

# ESCAPE

## Building the infrastructure for the next generation astronomy



# Challenges

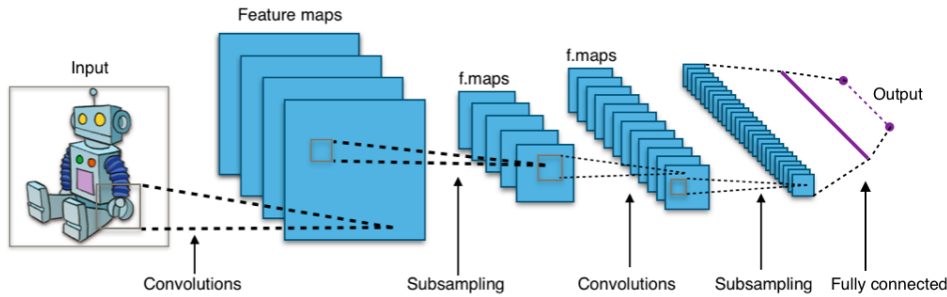
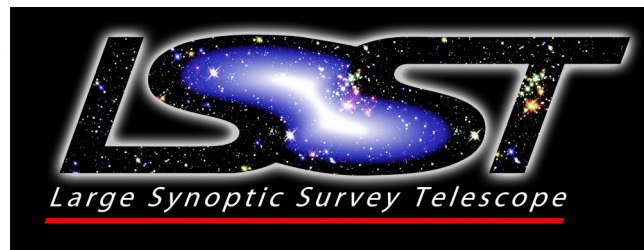


Image from Wikipedia



Big Data by Nick Youngson CC BY-SA 3.0 Alpha Stock Images



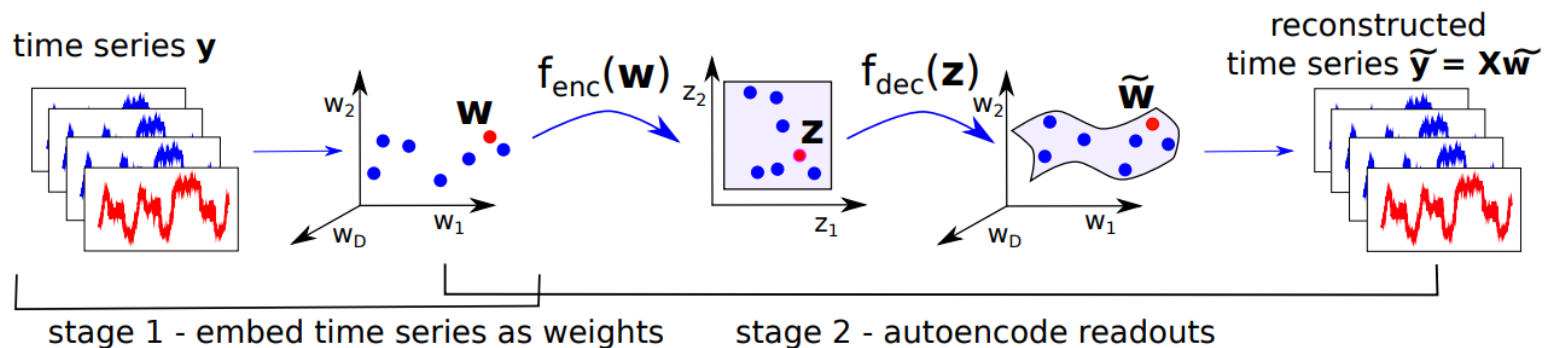
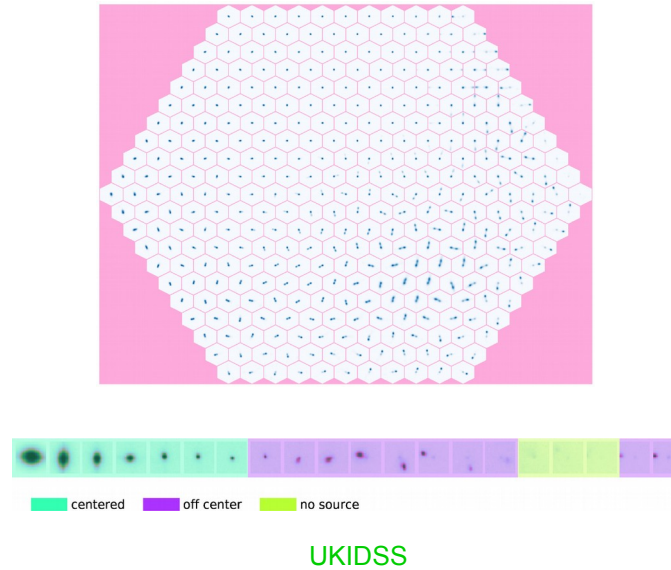
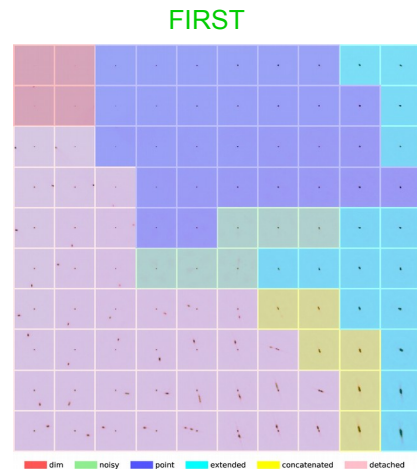


## ESCAPE: European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

- Accessibility to huge amount of data provided by research infrastructures and facilities
- Bring together partners from astronomy and particle physics
- Deliver solutions to ensure integration of data, tools, services and software
- Build standards and ensure interoperability

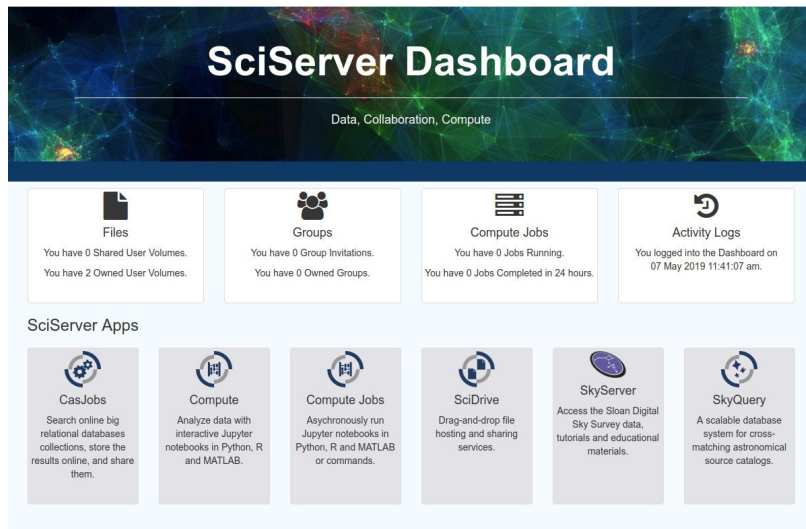
# Dimensionality reduction

The story so far...



# Challenges: data products

Working with catalogs is a simple task:



**SciServer Dashboard**  
Data, Collaboration, Compute

**Files**  
You have 0 Shared User Volumes.  
You have 2 Owned User Volumes.

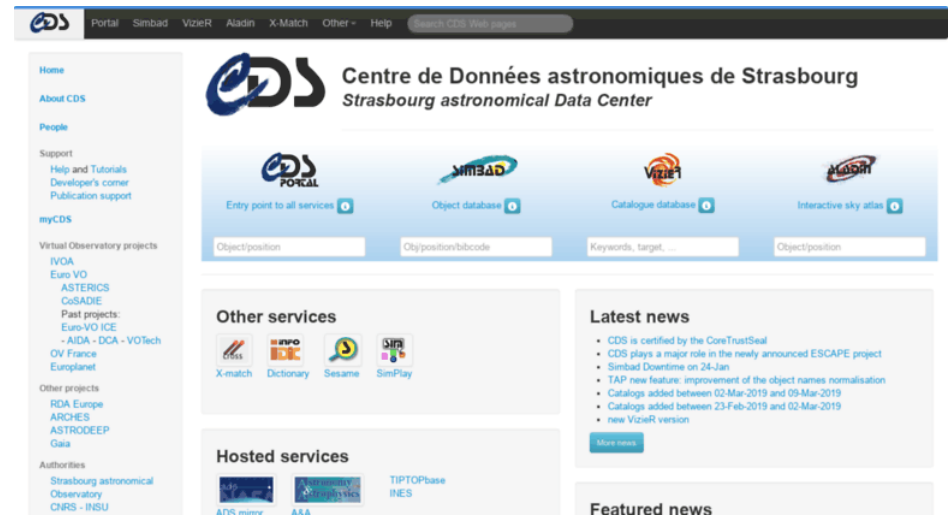
**Groups**  
You have 0 Group Invitations.  
You have 0 Owned Groups.

**Compute Jobs**  
You have 0 Jobs Running.  
You have 0 Jobs Completed in 24 hours.

**Activity Logs**  
You logged into the Dashboard on  
07 May 2019 11:41:07 am.

**SciServer Apps**

- CasJobs**  
Search online big relational databases collections, store the results online, and share them.
- Compute**  
Analyze data with interactive Jupyter notebooks in Python, R and MATLAB.
- Compute Jobs**  
Asynchronously run Jupyter notebooks in Python, R and MATLAB or commands.
- SciDrive**  
Drag-and-drop file hosting and sharing services.
- SkyServer**  
Access the Sloan Digital Sky Survey data, tutorials and educational materials.
- SkyQuery**  
A scalable database system for cross-matching astronomical source catalogs.



**CDS** Centre de Données astronomiques de Strasbourg  
*Strasbourg astronomical Data Center*

Portal Simbad Vizier Aladin X-Match Other - Help

**Other services**

- CDS X-match
- IRAP Dictionary
- ESO Sesame
- DSO SimPlay

**Hosted services**

- ATP mirror
- AA
- TIPTOPbase
- INES

**Latest news**

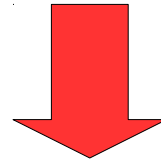
- CDS is certified by the CoreTrustSeal
- CDS plays a major role in the newly announced ESCAPE project
- Simbad Downtime on 24-Jan
- TAP new feature: improvement of the object names normalisation
- Catalogs added between 02-Mar-2019 and 09-Mar-2019
- Catalogs added between 23-Feb-2019 and 02-Mar-2019
- new Vizier version

**Featured news**

➔ Problems start with images and spectra!

# Some “simple” tasks...

1. Given the coordinates, download 28x28 pixel<sup>2</sup> images for all the quasars in SDSS.
2. Download some hundreds of thousands of images from FIRST/UKIDSS.
3. Download all the HARPS spectra from ESO archive.



Obtaining data  
products can be a  
not easy task

# Task 3 - Solutions?

- Python script available on request, but frequent crashes experienced → required two months of attempts
- Alternative: direct shipping of hard drive



**Asking friends and colleagues for data and support can be helpful, but it is not what standardization and the ESCAPE paradigm are about!**

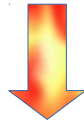


# HARPS

HARPS: the High Accuracy Radial velocity Planet Searcher at the  
ESO La Silla 3.6m telescope

267.487 high resolution spectra

→ multiple observations



7.535 unique spectra

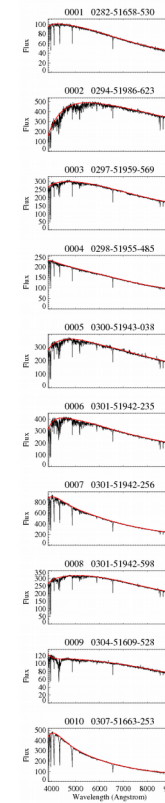
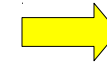
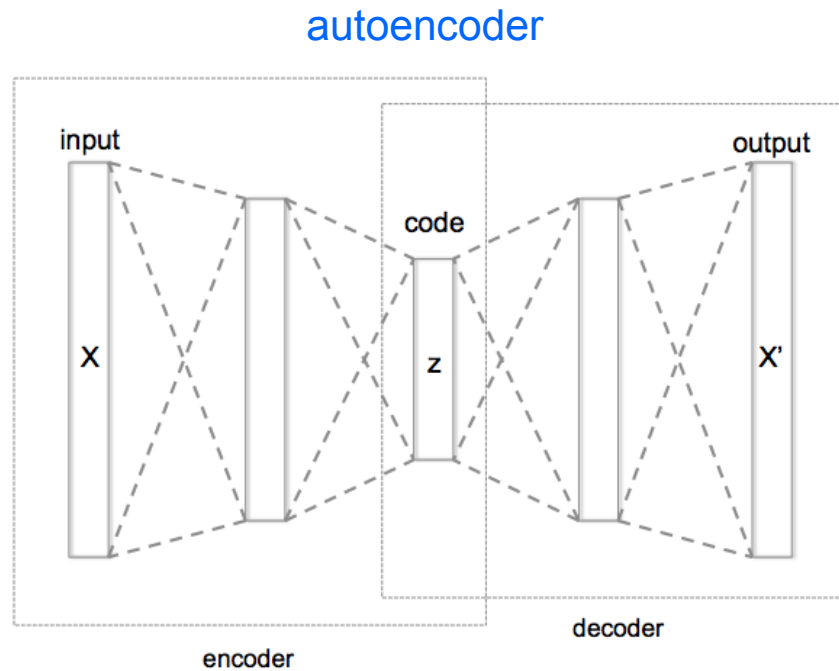
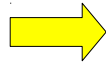
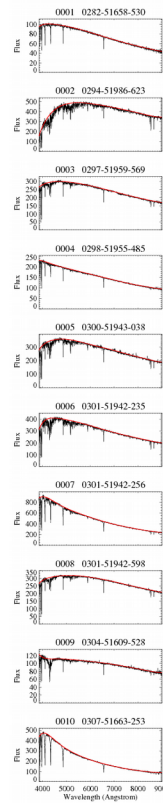


Image from Wikipedia

# First developments

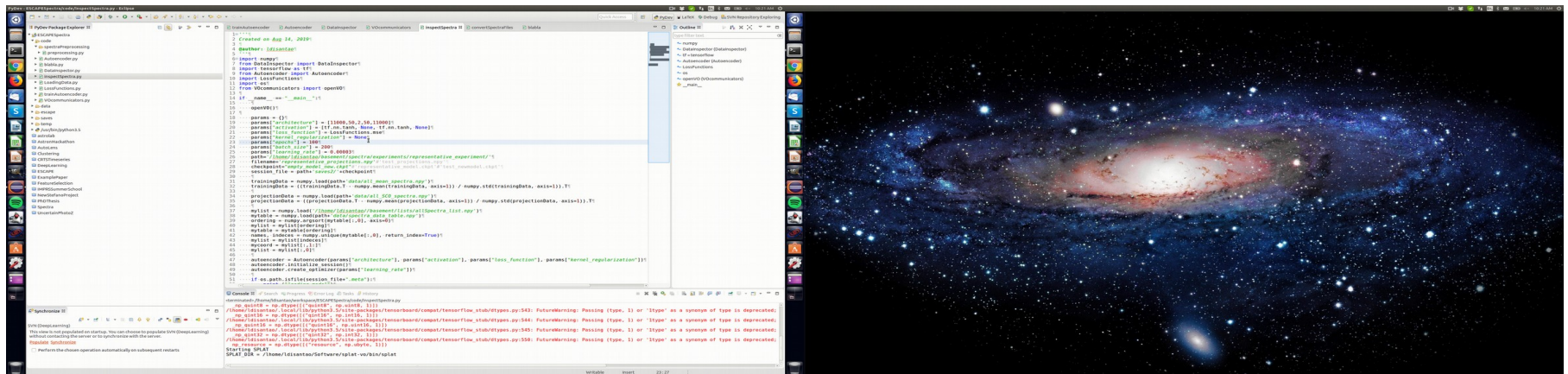
Development of a prototype for:

Dimensionality reduction and analysis of spectra



representation

# Prototype



# Conclusions

- ESCAPE project is going to be a step to build a new infrastructure for data-intense astronomy.
- A lot of work to do:
  - × data products access
  - × building standards
  - × bringing code to the data
- Development of a first prototype → big potential and future integration in web services → allowing to do a lot of science.
- Final question: are we ready for machine learning and big data?