Advanced European Network of E-infrastructures for Astronomy with the SKA

Visualization Tools and Scientific Gateway

Alessandro Costa,

Ug Becciani, Eva Sciacca, Fabio Vitello,

INAF - Catania Astrophysical Observatory

Outline

- Current Interaction Models and Tools
- Data Visualization
- Scientific Gateway

H2020 AENEAS Outputs

Work Package 5 focus on the production of **recommendations** for the design of **user interfaces** for **data processing, reprocessing, analysis and visualization** for the ESRC.

- Evaluation of tools currently in use in Radio-astronomy.
  - Packages such as CASA or AIPS or software used in LOFAR, ASKAP, ecc.
  - Tools for data processing, data imaging, visualization
  - Tools for post-processing (e.g. source finding)
- Evaluation of the user survey results (D5.1) and gap analysis (D5.2)
  - “Local” computing ⇒ “Distributed” computing
- Production of **recommendations** for the design of **user interfaces** that should facilitate the distributed processing foreseen for ESDC ⇒ *Science Gateway*

Current User Interaction Models and Tools

Radio astronomy packages

- CASA
- Miriad
- AIPS
- LOFAR software stack
- MeerKAT pipelines
- ASKAPsoft
- Python tools
- …

Data Post-Processing

- Source extraction tools (point-like sources and extended ones)
- Cross matching tools

Data Processing (SDP?)

1. Flag the bandpass and flux calibrator;
2. Find the solution for the bandpass and flux calibration;
3. Apply the solution on the secondary calibrator (the gain calibrator);
4. Flag the gain calibrator;
5. Find the gain solution;
6. Apply the calibration on the data;
7. Flag the data.
8. (Optional) Direction-dependent Calibration

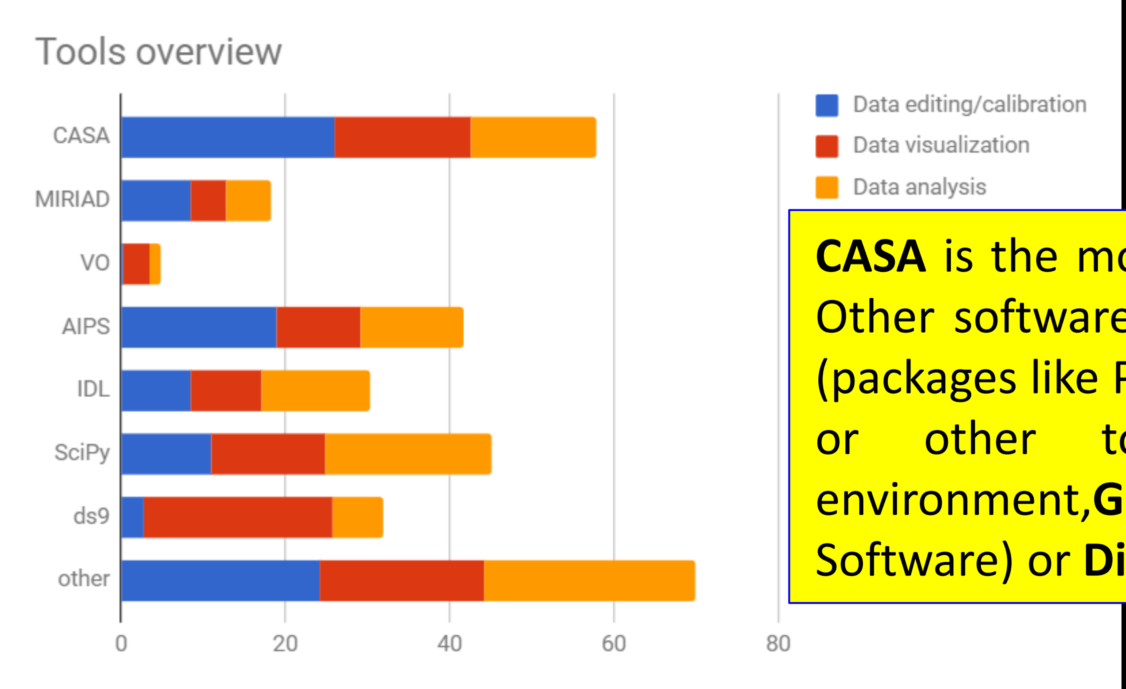
Data Imaging

- *tclean* task
- WSClean
- mfclean
- …

Data Visualization

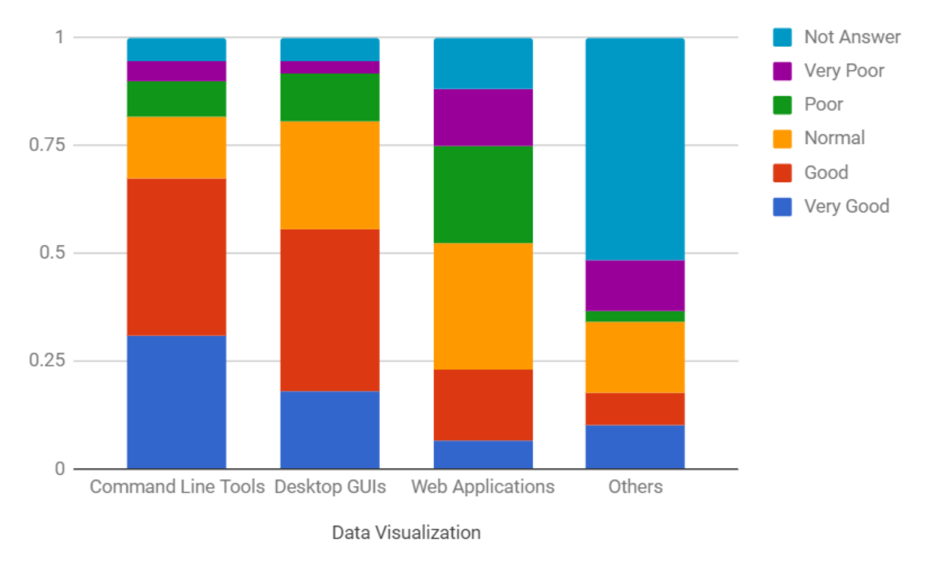
- SAOImage DS9
- Aladin, TOPCAT, VisIVO
- Karma
- Casaviewer
- Python tools

Actual Interaction Models

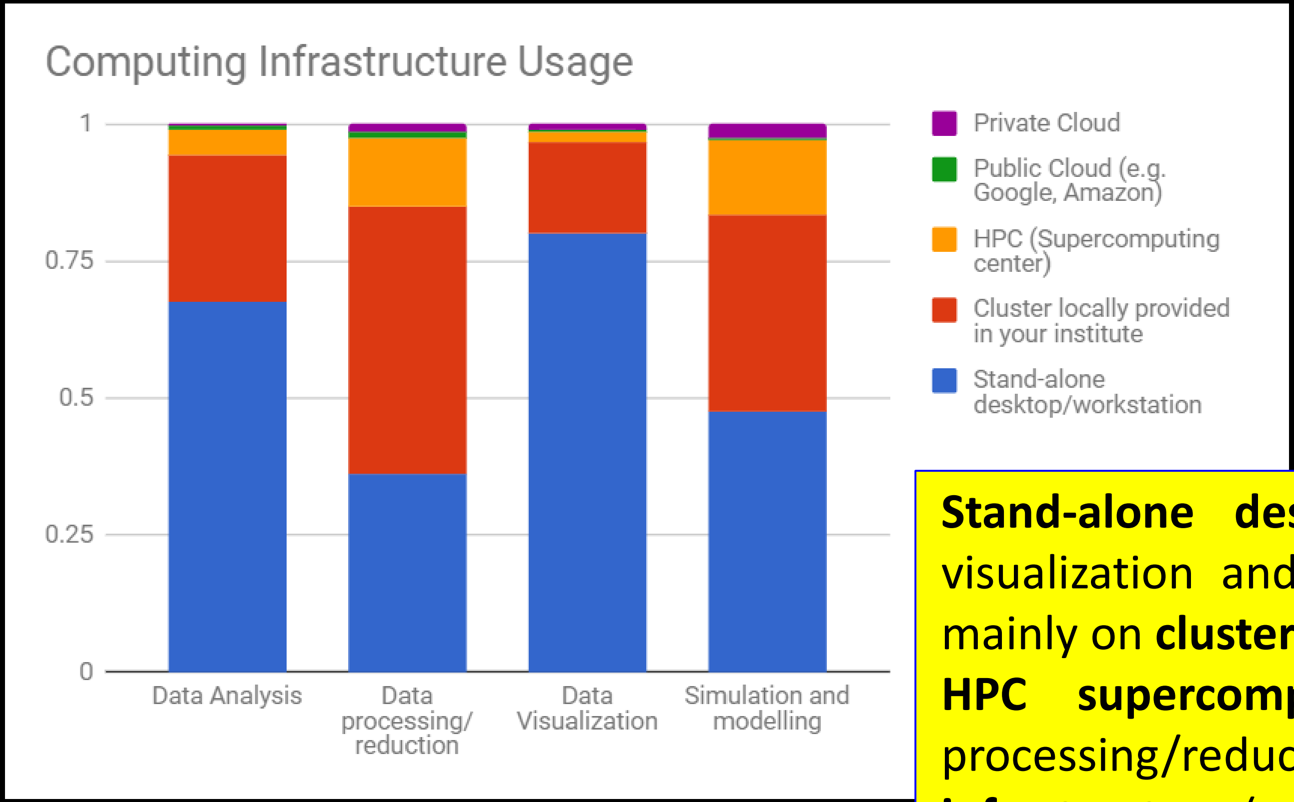


CASA is the most commonly used tool followed by **SciPy** and **AIPS**. Other software may include **user codes** or **libraries** e.g. in **Python** (packages like ParselTongue and libraries such as AstroPy) or Matlab, or other tools such as the **LOFAR data reduction** environment,**GILDAS** (Grenoble Image and Line Data Analysis Software) or **Difmap**.

Command line tools are the most commonly used mainly for all the data processing tasks. **Desktop GUIs** are mainly preferred for data visualization and analysis and. Other interaction models may include **scripting** and **batch interfaces** or **text-based user interfaces**.



Actual Interaction Models



Stand-alone desktop/workstation are mainly preferred for data visualization and analysis while data reduction tasks are processed mainly on **clusters locally provided** by the hosting research institutes.
HPC supercomputing centers are mainly exploited for data processing/reduction and simulation and modeling while **Cloud infrastructure** (public or private) have currently very limited usage.

"The purpose of (scientific) computing is insight, not numbers."

Richard Hamming

Why In-*sight*?

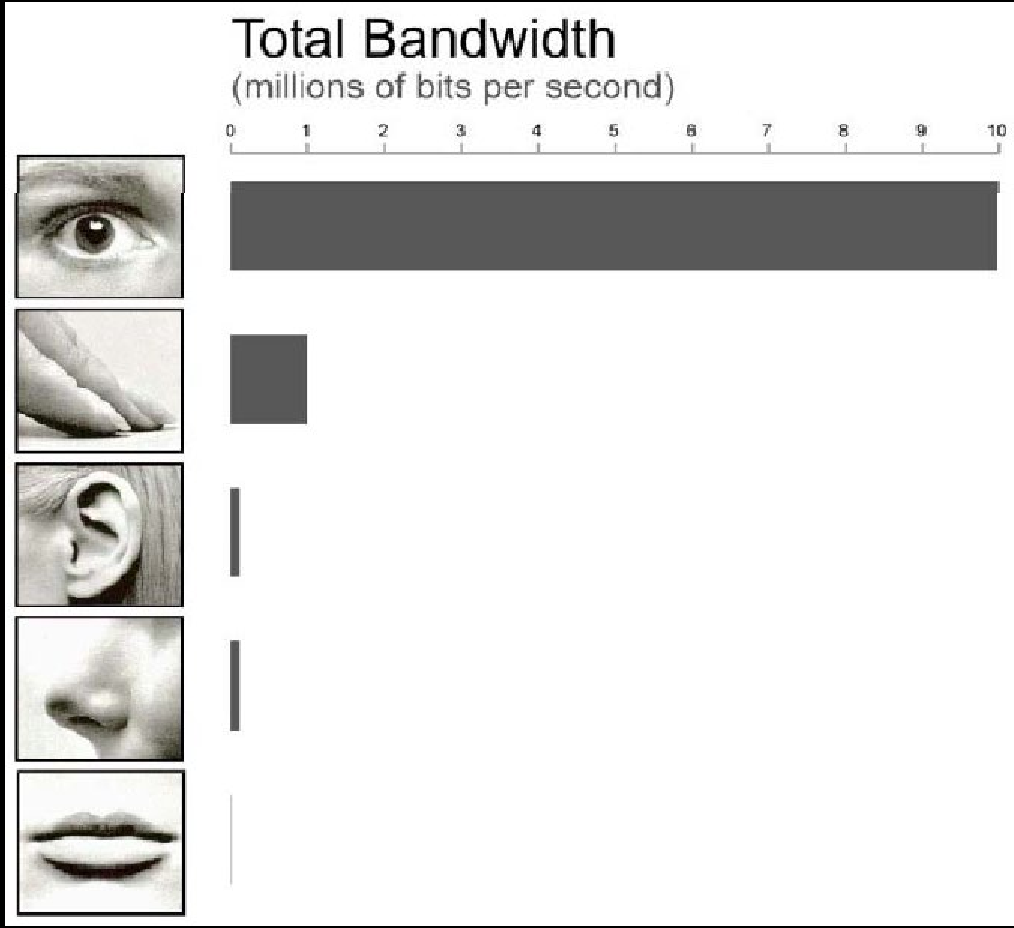


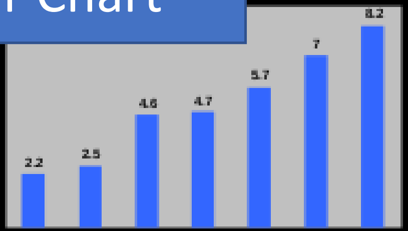
image source: [Few 2006]

- Figures are **richer**; provide more information with less clutter and in less space.
- Figures provide the 'gestalt' effect: they give an overview; **make structure more visible**.
- Figures are **more accessible**, easier to understand, **faster to grasp**, more comprehensible, **more memorable**, more fun, and less formal.

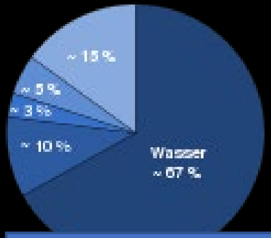
list adapted from: [Stasko et al. 1998]

Why not use a Diagram?

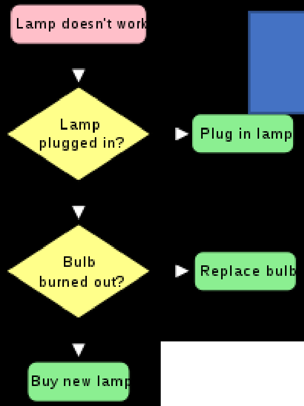
Diagram: a <u>symbolic</u> <u>geometric</u> <u>representation</u> of information



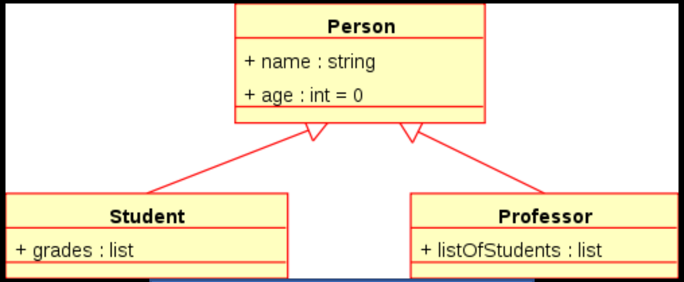
Bar Chart



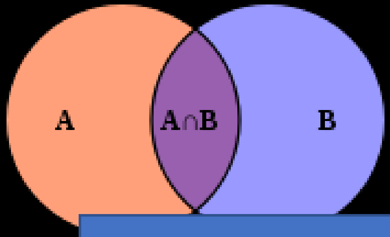
Pie Chart



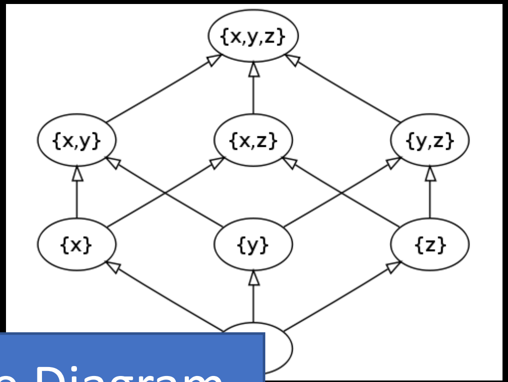
Flow Chart



Class Diagram



Venn Diagram

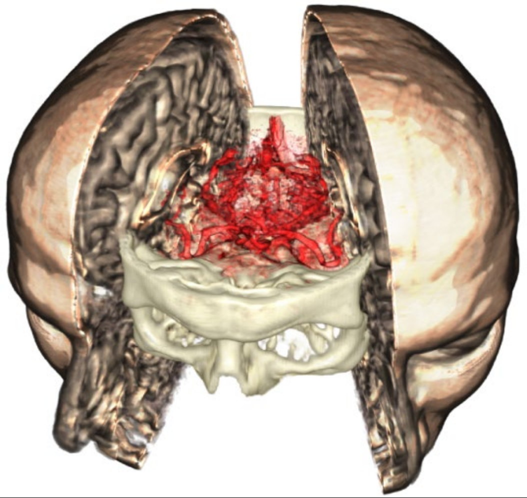


Hasse Diagram

Why not use a Diagram?

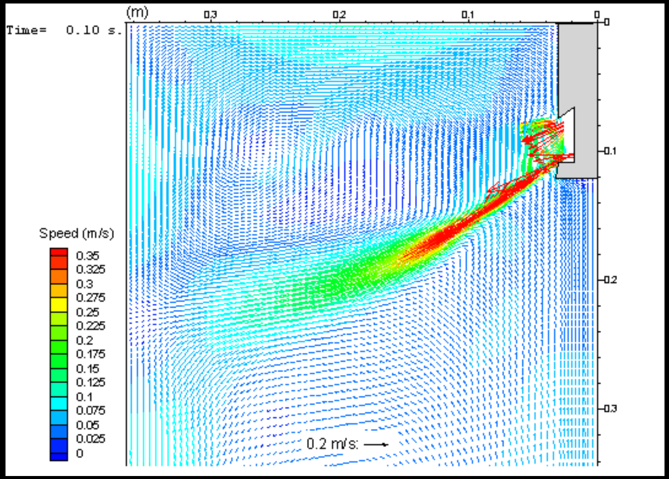
Visualization is a superset of Diagrams: While each diagram is a visualization, not all visualizations are diagrams.

It may be realistic and not symbolic:

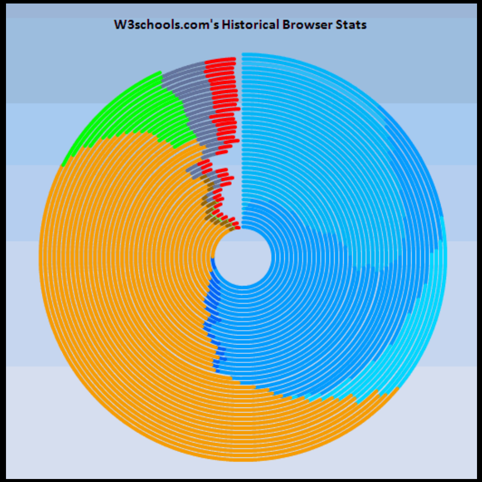


[Rößler et al. 2008]

It may use an animation and not be purely geometric:



It may be for interactive exploration and not only for representation:

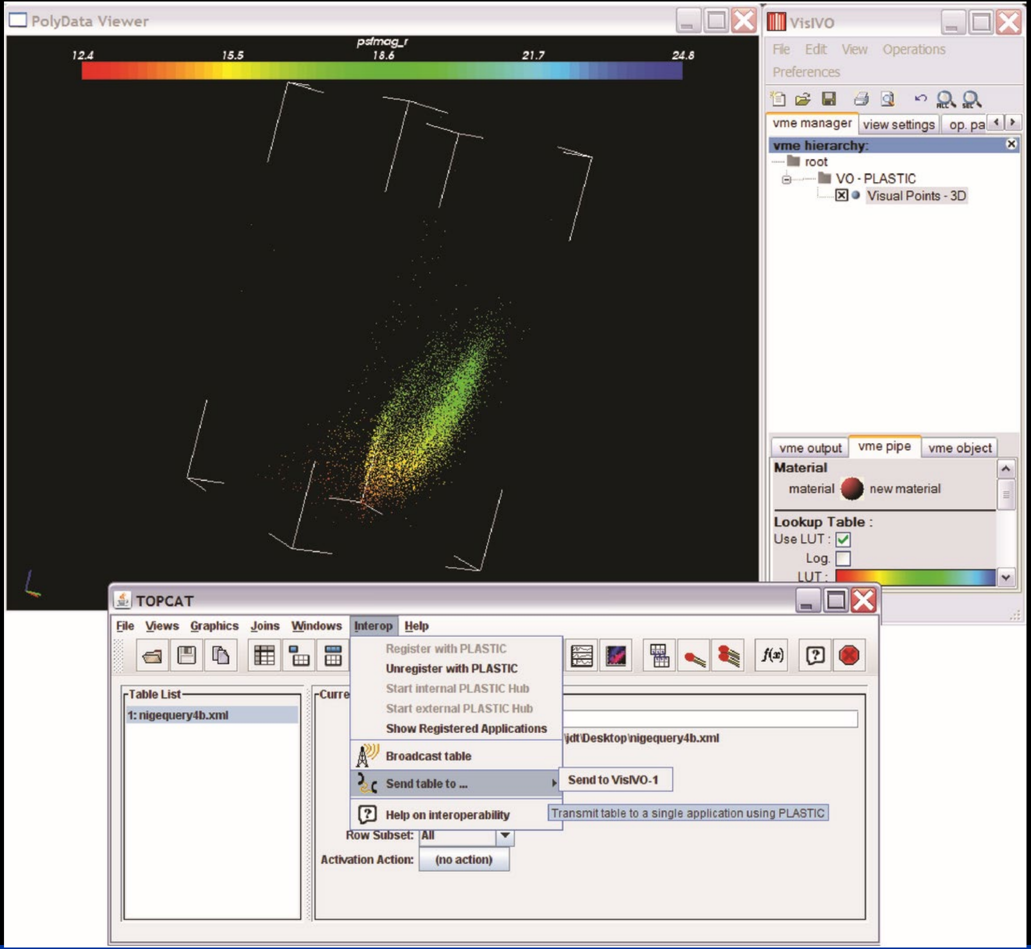


VisIVO Desktop

VisIVO was **first experience** of an immersive Visualisation and Data Analysis Tool in astrophysics developed as a collaboration between the Italian National Institute for Astrophysics Astrophysical Observatory of Catania and CINECA.

Supports **different kinds of file formats**: VOTables, FITS, HDF5, ASCII, raw binaries, GADGET, etc..

The capabilities of VisIVO are extendable through an application **interoperability** protocol called PLASTIC (Platform for Astronomy Tool Interconnection) to leverage the abilities of different desktop applications in a seamless way.



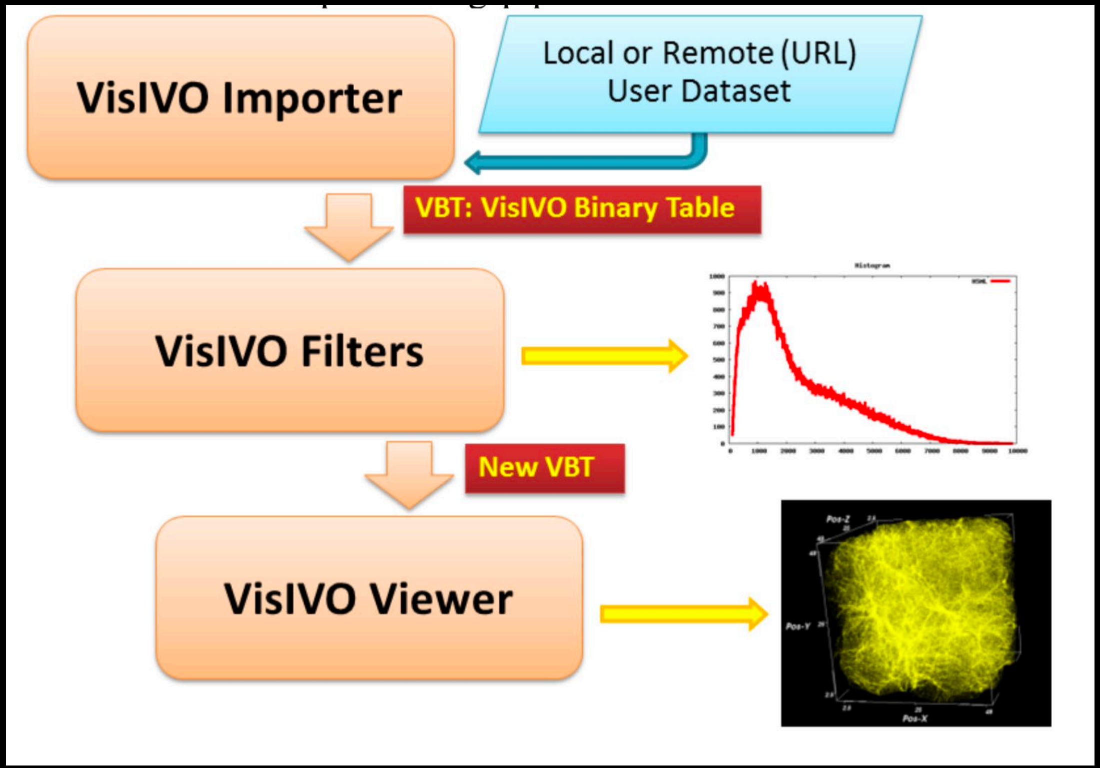
2005	2009	2010	2011	2013	2013	2014	2015	2017

VisIVO Server

VisIVO Server is an **open source** collection of visualization modules for **fast rendering** of 3D views of astrophysical datasets.

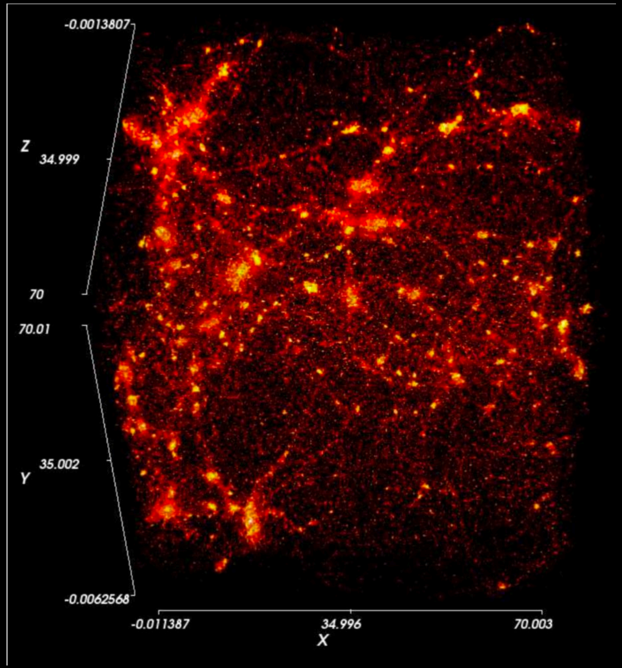
VisIVO Server is built upon the VisIVO Desktop functionality and supports Unix platforms.

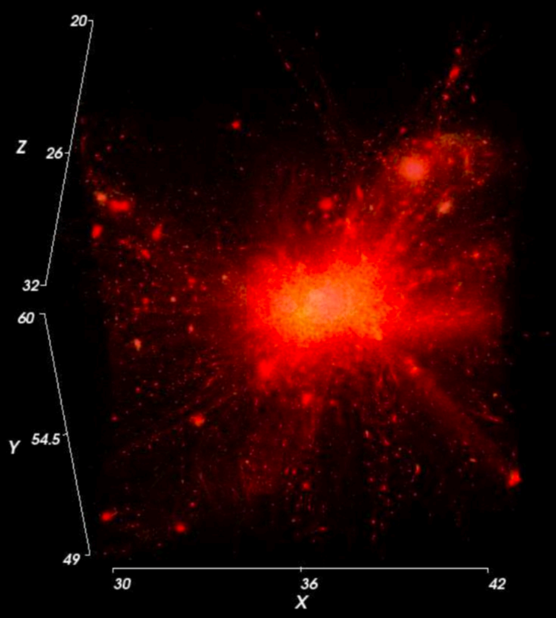
The defining characteristic of VisIVO Server compared to VisIVO Desktop is support for very **large-scale datasets**; no fixed limits (in principle) are imposed for visualization.

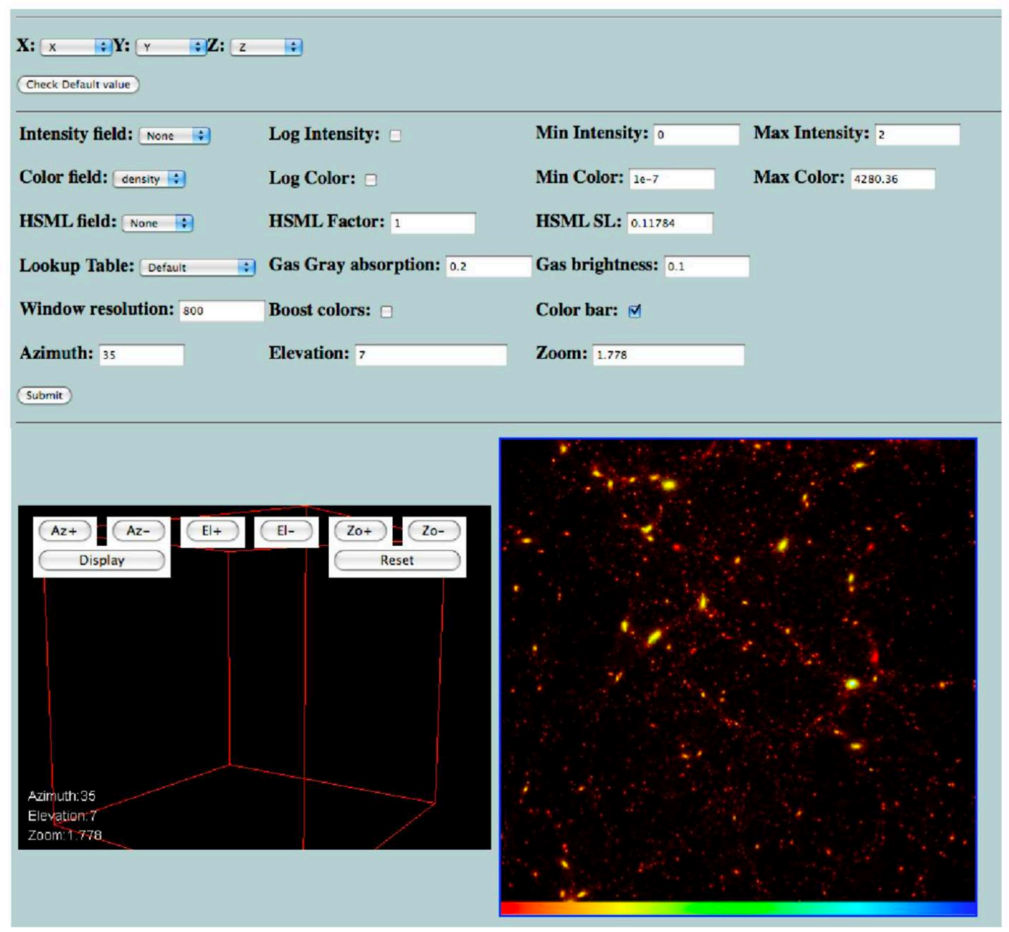


VisIVO Web

The main goal of VisIVO Web is to provide an **easy to use** service for accessing the full functionality of VisIVO Server through a www portal.







2005 | 2009 | 2010 | 2011 | 2013 | 2013 | 2014 | 2015 | 2017

VisIVO Library

The VisIVO Library was developed to port VisIVO on the **gLite GRID**.

It is written in **C++** language and allows a job running on a grid node (or HPC system) to produce a set of **images or movies** directly using VisIVO with its internal data arrays without the need to produce intermediate files.

This is particularly important when running on the grid and running large simulations, where the user wants to have a **quick look to the results** during the data production phase.

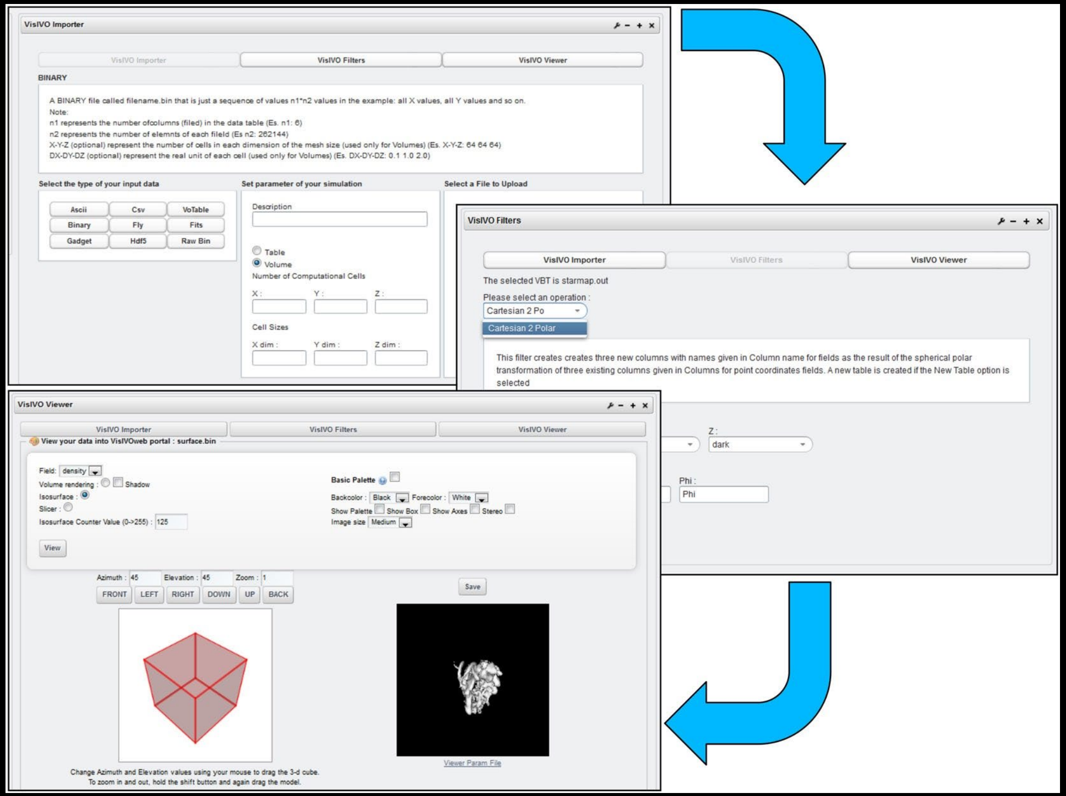
```
#include "visivo.h"
#include <string.h>
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>


#define NB 16777
#define NVOL 262144

int main(int argc, char*argv[])
{
int errorCode;

char filename[256];

//*********************************
//*********************************
//********************************* VisIVOImporter
VisIVOImporter envVI1;

errorCode=VI_Init(&envVI1);
errorCode=VI_SetAtt(&envVI1,VI_SET_FFORMAT,"ascii");
errorCode=VI_SetAtt(&envVI1,VI_SET_FILEPATH,"mrvbt16.ascii");
errorCode=VI_SetAtt(&envVI1,VI_SET_OUTFILEVBT,"mrvbt16.bin");

VI_Import(&envVI1);
//********************
//********************
//******************** VisIVOFilter
VisIVOFilter envVF1;

char operation[256];
strcpy(operation,"pointproperty");

errorCode=VF_Init(&envVF1);
errorCode=VF_SetAtt(&envVF1,VF_SET_OPERATION,operation);
errorCode=VF_SetAtt(&envVF1,VF_SET_FILEVBT,"mrvbt16.bin");
errorCode=VF_SetAtt(&envVF1,VF_SET_RESOLUTION,"32 32 32");
errorCode=VF_SetAtt(&envVF1,VF_SET_POINTCOLUMNS,"X Y Z");
errorCode=VF_SetAtt(&envVF1,VF_SET_APPEND,"");
errorCode=VF_SetAtt(&envVF1, VF_SET_OUTCOL,"density");
```

2005 | 2009 | 2010 | 2011 | 2013 | 2013 | 2014 | 2015 | 2017

VisIVO Science Gateway

The VisIVO Science Gateway is designed as a **workflow enabled portal** that is wrapped around WS-PGRADE providing **visualization and data management** services to the scientific community by means of an **easy-to-use** graphical environment for accessing the full functionality of VisIVO Server.

Complex workflows can be created and executed on a **variety of infrastructures** (e.g. clouds, desktop and service grids or supercomputers) to obtain comprehensive exploration and analysis of large-scale astrophysical datasets. The gateway offers role-based authorization modules and supports secure login.



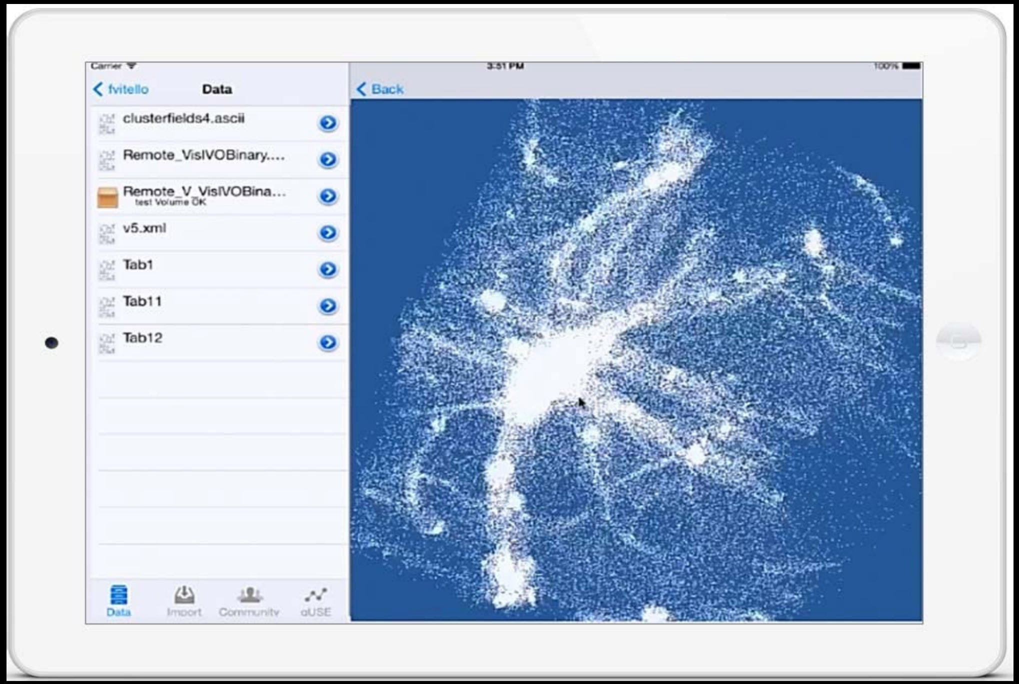
2005	2009	2010	2011	2013	2013	2014	2015	2017

VisIVO Mobile

The VisIVO Mobile application allows **iPad devices** to exploit VisIVO Science Gateway functionalities to **access large-scale astrophysical datasets** residing on a server repository for analysis and visual discovery.

The application **produces customized visualizations (images or movies) on DCIs** through interactive widgets submitting VisIVO-based workflows.

The VisIVO Mobile application **extends native WS-PGRADE/gUSE** utilities to create, configure and submit a workflow from scratch directly from the application.

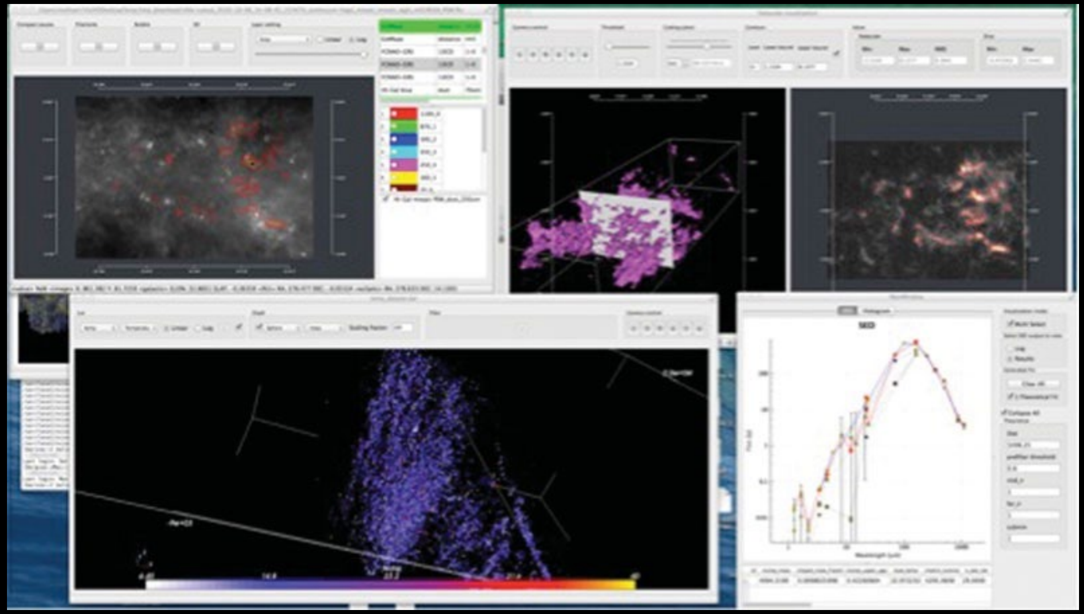


VIALACTEA Visual Analytic Client

The tool has been developed within the VIALACTEA project, founded by the 7th Framework Programme of the European Union.

Visual analytics environment allows to easily conduct research activities for multidimensional data and information visualization. It provides real-time data interaction to carry out complex tasks for multi-criteria data/metadata queries on subsamples selection and further analysis.

Visual analytics combine **data mining** algorithms and **advanced analysis techniques** with highly **interactive visual interfaces** offering scientists the opportunity for in-depth understanding of **massive**, **noisy**, **and high-dimensional data**.



Ref:
Vitello et al (2018)
[10.1088/1538-3873/aac5d2](10.1088/1538-3873/aac5d2)

NEANIAS Project

Call: H2020-INFRAEOSC-2018-2020 (Implementing the European Open Science Cloud)

Topic: INFRAEOSC-02-2019 (Prototyping new innovative services)

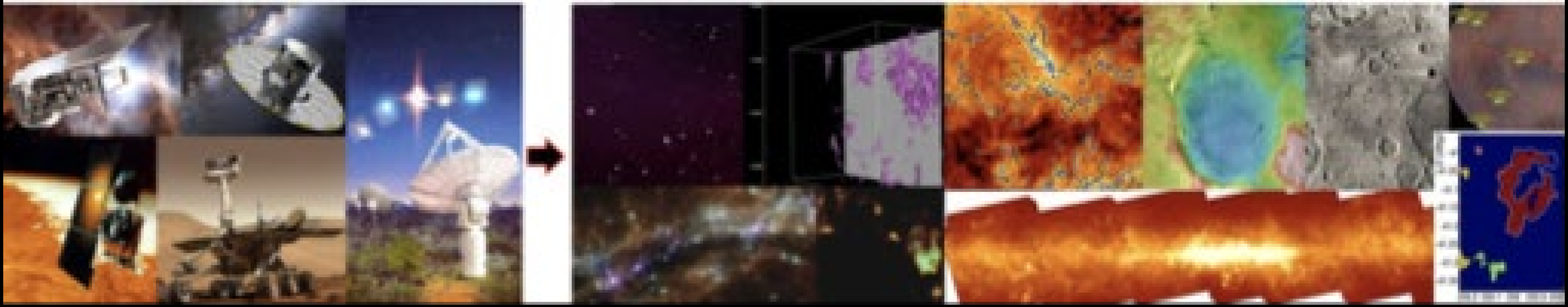
Project Duration: 36 months (starting on November 2019)

Partners: 20 **Overall budget:** € 5.597.025,00

Coordinator: National and Kapodistrian University of Athens - NKUA

PI: Mema Dimitra-Isidora Roussopoulou, Associate Professor of Computer Science and head of the Distributed Systems Research Group.

NEANIAS Space thematic Services



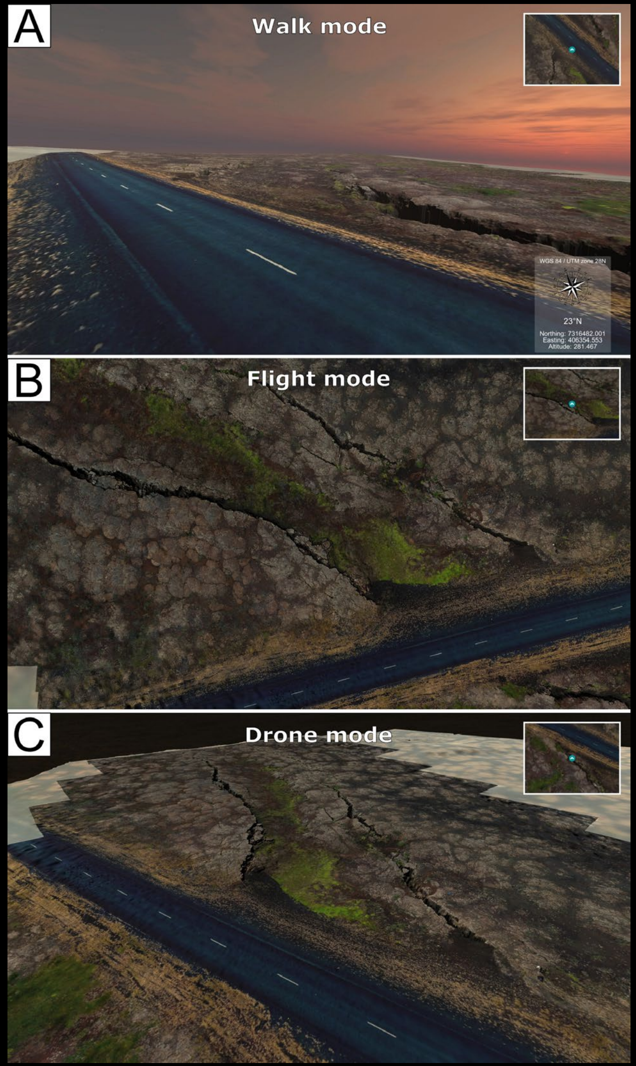
From Space data to Space products developing the following services:

- **S1 FAIR Data Management and Visualization service**.
- **S2 Map Making and Mosaicing of Multidimensional Space Images service**.
- **S3 Structure Detection on Large Scale Maps with Machine Learning service (MoU with University of Malta)**.

deployed on EOSC Hub.

VR Tools

Recently the VisIVO development has been exploited for the experimentation of cutting-edge interactive visualization technologies for the improvement of teaching and scientific dissemination.
This work is being carried out as a knowledge transfer from astrophysical sciences to geological sciences in the context of an international collaboration to innovate teaching, learning and dissemination of earth sciences, using virtual reality.



Different Classes of Users

The ESRC is expected to offer its services to 3 main classes of potential users: (D5.8)

- “Principal Investigators”
  - ..and their co-I, research group ….
  - During telescope commissioning phase and “early science”
    - expert team with solid tech. know-how
  - in the latter stages
    - Member of the KSP collaborations
- “Archive Users”
- “Software Developers”

User Interaction Model from Local to Distributed Analysis

Survey results show that users run their tools on **local resources** that are under their control.

Remote and distributed resources requires web interfaces that should be designed to facilitate the access to the computing infrastructures and give the flexibility to include command-line tools and smooth integration with local desktop environments.

The SKA REGIONAL CENTRE REQUIREMENTS document (SKA-TEL-SKO-0000735) defines a distributed framework, provided by the SKA Regional Center Alliance, for accessing the science data products. In particular, it states:

"Access to SKA science data products, as well as the tools and processing power necessary to fully exploit the science potential of those products, is provided via a Science Gateway. Access to science data products is irrespective of a SKA user's geographical location, or whether their local region or country hosts an SRC".

User Interface Recommendation

Main recommendation: expose the ESRC processing system using a Science Gateway.

The term "Science Gateway (SG)" describes a whole class of **interfaces** to scientific processing and underlying computing infrastructures.

A Science Gateway is:

- Unique and community-specific set of tools, applications and data collections.
- Integrated via a web portal

A Science Gateway provides:

- Resources and services
- of distributed computing infrastructures

User Interface Recommendation

A Science Gateway for an SRC shall provide services

Distributed on 3 “Activity Departments”

- Information and Science Exploitation
- Computation
- Archive

“Activity Departments”

A Science Gateway for an SRC shall provide services

Distributed on 3 “Activity Departments”

- Information and Science Exploitation
  - Knowledge distribution
  - Communication to and from the telescope
  - Supporting the users during the stage of their projects
- Computation
- Archive

“Activity Departments”

A Science Gateway for an SRC shall provide services

Distributed on 3 “Activity Departments”

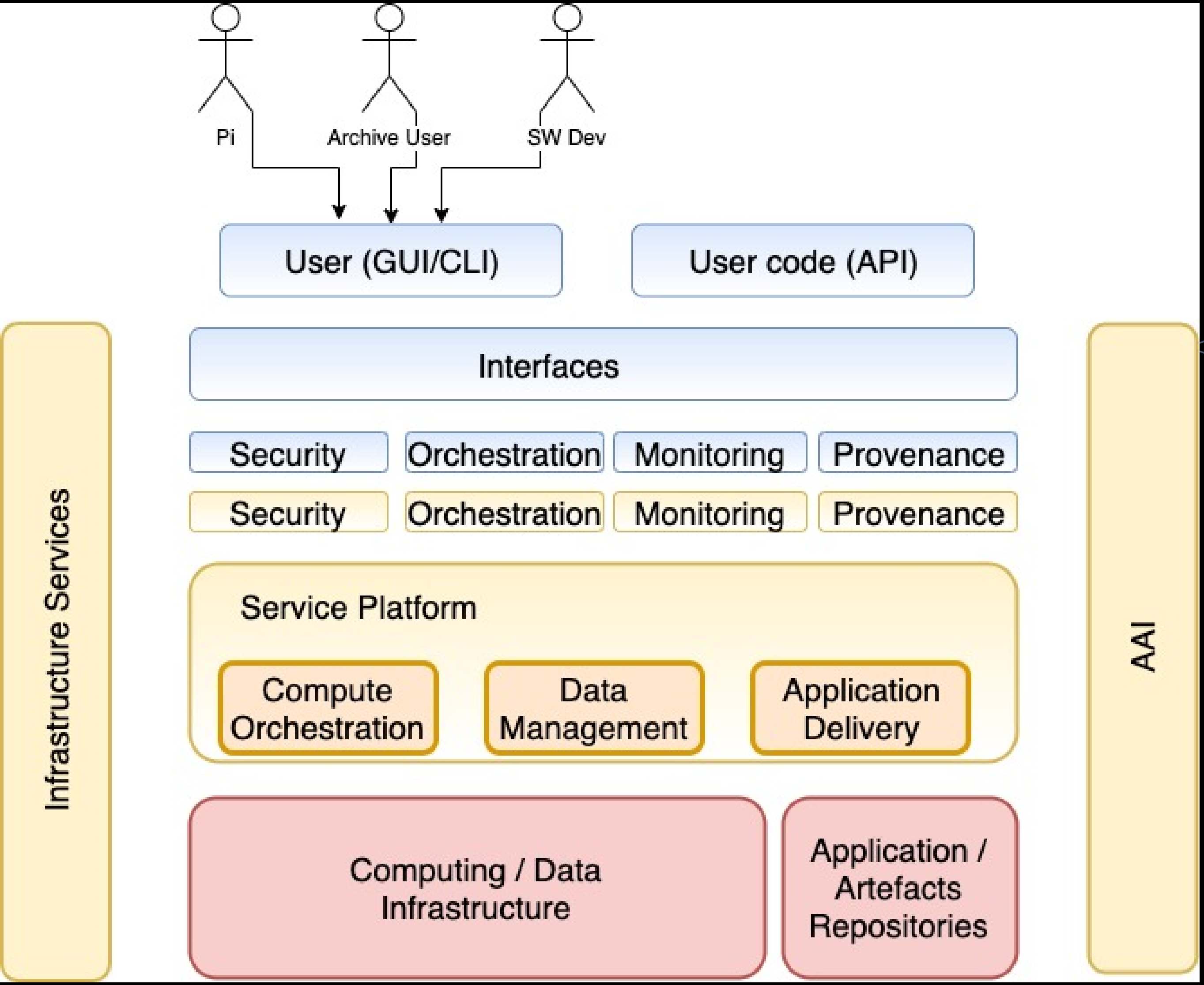
- Information and Science Exploitation
- Computation
  - **Computational Resources Seamless Access**
  - Development and maintenance of analysis tools
  - Code repository
- Archive

"Activity Departments"

A Science Gateway for an SRC shall provide services
Distributed on 3 "Activity Departments"

- Information and Science Exploitation
- Computation
- Archive
  - Data Storage resources and archive content
  - Access to metadata and the data
  - Query interfaces
  - Archive interoperability and compliance with common (i.e VO) standards

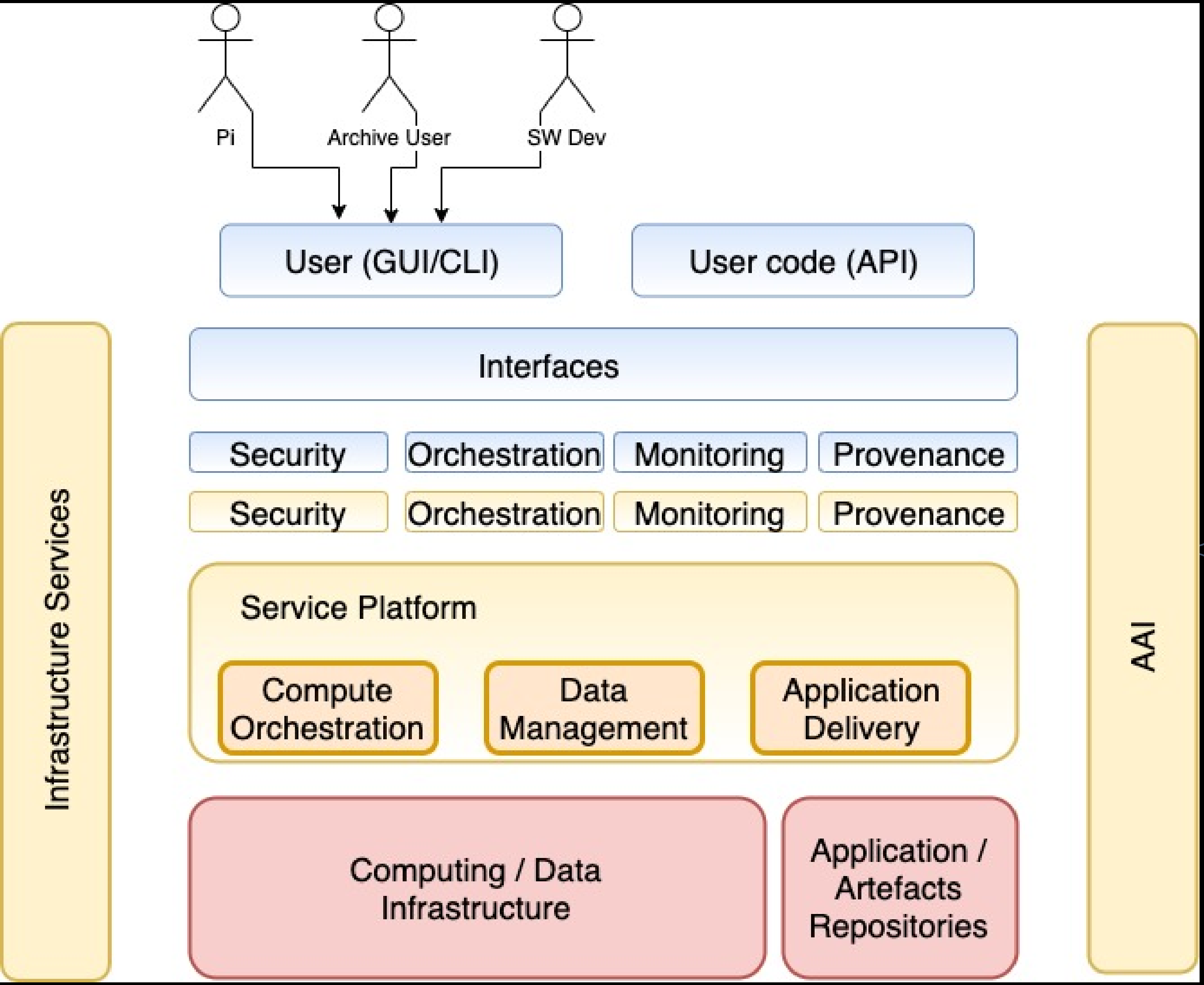
Science Gateway Features



Users typically interact through ***Web-based Graphical User Interfaces*** **(GUIs)**. Others may prefer ***Command-Line Interfaces*** **(CLIs)** that offer more control and from which it is easier to automate repetitive tasks.

External applications can utilize **Application Programming Interfaces (APIs)**. Additionally, CLIs can also be used in scripts and programs to automate tasks or integrate with other systems.

Science Gateway Features



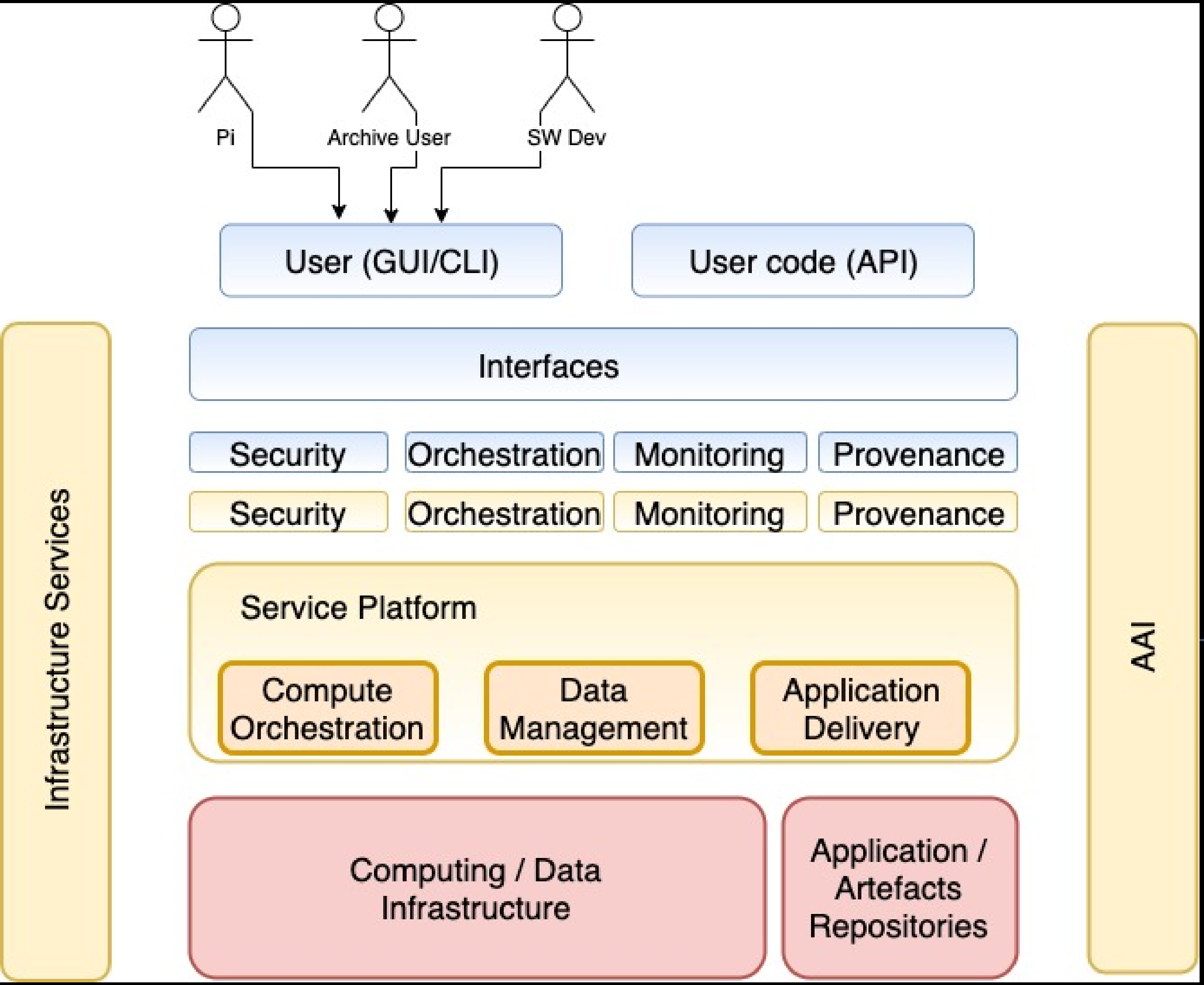
Security to enforce authenticated and authorized access to data, applications, computing resources, communities, information, and SG functions.

Orchestration to integrate and coordinate data, computing, and community management to instante and set up applications. Such functions are typically provided by a Workflow Management System (WMS).

Monitoring to collect and store the status of resources and different parts of the system, keeping track of the events and actions performed via the SG by users and automated processes.

Provenance to collect and provide lineage information about the actions performed by the SG and its users.

Science Gateway Features



Compute & Orchestration provides the scheduling, executing, and managing of various pipelines applied to a pool of resources to perform data analysis, processing or visualization offering transparent, scalable and interoperable computing management.

Data Management to support interfaces with archive related tasks such as acquisition, storage, retrieval, transport, organization, replication, curation, integration, and aggregation of data and metadata.

Application Delivery to support interfaces with application and/or artefact repositories including repositories of software codes (e.g. GitHub), containers that are able to package the software and all its dependencies (e.g. Docker Hub), pipelines and scientific workflows.

Evaluation of SG technologies

	gUSE/ WS-PGRADE	Galaxy	HubZero	Catania SG Framework	Apache Airavata	Agave Platform
GUI/CLI	Yes	Yes	Yes	Yes	No	No
API	Partial	Yes	Yes	Partial	Partial	Yes
Coordination	Yes	Yes	Yes	Partial	Yes	Yes
Security	Partial	Partial	Partial	Yes	Partial	Yes
Monitoring	Partial	Yes	Partial	Yes	Partial	Yes
Provenance	No	Yes	Yes	No	Partial	Partial
Data Management	Yes	Yes	Yes	Yes	Partial	Yes
Computing Management	Yes	Yes	Yes	Yes	Yes	Yes
Community Management	Yes	Yes	Yes	Yes	Partial	Partial

Web-based science gateway frameworks include: gUSE/WS-PGRADE, Galaxy, HubZero and the Catania Science Gateway Framework.

APIs and libraries based science gateway frameworks include: Apache Airavata and the Agave Platform, which aim at reducing the effort on the developer side while enabling to apply novel user interface technologies and frameworks.