AENEAS

H2020 -731016

PI: M. Wise (ASTRON)

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

Marcella Massardi

(INAF- IRA \ Italian node of the European ARC)

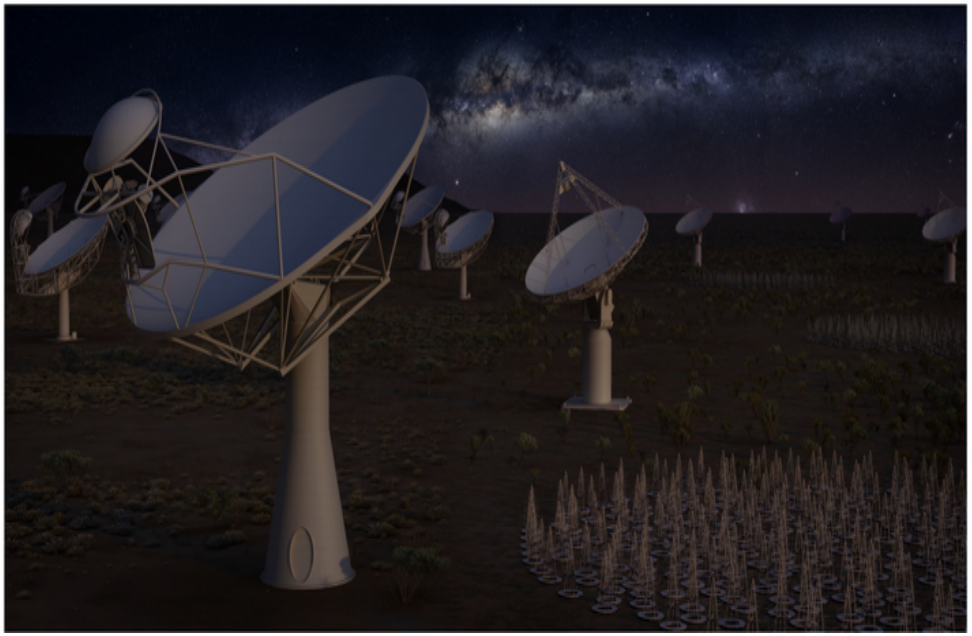
EUROPEAN ARC

ALMA Regional Centre || Italian

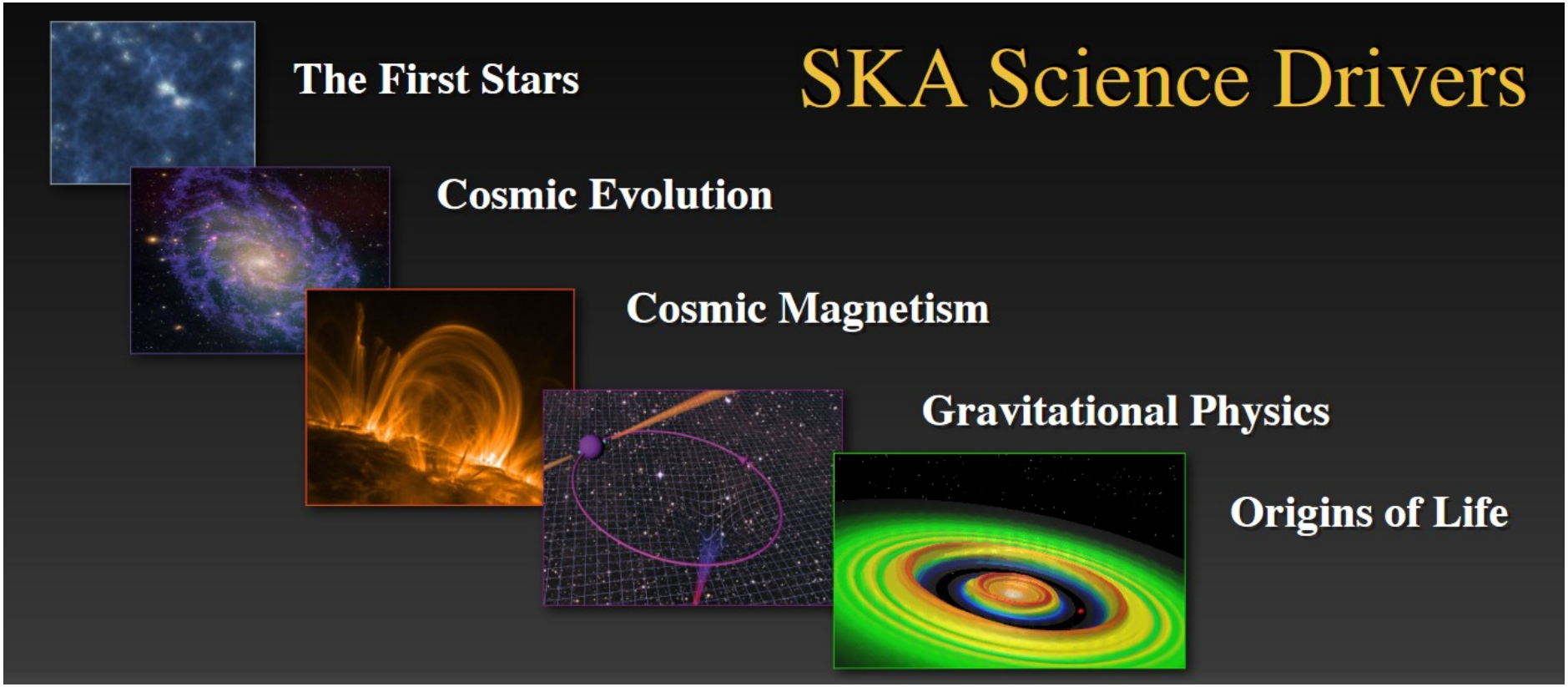
The AENEAS (Advanced European Network of E-infrastructures for Astronomy with the SKA) project has been funded in the Horizon 2020 Work Programme call "Research and Innovation Actions for International Co-operation on high-end e-infrastructure requirements" supporting the Square Kilometre Array (SKA).

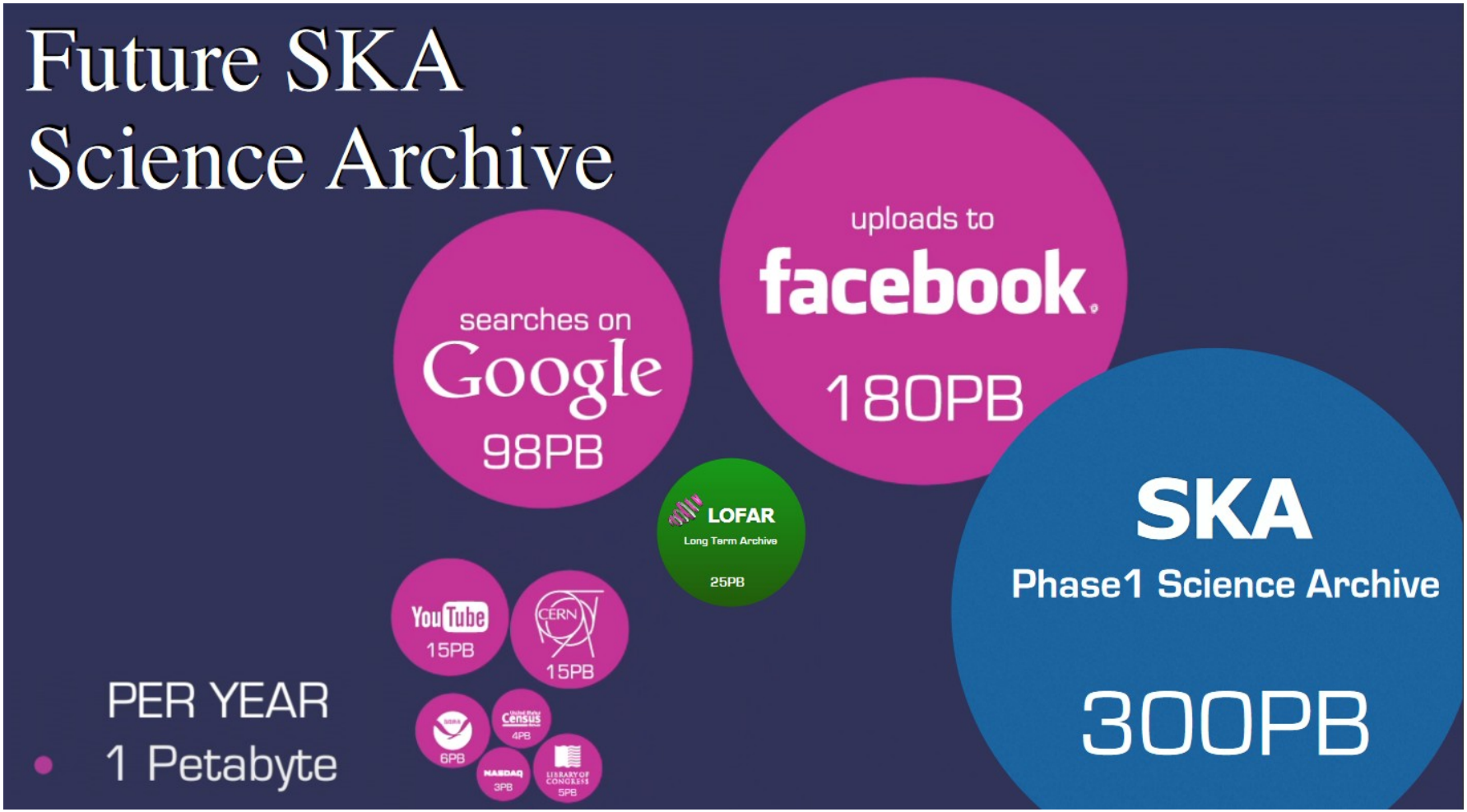
SKA will be:

- the world's largest radio telescope capable of transformational science and discoveries impossible with current facilities.
- built over two sites in Australia and Africa, over a million square metres of collecting area through many thousands of connected radio antennas.
- constructed in two phase with existing technology and then upgraded
- one of the priorities for the scientific communities
- one of the biggest challenge in data management, computing, networking

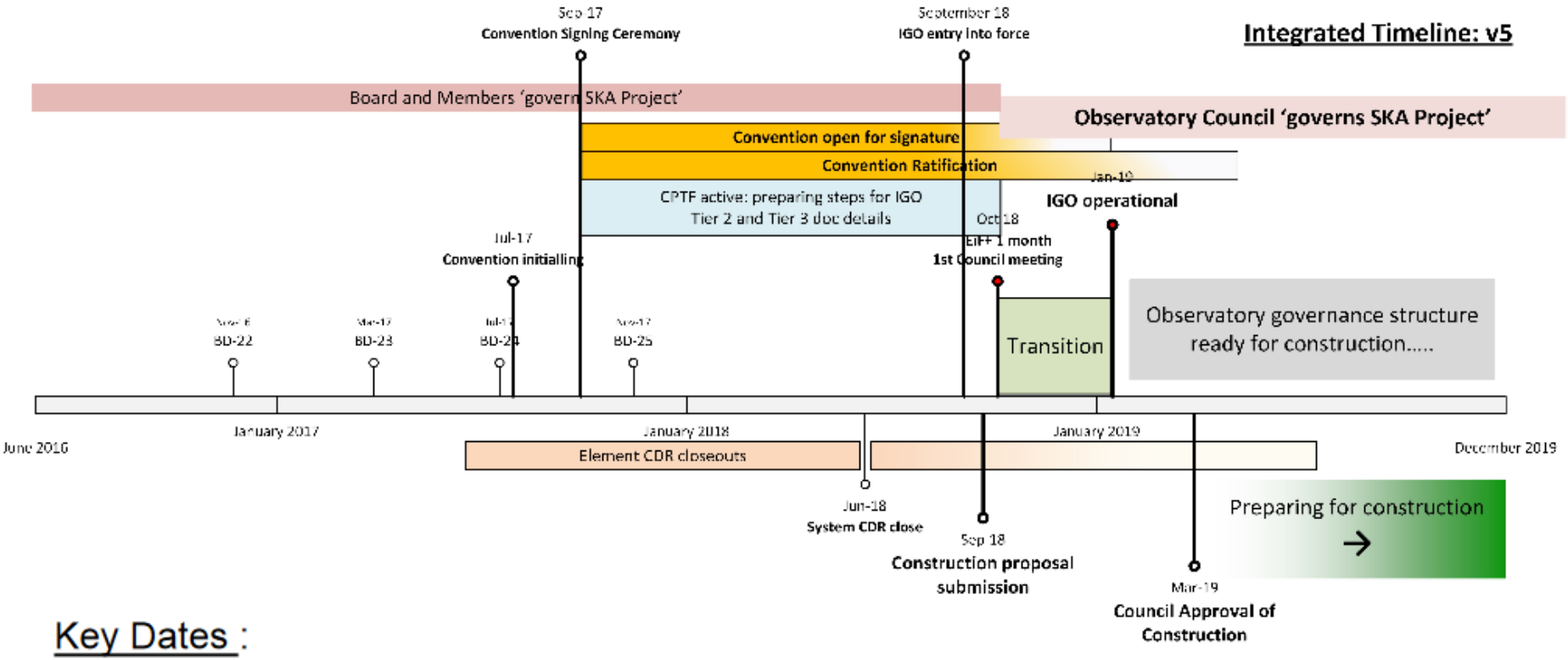


Like the SKA itself, a coordinated, global effort will be required to solve these challenges and fully realize the ground-breaking scientific potential of the project.


SKA Science Drivers
The First Stars
Cosmic Evolution
Cosmic Magnetism
Gravitational Physics
Origins of Life

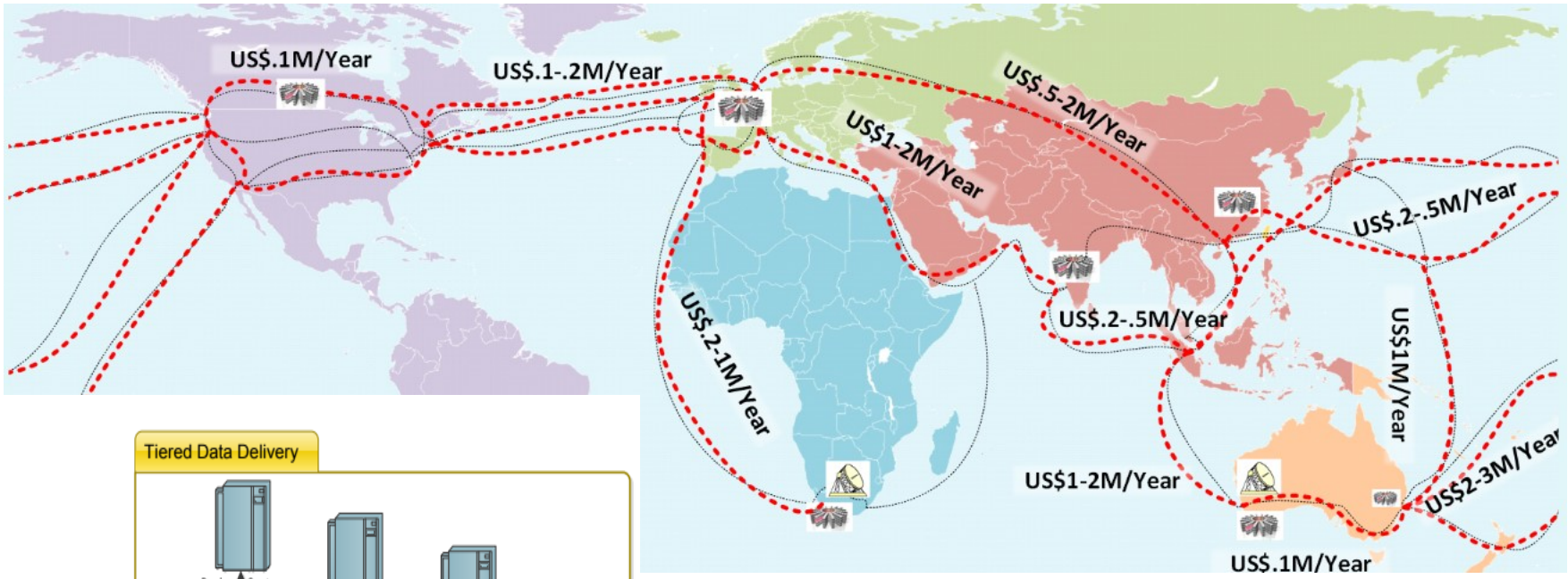


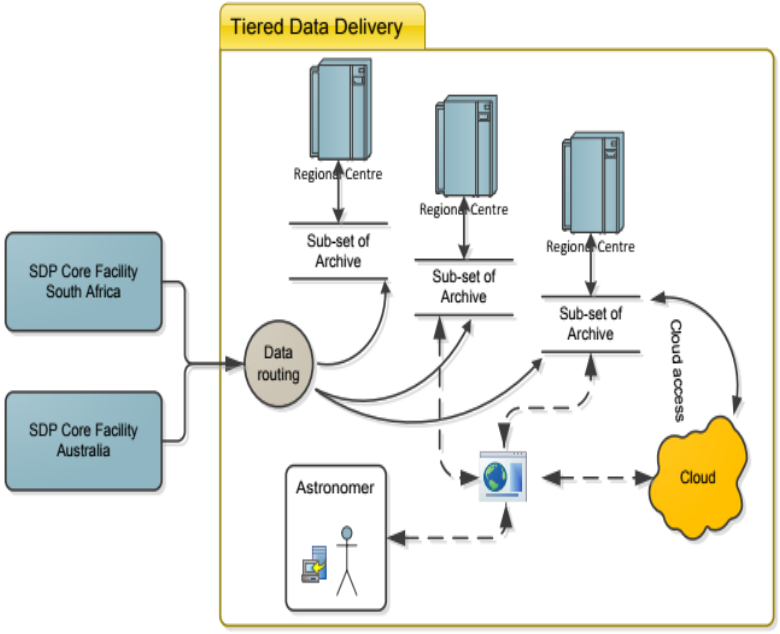
SKA is expected to produce an archive of standard data products with a growth rate on the order of 50—300 petabytes per year



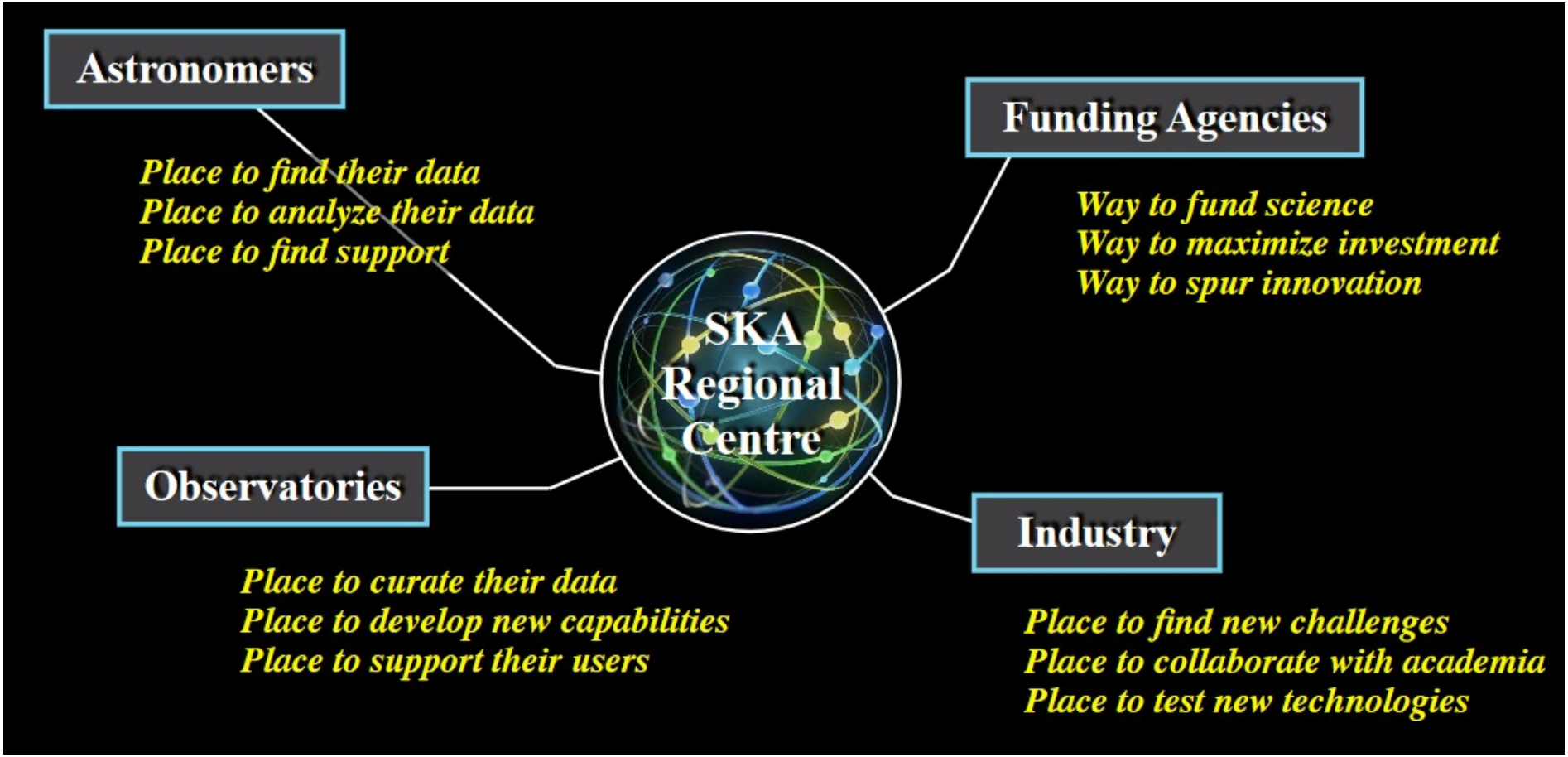
Key Dates :

- IGO in operation: early 2019
- Design process 'complete' ~mid/late 2018
- IGO Council approves construction: early/mid 2019
- SKA1 construction procurement begins: ~late 2019





Data distribution will be based on a tiered structure.
Raw data will not be transferred!
Archive of observatory products (images, cubes, timelines…) could be fully or partially copied to the nodes


Astronomers
Place to find their data
Place to analyze their data
Place to find support
Funding Agencies
Way to fund science
Way to maximize investment
Way to spur innovation
SKA Regional Centre
Observatories
Place to curate their data
Place to develop new capabilities
Place to support their users
Industry
Place to find new challenges
Place to collaborate with academia
Place to test new technologies

The objective of the AENEAS project is to develop a concept and design for a distributed, federated European Science Data Centre (ESDC) to support the astronomical community in achieving the scientific goals of the Square Kilometre Array (SKA).

The effort leads towards a Global SKA Alliance.

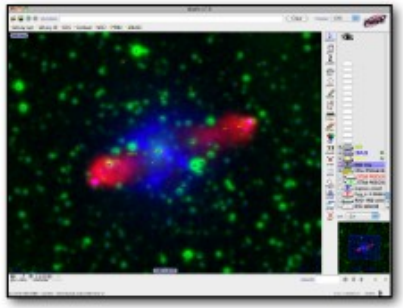
AENEAS is evaluating the management model to suggest for the European SDC (centralized node vs geographycally distributed network).

Costs, efforts and models are being planned evaluating the current facilities, the gaps to approach the SKA, and the future development.

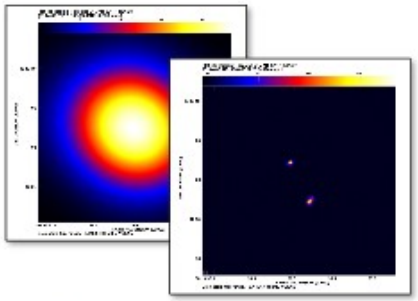
Regional Centre Functionality

Data Discovery

- Observation database
- Quick-look data products
- Flexible catalog queries
- Integration with VO tools
- Publish data to VO



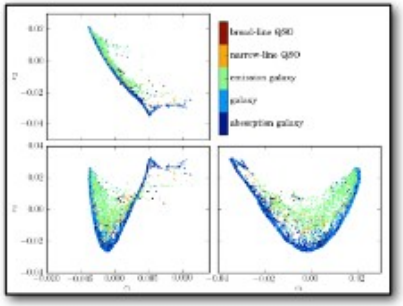
Data Processing



- Reprocessing
- Calibration and imaging
- Source extraction
- Catalog (re-)creation
- DM searches

Data Mining

- Multi-wavelength studies
- Catalog cross-matching
- Transient classification
- Feature detection
- Visualization



AENEAS Structure and goals

- WP1: Project Management
- WP2: Governance Structure and Business Models
- WP3: Computing and Processing Requirements
- WP4: Data Transport and Optimal European Storage Topologies
- WP5: Data Access and Knowledge Creation
- WP6: User Services

*This work package (WP5) is focused on the **interface between a distributed European SKA Data Centre (ESDC) and a distributed body of end users** whose goal is the exploitation of SKA data for knowledge creation. WP5 will therefore study the design of "user interaction models" that could be implemented for the ESDC*

Task 5.1 Survey of existing user interaction models for large-scale radio astronomy facilities and **integration of WP5 outputs into consolidated ESDC design study** (responsible M. Massardi)

Task 5.2 Recommendations for the design of user interfaces for data discovery, access, and retrieval (responsible R. Smareglia)

Task 5.3 Recommendations for the design of user interfaces for data processing, reprocessing, analysis, and visualization (responsible A. Costa)

Task 5.4 Integration with VO Interoperability Framework (responsible C. Knapic)

Task 5.5 Recommendations for the resourcing of an ESDC user interaction model (responsible J. Brand)

Task 5.6 Recommendations for a plan of user community formation and knowledge distribution (responsible M. Massardi)

Deliverable (number)	Deliverable name	Work package number	Short name of lead participant	Type	Dissemination level	Delivery date (in months)
D5.1	Survey report	5	INAF	R	PU	18
D5.2	Gap analysis	5	INAF	R	PU	18
D5.3	Design recommendations #1	5	INAF	R	PU	24
D5.4	Design recommendations #2	5	INAF	R	PU	24
D5.5	Applicability of VO framework	5	INAF	R	PU	28
D5.6	User interaction model resourcing	5	INAF	R	PU	28
D5.7	Growing the ESDC community	5	INAF	R	PU	28
D5.8	Final integration of WP5 materials	5	INAF	R	PU	34

Questions for Astronomical facilities

The Square Kilometre Array will enable transformational science across a wide range of research areas. By the same token, the large scale, rate, and complexity of data the SKA will generate present challenges in data management and computing that are similarly world-leading. Based on current projections, the SKA Observatory, once operational, is expected to produce an archive of standard data products with a growth rate on the order of 300 petabytes per year. Although the challenges associated with populating and maintaining the SKA science archive are already impressive, these data products actually represent only the first part of the full science extraction chain. Any further processing and subsequent science extraction by users will require significant, additional scientific, computing and storage resources in the form of a federated, global network of SKA Regional Centres.

Sent to the responsibles of LOFAR, MWA, ATCA, JVLA, PdBI, VLBA ...
VLBI Networks

Services provided and interaction policies

1. How do you define a user of your facility and what are the requirements to be a user?

Your answer

2. Please provide a summary of how users interact with your facility and the staff at your facility to generate science outputs

Your answer

3. Please list the services your facility provides (including observing, online tools, archive, user support...)?

Your answer

4. Can you provide links to facility policies that are relevant for users (e.g. proposal submission, time allocation policies, data access policies, policies describing the level of support from the facility for users to extract science from their data, procedures for publications and outreach)

Quantification of user community and products

6. How many users do your facility have?

- ◯ <100
- ◯ 100-500
- ◯ 500-1000
- ◯ 1000-2500
- ◯ >2500

7. How many proposals per call did you receive, on average in the last 3 calls for proposals?

	<50	50-100	100-500	500-1000	>1000
Proposals	◯	◯	◯	◯	◯

8. How many of these are accepted (including fillers) and, how many of the accepted are observed (including fillers) ?

	<30%	30-50%	50-70%	70-90%	>90%
Accepted	◯	◯	◯	◯	◯
Observed	◯	◯	◯	◯	◯

11. What do you estimate is the most resource-consuming aspect of supporting users of your facility, and why?

Your answer

12. Rate the driver for the cost of your user interaction model (1=strong, 5=weak)?

	1	2	3	4	5
data storage	○	○	○	○	○
computing	○	○	○	○	○
number of projects	○	○	○	○	○
number of staff required	○	○	○	○	○
number of users	○	○	○	○	○
software and analysis tools	○	○	○	○	○
Other(specify)....	○	○	○	○	○

13. How many FTE staff does your facility employ (or utilize) to implement your facility's user interaction model (include FTEs associated with computing centres that support your facility in an in-kind fashion, if appropriate) divided in the following categories?

<10	10-20	20-50	50-100	>100

Analysis of cost drivers

Staff skills and activities

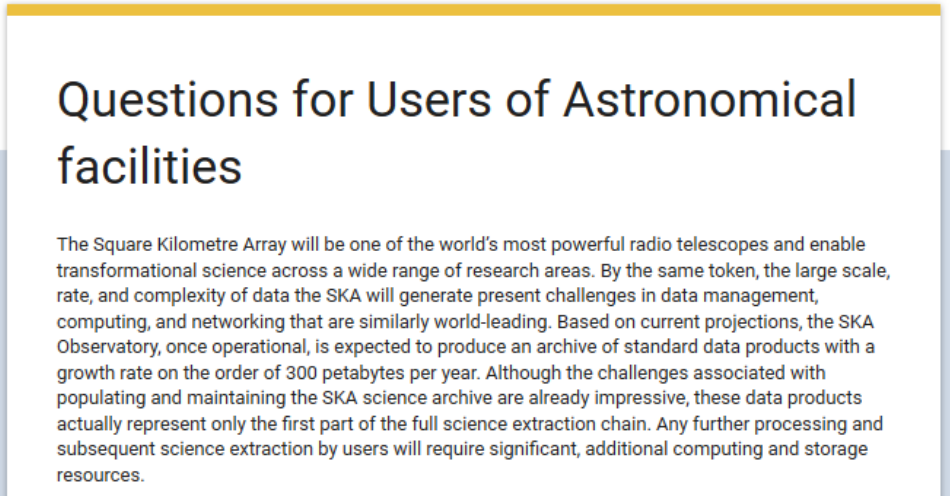
16. Rate the following skills that the people involved in your facility's user interaction need (1=necessary, 5=less important)

	1	2	3	4	5
Knowledge of the facility itself (available instruments, observing procedures, calibration requirements)	☐	☐	☐	☐	☐
Knowledge of the science carried out with the facility	☐	☐	☐	☐	☐
Knowledge of the software required to observe, calibrate, image	☐	☐	☐	☐	☐
Knowledge of the use of large-scale computing facilities	☐	☐	☐	☐	☐
Knowledge of managing big dataset	☐	☐	☐	☐	☐

17. What fraction of the total FTE is used for each of the following activity?

	<10%	10-20%	20-30%	30-40%	40-50%	50-60%	60-70%	70-80%
Archive maintenance	○	○	○	○	○	○	○	○
Data processing and quality assessment	○	○	○	○	○	○	○	○

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

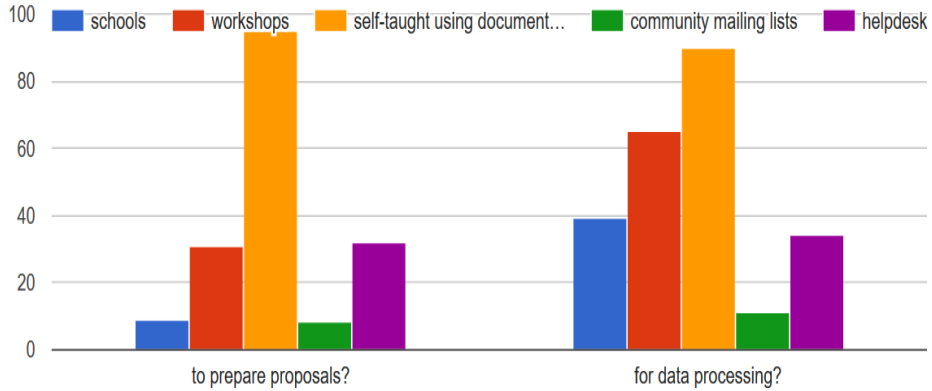


Questions for Users of Astronomical facilities

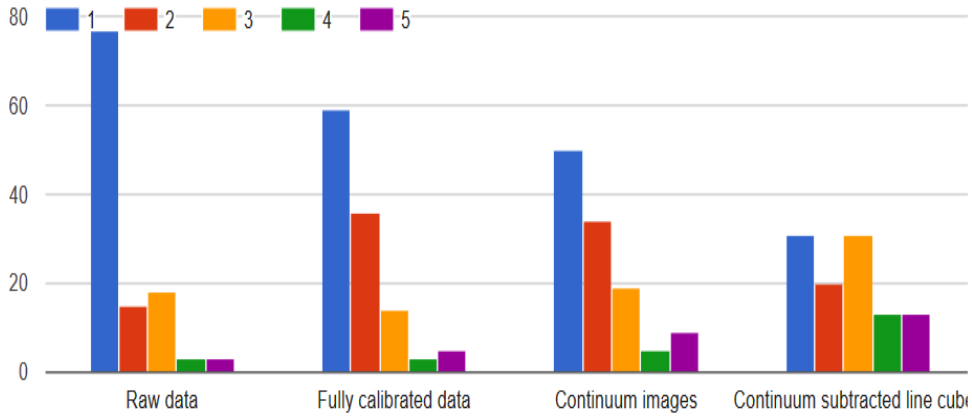
The Square Kilometre Array will be one of the world's most powerful radio telescopes and enable transformational science across a wide range of research areas. By the same token, the large scale, rate, and complexity of data the SKA will generate present challenges in data management, computing, and networking that are similarly world-leading. Based on current projections, the SKA Observatory, once operational, is expected to produce an archive of standard data products with a growth rate on the order of 300 petabytes per year. Although the challenges associated with populating and maintaining the SKA science archive are already impressive, these data products actually represent only the first part of the full science extraction chain. Any further processing and subsequent science extraction by users will require significant, additional computing and storage resources.

Sent to the user communities of the astronomical facilities and to members of SKSP

What is your preferred way of training...



Rate what would you like to find in an archive (1=necessary, 5=useless)



The questionnaire depicts a community that is not used to ask for any help...but used to handle its own raw data

SKA will require that users
- UNDERSTAND the limits
- TRUST the products
- “BRING THEIR CODE TO THE DATA”

SDC will be the first point of contact
Has to demonstrate
- TRUSTABILITY
- FLEXIBILITY
- EXPERTISE

and hide the difficulties...

IT REQUIRES TIME….
-How would be SKA day0?
-What are we doing in the meantime?

RESILIENCE

[ri-zil-yuhns]

The power to be
able to recover readily
from adversity or challenge.

Archive management, Data processing and visualization interfaces

AENEAS INTERNAL NOTE N. XXX

User Interfaces for Data Processing, Reprocessing, Analysis, and Visualization for the European SRC

E. Sciacca, F. Cavallaro, S. Riggi, A. Costa - INAF OACT

Archive management, Data processing and visualization interfaces

ARI aims to evaluate the feasibility and effort of using the now existing ALMA Imaging Pipeline to image the already existing calibrated science data of Cycles 0-4 as well as to ingest the resulting products into the ALMA Science Archive

Such a complete set of imaging products would be highly relevant for all science-cases and would
● dramatically improve the user-experience of archival research
● allow to show complete previews, easing the access to the Archive also to non expert archive data-miners
● be usable for Virtual Observatory services, or could be sent to visualization or analysis tools like CARTA or ADMIT.

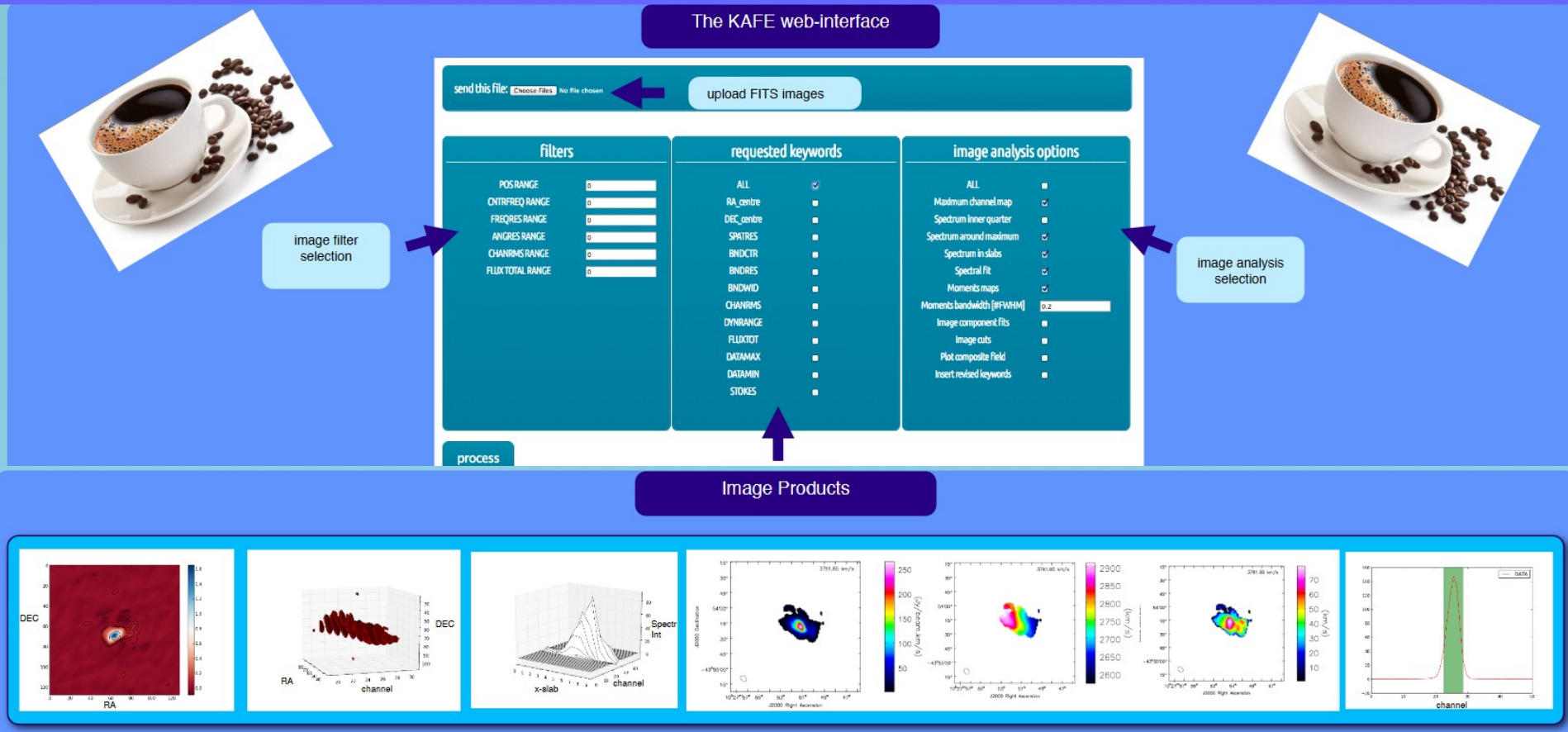
We run the full process on CHIPP successfully!!
See talk by A. Giannetti for details

Archive management, Data processing and visualization interfaces

KAFE - a FITS-image analysis tool

Sandra Burkutean, Andrea Giannetti, Elisabetta Liuzzo, Francesco Bedosti, Marcella Massardi,
Jan Brand, Kazi Rygl, Rosita Paladino, Matteo Bonato, Claudia Mancuso

Hosted by the AI2 (TS) webserver - See talk by S. Burkutean for details



INAF & the European SDC: why (my personal view)?

What do we have:
- A community of >200 researchers (>80 signed the white book in 2014)
- Knowledge & capabilities distributed in several institutes
  (i.e. taking expertise from different fields)
- Link to HPC/network facilities

Needs for any European node:
- collaborative approach
- long term plans for recruitments
- new professional figures dedicated to support
- long term investments in HW/SW

What can we get:
- A possibility to maximize our exploitation of SKA scientific results
- A huge return in scientific and technological skill
- A coordinated network to approach big-data/ICT future challenges
- A trained generation of astronomers that could use SKA & future facilities
- ...