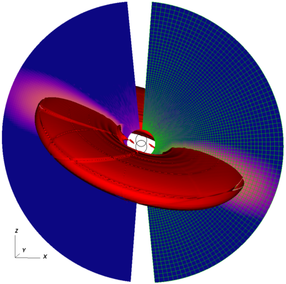
Every mass element is represented by a small light-bulb, the light being proportional to the mass, and the total light along the x and y axes is measured by a combination of a photocell and a galvanometer. The measured values represent the components of the gravitational force. The components ponents are obtained by adding up the attractions due to individual mass elements, each multiplied by the cosine of the corresponding projection angle.

modern astronomical simulations run on the computer:

Grid-based simulations:

Here, space is divided into cells, for instance a space-filling set of small cubes. For each cell, properties such as density and temperature are tracked while they change as the simulation runs.

In more complex simulations, the grid itself can also change in ways that are adapted to making the simulation more efficient (*adaptive grid*).



The evolution of an inclined disk around an eccentric binary was simulated using the grid-based code ATHENA++. Colors show a slice of density along the *xz* plane, with red contour showing a 3D isosurface. The orbits of the binary stars are visible in the center.
Credit: Rabago et al. (2022).

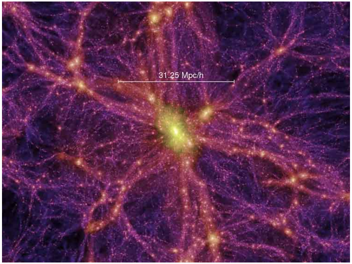
Simulations

modern astronomical simulations run on the computer:

N-body simulations:

A N-body simulation is a simulation of massive particles under the influence of physical forces, usually gravity and sometimes other forces. Depending on the simulation, a point particle could represent a lump of gas, or a star, or a group of a few 10^4 or 10^5 stars in a galaxy, or a lump of dark matter.

Every particle is characterized by a position in space (and potential other properties) which change under the physical forces as the simulation runs.

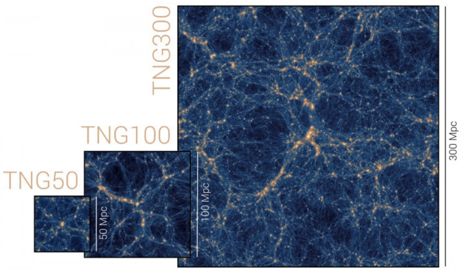


The Millennium Simulation simulates the universe until the present state, where structures are abundant, manifesting themselves as stars, galaxies and clusters. The dark matter density field is colour-coded by density and local dark matter velocity dispersion. Credit: Springel et al. (2005).

Cosmological simulations:

There are many N-body simulations in cosmology. One is IllustrisTNG, a series of magnetohydrodynamical simulations of galaxy formation.

The smallest but most detailed of the TNG runs, TNG50, follows the fate of a cube that, in the present universe, has a side-length of 50 Mpc. Within this volume, dark matter is represented by 10 billion point particles. 10 billion point particles are representing gas.

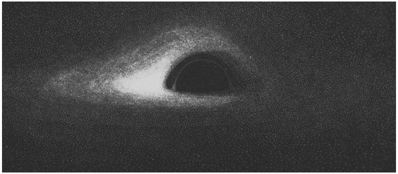


All the results and data from the three IllustrisTNG simulation volumes – TNG50, TNG100 and TNG300 – are publicly available (`http://www.tng-project.org/`). The data volume sums up to **more than one Petabyte**.

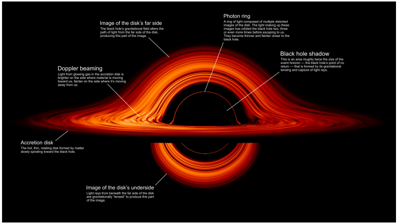
Simulations

relativistic raytracing now and then

Raytracing is the simulation of the path of light through spacetime. This rendering technique can simulate physically accurate reflections, refractions, shadows, and indirect lighting. Also different spacetime geometries can be taken into account to for example generate realistic images of the neighborhood of a black hole, such as the accretion disk.

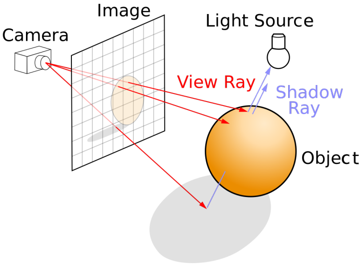


Visualization of a black hole, calculations done by using a computer, data points drawn by hand, J.-P. Luminet, (1979)



Visualization of a black hole by using a raytracing algorithm, NASA (2019)

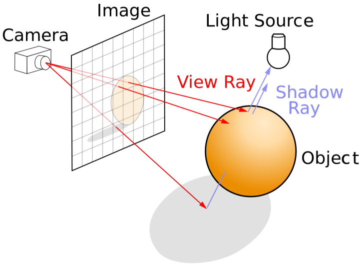
The Raytracing Algorithm



An image is constructed by extending rays through the image plane into a scene, bouncing them off objects and towards light sources to approximate the color value of pixels. Scenes in raytracing are described mathematically.
Light rays might be bent by non-Euclidean spacetime geometry.

Simulations

The Raytracing Algorithm



An image is constructed by extending rays through the image plane into a scene, bouncing them off objects and towards light sources to approximate the color value of pixels. Scenes in raytracing are described mathematically.
Light rays might be bent by non-Euclidean spacetime geometry.

Sending rays in the other direction than light travels in reality might at first be counterintuitive, but doing so is many orders of magnitude more efficient. As most light rays from a given light source do not make it into the viewer's eye, a "forward" simulation could potentially waste a tremendous amount of computation on light paths that are never recorded.

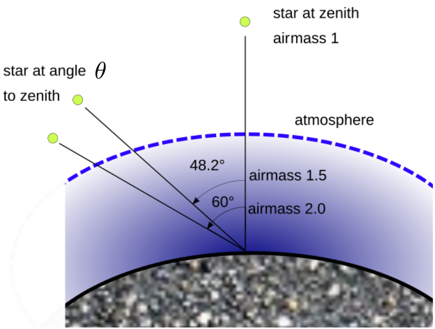
Survey Strategy

Observation planning (efficiently scheduling astronomical observations) is a very important topic in observational astronomy:

- visibility duration (location, season)
- moon influence
- pointing
- multiple objects per field
- weather
- target (object) characteristics
- exposure time necessary

Survey Strategy

In astronomy, **airmass** is a measure of the amount of air along the line of sight when observing through Earth's atmosphere.
The closer your target is to the horizon, the more air you have to look through, and the more degraded your view gets. The amount of air directly overhead is called **one airmass**.



If θ is the angle between zenith and the observed star (or other object), then a simple model of airmass X is $X = 1/cos(\theta)$.

Survey Strategy

Motivation
Course Logistics
What is Astroinformatics?
Astronomical Data
Tools
Survey
Outlook

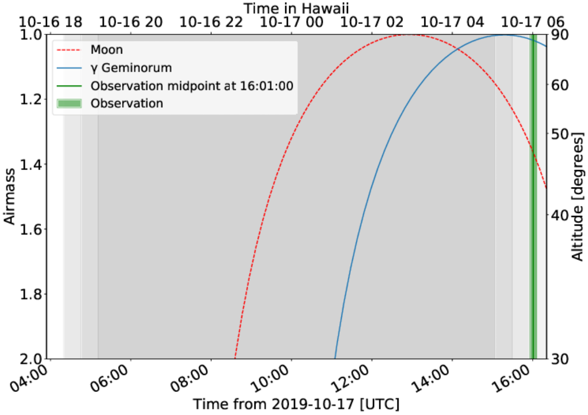


Illustration of time vs. airmass/altitude of the observation at 6:01 UTC -10 on the 17th of Oct. 2019 from Mauna Kea Observatory (green vertical region), for γ Geminorum (solid blue line) and the moon (red dashed). Exposure started at the very end of observing time, close to sunrise.

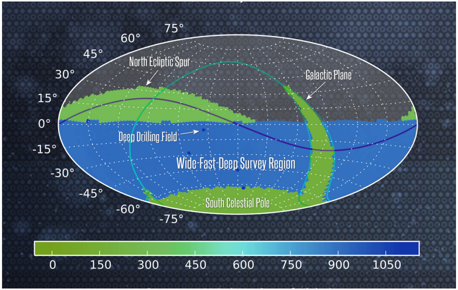
Survey Strategy

What is the case for individual observations or small surveys is also the case for large all-sky surveys:

Survey strategies needs to be **optimized**.

The LSST Survey

- 10-year photometric *ugrizy* survey (near-UV, optical, near-IR)
- depth of r∼27.5 mag
- 1000 images/night = 15 TB/night, 10 million transients/night
- start of operations: 2024



LSST survey strategy, number of visits incl. sub-surveys. (credit: www.lsst.org)

The LSST Survey

Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

results in the following **constraints** for the LSST Survey Strategy:

The LSST Survey

Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

results in the following **constraints** for the LSST Survey Strategy:

Cosmological parameter estimation by many techniques requires uniform coverage of 18,000 square degrees of sky. Obtaining accurate photometric redshifts in every field requires a specified number of visits in each filter.

Survey Strategy

The LSST Survey

Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

results in the following **constraints** for the LSST Survey Strategy:

Weak lensing shear measurements benefit from allocating times of best seeing to observations in the r and i bands. Maximizing signal-to-noise ratios requires choosing the next filter based upon the current sky background.

Survey Strategy

The LSST Survey

Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

results in the following **constraints** for the LSST Survey Strategy:

Supernova cosmology requires frequent, deep photometry in all bands, with z and y observations even during dark time.

The LSST Survey

Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

results in the following **constraints** for the LSST Survey Strategy:

Detecting the motion of solar system objects and transients, characterizing variability on various timescales, and acquiring the best proper motions and parallaxes place further demands upon the distribution of revisit intervals and observation geometries to each point on the sky.

Survey Strategy

The LSST Survey

Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

results in the following **constraints** for the LSST Survey Strategy:

Detecting the motion of solar system objects and transients, characterizing variability on various timescales, and acquiring the best proper motions and parallaxes place further demands upon the distribution of revisit intervals and observation geometries to each point on the sky.

⇨ **complex requirements** where a strategy is chosen and then observations are scheduled by **optimization and machine learning**

Data Access

In order to do science, we need (in most cases) access to astronomical data.

For this, Astroinformatics entails

development of services and tools to enable users to access the data

using those services to access the data

Data Access

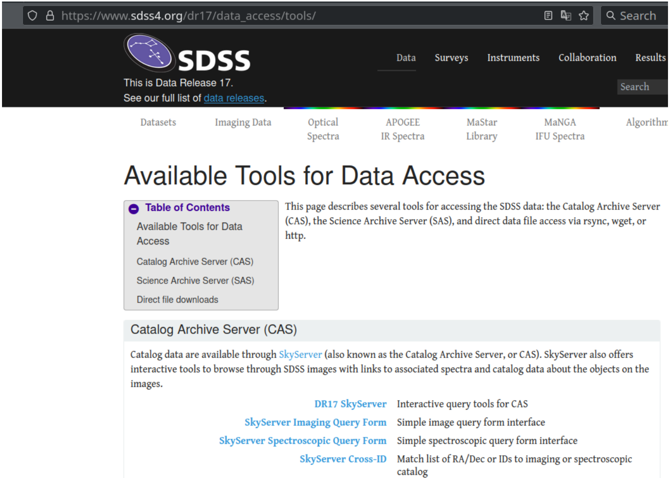
Some astronomical survey data are shared via the observatory that acquired them, for example, the all-sky data acquired with Planck, Wilkinson Microwave Anisotropy Probe, and COBE.

Other surveys serve their own data. Examples include the SDSS, 2MASS, UKIDSS.

Surveys as Astronomical Data Sources

Example: SDSS

`https://www.sdss4.org`



https://www.sdss4.org/dr17/data_access/tools/

SDSS

This is Data Release 17.
See our full list of data releases.

Data | Surveys | Instruments | Collaboration | Results

Search

Datasets | Imaging Data | Optical Spectra | APOGEE IR Spectra | MaStar Library | MaNGA IFU Spectra | Algorithm

Available Tools for Data Access

Table of Contents

This page describes several tools for accessing the SDSS data: the Catalog Archive Server (CAS), the Science Archive Server (SAS), and direct data file access via rsync, wget, or http.

Catalog Archive Server (CAS)

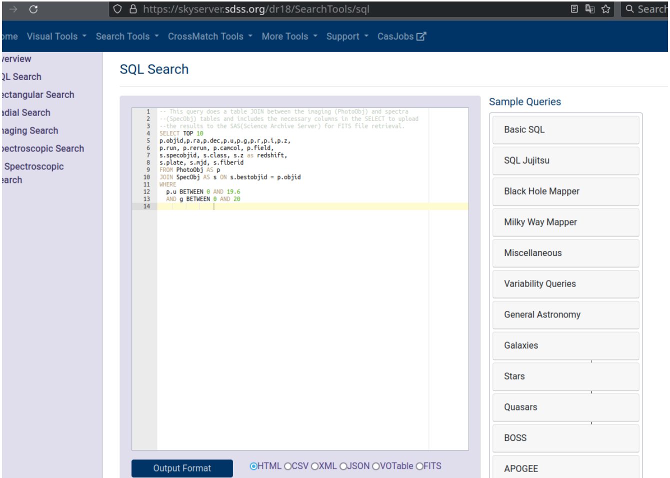
Catalog data are available through SkyServer (also known as the Catalog Archive Server, or CAS). SkyServer also offers interactive tools to browse through SDSS images with links to associated spectra and catalog data about the objects on the images.

DR17 SkyServer	Interactive query tools for CAS
SkyServer Imaging Query Form	Simple image query form interface
SkyServer Spectroscopic Query Form	Simple spectroscopic query form interface
SkyServer Cross-ID	Match list of RA/Dec or IDs to imaging or spectroscopic catalog

Motivation
Course Logistics
What is Astroinformatics?
Astronomical Data
Tools
Survey
Outlook

Example: SDSS

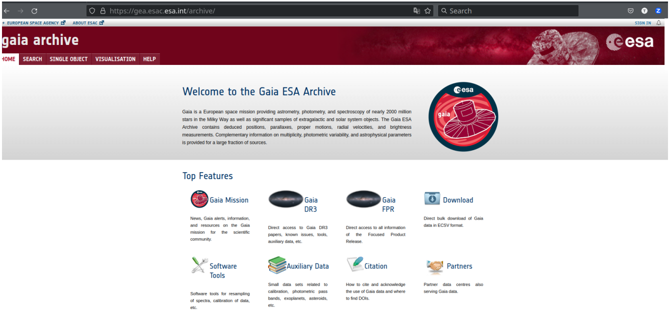
`https://skyserver.sdss.org`



Surveys as Astronomical Data Sources

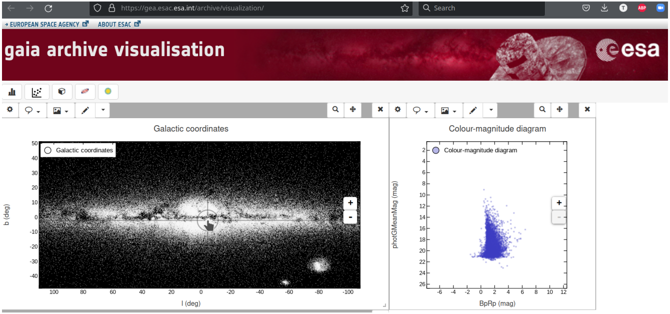
Example: Gaia

`https://gea.esac.esa.int/archive/`



Example: Gaia

`https://gea.esac.esa.int/archive/visualization`



Services

With a lot of different surveys existing, a **common scripted interface** to tie all these services together is a good way to make the data more accessible.

It also provides **reproducibility** of analysis done for publications. A centrally maintained library also safeguards against lost links to data archives, moving some of the responsibility for maintaining long-term reproducibility from each individual researcher to the broader community.

Services

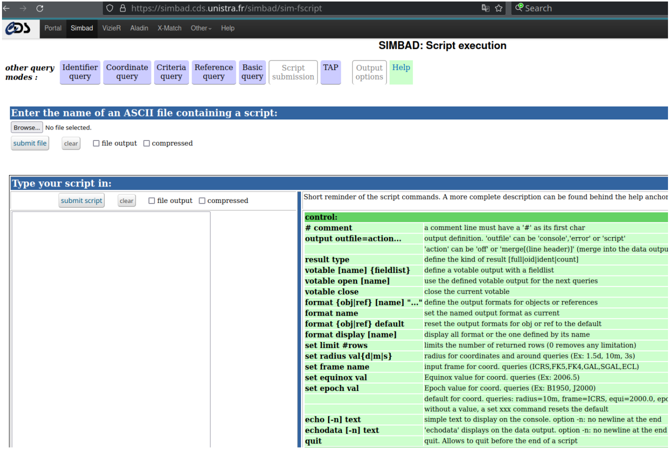
The SIMBAD astronomical database provides basic data, cross-identifications, bibliography and measurements for astronomical objects outside the solar system. SIMBAD can be queried by object name, coordinates and various criteria. Lists of objects and scripts can be submitted. Links to some other on-line services are also provided.

The **SIMBAD** database currently contains information for about 13,000,000 astronomical objects (stars, galaxies, planetary nebulae, clusters, novae and supernovae, etc.).

The only astronomical objects specifically excluded from SIMBAD are Solar System bodies (planets, satellites, asteroids, comets)

Services

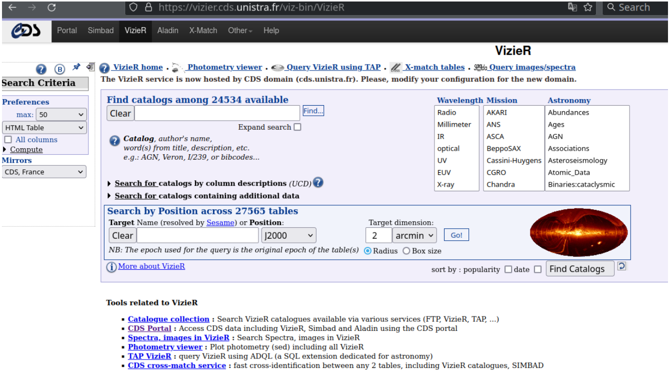
`https://simbad.unistra.fr/simbad/`



Services

The **VizieR** service provides mostly catalogs associated with papers or authors.

`https://vizier.cds.unistra.fr/viz-bin/VizieR`



Community brokers

The newest generation of large survey telescopes such as the Vera C. Rubin Observatory (which will conduct the 10-year Legacy Survey of Space and Time (LSST)) are designed to produce large streams of transient alerts from difference imaging. They will be processed, annotated and classified by a new generation of astronomical software for time-domain surveys, called **alert brokers**.

Community brokers

The newest generation of large survey telescopes such as the Vera C. Rubin Observatory (which will conduct the 10-year Legacy Survey of Space and Time (LSST)) are designed to produce large streams of transient alerts from difference imaging. They will be processed, annotated and classified by a new generation of astronomical software for time-domain surveys, called **alert brokers**.

Brokers are accessible over the internet, providing real-time information crucial for:

- follow-up
- observations depending on accurate & up-to-date light curve information

Community brokers

The newest generation of large survey telescopes such as the Vera C. Rubin Observatory (which will conduct the 10-year Legacy Survey of Space and Time (LSST)) are designed to produce large streams of transient alerts from difference imaging. They will be processed, annotated and classified by a new generation of astronomical software for time-domain surveys, called **alert brokers**.

Brokers are accessible over the internet, providing real-time information crucial for:

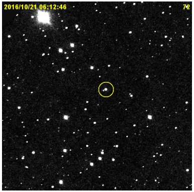
- follow-up
- observations depending on accurate & up-to-date light curve information

Brokers are currently developed tested with the Zwicky Transient Facility (ZTF). Usually, brokers can be accessed from a website as well as a Python interface.

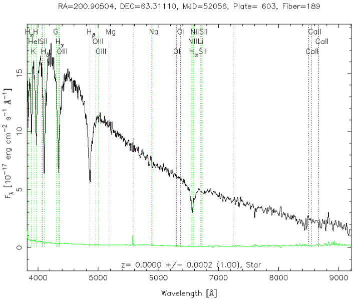
Astronomical Data

astronomical data typically comes in one of three forms:

images



spectra



light curves

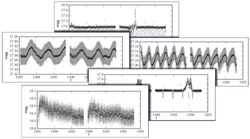


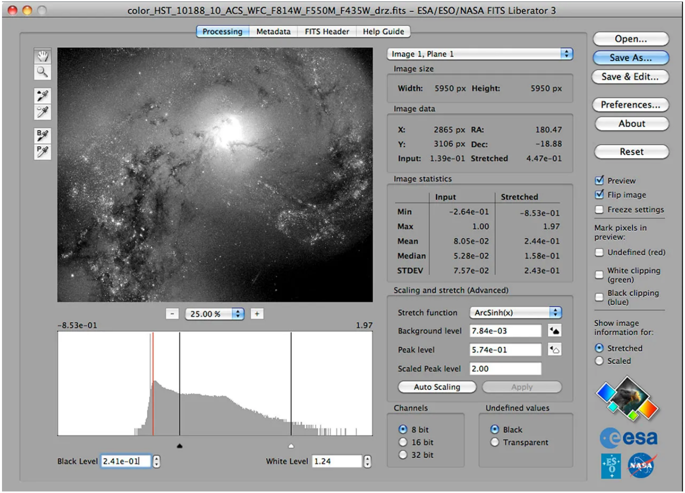
Image Data

An image can be seen as 2D array of pixels, where each pixel value denotes a brightness value. In an ordinary color image, each pixel will have three brightness values for the contribution from red, green, and blue (RGB).

As astronomers use many specialist filters, astronomical images can have even more color values per pixel. They are typically displayed as one grey-scale image per channel.

A commonly used file format is FITS (Flexible Image Transport System). Such files can be opened with specialized software or can be read in with your own code.

Image Data

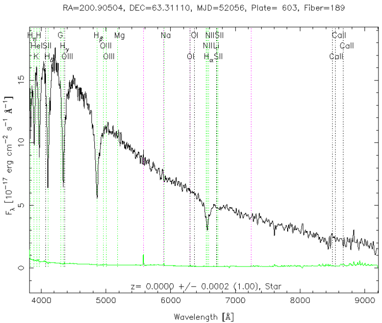


color_HST_10188_10_ACS_WFC_F814W_F550M_F435W_drz.fits – ESA/ESO/NASA FITS Liberator 3
Processing
Metadata
FITS Header
Help Guide
Image 1, Plane 1
Image size
Width: 5950 px Height: 5950 px
Image data
X: 2865 px RA: 180.47
Y: 3106 px Dec: -18.88
Input: 1.39e-01 Stretched 4.47e-01
Image statistics
Input
Stretched
Min -2.64e-01 -8.53e-01
Max 1.00 1.97
Mean 8.05e-02 2.44e-01
Median 5.28e-02 1.58e-01
STDEV 7.57e-02 2.43e-01
Scaling and stretch (Advanced)
Stretch function ArcSinh(x)
Background level 7.84e-03
Peak level 5.74e-01
Scaled Peak level 2.00
Auto Scaling
Apply
Channels
8 bit
16 bit
32 bit
Undefined values
Black
Transparent
Open...
Save As...
Save & Edit...
Preferences...
About
Reset
Preview
Flip image
Freeze settings
Mark pixels in preview:
Undefined (red)
White clipping (green)
Black clipping (blue)
Show image information for:
Stretched
Scaled
esa
25.00 %
-8.53e-01
1.97
Black Level 2.41e-01
White Level 1.24

Spectra

Spectra show us how the energy of the light emitted by an object is distributed among the different possible wavelengths. Such simple spectra are one-dimensional: for each wave-length value, we know the contribution of light from that particular wavelength region.

Spectra are typically stored as tables with the columns being wavelength and intensity. They are plotted as intensity over wavelength.

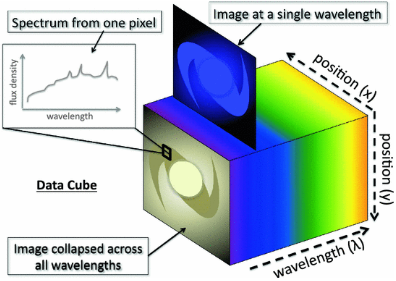


A SDSS spectrum. The plot includes additional information, such as the object, time of observation, the plate and fiber the spectrum was taken. In green, the positions of specific spectral lines are highlighted.

Data Cubes

A data cube can be thought of as enhanced version of an image where each pixel contains not a single brightness value (or three for RGB), but a whole spectrum received from the region of the sky within that pixel.

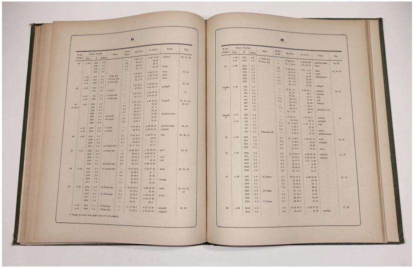
An example for a data cube is data from **Integral Field Spectroscopy (IFS)**. IFS combines spectrographic and imaging capabilities: take a single exposure of a 2D region and obtain spatially resolved spectra.



Integral Field Spectroscopy, schematic. (Credit: Ch. M. Harrison (2016))

Catalog Data

Catalogs group astronomical data together in the form of tables containing data of multiple objects, or of the same object at multiple times, or both. The first astronomical catalogs were books containing coordinates, names and magnitudes of objects.

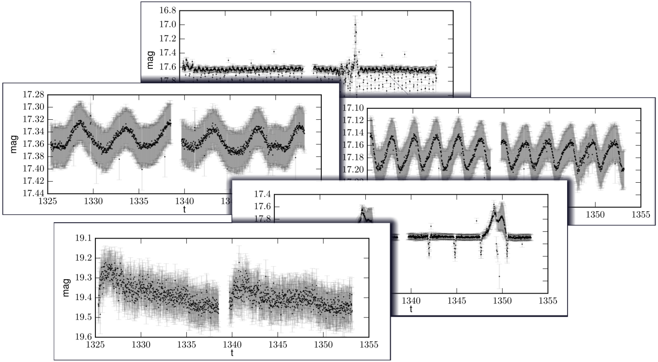


The catalog from the "Bonner Durchmusterung" (survey by Bonn observatory) from 1862

Catalog Data

Catalogs group astronomical data together in the form of tables containing data of multiple objects, or of the same object at multiple times, or both.

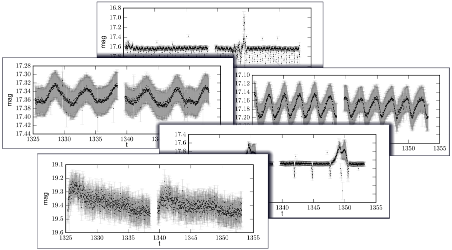
Photometry of an object taken at multiple times is called a **light curve**.



Catalog Data

Catalogs group astronomical data together in the form of tables containing data of multiple objects, or of the same object at multiple times, or both.

Photometry of an object taken at multiple times is called a **light curve**.



Light curves are given in tabular format, typically containing the columns time, magnitude, magnitude error and in the case of multiple photometric bands also indicating the band.
Light curves can be saved as individual tables each containing data from a single object, or as part of a larger table containing the light curves for multiple objects, where as an additionl column an object ID is given.

Other Astronomical Data

The list so far is not complete.

For instance, in **interferometric imaging**, for trying to reconstruct an image by combining coherently the measurements of different telescopes (aperture synthesis), the raw data will be time-stamped data from the single telescopes, and the between those telescopes.

Other Astronomical Data

The list so far is not complete.

For instance, in **interferometric imaging**, for trying to reconstruct an image by combining coherently the measurements of different telescopes (aperture synthesis), the raw data will be time-stamped data from the single telescopes, and the between those telescopes.

What we have seen in general is that **providing and accessing**, as well as **processing** nowadays astronomical data requires some amount of informatics background.

Tools

Typically, we need some software tools to work with astronomical data. Here we can distinguish three types:

Application software is software written for a specific set of tasks. We might, for example, have software to load data tables (acting similar to speadsheet programs), to load astronomical images, and so on.

A **programming language** is a tool for writing your own custom applications.

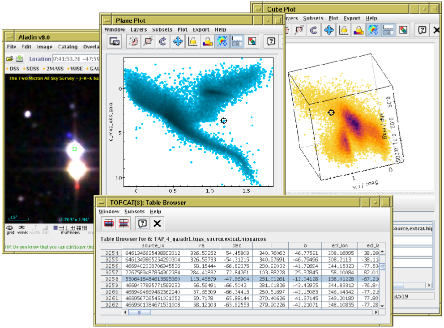
A **data access language** allows for selecting data from (usually remote) databases.

Application Software

commonly used software include the following:

TOPCAT is a standard tool for viewing, manipulating and plotting tables, including cross-matching catalogs. Its main advantage is that it allows for fast access to large datasets (millions of rows/hundreds of columns) - tables which cannot efficiently explored with non-astronomical standard tools.

TOPCAT is available at `https://www.star.bristol.ac.uk/mbt/topcat/`



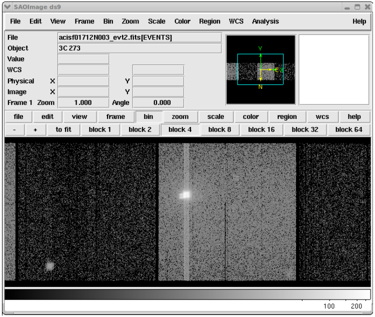
Application Software

commonly used software include the following:

SAOImage DS9 is an astronomical imaging and data visualization application. DS9 supports FITS images and tables, region manipulation, and many scale algorithms and colormaps. The highly configurable and extensible software can communicate with external analysis tools.

DS9 is available at

`https://sites.google.com/cfa.harvard.edu/saoimageds9`



Data Access and Programming Languages

Commonly used **programming languages** in astronomy are:

- Python
- C/C++
- IDL
- FORTRAN

Data Access and Programming Languages

Commonly used **programming languages** in astronomy are:

- Python
- C/C++
- IDL
- FORTRAN

Python has become the language of choice for astronomers and astrophysicists working with data analysis and visualization.

Python can be combined with C++ for efficiency. Also code in FORTRAN (usually older code) can be integrated.

Why use Python?

The **Python ecosystem** provides a single environment that is sufficient for the vast majority of astronomical analysis. It does so on several levels:

Why use Python?

The **Python ecosystem** provides a single environment that is sufficient for the vast majority of astronomical analysis. It does so on several levels:

- open source, which means it is free (as opposed to proprietary languages (like IDL) which require you to buy a license)
- powerful language, many scientific libraries available.
- strong set of 3rd party analysis tools that are professionally and actively developed.
- robust methods for binding with C, C++ and FORTRAN libraries (speed, legacy)
- standard library support for web, GUI, databases, process management, etc.
- very active user and developer community

Data Access and Programming Languages

In contrast to programming languages which enables us to write software to analyze data, data access languages (sometimes also known as *database languages*) enable us to select data from large data sources, usually databases.

Data Access and Programming Languages

Example: Gaia

In addition to the bulk download of data, one can write Astronomical Data Query Language (ADQL) queries which are similar to SQL queries.

example use case: Cone search sorted by angular separation
retrieve a sample of filtered sources (brighter than G=20.5 mag and with a parallax measurement) in a circular region centred on the LMC [(R.A., Dec) = (81.28, -69.78) deg] with a search radius of 5 arcmin. The output should be ordered by the angular separation, from small to large.

Target table: `gaiadr3.gaia_source`

Query:

```
SELECT *, DISTANCE(81.28, -69.78, ra, dec) AS ang_sep
FROM gaiadr3.gaia_source
WHERE DISTANCE(81.28, -69.78, ra, dec) < 5./60.
AND phot_g_mean_mag < 20.5
AND parallax IS NOT NULL
ORDER BY ang_sep ASC
```

Survey

A rapid survey to better understand your preknowledge:

Linux/Unix
Experience?

?

Survey

A rapid survey to better understand your preknowledge:

Linux/Unix
Experience?

Programming
Experience?
Languages?

Survey

A rapid survey to better understand your preknowledge:

Linux/Unix
Experience?

Programming
Experience?
Languages?

Expectations
Projects?
Plan attending other courses?

Survey

A rapid survey to better understand your preknowledge:

Linux/Unix
Experience?

Programming
Experience?
Languages?

Expectations
Projects?
Plan attending other courses?

Questions?

An Outlook: Using Linux

We have seen which topics fall under the relatively broad umbrella of *astroinformatics*.

In the next lecture we will see how to use the **Linux command line** for manipulating files.