Science Digital @ UNGA 75

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

Time: 12:00 - 16:00 (CET)

# Open Science: a **new** concept?

- Too many **adjectives** for science:

  excellent, high quality, trustable, ... Open

Open Science: a **new** concept?

- Too many **adjectives** for science:

  excellent, high quality, trustable, ... Open

  - **Let's go back 383 years in time...**

DISCOURS DE LA METHODE

RENÉ DESCARTES

Scientific Reproducibility is a fundamental principle of the Scientific Method, a process established in the 17th century that marked the beginning of modern science and laid the foundations for the Philosophy of Science
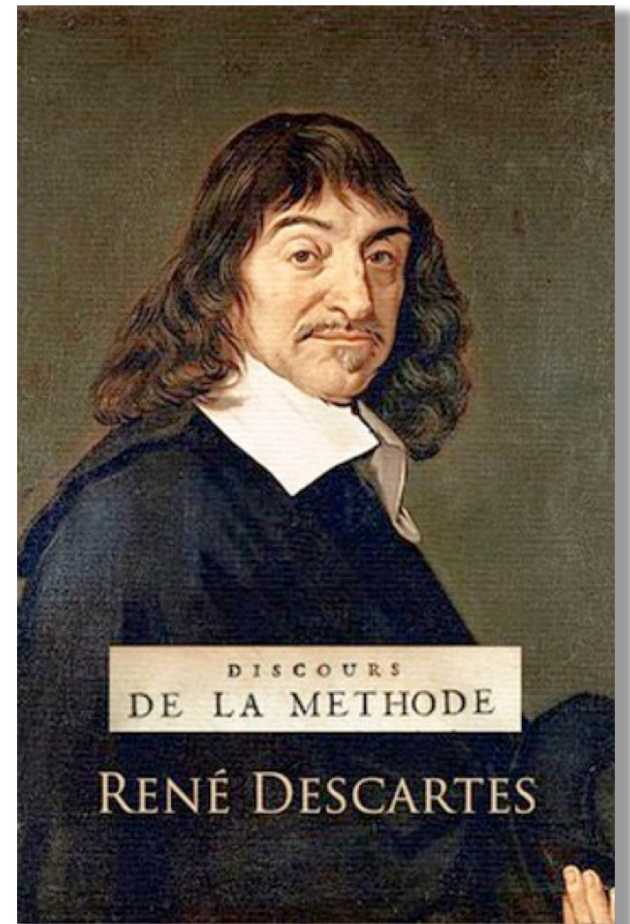
- Science = Scientific Method = Reproducible = Open!

Open Science: but then we already follow it, right?

- **We are scientists! We (want to) follow the Scientific Method!**

Open Science: but then we already follow it, right?

- **We are scientists! We (want to) follow the Scientific Method!**

nature International weekly journal of science   **2016**

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive › Specials and supplements archive › Challenges in irreproducible research

SPECIAL   ► See all spe

CHALLENGES IN IRREPRODUCIBLE RESEARCH

Science moves forward by corroboration – when researchers verify others' results. Science

Questionnaire on reproducibility (1500 scientists)

- 70% of researchers have tried and failed to reproduce another scientist's experiments
- > 50% have failed to reproduce their own ones!
  - Chemistry: 90% (60%)
  - Biology: 80% (60%)
  - Physics and engineering: 70% (50%)
  - Medicine: 70% (60%)
  - Earth and environmental science: 60% (40%)

Open Science: but then we already follow it, right?

- **We are scientists! We (want to) follow the Scientific Method!**

Questionnaire on reproducibility (1500 scientists)

- 70% of researchers have tried and failed to reproduce another scientist's experiments
- > 50% have failed to reproduce their own ones!
  - Chemistry: 90% (60%)
  - Biology: 80% (60%)
  - Physics and engineering: 70% (50%)
  - Medicine: 70% (60%)
  - Earth and environmental science: 60% (40%)

**Aha!
So you don't empathise?**

Open Science: but then we already follow it, right?

- **We are scientists! We (want to) follow the Scientific Method!**

nature — International weekly journal of science — 2016

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive > Specials and supplements archive > Challenges in irreproducible research

SPECIAL ▶ See all spe

CHALLENGES IN IRREPRODUCIBLE RESEARCH
Science moves forward by corroboration – when researchers verify others' results. Science

**Aha!
So you don't empathise?**

Questionnaire on reproducibility (1500 scientists)
- 70% of researchers have tried and failed to reproduce another scientist's experiments
- > 50% have failed to reproduce their own ones!
  - Chemistry: 90% (60%)
  - Biology: 80% (60%)
  - Physics and engineering: 70% (50%)
  - Medicine: 70% (60%)
  - Earth and environmental science: 60% (40%)

Overly Honest Method
@OverlyHonestly

Maybe with this?

You can download our code from the URL supplied. Good luck downloading the only postdoc that can get it to run, though #OverlyHonestMethods

Open Science: then what happened since 1637?

- **Moving beyond the PDF**

40% Knowledge Burying in paper publication =

Rest **I**n **P**aper

(S. Bechhofer 2011, Research Objects: Towards Exchange and Reuse of

Digital Knowledge)

http://www.clipartkid.com/rip-cliparts/

Open Science: then what happened since 1637?

- **Moving beyond the PDF**

40% Knowledge Burying in paper publication =

Rest **I**n **P**aper

(S. Bechhofer 2011, Research Objects: Towards Exchange and Reuse of Digital Knowledge)

In practice

http://www.clipartkid.com/rip-cliparts/
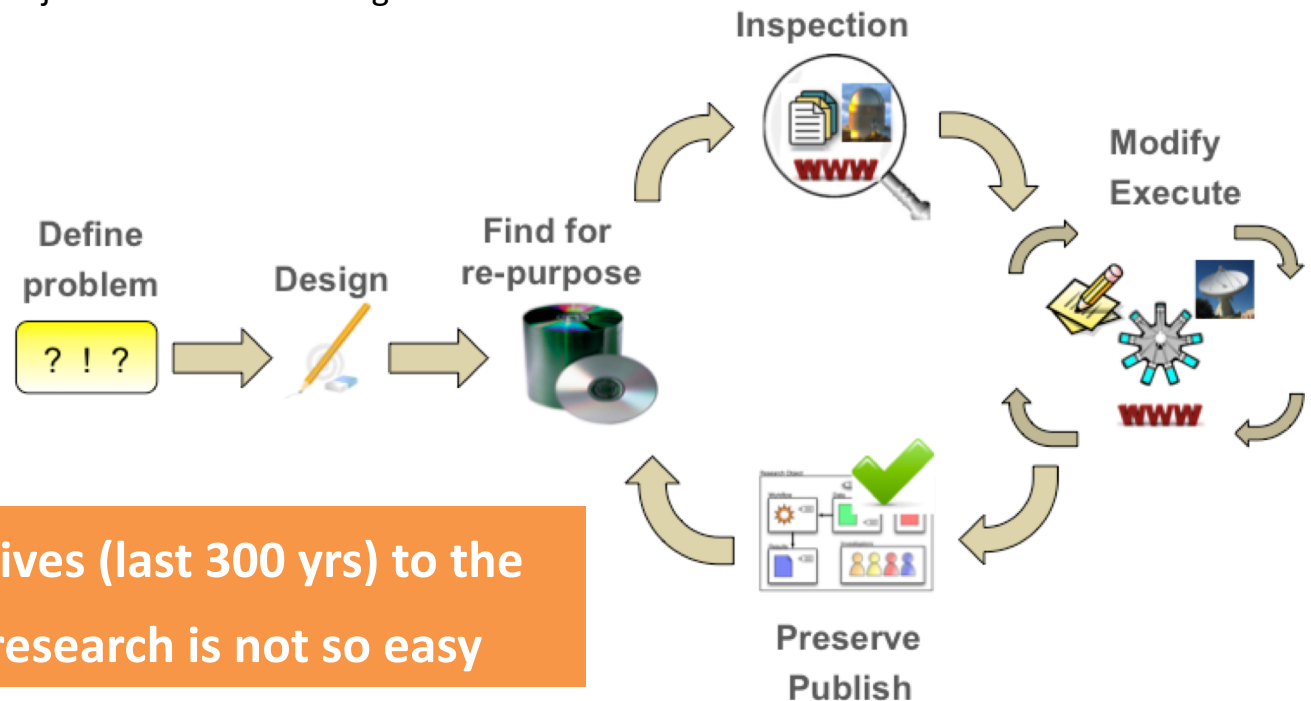
Moving from narratives (last 300 yrs) to the actual output of research is not so easy

Open Science for sustainability and inclusiveness

Open Science represents an approach to research that is collaborative, transparent and accessible

Open Science definition, European Commission, 2017, doi: 10.2777/75255

Remember: Open Science started bottom-up

(DORA, Metric Tide, Leiden Manifesto, Altmetrics, etc)

Open Science embodies the need to transform, open and democratize the entire knowledge generation to ensure that every scientific challenge is faced and really drives and allows the achievement of the United Nations Sustainable Development Goals

UNESCO and Open Science [1]

[1]https://en.unesco.org/news/unesco-launches-global-consultation-develop-standard-setting-instrument-open-science

Open Science for sustainability and inclusiveness

Science hidden behind paywall barriers

- 🚫 Free access to research sources to the whole scientific community = limitations to science progress

- OS = Data and results more accessible and reliable

- OS = Promotion of scholarly exchange of ideas

- OS = Avoid duplication

Acceleration of knowledge transfer to Society, pandemics, sanitary crisis

- Speed up building of skills

- Teaching, e.g. how to access public archives, fostering collaborative practices

- Citizen science

Open Science for sustainability and inclusiveness

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.

Open Science
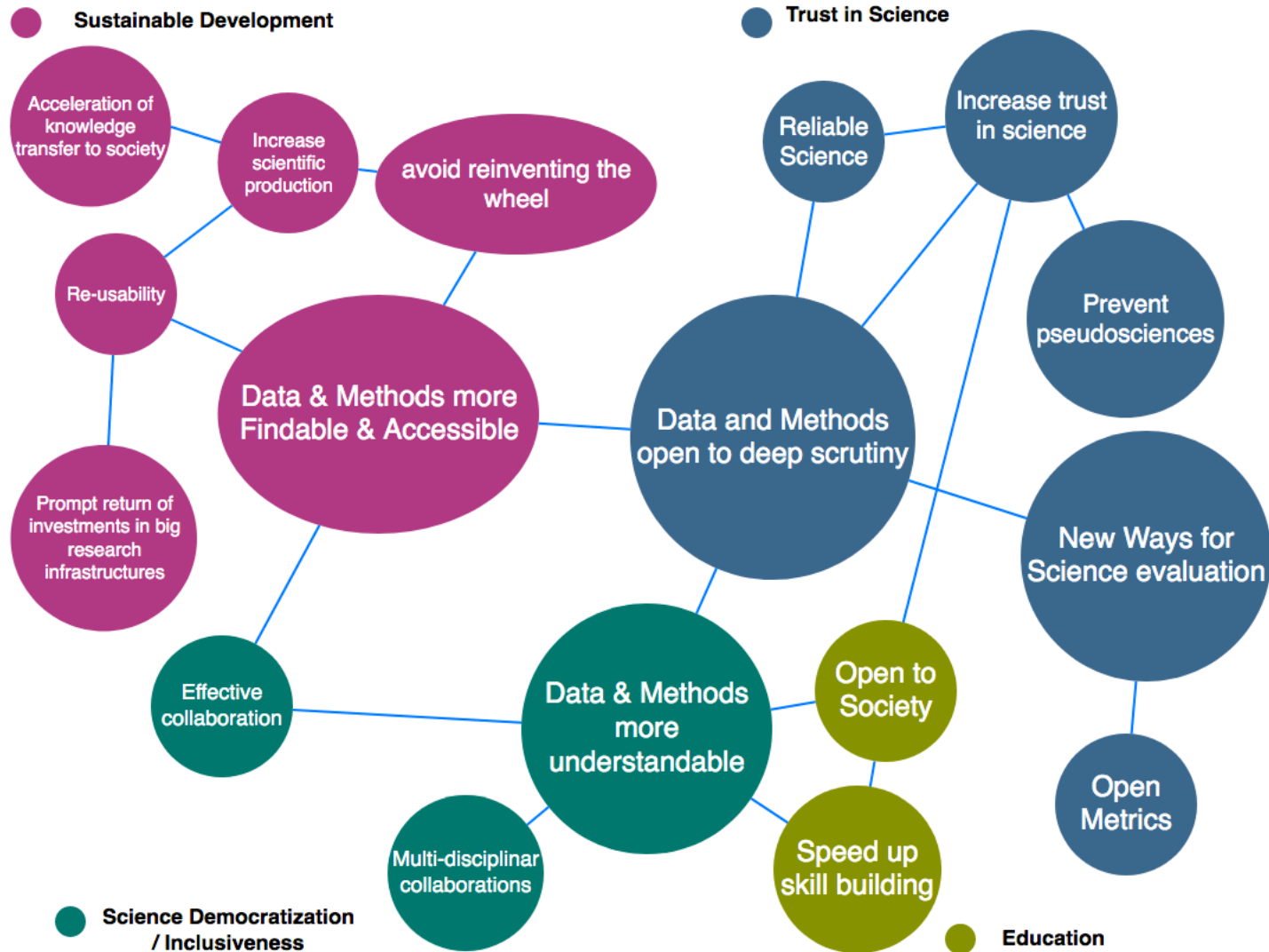
Open Science in the Big Data era

We are in a race to exploit ever larger datasets

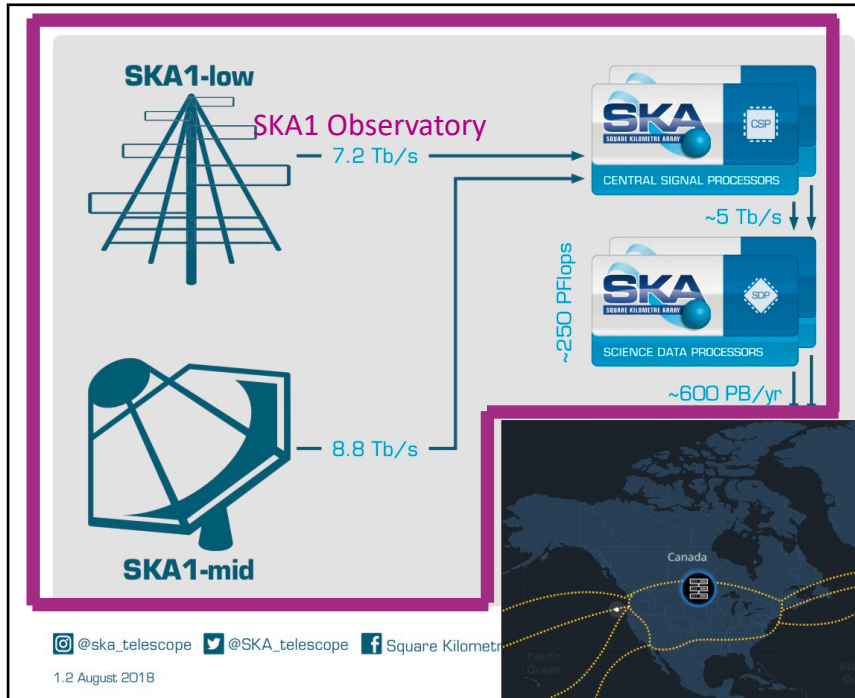**in our quest for "efficiency" we risk forgetting about reproducibility**

The era of Big Data is beginning across sciences

Today is the time to ask

What kind of research want mega-science infrastructures to do in tomorrow's future?

The Square Kilometre Array

SKA Fact sheets. August 2018. skatelescope.org

Credits: skatelescope.irg

Credits: AENEAS project

The SKA Regional Centre Network

Access to data products, tools and processing power to generate and analize Advanced Data Products (ADPs)
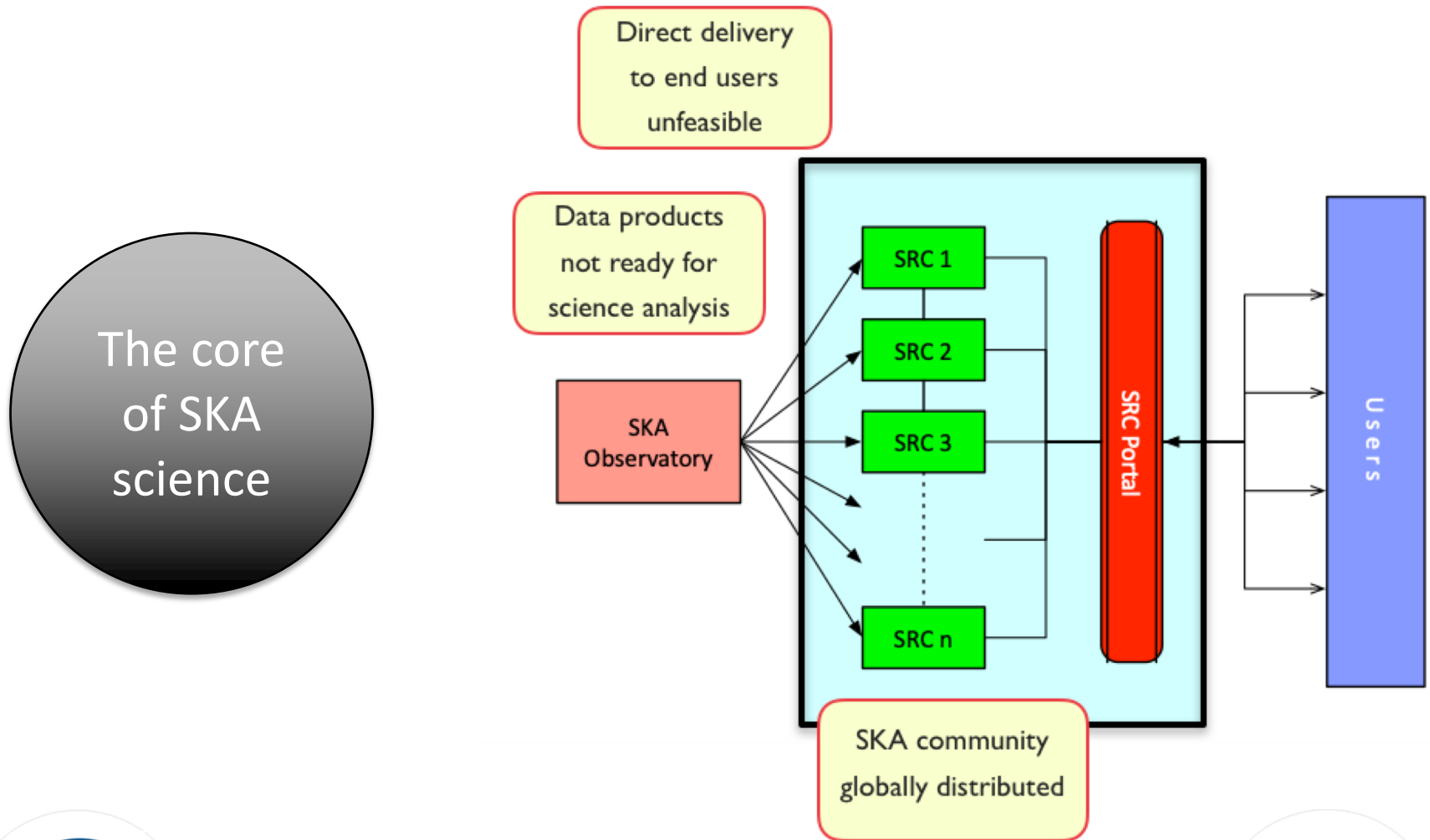
The SKA Regional Centre Network

Access to data products, tools and processing power to generate and analize Advanced Data Products (ADPs)
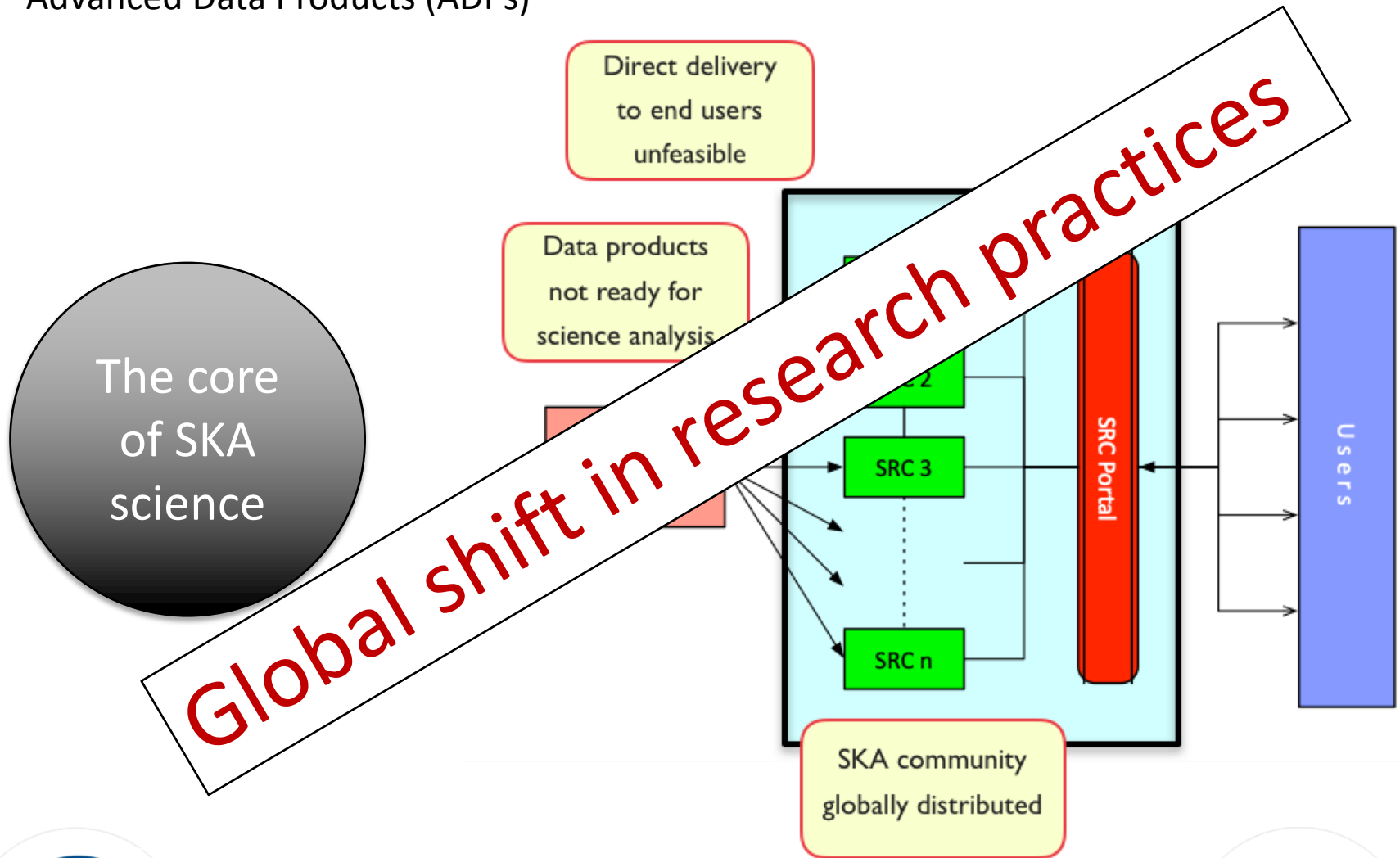
SKA duty: a role model

Is the current measure of excellence perverting Science?
A Data deluge is coming, it is time to act

Lourdes Verdes-Montenegro
Instituto de Astrofísica de Andalucía (IAA-CSIC)

CSIC

Session Organiser: William Garnier (SKAO)
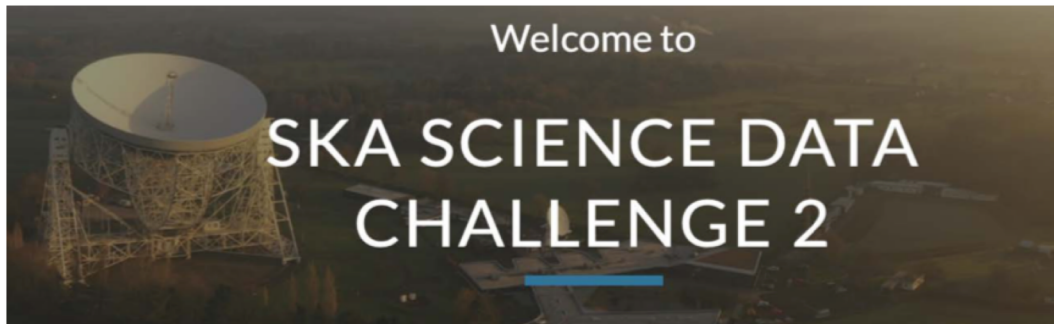(Submitter and Manager)

SKA
SQUARE KILOMETRE ARRAY

Theme #3 Science policy and transformation of research practice

#ESOF2018 @ESOF_eu ESOF.eu

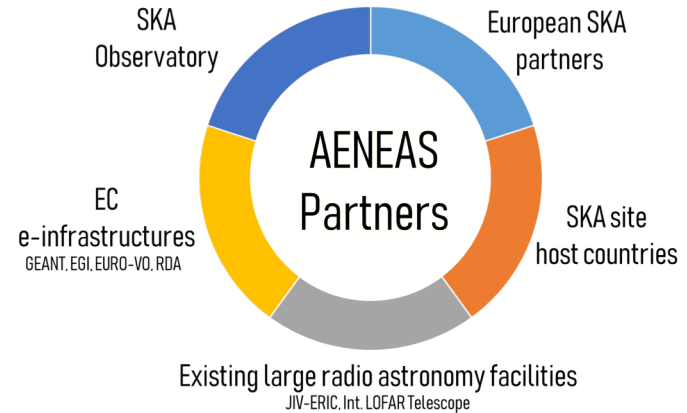Adoption of Open Science values

Reproducibility as a metric of success

Welcome to
SKA SCIENCE DATA CHALLENGE 2

Applying Open Science to build a federated network of European SRCs

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

2017-2019
http://www.aeneas2020.eu

SKA Observatory — European SKA partners — SKA site host countries — Existing large radio astronomy facilities (JIV-ERIC, Int. LOFAR Telescope) — EC e-infrastructures (GEANT, EGI, EURO-VO, RDA)

AENEAS Partners

SKA is contributing to build the European Open Science Cloud (EOSC)

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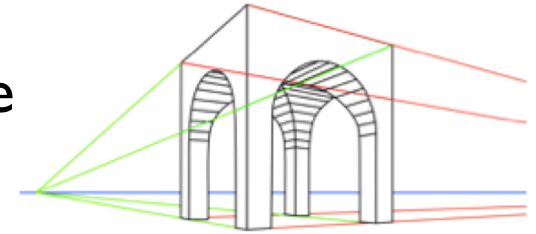ESFRI research infrastructures

2019-2022
https://projectescape.eu/

Addressing the Open Science challenges shared by ESFRI facilities and other pan-European research infrastructures in Astronomy and Particle Physics

IAA

Instituto de Astrofísica de Andalucía, IAA-CSIC
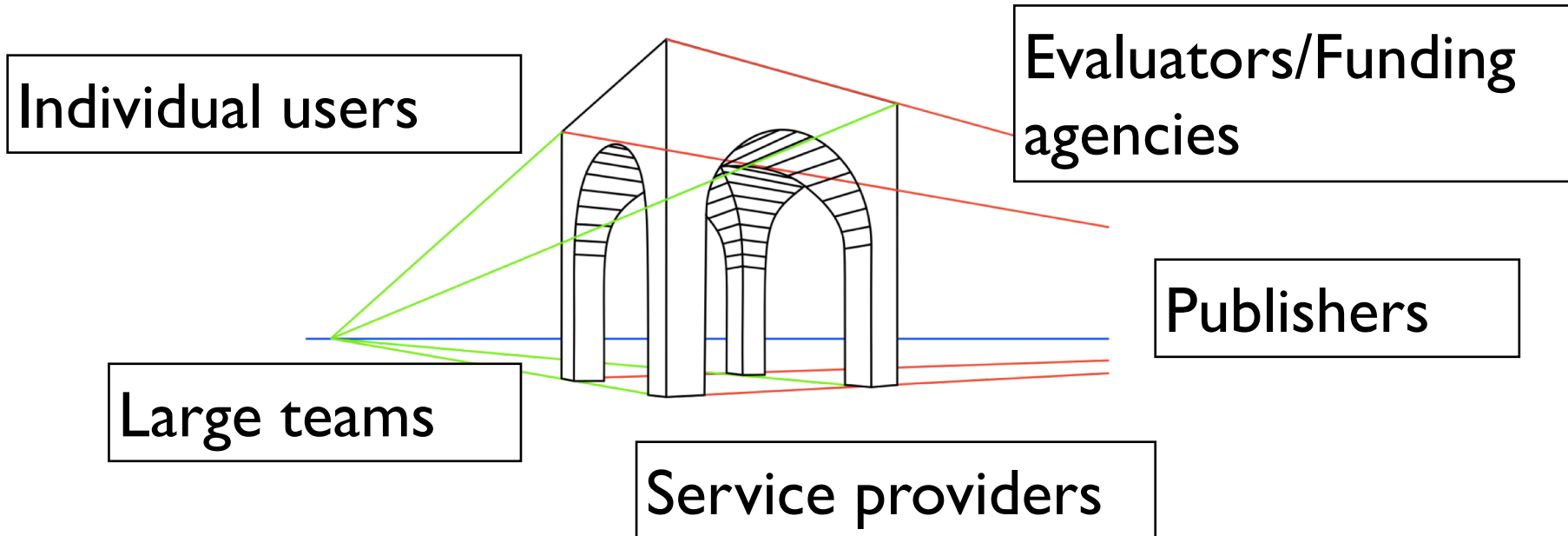
EXCELENCIA SEVERO OCHOA

The challenge from different perspectives

Implementation of Open, reproducible science

is challenging, even more in this new
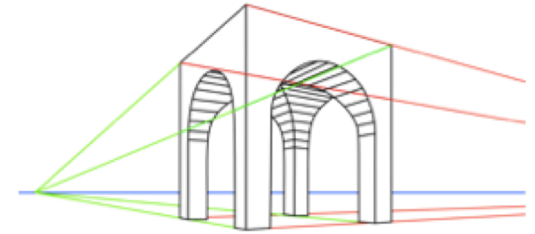
framework:

new roles ➡ **new perspectives**

Individual users

Large teams

Service providers

Evaluators/Funding agencies
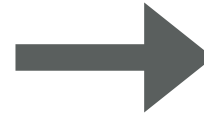
Publishers

Data to the desktop: "individual scientist"

About trust

- I have the best code, which I know how to use and can do special things

- I do not trust any pipeline that you made
    - partly because I know better how to do it
    - partly because I read the news and there is a reproducibility crisis
    - well, and I can hardly reproduce the results of my own papers some years later...

- In general I want full control of the software and of the computational environment

# Computation to data, providers perspective: Data Centres

About technology
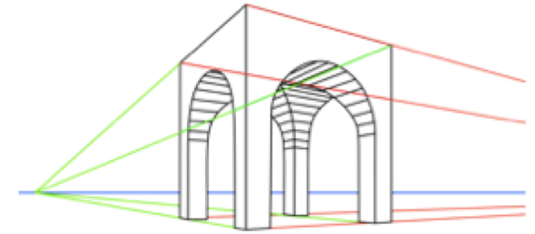
- We need to install your software in our platform. Can we trust it? Can we run it? Environment, dependencies, etc

- Hey, we are offering services to the community, computation + tools. We would be grateful if you allow us to share it with other users (with proper credit)

- Mmmm, sharing is great, but, **putting the software in the platform is not enough**: you need to provide the context for people to be able to rerun the software on the same or other data

Large alliances of scientists

About metrics of research careers
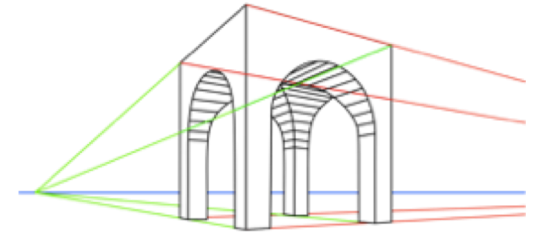
- We have tools to generate Advanced Data Products, and we will put them there where the storage and computation is (Data Centres)

- But... we put effort on it, what would we gain if we make the *additional effort* to make it reusable? If we make it, then we will pave the way to competitors

- Well, maybe we will share in 4 yrs time (PhD typical time)
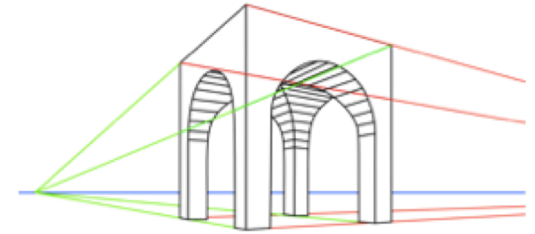
Publishers

Publishing models
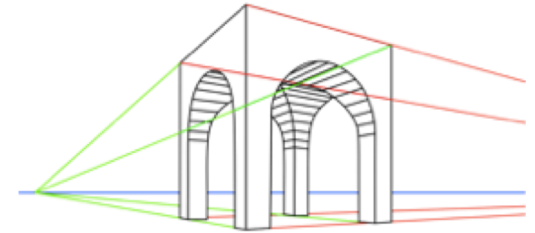
- Will we need **different profiles of referees** to evaluate the scientific discussion together with the data quality and the methods (aka. Reproducibility)?

- If the data and the methods (tools) will be in Data Centres, **will our referees need to become a "user" of the Data Centres** to be able to validate a paper?

- Will we be able to engage **so many referees** as may be needed?

- Will we need to validate the data, the tools, and the scientific analysis **separetely**?

IAA

EXCELENCIA
SEVERO
OCHOA

Policy makers / funding agencies

Evaluation

- How to measure reproducibility?

- How to weight it and/or aggregate with other indicators?

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:

Open Science & e-Science

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

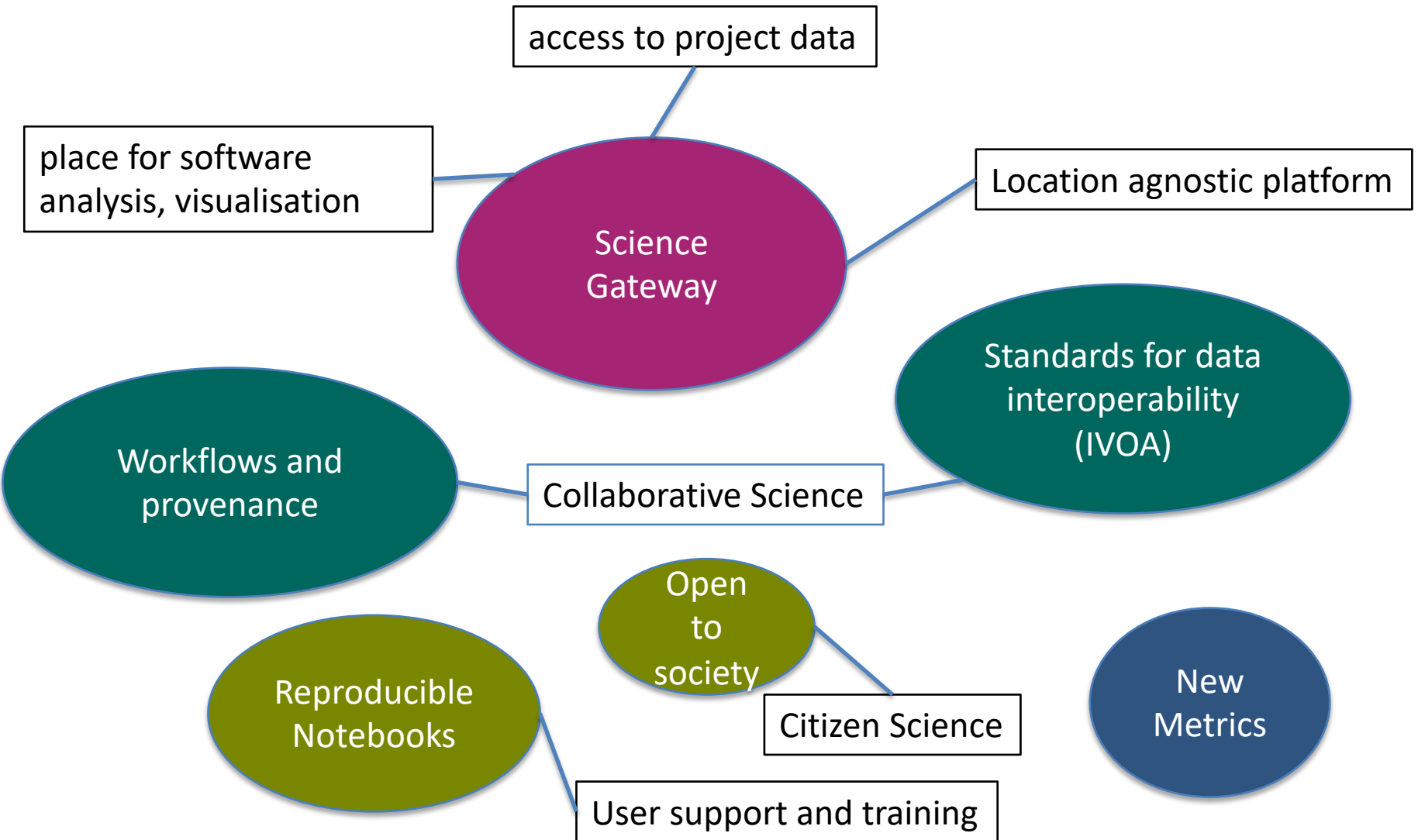## Open Science is the Aim and also the Mean

EXCELENCIA
SEVERO
OCHOA

Supporting Open Science in the SKA Regional Centres

A network of interoperable SRCs should provide:

– access to project data / storage capacity for archiving SKA data

– place for software analysis, modelling, visualisation, algorithm development / computational capacity, validation

– All this through a platform that

– is transparent and a location agnostic interface for users

– enables collaborative science

– User support (to enable scientific exploitation of the SKA data), training the new generation of radioastronomers, citizen science

Antonio Chrysostomou's talk

Key ingredients of the SRCs to support Open Science

access to project data

place for software analysis, visualisation

Science Gateway

Location agnostic platform

Standards for data interoperability (IVOA)

Workflows and provenance

Collaborative Science

Reproducible Notebooks

Open to society

Citizen Science

New Metrics

User support and training

SKA: a mega-science infrastructure engaged with trustable, reproducible science, going beyond numbers of papers/citations: pursuing the Scientific Method

Difficult? yes!
But this word doesn't seem to intimidate those aiming to build the SKA

SKA: a mega-science infrastructure engaged with trustable, reproducible science, going beyond numbers of papers/citations: pursuing the Scientific Method

THANK YOU FOR YOUR ATTENTION

Difficult? yes!
But this word doesn't seem to intimidate those aiming to build the SKA

Instituto de Astrofísica de Andalucía, IAA-CSIC

EXCELENCIA SEVERO OCHOA