Astronomy: pioneer in Open/FAIR Data

IVOA (a case of study for the EC [1])

• Established in 2002
• Developing standards required to make data FAIR
• Open and Inclusive framework:
  – Anyone can publish data / develop a VO tool

Astronomy: pioneer in Open/FAIR Data

- Data are preserved in archives
- Published after an embargo period
- Culture of re-using data

Enhancing the scientific returns from investments in astronomical infrastructures

Source: M. Romaniello’s talk “The VO-Service at ESO”. ESO Telescope Bibliography

Source: 10.1051/epjconf/201818610003
Understandable Software for supporting Scientific Reproducibility

- **Open Notebooks:** [https://doi.org/10.5281/zenodo.2631868](https://doi.org/10.5281/zenodo.2631868)

- **Scientific workflows:** networks of analytical steps [...] including computationally intensive jobs on HPC ([https://doi.org/10.1002/cpe.994](https://doi.org/10.1002/cpe.994))

**5. Results**
In this section we present the results of our H\(_1\) analysis, first for the group as a whole and then for each galaxy.
Opening all involved elements: Research Object

- input and output examples
- annotations (human/machine readable)
- metadata: data + software versión + config. parameters, execution environment, description of main steps, etc
- interoperability
Big Data science: The Square Kilometre Array case
The Square Kilometre Array Observatory

Open key questions in Astrophysics, Astrobiology and Fundamental Physics

- Formation of the 1st galaxies in a dark Universe dominated by atomic gas
- Evolution of the atomic gas and star formation till the current epoch
- Strong Field Tests of Gravity Using Black Holes
- Active Galactic Nuclei and the Galactic Centre
- Extrasolar planets (proto-planetary disks, biomarkers)

SKA1-Mid
the SKA’s mid-frequency telescope
- Location: South Africa
- Frequency range: 350 MHz to 15.4 GHz with a goal of 24 GHz
- 197 dishes (including 64 MeerKAT dishes)
- Maximum baseline: 150km

SKA1-Low
the SKA’s low-frequency telescope
- Location: Australia
- Frequency range: 50 MHz to 350 MHz
- 131,072 antennas spread between 512 stations
- Maximum baseline: ~65km

SKAO
- 2024: Commissioning
- 2026: Science verification
- Q2 2029: End of construction
The Square Kilometre Array “case”

The SKA Regional Centres, the core of the SKA Science
The Challenge: extraction of Scientific Knowledge

Huge and complex data volumes
Large teams distributed globally

A shared challenge for data-intensive research

Computing / storage / network / human resources will be needed:

• Efficient exploitation of Distributed Computing Infrastructures
• Large international alliances of scientists
  • Tools to enhance scientific collaboration
  • Platforms to share data, methods and knowledge

Open Science is the Aim and also the Mean
The SKA and Open Science

Adoption of Open Science values

“Open Science, based on the precept of making scientific research collaborative, transparent and accessible to all, is rooted in SKA’s foundational principles. So is the related concept of scientific reproducibility, a fundamental aspect of the modern Scientific Method since the 17th century allowing independent teams to have access to methodology and tools to be able to confirm experiments and validate results.”

Reproducibility as a metric of success

“Reproducibility of SKA science data products. This metric will measure how complete the workflow description is that is linked to each SKA data product. [...] must reflect completeness of the provenance information for each data product and accessibility of the software used. This is related to how well SKA science data products adhere to the FAIR principles.”
Sustainable development goals

The SKAO: A global Research Infrastructure for the 21st Century and beyond

Open Science for sustainability and inclusiveness: the SKA role model

Lourdes Verdes-Montenegro, Susana Sánchez
IAA Severo Ochoa Centre of Excellence (CSIC)

Tuesday 29th September 2020

Credits: UNESCO
The SKA, Open Science & SDGs

### Science hidden behind paywall barriers = limitation to science progress
- Free access to research sources to the whole community, avoiding reinvention
- Data and results more accessible and reliable
- Promotion of **scholarly exchange** of ideas

### Promote equity, diversity and inclusion: All previous items +
- A tool enabling an **objective evaluation** of work
- Barriers are even more emphasized to scientist women in places where their **contribution tend to be ignored or anonymized**

- Speed up building of **skills**
- **Teaching**, e.g. access to public archives, **fostering collaborative** practices
- **Citizen** science

**Accelerated knowledge transfer to Society, pandemics, sanitary crisis**
ESCAPE

Consortium:

ASTRON, CWI, FAU, SKAO, FAIR, ILARION, CNRS, LAPP, CSIC, GSI

Budget: 15.98 M€
From Feb. 2019 until Jan. 2023 (extended)
Coordinator: CNRS-LAPP

Source: https://projectescape.eu/sites/default/files/2022-04-12%20%E2%80%94%20SCiMMA%20Webinar.pdf
Toward a Spanish SKA Regional Centre fully engaged with Open Science

http://dx.doi.org/10.1117/1.JATIS.8.1.011004
espSRC: Supporting the Spanish Community

- **>20 data analysis projects:**
  - SKA related & non-SKA

- **SKA precursors proposals**
  - E.g.: 2 PI MeerKAT proposals

- **Open Science Training**
  - 1st SKAO – OS School
  - Droplets

- **SKA Data challenges:**
  - SDC2:
    - Spanish team 5\textsuperscript{th}/40 +
    - Gold Medal on Reproducibility
  - SDC3:
    - Spanish team led by IFCA
    - Collaboration with CESGA

esSRC: Supporting the SRC network development

**SRC Steering Committee Working Groups**
- Design phase (IAA-CSIC, BSC, OAN, IFCA, ICE/IECC-CSIC, Univ. Valencia, RedIRIS)

**Builder of testbeds for technologies**
- Mini-SRCNet demonstrator
- Data Management system (SKAO Data Lake)

**CORAL TEAM**
Lead by IAA-CSIC = Product Owner + Scrum Master
Conclusions

- Astronomy is **Pioneer** in Open Science: Spanish VO
- Involvement in EOSC through e.g. ESCAPE H2020 Project:
  - IFAE, UCM, INTA-Spanish VO, IAA-CSIC
- Astronomy is facing an extreme Big Data challenge: the SKA Observatory
  Principles aligned with those of EOSC, as an ESFRI is involved in
  EOSC projects and its community engaged with OS

**Scientific infrastructures are key for implementing OS**

With financial support from