

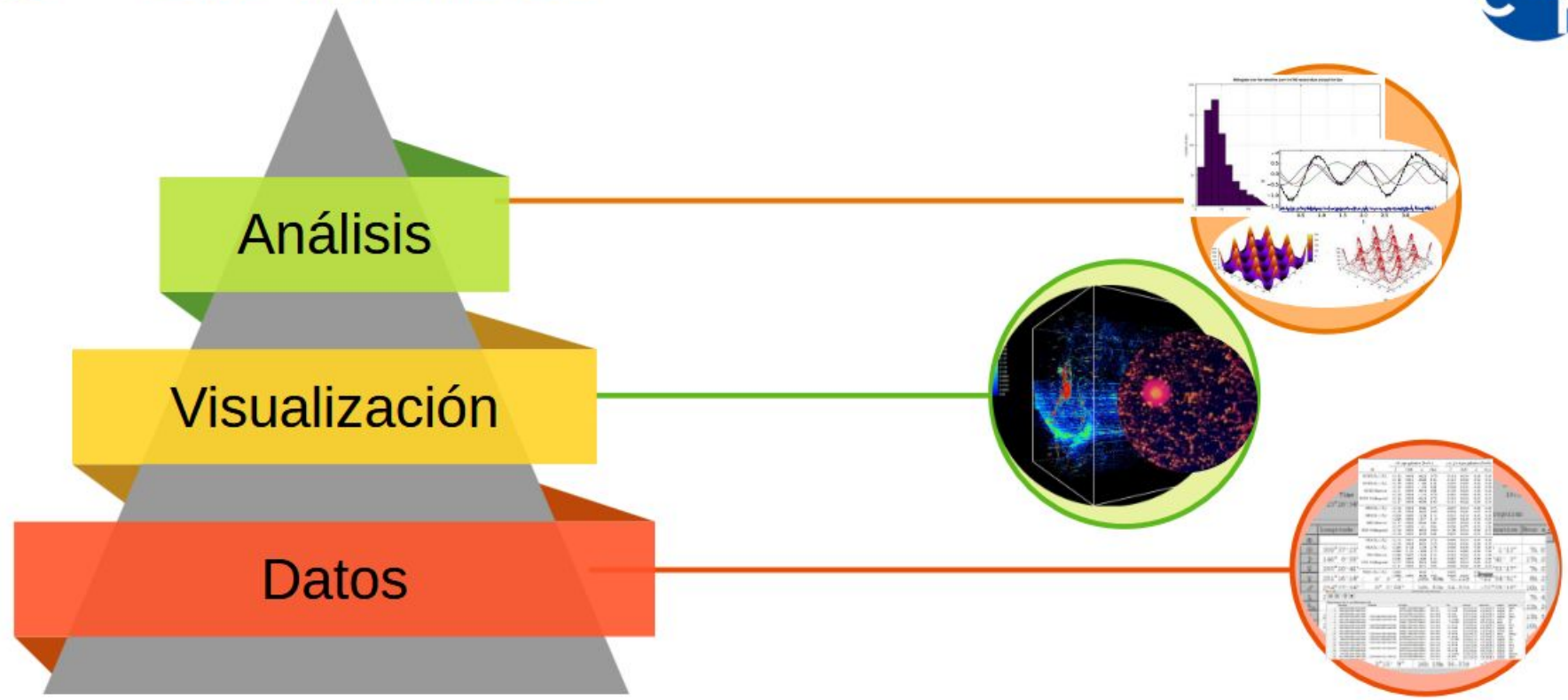
Pluralismo
Compromiso
Inclusión



Procesamiento de datos

Licenciatura en Astronomía

Data warehouses & surveys



Base de datos centralizada, integrada y consolidada.

• Data Warehouse en astronomía

- Diversos intentos en el pasado.
- Mayores desafíos:
 - (1) integración de datos provenientes de diferentes fuentes (datos heterogéneos).
 - (2) preprocesamiento de alto nivel
 - (3) herramientas de visualización.

Mismo problema tuvo que abordar la industria hace 60 años atras.

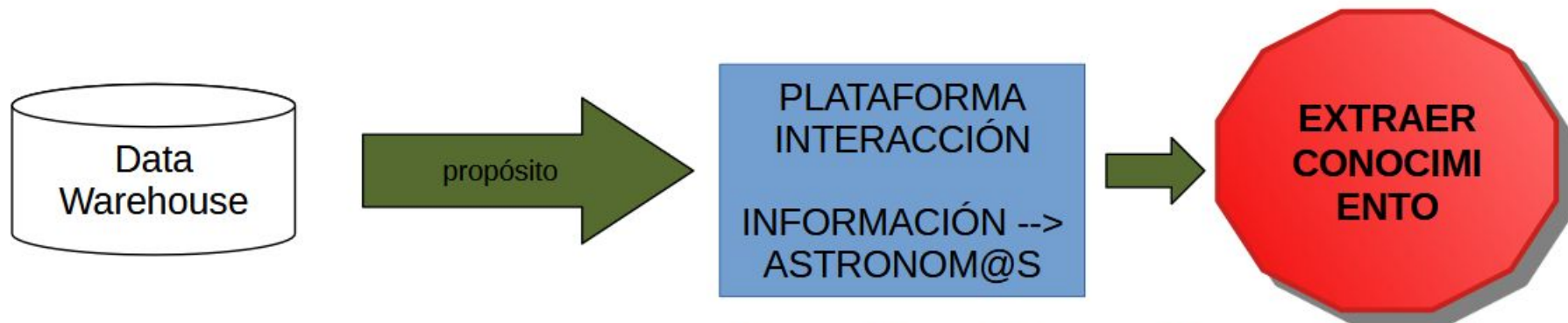


Table 1. Big Data 3V characteristics in astronomical sky surveys.

Sky Survey	Volume	Velocity	Variety
SDSS <i>Sloan Digital Sky Survey</i>	50 TB	200 GB per day	images, catalogs, redshifts
GAIA	100 TB	40 GB per day	more than 100 parameters
Pan-STARRS <i>Panoramic Survey Telescope and Rapid Response System</i>	5 PB	5 TB per day	images, catalogs
LSST <i>Large Synoptic Survey Telescope</i>	60 PB	10 TB per day	images, catalogs
SKA <i>Square Kilometer Array</i>	3 ZB	150 TB per day	images, catalog, redshifts

Notes:

The column Volume refers to raw data produced at the end of the experiment.

Values regarding Pan-STARRS, LSST, and SKA surveys refer to expected Volume and Velocity values.

Garofalo et al. 2016



Problemática actual: Nuestra incapacidad de procesar correctamente toda esta información, ha llevado a un decrecimiento en la generación de nuevo conocimiento en astronomía, a pesar de que cada año el volumen de datos disponibles se duplica! (seguimos apilando datos).

Data Warehouses en astronomía: iniciativas



Data Mining & Exploration


We make science discovery happen

- Home
- Software Services
- Science Cases
- Publications
- Education & Lectures
- Who's who

Latest News and Events

May 14-17, 2019
SAN 2019
LXXII Congresso della SAI
Rome, Italy

June 4-7, 2019
EC Meeting 2019
Helsinki, Finland




Welcome on DAME official web portal!

! The DAME website has changed the home address. The new address is [HERE](http://dame.oacn.inaf.it/). Please use this address in the future!

Nowadays, many scientific areas share the same need of being able to deal with massive and distributed datasets and to perform complex knowledge extraction tasks. DAME (Data Mining & Exploration) is a general purpose, web-based, distributed data mining infrastructure specialized in Massive Data Sets exploration with machine learning methods.

Initially fine tuned to deal with astronomical data only, DAME has evolved in a general purpose platform program, hosting a cloud of applications and services useful also in other domains of human endeavor. DAME is an evolving platform and new services as well as additional features are continuously added. The work for architecture of DAME can also be extended in public applications, hereby:


STraDiWA



A web application for the research on sky transient objects based on interactive tools to generate and process astronomical simulated images.

[Go to the service](#)


VOGClusters



A web application specialized for data and text mining on globular clusters.

[Go to the service](#)

ViaLactea



Local site hosting an overview of data mining tools designed for the FIT Programme.

[Go to the service](#)



Centre de Données astronomiques de Strasbourg



Point d'entrée vers tous les services



Base de données d'objets



Base de données de catalogues



Atlas interactif du ciel

Autres services



X-match



Diccionario



Sesame



SimPlay

Actualités

- Aladin Lite v1 talk at ADASS 2022
- Aladin Lite v5 release
- DECaPS DR2 HPS available
- Aladin Desktop v12 release
- Catalogs added between 18-Feb-2023 and 25-Feb-2023
- Catalogs added between 11-Feb-2023 and 18-Feb-2023
- Catalogs added between 04-Feb-2023 and 11-Feb-2023

[Plus de news](#)

Services de nos partenaires



ADS



A&A



TPTOPbase

A la une

Centro de datos CDS:

<https://cdsweb.u-strasbg.fr/index-fr.gml>

SIMBAD: identificación y bibliografía.

VizieR: catálogos y tablas astronómicas.

Aladin: software de visualización y análisis.

Data Warehouses en astronomía: iniciativas



Starlink SPLAT-VO

DAVID SPLAT

The VO elements of SPLAT are now being developed by the Göttingen Astrophysics Virtual Observatory (GAVO) in co-operation with the Astrophysics Institute of the University of Sciences of the Czech Republic. There is a webpage describing this effort at:

• <http://www.gvo.uni-goettingen.de/splat/>

Earlier releases of SPLAT may become available. A mailing list for the discussion of these developments is also available at:

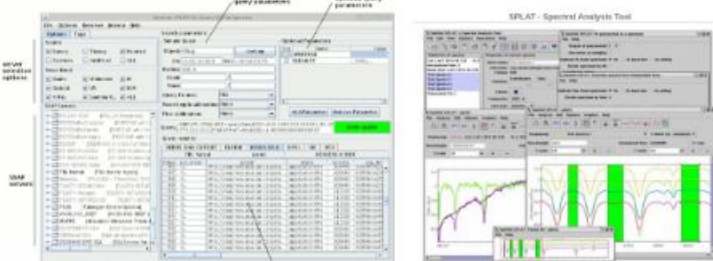
• <http://lists.gvo.uni-goettingen.de/mailman/listinfo/splat-users>

Starlink SPLAT-VO



SPLAT has been extended to include facilities that allow it to work as part of the Virtual Observatory (VO). These facilities (and a few others) have, so far, been used for querying and controlling spectra from SSAP servers and for interfacing with VO tools, such as TOPCAT, on your desktop using SAMF.

The SSAP facilities have been developed by Mark Taylor and the SSAP users by myself and Mark Taylor. The other changes to the extension available for use for the work of the Göttingen group are some of the effort described at DAVID SPLAT.



SPLAT-VO:

<http://star-www.dur.ac.uk/~pdraper/splat/splat.html>

Herramienta de análisis espectral que puede bajar y analizar espectros desde internet y otras bases de datos.

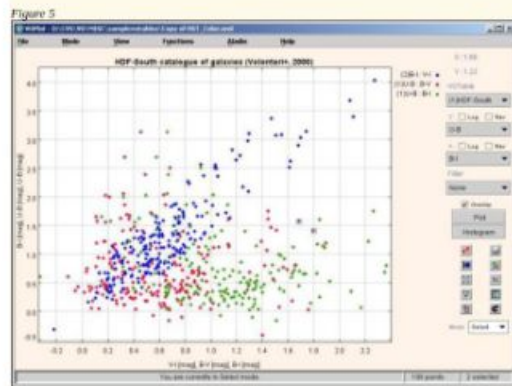
VOPlot – The VOTable plotting utility (Version 1.2.1)



Introduction

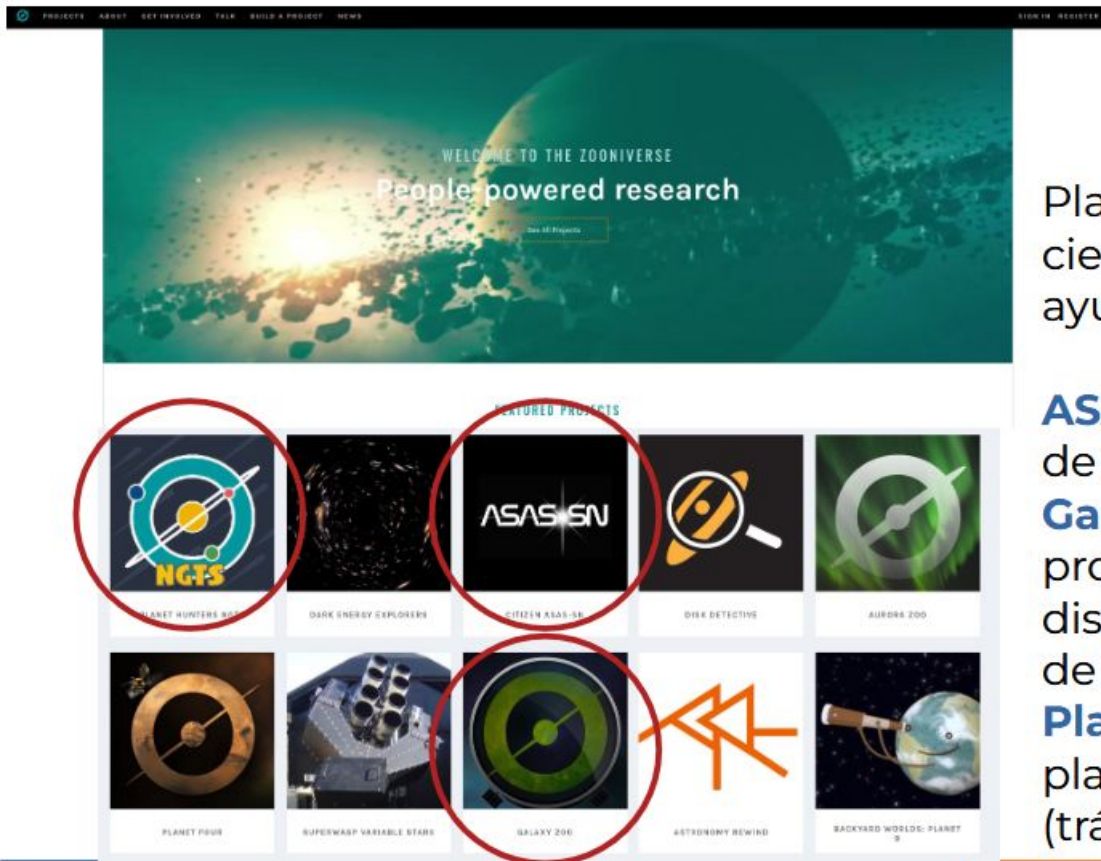
VOPlot (VOTable Plotting tool) is an applet for plotting different astronomical graphs using data stored in VOTable format. VOPlot is available in desktop version and a web-based version. The web-based version is integrated with the [VizieR Catalogue Service](http://vizier.cds.fr) and can be used to plot any catalogue by selecting output layout as "Plot (VOPlot)". The desktop version can be downloaded from <http://vo.iucaa.ernet.in/~voi/voplot.htm>.

A sample plot with two VOTables loaded is shown in figure 5 below.



http://cdsarc.u-strasbg.fr/vizier/VOPlot/#_Introduction

• Data Warehouses en astronomía: iniciativas



<https://www.zooniverse.org/>

Plataforma que reúne diversos proyectos científicos y que permite que voluntarios ayuden en el análisis de los datos.

ASAS-SN: Clasificación de curvas de luz de estrellas variables inusuales.

Galaxy-Zoo: Imágenes de campo profundo de diversas galaxias desde distintos catálogos. Se deben clasificar de acuerdo a su morfología.

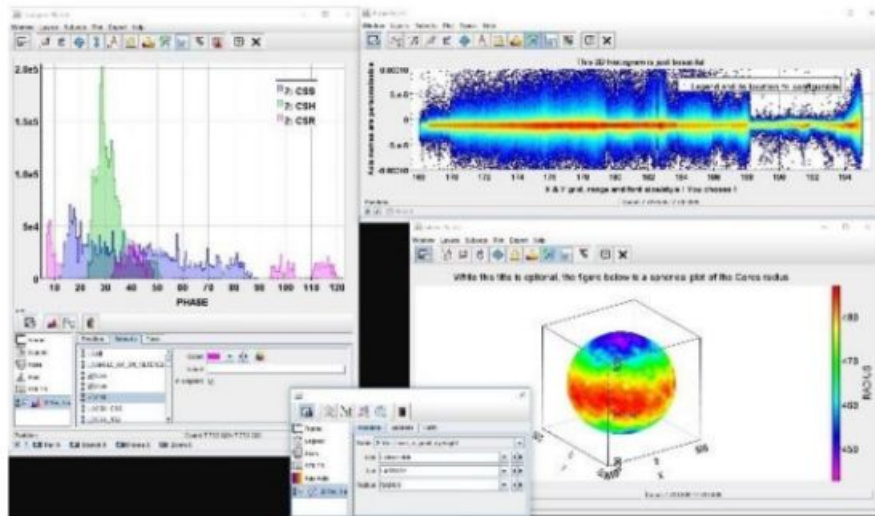
Planet Hunters: Descubrir nuevos planetas usando datos de surveys (tránsito).



TOPCAT

What is TOPCAT?

TOPCAT is an interactive graphical viewer and editor for tabular data. Its aim is to provide most of the facilities that astronomers need for analysis and manipulation of source catalogues and other tables, though it can be used for non-astronomical data as well. It understands a number of different astronomically important formats (including FITS, VOTable and CDF) and more formats can be added. It is especially good at *interactive exploration of large (multi-million row, lots of columns) tables*.



Provee diversas herramientas para manipular y analizar grandes volúmenes de datos provenientes de catálogos y tablas.

Diversos tipos de gráficos

Parámetros estadísticos (media, SD, varianza, cuantiles, etc..)

Cross-matching (datos en Vizier o SIMBAD).

Table 2.1. Virtual Observatories Worldwide

Name	Internet Address	Main Focus
International Virtual Observatory Alliance (IVOA) ¹	http://www.ivoa.net/	#Collaboration #Promotion #Developing Standards #Tools and Services #FITS #Images #Data Center #Visualization #Statistics #Spectroscopy #Photographic #DFBS ² #Images #Search #Theoretical Queryable Data ³ #SkyMapper ⁴ #Promotion #Secure Data ⁵ #Data Center #Search #Data Integration #Collaboration #Images #Data Center ⁶ #Data Center ⁷ #FITS search #Jupyter Notebooks ⁸ #Education #Images #Spectra (Search and Visualization) #Promotion #Data Center #Teaching #Services ⁹ #Collaboration #Morphological Parameters of Galaxy Images #Visualization #Statistical #FITS #Promotion #Promotion #Star Evolution #Variable Stars #Experimental Astronomy #Spectro-UV ¹⁰ #Multi-wavelength #Promotion #Data Center - Spectra Photometric and Theoretical ¹¹ #Services #Data Mining (Classification) #Promotion #Data Center #Services #Collaboration #Promotion #Standards #Tools and Services ¹²
Nuevo Observatorio Virtual Argentino (NOVA)	http://nova.iafe.uba.ar/	
Armenia Virtual Observatory (ArVO)	https://www.aras.am/Arvo/main.htm	
AstroGrid	http://www.astrogrid.org/	
Australian Virtual Observatory	http://www.asvo.org.au/	
Brazilian Virtual Observatory (BRAVO)	http://bravo.iag.usp.br/	
Chinese Virtual Observatory (China-VO)	http://www.china-vo.org/	
Canadian Virtual Observatory (CVO)	http://www.cadc-ccda.hia-ihp.nrc.ca/cvo/	
Chile Virtual Observatory (CHIVO)	http://www.chivo.cl/	
The European Virtual Observatory EURO-VO	http://www.euro-vo.org/	
France Virtual Observatory	http://www.france-vo.org/	
German Astrophysical Virtual Observatory (GAVO)	http://www.g-vo.org/	
Hungarian Virtual Observatory (HVO)	http://hvo.elte.hu/	
India Virtual Observatory	http://voi.iucaa.res.in/	
Japanese Virtual Observatory (JVO)	http://jvo.nao.ac.jp/	
Korean Virtual Observatory	http://kvo.nasi.re.kr/	
Russian Virtual Observatory (RVO)	http://www.inr.su/eng/rvo/	
South African Virtual Observatory	http://www.svo.ac.za/	
Spanish Virtual Observatory	http://vo.inta-csic.es/main/index.php	
Italian Virtual Observatory	http://vobs.astro.it/	
Ukrainian Virtual Observatory	http://www.ukr-vo.org/	
United States Virtual Observatory Alliance (USVOA)	http://usvoa.cfa.harvard.edu/	



Spanish Virtual
Observatory



EURO VO



Machin, J. 2017

• Data Warehouses en astronomía: desafíos

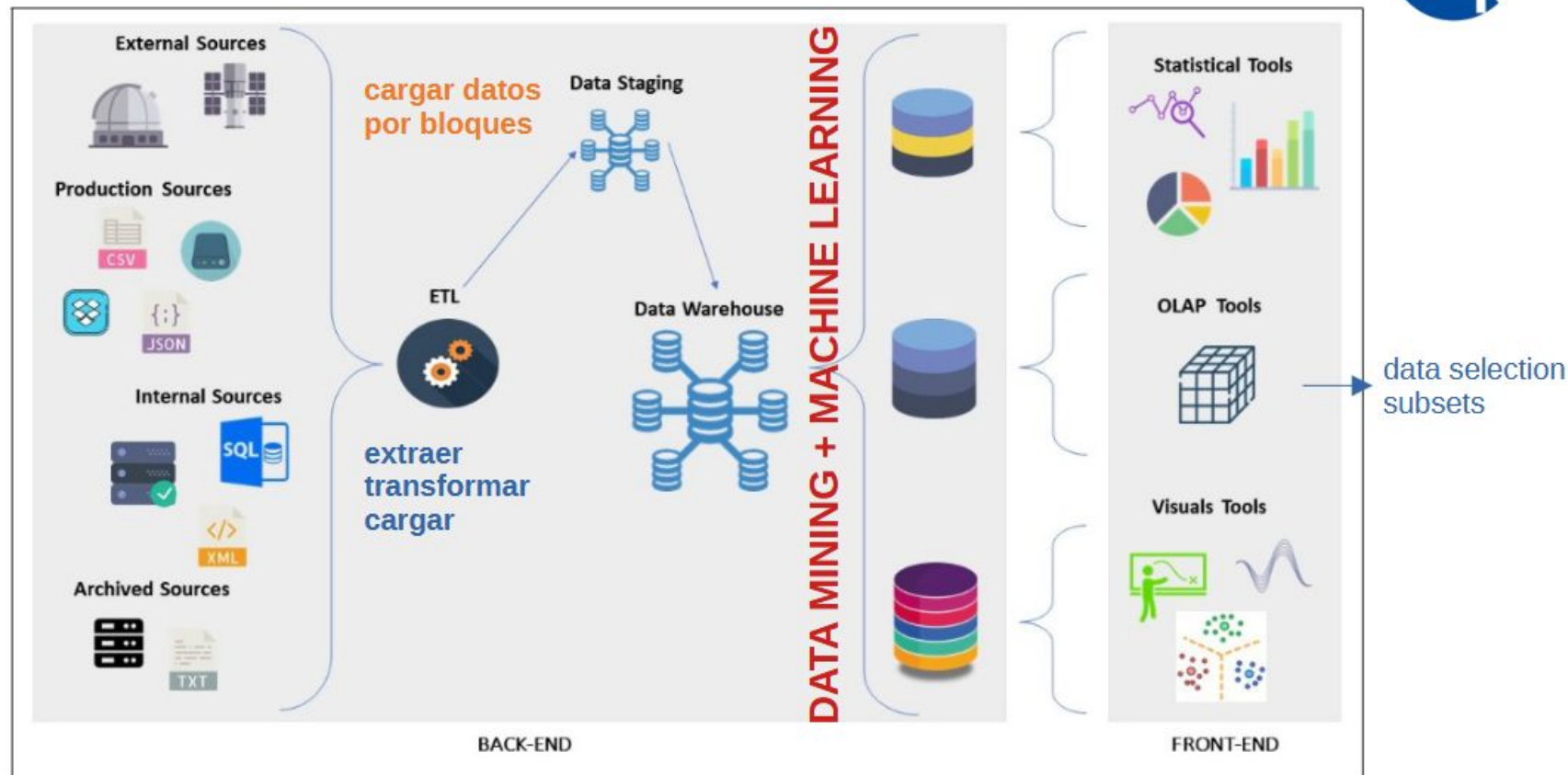
- **Integración:** Datos 'dispersos' en distintas plataformas.
- **Actualización:** muchas de las actuales iniciativas no se encuentran actualizadas.
- **Plataformas 'poco amigables':** interfaces poco intuitivas.
- **Falta de herramientas de visualización**
- **Formato de datos restringido:** requieren de habilidades de programación y algoritmos complejos para resolver algunos problemas específicos.
- **Librerías de terceras fuentes:** requeridas en algunos casos (estas pueden estar descontinuadas o no tienen soporte continuo).

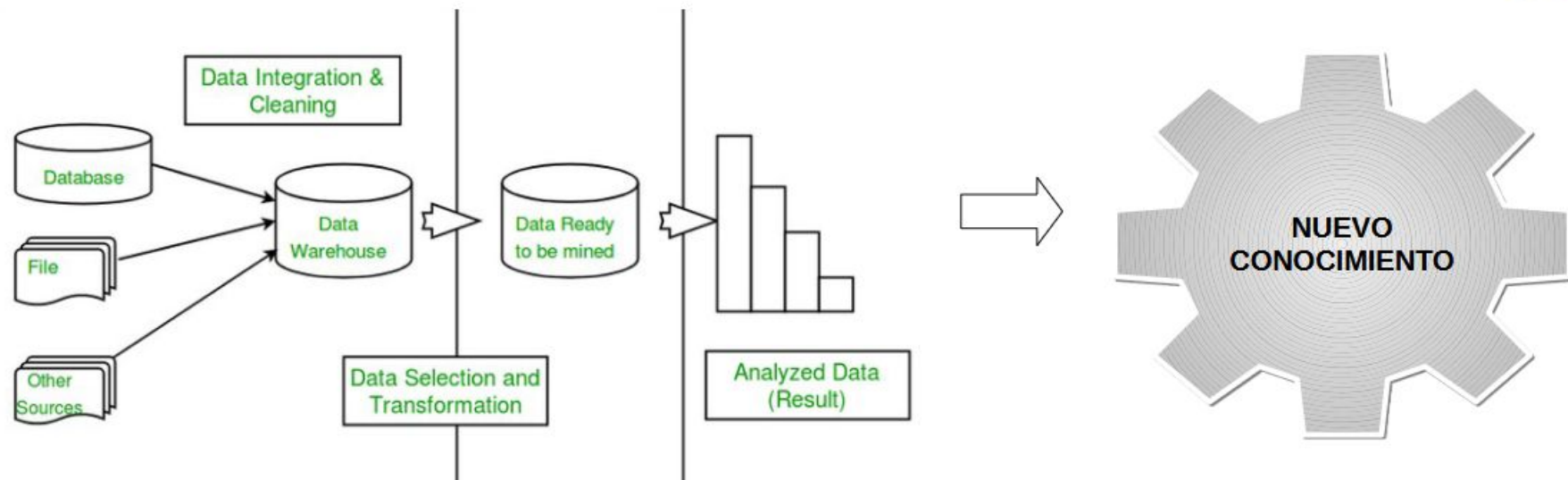
	Aladin	Simbad	Vizer	TopCat	Astropy	VOStat	SkyView	AstroML
Query Search	✓	✓	✓	✓	✓	×	✓	✓
Visualization	✓	×	×	✓	✓	✓	✓	✓
Data Integration	✓	✓	✓	✓	×	×	×	×
Machine Learning	×	×	×	×	✓	✓	×	✓
Data Sets Editions	✓	✓	✓	✓	✓	×	×	×
Features Extraction	×	×	×	×	×	×	×	×
Multi Catalog Feature Search	×	×	×	×	×	×	×	×

Principales diferencias entre algunos software de astronomía (Machin, 2017).



Data Warehouse: ejemplo de arquitectura





Data Mining:

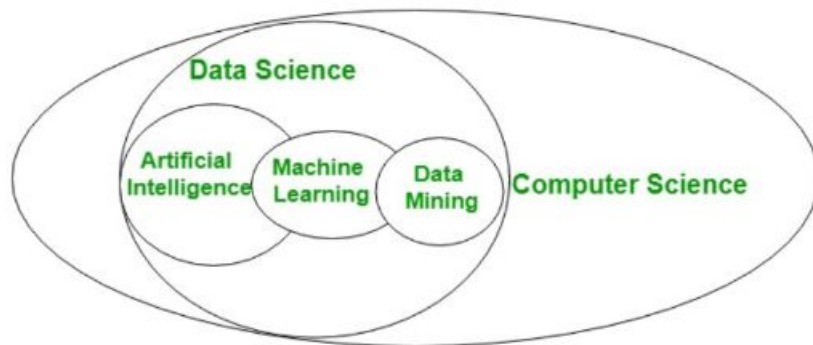
Extraer información útil desde grandes bases de datos y descubrir tendencias y patrones usando técnicas de clustering, clasificación y algoritmos de regresión.

Machine Learning:

Librerías de algoritmos que permiten al computador ir 'aprendiendo' patrones ó tendencias en al análisis de grandes volúmenes de datos.

En el Warehouse trabaja en conjunto con el data mining.

Propósito: Hacer predicciones respecto a nuevos conjuntos de datos.





Estamos
comprometidos con la
Región de Coquimbo



4 AÑOS ACREDITADA
GESTIÓN INSTITUCIONAL | DESDE DICIEMBRE 2017
DOCENCIA DE PREGRADO | HASTA DICIEMBRE 2021
VINCULACIÓN CON EL MEDIO

www.ucentral.cl