Astroinformatics I (Semester 1 2025)

Astronomical Survey Data

Nina Hernitschek
Centro de Astronomía CITEVA
Universidad de Antofagasta

April 28, 2025

Motivation

One of the cornerstones of research is **reproducibility**.

Reproducibility strongly depends on the availability of data.

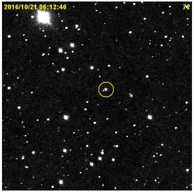
surveys

data services

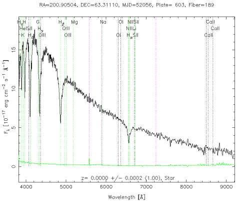
Astronomical Data

astronomical data typically comes in one of three forms:

images



spectra



light curves

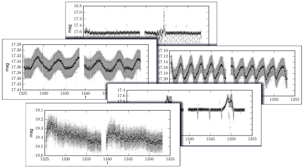


Image Data

Astronomers take images using telescopes and suitable instruments (such as cameras) attached to those telescopes.
Public versions of such images are detailed and colorful, contributing significantly to the fascination of the general public with astronomy.
The underlying **science images**, however, look differently.

Image Data

Astronomers take images using telescopes and suitable instruments (such as cameras) attached to those telescopes.
Public versions of such images are detailed and colorful, contributing significantly to the fascination of the general public with astronomy.
The underlying **science images**, however, look differently.

Two images of the open star cluster Westerlund 2.
Left: Image taken by the Hubble Space telescope, credit: The Hubble Heritage Team (STScI/AURA).
Right: Image taken through an R (red) filter in April 2017 with the 2 m Faulkes Telescope (Las Cumbres Observatory, Siding Spring, Australia).

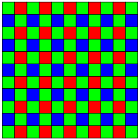
Image Data

In the following, we will take a look at how to work with astronomical imaging data.

Image Data

there are many ways in which professional astronomical images differ from your "everyday" photo:

the colors: Professional astronomical images are greyscale ("black-and-white") images. One reason for this is that (analog and digital) images are in fact greyscale (intensity only) images.
Cameras as in your smartphone or your digital camera only produce colour images because of an array of filters installed in front of the array of light-detecting sensor pixels. Software then converts this raw image (*.raw format) into an RGB image by interpolation.

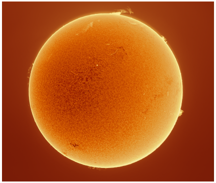


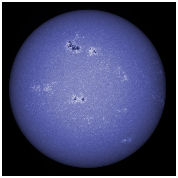
Bayer mask pattern (schematic): An array of filters installed in front of detector pixels to enable the quick creation of a colour image.

Image Data

For astronomical images, a **fixed filter mask is impractical** for several reasons:

The most obvious drawback of this is that photons are split up between different filters, reducing resolution.
Another problem is that color interpolation means that some of the color information gets lost.
A third problem is that in astronomy, different combinations of filters are used depending on the science case - not only red, green, blue.

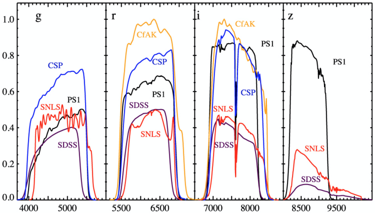


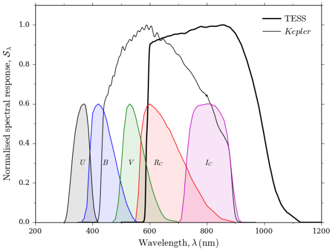


The Sun through a Hα filter (left) and a Calcium K filter (right). The Hα filter highlights the Sun's chromosphere. The Calcium K filter shows a layer of the Sun that is slightly lower and cooler than the layer viewed in Hα.
credit: Lunt Solar Systems

Image Data

Some of the filters used by astronomers cover a wider wavelength range, while others might capture just a particular spectral line (**broad-band filters vs. narrow-band filters**).
Most astronomical objects change only very slowly. Using different filters in succession, taking an image first through one filter, then through the next, allows to use the full resolution.

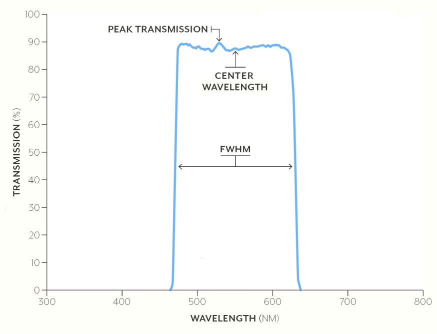




Normalized spectral response curves from different surveys.
left: comparing Pan-STARRS (PS1), Carnegie SN Project (CSP), CfA-Keplercam (CfAK), SN Legacy Survey (SNLS) credit: Scolnic et al. (2015)
right: comparing Kepler and TESS, along with the standard Johnson-Cousins photometric system. credit: Lund et al. (2016).

Image Data

Definitions of bandpass filter performance parameters



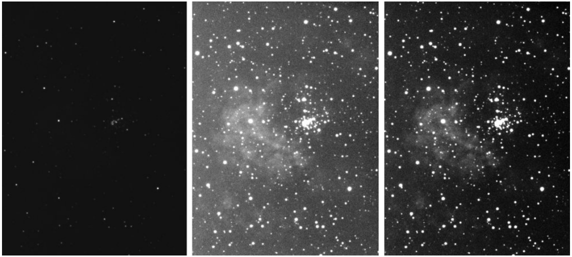
Peak Transmission is the maximum transmission of the filter (%). Peak transmission is a function of the reflectivity of the surfaces.

Center Wavelength is the midpoint of the transmitting region (nm).

Full-Width at Half-Maximum (FWHM) is the width of the transmitting region at half of the maximum transmission value (nm). FWHM generally increases with wavelength.

Image Data

We have seen that professional astronomical images differ from ”everyday” photos regarding the filters used and their technical implementation.
Another important aspect is the **brightness scale** of the image.
FITS (Flexible Image Transport System) images typically have 16-bit brightness values: each pixel can take on values from 0 to 216 = 65 536.
This range can typically not be displayed on a RGB computer display. Also, this range on a linear scale will not automatically show interesting details.



The Westerlund 2 image taken through an R (red) filter with the 2 m Faulkes Telescope with different scaling. Left: Linear scaling from 0 to 65536. Center: Linear scaling from 4572 to 6002. Right: Square scaling from 4572 to 6002.

Image Data

When working with astronomical images, it is important to keep in mind **PSF and noise**.

To get an idea on what they are, we zoom into the image of Westerlund 2.

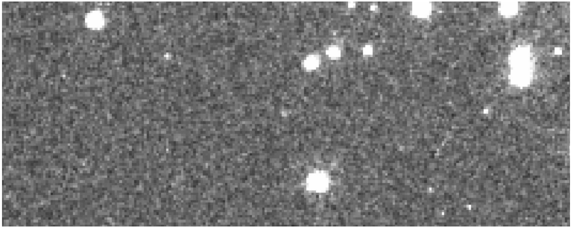
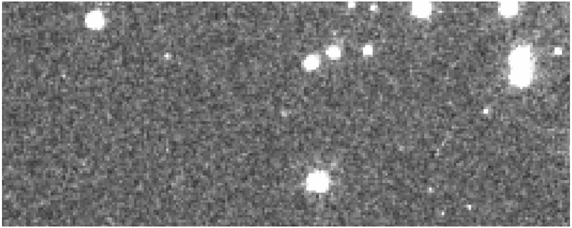


Image Data

When working with astronomical images, it is important to keep in mind **PSF and noise**.

To get an idea on what they are, we zoom into the image of Westerlund 2.



The disk-shaped bright objects in the image are stars. Why do they look like that?

Image Data

A few estimates:

Westerlund 2 is at a distance of $\sim$20 000 light-years. At that distance, even a supergiant with 1500 $R_\odot$ is under an angle of only 0.002". Each image pixel has a side length of 0.1". If our image were a true map showing the exact direction of light, they would fall within a single pixel. Instead, stars appear in the image as disks.

Image Data

A few estimates:

Westerlund 2 is at a distance of ∼20 000 light-years. At that distance, even a supergiant with 1500 $R_\odot$ is under an angle of only 0.002”. Each image pixel has a side length of 0.1”. If our image were a true map showing the exact direction of light, they would fall within a single pixel. Instead, stars appear in the image as disks.

This happens because of **diffraction**, an effect of light passing through an opening (the aperture of the telescope). The result is a diffraction pattern that makes the image of a point source a disk with concentric rings.

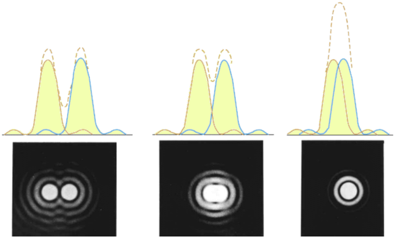
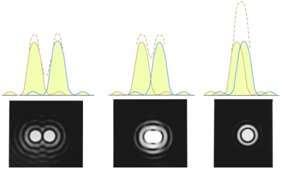


Image Data

A few estimates:

Westerlund 2 is at a distance of ∼20 000 light-years. At that distance, even a supergiant with 1500 $R_\odot$ is under an angle of only 0.002”. Each image pixel has a side length of 0.1”. If our image were a true map showing the exact direction of light, they would fall within a single pixel. Instead, stars appear in the image as disks.

This happens because of **diffraction**, an effect of light passing through an opening (the aperture of the telescope). The result is a diffraction pattern that makes the image of a point source a disk with concentric rings.



The function that defines the telescope’s brightness distribution when imaging a point source is called the **point-spread function (PSF)**.

Image Data

We take another look at the image of Westerlund 2, this time at a part of the image that is not stars, but background. The background is not uniformly black, but mottled grey which becomes obvious with zooming in:

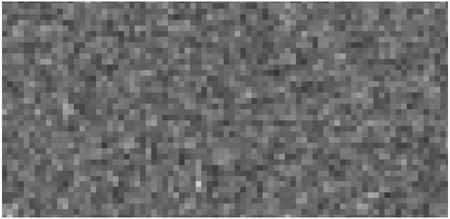
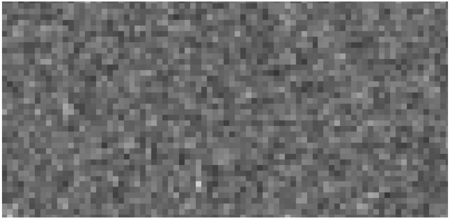


Image Data

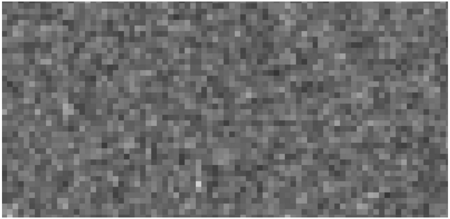
We take another look at the image of Westerlund 2, this time at a part of the image that is not stars, but background. The background is not uniformly black, but mottled grey which becomes obvious with zooming in:



With what we have seen regarding the PSF, the small-scale variation from pixel to pixel cannot be explained by distant, unresolved astronomical objects.

Image Data

We take another look at the image of Westerlund 2, this time at a part of the image that is not stars, but background. The background is not uniformly black, but mottled grey which becomes obvious with zooming in:



With what we have seen regarding the PSF, the small-scale variation from pixel to pixel cannot be explained by distant, unresolved astronomical objects.

Instead, the variability is **noise**.

Image Data

In addition to noise, many astronomical images show traces of cosmic ray particles, like in the following image taken with the Hubble Space telescope.

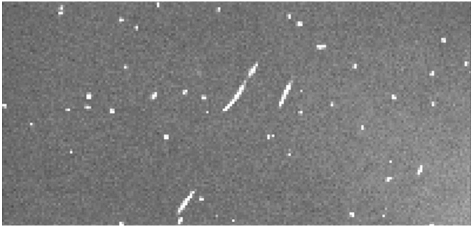


Image credit: NASA and ESA

Image Data

In addition to noise, many astronomical images show traces of cosmic ray particles, like in the following image taken with the Hubble Space telescope.

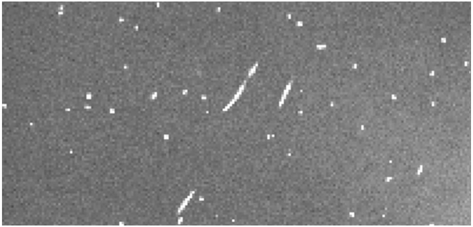


Image credit: NASA and ESA

⇒ raw science images must be preprocessed in a step called **data reduction**

Data Reduction for Image Data

Data reduction is the term for the extraction of astronomical data.

Data Reduction for Image Data

Data reduction is the term for the extraction of astronomical data.

In the case of image data:
raw images $\Rightarrow$ reduced images $\Rightarrow$ light curves

For large surveys, this nowadays happens automatically in a **data reduction pipeline**. Nevertheless it is very useful to know the underlying steps to understand the data produced, and to be able to debug in case of something going wrong. Also, for small observing programs, data reduction often is still done manually, assisted by software.

Data Reduction for Image Data

Data reduction is the term for the extraction of astronomical data.

In the case of image data:
raw images ⇒ reduced images ⇒ light curves

For large surveys, this nowadays happens automatically in a **data reduction pipeline**. Nevertheless it is very useful to know the underlying steps to understand the data produced, and to be able to debug in case of something going wrong. Also, for small observing programs, data reduction often is still done manually, assisted by software.

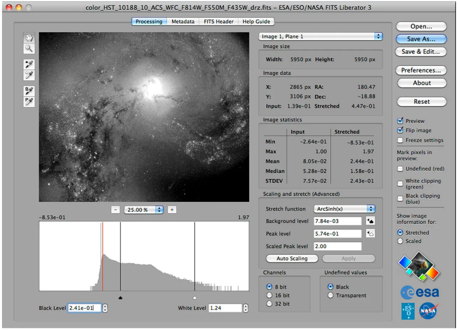
Typical software used for this are IRAF and Source Extractor (SExtractor). In the following, we will demonstrate the general steps and their purpose.

Data Reduction for Image Data

Astronomical images are stored in the FITS (Flexible Image Transport System) format(`*.fits` or `*.fit`). A header is included which gives information such as the properties of the CCD pixels, RA and Dec coordinates for where the telescope was pointing when it took the images.

Data reduction software such as SExtractor, IRAF, Cloudmaker... can then incorporate the header values into the data reduction process.

In addition to imaging data, FITS files can store spectra and data tables.



Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Raw science frame refers to a single exposure image of the object without any post-processing applied to it.

Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Raw science frame refers to a single exposure image of the object without any post-processing applied to it.

Flat frame refers to images taken of a homogeneous surface.

- Using a light panel: Point the telescope up and place a light panel on top of it, or use the dome lights.
- Using the sky: Called *Sky Flats*, this is done when the sky is evenly lit right after sunset or, even better, right before sunrise.

Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Raw science frame refers to a single exposure image of the object without any post-processing applied to it.

Flat frame refers to images taken of a homogeneous surface.

- Using a light panel: Point the telescope up and place a light panel on top of it, or use the dome lights.
- Using the sky: Called *Sky Flats*, this is done when the sky is evenly lit right after sunset or, even better, right before sunrise.

Dark frame refers to images taken when no light can reach the sensor. Dark frames help reduce the noise in our data, although the low thermal noise of new sensors makes them less crucial.

Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Raw science frame refers to a single exposure image of the object without any post-processing applied to it.

Flat frame refers to images taken of a homogeneous surface.

- Using a light panel: Point the telescope up and place a light panel on top of it, or use the dome lights.
- Using the sky: Called *Sky Flats*, this is done when the sky is evenly lit right after sunset or, even better, right before sunrise.

Dark frame refers to images taken when no light can reach the sensor. Dark frames help reduce the noise in our data, although the low thermal noise of new sensors makes them less crucial.

Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Raw science frame refers to a single exposure image of the object without any post-processing applied to it.

Flat frame refers to images taken of a homogeneous surface.

- Using a light panel: Point the telescope up and place a light panel on top of it, or use the dome lights.
- Using the sky: Called *Sky Flats*, this is done when the sky is evenly lit right after sunset or, even better, right before sunrise.

Dark frame refers to images taken when no light can reach the sensor. Dark frames help reduce the noise in our data, although the low thermal noise of new sensors makes them less crucial.

Dark for flat refers to a dark image taken with the same exposure time as the single flat field image.

Data Reduction for Image Data

Before we get started with data reduction, some terms regarding **calibration frames**:

Raw science frame refers to a single exposure image of the object without any post-processing applied to it.

Flat frame refers to images taken of a homogeneous surface.

- Using a light panel: Point the telescope up and place a light panel on top of it, or use the dome lights.
- Using the sky: Called *Sky Flats*, this is done when the sky is evenly lit right after sunset or, even better, right before sunrise.

Dark frame refers to images taken when no light can reach the sensor. Dark frames help reduce the noise in our data, although the low thermal noise of new sensors makes them less crucial.

Dark for flat refers to a dark image taken with the same exposure time as the single flat field image.

Dark for science refers to a dark image taken with the same exposure time as the raw science frame image.

Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

1. Dark Subtraction and Flat Fielding Correction

The dark subtraction and the flat fielding correction are applied to the *raw science frame* according to the following equation:

$$\textit{science frame} = \frac{\textit{raw science frame} - \textit{master dark}}{\textit{master flat} / \langle \textit{master flat} \rangle}$$

where *science frame* refers to the our result, the *raw science frame* after dark subtraction and flat fielding correction; *master dark* refers to the median image of all the *dark for science* images; *master flat* refers to the median image of all the flat images after *dark for flats* subtraction and $\langle$*master flat*$\rangle$ is the mean value of the *master flat* image.

Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

1. Dark Subtraction and Flat Fielding Correction

2. Cosmic Ray Rejection
Correctly identifying and masking cosmic rays is a crucial part of data reduction, since it may otherwise lead to spurious detections. The cosmic ray rejection can be performed with various methods, of them some being performed by machine learning.

Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

1. **Dark Subtraction and Flat Fielding Correction**

2. **Cosmic Ray Rejection**

3. **Background Subtraction**
To accurately measure the photometry and morphological properties of astronomical sources, one requires an accurate estimate of the background, which can be from both the sky and the detector. Similarly, having an accurate estimate of the background noise is important for determining the significance of source detections and for estimating photometric errors. Once the background is estimated, it is subtracted from each science image.

Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

1. **Dark Subtraction and Flat Fielding Correction**

2. **Cosmic Ray Rejection**

3. **Background Subtraction**

4. **Astrometric Calibration**
If for any reason the coordinates of the object are lost. astrometric calibration can recover them.

Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

1. **Dark Subtraction and Flat Fielding Correction**
2. **Cosmic Ray Rejection**
3. **Background Subtraction**
4. **Astrometric Calibration**
5. **Stacking**
Stacking consists of combining several astronomical images. In order to do that, images have to be aligned one respect to each other.

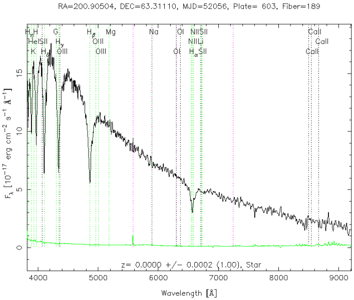
Data Reduction for Image Data

We are now ready to follow the Data Reduction Steps.

1. **Dark Subtraction and Flat Fielding Correction**
2. **Cosmic Ray Rejection**
3. **Background Subtraction**
4. **Astrometric Calibration**
5. **Stacking**
6. **Photometric Calibration**
Astronomical instruments deliver electron counts which are proportional to photon counts. From them, instrumental magnitudes can be calculated. Photometric calibration, which is the conversion to standard magnitudes, requires the image of a standard star that has gone through all the data reduction steps we mentioned so far.

Spectra

Photometric data is taken through a filter (or multiple filters), covering a certain wavelength range and integrating the light over it.



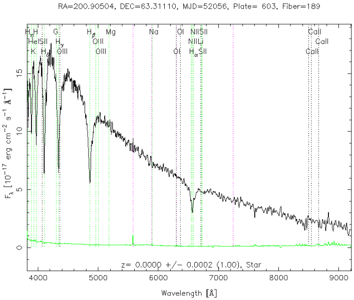
A SDSS spectrum. The plot includes additional information, such as the object, time of observation, the plate and fiber the spectrum was taken. In green, the positions of specific spectral lines are highlighted.

Spectra

Photometric data is taken through a filter (or multiple filters), covering a certain wavelength range and integrating the light over it.

In contrast, spectra show us how the energy of the light emitted by an object is distributed among the different possible wavelengths. Such 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.



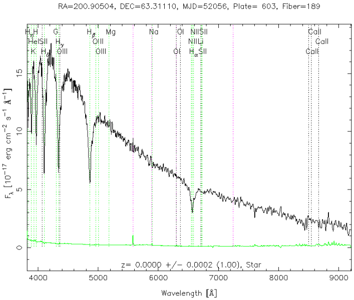
A SDSS spectrum. The plot includes additional information, such as the object, time of observation, the plate and fiber the spectrum was taken. In green, the positions of specific spectral lines are highlighted.

Spectra

Photometric data is taken through a filter (or multiple filters), covering a certain wavelength range and integrating the light over it.

In contrast, spectra show us how the energy of the light emitted by an object is distributed among the different possible wavelengths. Such 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.

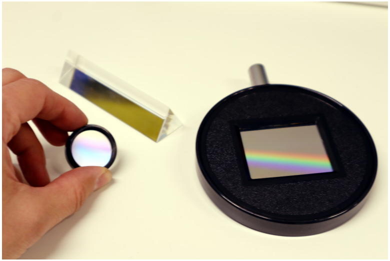


A SDSS spectrum. The plot includes additional information, such as the object, time of observation, the plate and fiber the spectrum was taken. In green, the positions of specific spectral lines are highlighted.

Spectra

Spectra are measured with a **spectrograph**.

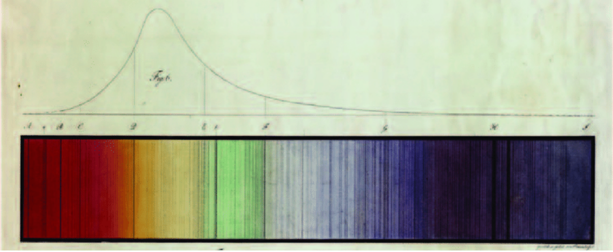
A spectrograph contains a dispersive element (or even more than one), which splits the incoming light into its rainbow colours or, in physics terminology, into its different wavelengths.



The three basic types of dispersive element, from left to right: transmission grating, prism, and reflective grating. Credit: Pössel (2020).

Spectra

The first spectra taken were those of the Sun.
In 1666, Isaac Newton showed that a glass prism could be used to split sunlight into a spectrum. Further studies by William Wollaston in 1802 revealed some black lines in the solar spectrum. More detailed observations by Joseph von Fraunhofer resulted in 574 of these lines being mapped by 1815. These lines are now known as the "Fraunhofer lines", are absorption lines from which elements in the Sun can be identified.



Fraunhofer's map of the solar spectrum (colored by hand, ca. 1814). Source: Deutsches Museum, Munich, map collection, StO 1107, cab. 39, shelf 03.

Spectra

A simple spectrometer is a setup like the following:

Engraving of Isaac Newton and prism experiment. Credit: Science Photo Library.

Spectra

A simple spectrometer is a setup like the following:

Engraving of Isaac Newton and prism experiment. Credit: Science Photo Library.

Modern spectrographs use multiple diffraction gratings, a CCD as detector (before: photographic paper), and a slit mask selects individual stars.

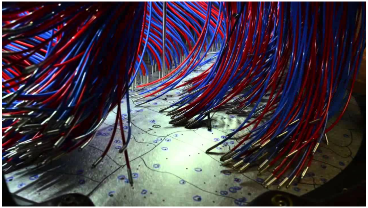
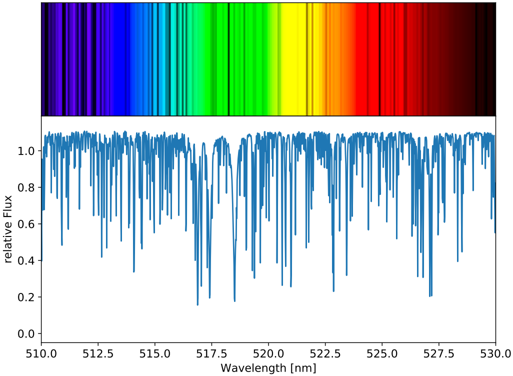


Plate in the multi-object, fiber-fed spectrograph for the Sloan Digital Sky Survey. Credit: SDSS

Spectra

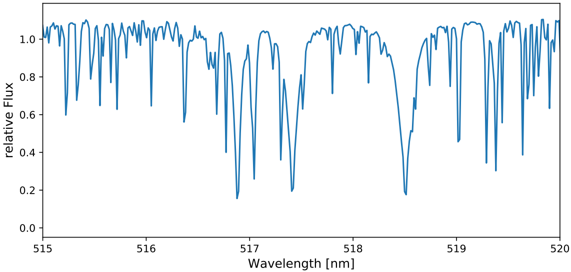
For each wavelength, we have a quantity indicating how much light is emitted in that particular wavelength region. Thus, spectra are typically plotted as curves.



Solar spectrum as a curve (bottom) and the reconstructed version of a slit spectrum image (top). IAG Solar Flux Atlas, Reiners et al. 2016

Spectra

The curve is quite complex with a **forest of absorption lines**. If we plot a much smaller wavelength interval, you can see the the lines themselves have characteristic shapes:



Spectra

When working with raw data from spectra, certain **data reduction** steps need to be taken.

Some of those are similar to the reduction of photometric data:

Flatfielding is again necessary to compensate for differing sensitivity of the instrument in different parts of the spectrum. This is more difficult for a spectrum than for an image, since for a true flat field, you would need a perfectly flat spectrum.
Instead, any well-known, preferably smooth calibration spectrum can be used to deduce the varying sensitivity.

Dark frames again can be used to take into account that the electronics of the detector will produce some spurious brightness in the image, which needs to be subtracted.

Spectra

Other reduction steps are specific for spectra:

Telluric lines are visible in the case of ground-based telescopes: Light absorption in the Earth's atmosphere creates telluric lines which need to be removed.

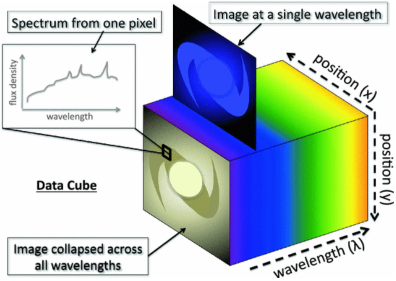
The spectra to analyse are expected to have an accurate **wavelength calibration**. However, small errors can affect the determination of the line spread function and column densities negatively. Therefore in each inclusion region, the synthetic spectrum being fitted can be adjusted to match the input spectrum, or the telluric lines can be used as a reference.

Image distortions can make spectral data reduction particularly challenging. A particularly complex case are Echelle spectrographs, where two kinds of spectral dispersion are combined.

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.

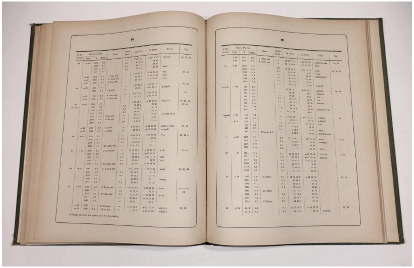
An example for a data cube is data from **Integral Field Spectroscopy (IFS)**. IFS combines spectrographic and imaging capabilities: take a single exposure of a 2D region and obtain spatially resolved spectra.



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.

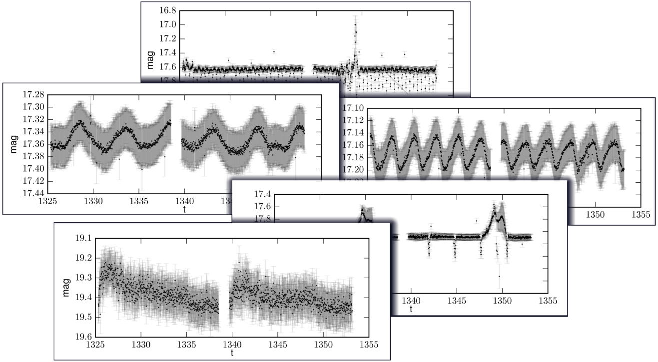


The catalog from the "Bonner Durchmusterung" (survey by Bonn observatory) from 1862

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.

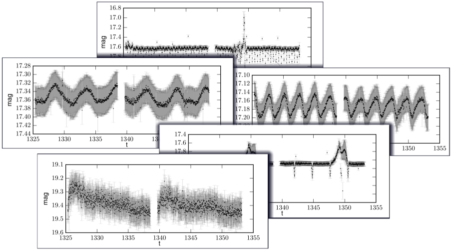
Photometry of an object taken at multiple times is called a **light curve**.



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.

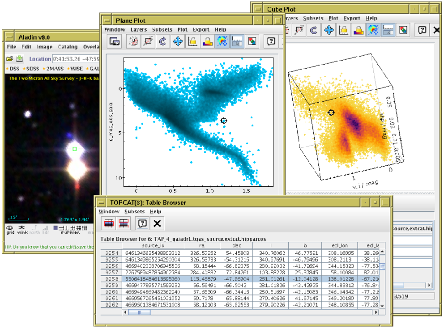
A **data access language** allows for selecting data from (usually remote) databases.

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/`



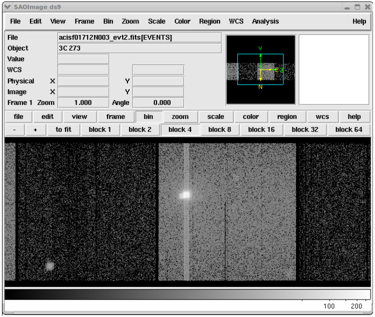
Application Software

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.

Surveys as Astronomical Data Sources

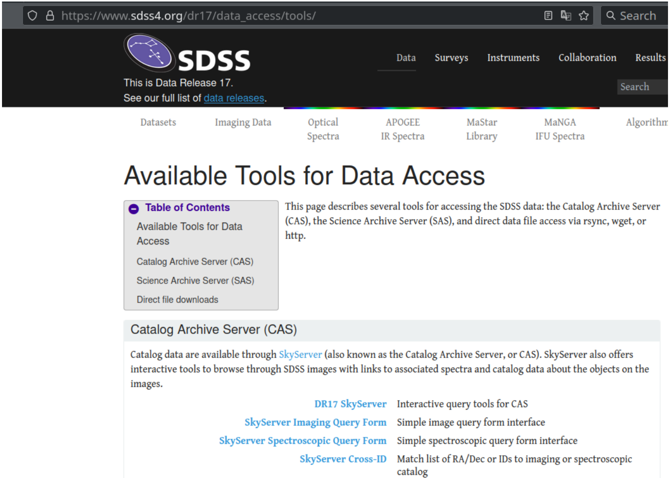
Astronomical surveys often share their full data sets. In some cases, these are shared via the observatory that acquired them, for example, the all-sky data acquired with Planck, Wilkinson Microwave Anisotropy Probe, and COBE.

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)

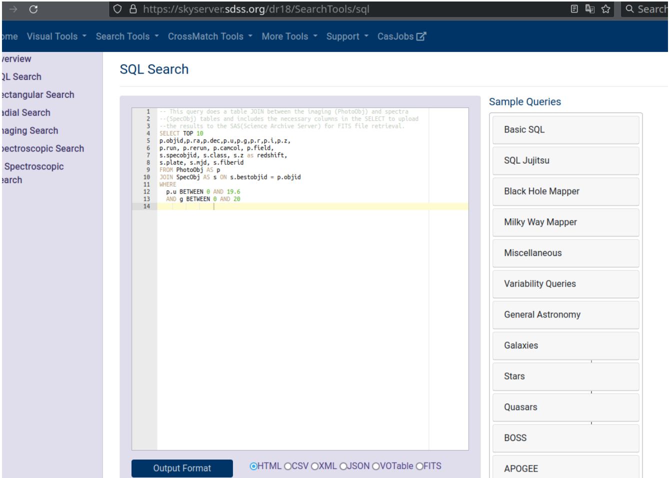
Catalog data are available through SkyServer (also known as the Catalog Archive Server, or CAS). SkyServer also offers interactive tools to browse through SDSS images with links to associated spectra and catalog data about the objects on the images.

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

Surveys as Astronomical Data Sources

Example: SDSS

`https://skyserver.sdss.org`



Surveys as Astronomical Data Sources

Example: SDSS

`sdss-access` provides Python utilities for downloading files from the Science Archive Server (SAS). Examples and more details can be found on the official documentation.

`https://www.sdss.org/dr18/software/packages/sdss-access/`

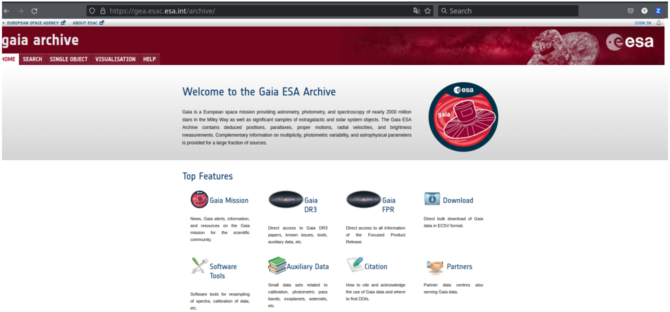
install `sdss-access` with:

```
pip install sdss-access
```

Surveys as Astronomical Data Sources

Example: Gaia

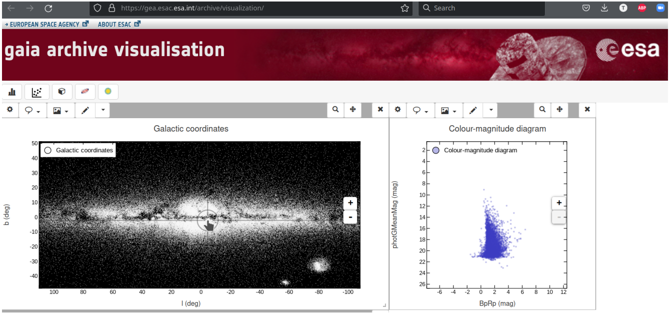
`https://gea.esac.esa.int/archive/`



Surveys as Astronomical Data Sources

Example: Gaia

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



Surveys as Astronomical Data Sources

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`

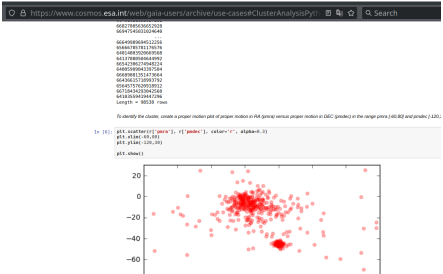
Query:
```
SELECT *, DISTANCE(81.28, -69.78, ra, dec) AS ang_sep
FROM gaiadr3.gaia_source
WHERE DISTANCE(81.28, -69.78, ra, dec) < 5./60.
AND phot_g_mean_mag < 20.5
AND parallax IS NOT NULL
ORDER BY ang_sep ASC
```

Example: Gaia

ESA Gaia Archive also provides a Python Astroquery package for easy access to the ESA Gaia Archive: `astroquery.gaia`.

In addition, a generic package is provided for accessing any TAP-compliant service: `astroquery.utils.tap`.

An example Jupyter notebook is provided:
`https://www.cosmos.esa.int/web/gaia-users/archive/use-cases#ClusterAnalysisPythonTutorial`



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

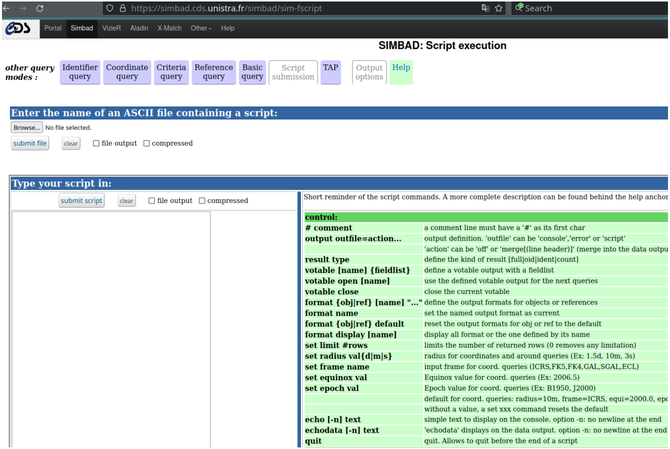
The SIMBAD astronomical database provides basic data, cross-identifications, bibliography and measurements for astronomical objects outside the solar system. SIMBAD can be queried by object name, coordinates and various criteria. Lists of objects and scripts can be submitted. Links to some other on-line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

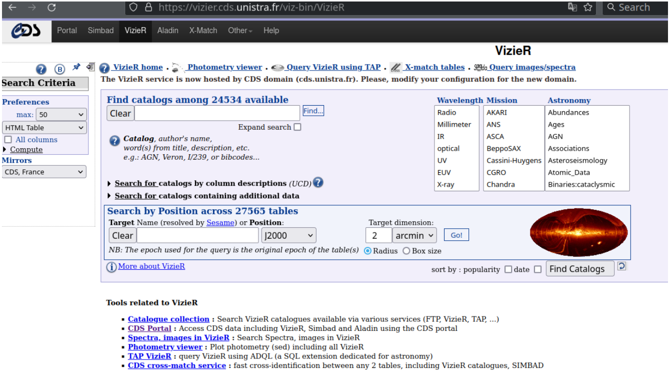
Services

`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

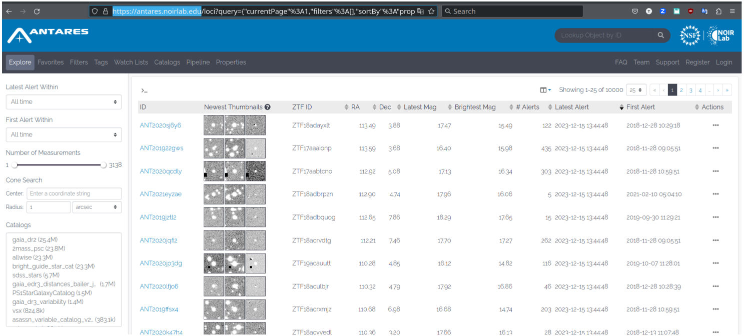
The newest generation of large survey telescopes such as the Vera C. Rubin Observatory (which will conduct the 10-year Legacy Survey of Space and Time (LSST)) are designed to produce large streams of transient alerts. They will be processed, annotated and classified by a new generation of astronomical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

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

Community brokers

Example: The ANTARES Broker

`https://antares.noirlab.edu/`



What is astroquery?

Astroquery as part of astropy is a set of tools for querying astronomical web forms and databases.

In addition to providing access to astronomical databases, astroquery provides data such as molecular and atomic properties, e.g. by the NIST Atomic Spectra Database, bibliographic databases such as the NASA Astrophysics Data System (ADS), or services that are computationally intensive or require regular updates, like solar system ephemerides provided by services like JPL HORIZONS, or the Minor Planet Center.

https://github.com/astropy/astroquery
https://astroquery.readthedocs.io/

What is astroquery?

These tools are built on the Python `requests` package, which is used to make HTTP requests, and `astropy`, which provides most of the data parsing functionality. astroquery modules generally attempt to replicate the web page interface provided by a given service as closely as possible, making the transition from browser-based to command-line interaction easy.

The complete documentation for `astroquery` can be found at

`http://astroquery.readthedocs.io/`.

Using astroquery

All astroquery modules are supposed to follow the same API. In its simplest form, the API involves queries based on coordinates or object names. Some simple examples, using SIMBAD:

```
from astroquery.simbad import Simbad

result_table = Simbad.query_object("m1")

print(result_table)
```

```
MAIN_ID      RA          DEC      ...     COO_BIBCODE      SCRIPT_NUMBER_ID
           "h:m:s"     "d:m:s"    ...
------- ---------- ---------- ... ------------------- ----------------
  M   1 05 34 30.9 +22 00 53 ... 1995AuJPh..48..143S                1
```

Using astroquery

All query tools allow coordinate-based queries:

```python
from astropy import coordinates
import astropy.units as u

# works only for ICRS coordinates:

c = coordinates.SkyCoord("05h35m17.3s -05d23m28s", frame='icrs')
r = 5 * u.arcminute

result_table = Simbad.query_region(c, radius=r)
print(result_table)
```

```
       MAIN_ID                RA        ...     COO_BIBCODE      SCRIPT_NU
                           "h:m:s"      ...
------------------------ ------------- ... ------------------- ---------
       NAME Ori Region    05 35 17.30  ...
                     ...           ... ...                 ...
2MASS J05353573-0525256 05 35 35.7755 ... 2020yCat.1350....0G
           V* V2114 Ori 05 35 01.6720 ... 2020yCat.1350....0G
Length = 3272 rows
```

Using `astroquery`

The following is an alphabetic **selection of available services** through `astroquery`; the full list can befound in the documentation:

ALMA Queries (`astroquery.alma`)
Atomic Line List (`astroquery.atomic`)
Besancon Queries (`astroquery.besancon`)
CADC (`astroquery.cadc`)
CASDA Queries (`astroquery.casda`)
Cologne Database for Molecular Spectroscopy (CDMS) Queries (`astroquery.linelists.cdms`)
DACE (`astroquery.dace`)
ESA Herschel Science Archive (`astroquery.esa.hsa`)
ESA HST Archive (`astroquery.esa.hubble`)
ESA ISO Archive (`astroquery.esa.iso`)
ESA JWST Archive (`astroquery.esa.jwst`)
ESA XMM-Newton Archive (`astroquery.esa.xmm_newton`)

Summary

We have seen a variety of data that can be present in astronomical surveys.

We have also seen that those data are usually not raw data, but certain data reduction steps were done, and are necessary for data you take during your observing runs.

Access to astronomical survey data is nowadays usually possible over the internet. Being familiar with queries helps us to select the data we are interested in for our scientific use case.