IAU 2012 Data Intensive Astronomy Symposium (Sp15)
Beijing, August 29th 2012

# e-Science for the Square Kilometre Array

Juan de Dios Santander Vela (IAA-CSIC)
on behalf of the AMIGA team

# Talk Overview

- The Square Kilometre Array (SKA)
- The SKA Challenge
- AMIGA & SKA
- e-Science Tools for the SKA
- SKA Computing Synergies
- Conclusions

# The Square Kilometre Array

# The Square Kilometre Array

- The embodiment of *The Hydrogen Array* concept
- Thousands of antennas, with up to **1 sq km collecting area**
  - Distributed across thousands of kilometres of terrain
- With **enormous simultaneous bandwidth** to increase survey speed
- Can be incrementally built

**A Continental Scale, Distributed Sensor Network**

# SKA Antennas

## Combination of Different Antenna Kinds

## Low-Frequency Aperture Arrays

- Sparse aperture arrays
- 70 – 450 MHz
- Multibeaming

**SKA1: 2016 -2019**

## Mid-frequency dishes

- 13m Gregorian-offset dishes
- 450 MHz – 3 GHz
- Surface accuracy to 10-25 GHz

# SKA Antennas


## Mid-Frequency Aperture Arrays

- Dense aperture arrays
- 200 – 500 MHz
- 200 $deg^2$ FoV

# SKA Antennas

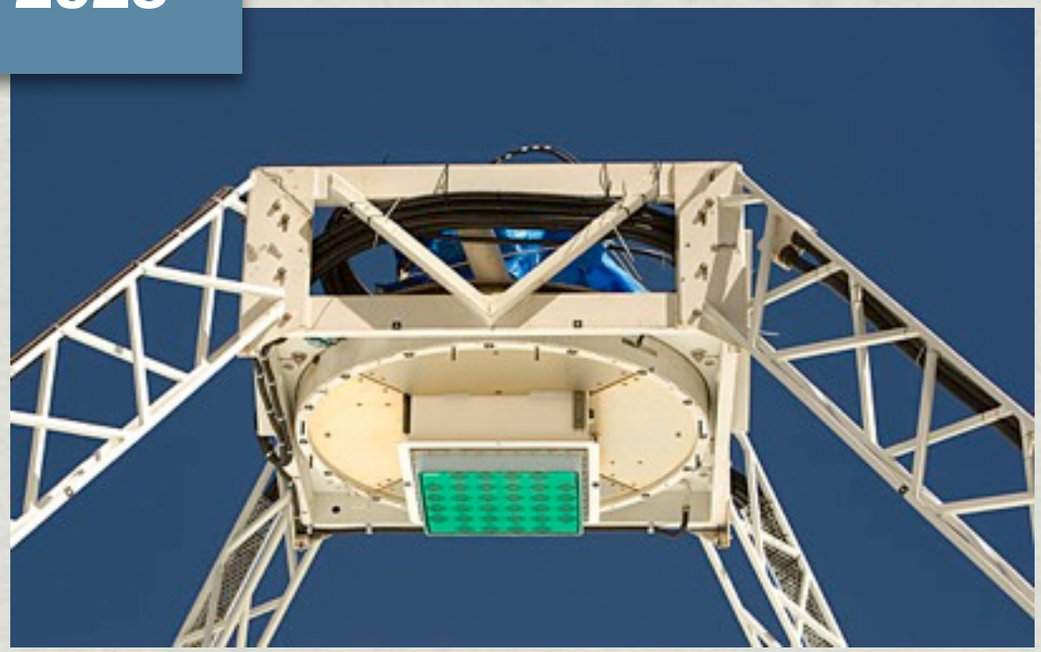
## Mid-Frequency Aperture Arrays

- Dense aperture arrays
- 200 – 500 MHz
- 200 deg$^2$ FoV

**SKA2: 2018 -2023**

## Focal Plane Arrays

- Multibeam Radio-Camera
- 12m antennas
- 700 MHz – 1.8 GHz
- Surface accuracy to 10 GHz



# SKA Site Selection

## South-Africa & Australia/New Zealand Joint Site

# SKA Site Selection

SKA2 AAs

SKA1&2 Mid


SKA2 Mid

SKA1 Low

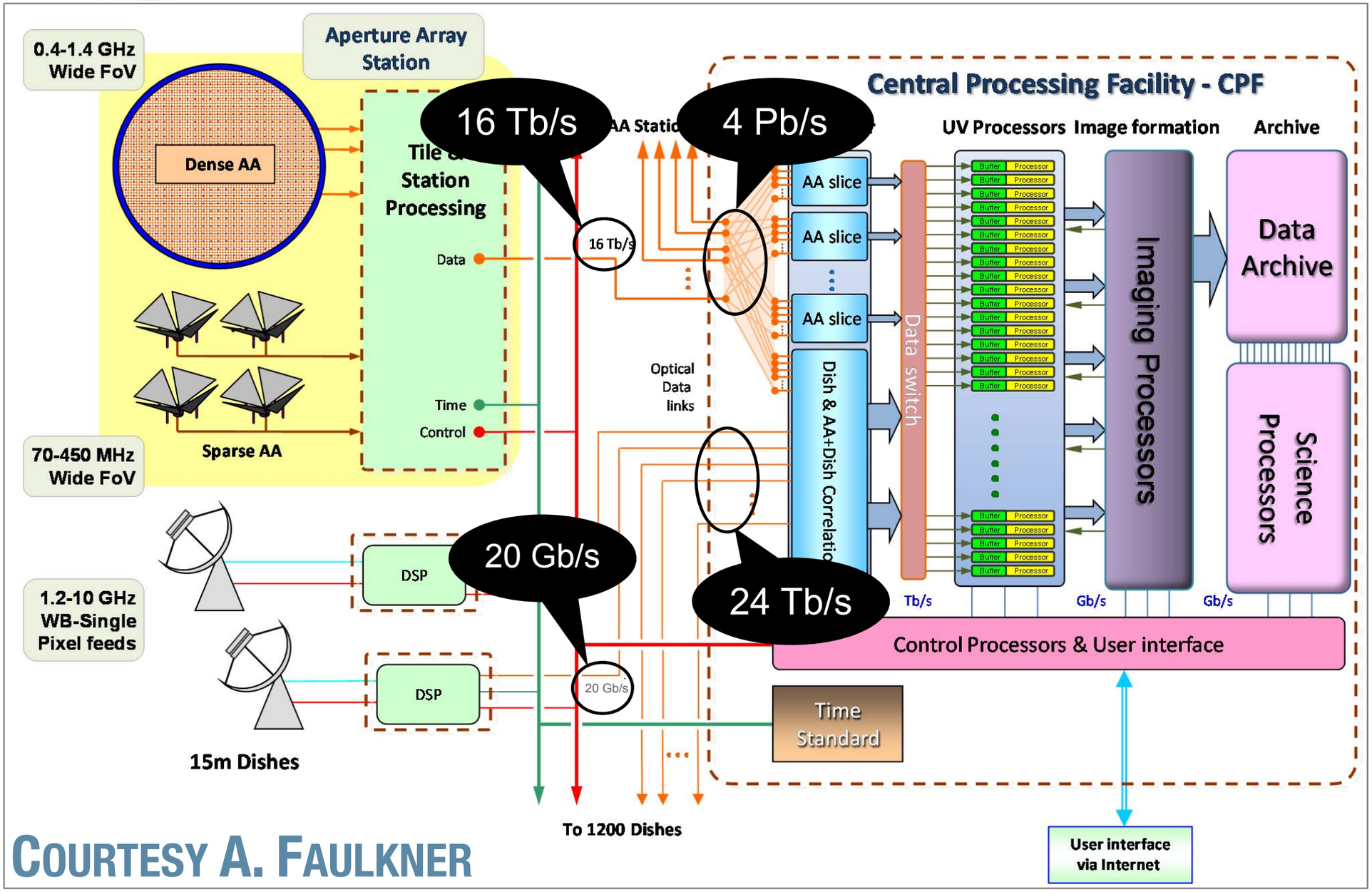
| SKA1 | | SKA2 | |
|---|---|---|---|
| SKA1_LOW | | SKA2_LOW | |
| SKA1_MID | | SKA2_MID | |
| SKA1_SURVEY | | SKA2_AA | |

SKA1 Survey

# The SKA Challenge

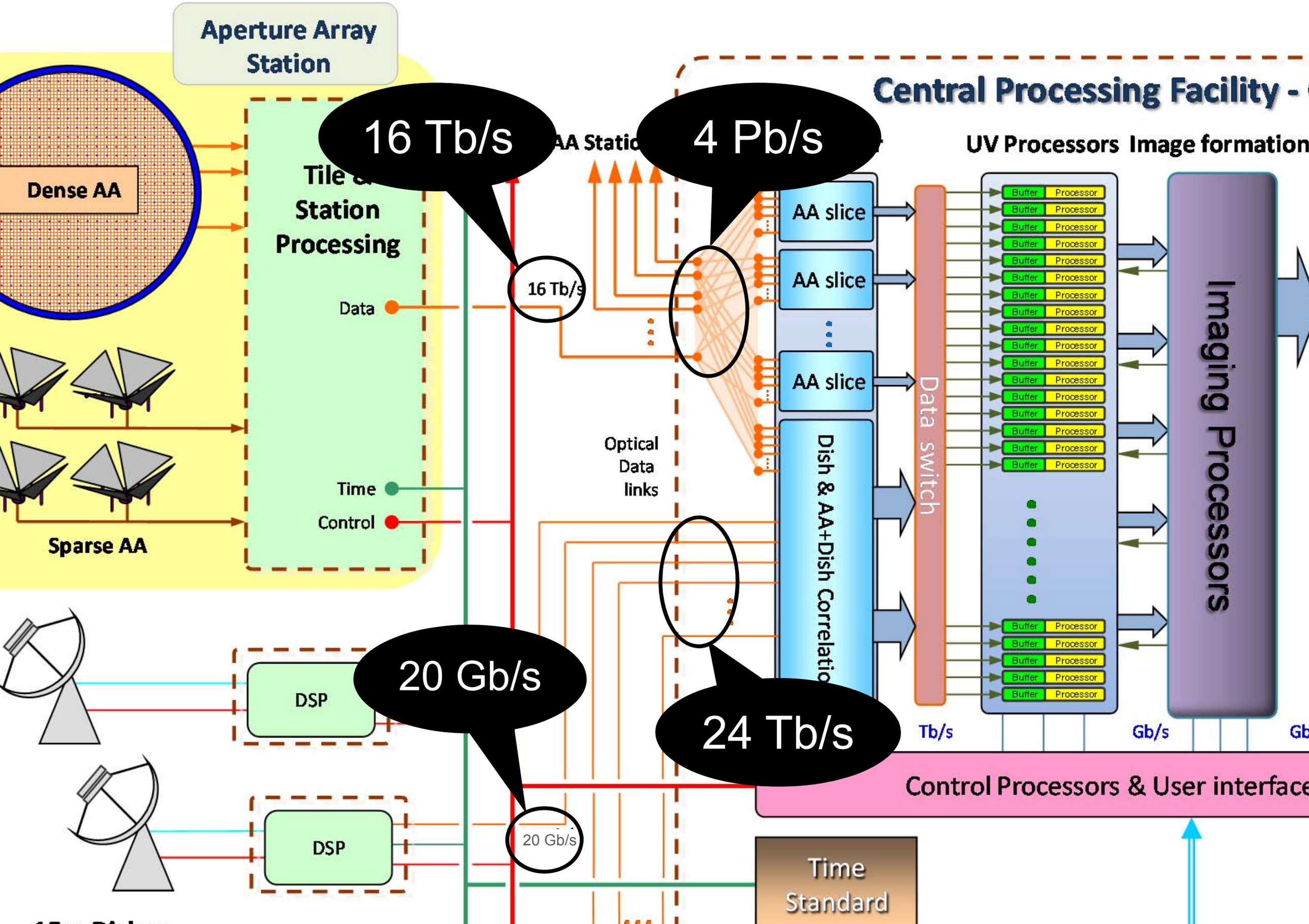
# MASSIVE DATA FLOW, STORAGE & PROCESSING

## $SKA_2$ wide area data flow

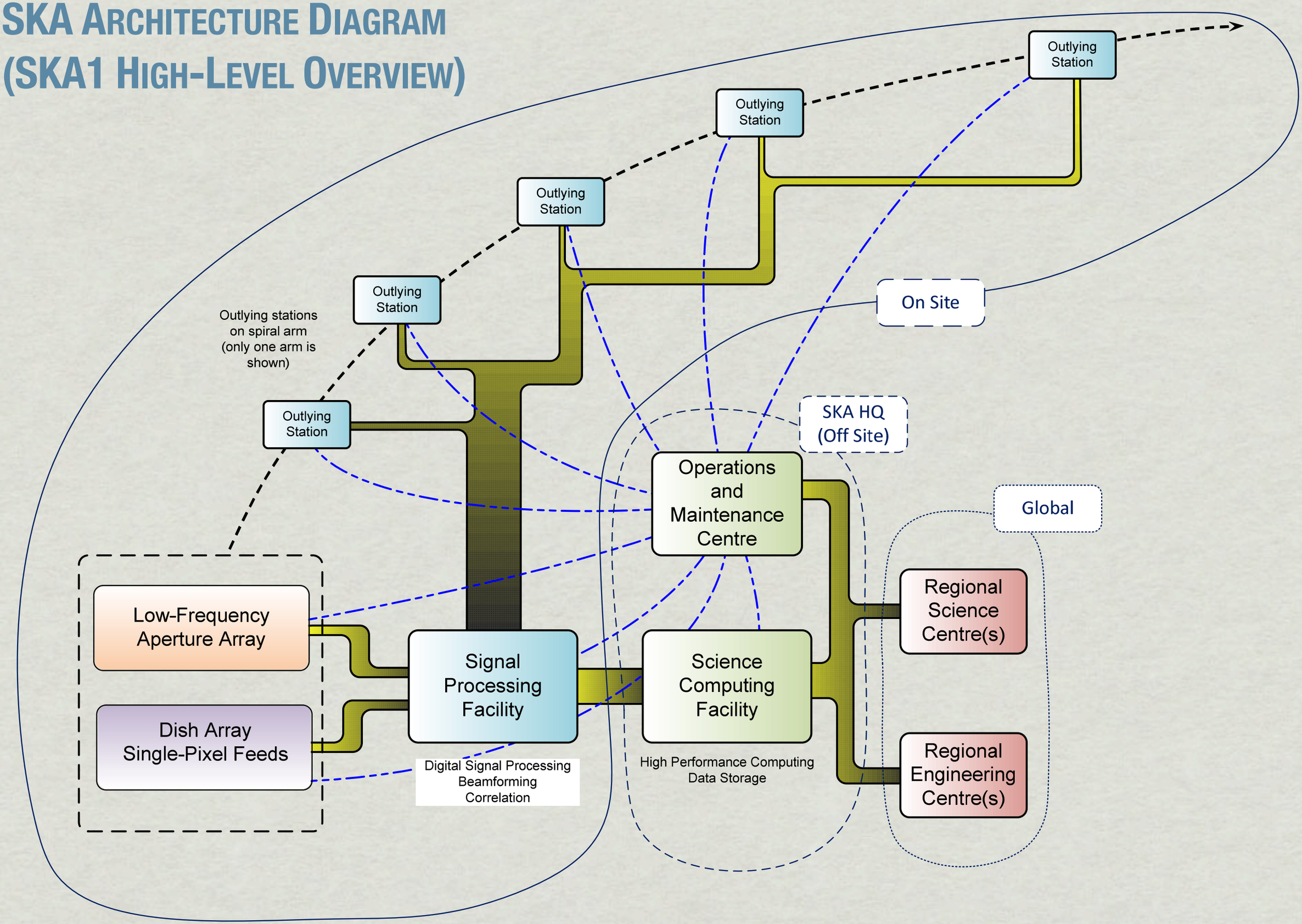



COURTESY A. FAULKNER

# SKA$_2$ wide area data flow



# SKA Architecture Diagram (SKA1 High-Level Overview)



ying stations
spiral arm
y one arm is
shown)

Outlying Station

Outlying Station

On Site

## Possible Science will Depend on Software Capabilities

SKA HQ (Off Site)

Operations and Maintenance Centre

Global

ncy
ay

y
eeds

Signal Processing Facility

Digital Signal Processing
Beamforming
Correlation

Science Computing Facility

High Performance Computing
Data Storage

Regional Science Centre(s)

Regional Engineering Centre(s)

## Software Driven

ying stations
spiral arm
y one arm is
shown)

Outlying Station

Outlying Station

On Site

SKA HQ (Off Site)

Operations and Maintenance Centre

Global

ncy
ay

y
eeds

Signal Processing Facility

Digital Signal Processing
Beamforming
Correlation

Science Computing Facility

High Performance Computing
Data Storage

Regional Science Centre(s)

Regional Engineering Centre(s)

INTERFACE TO THE USERS

# But, how can we do **our** science, then?

# AMIGA & SKA

# AMIGA

**A**nalysis of the interstellar **M**edium of **I**solated **GA**laxies

- Multi-wavelength, multi-object study on isolated galaxies with strict isolation criteria
- Careful curation of data
- Very careful processing of new parameters from
  - Group's own observation programs and data reduction
  - Literature table scanning
  - Virtual Observatory table harvesting and parsing
- Emphasis on marrying astronomy and computer science, and buy-in of the VO

E-SCIENCE USERS

# AMIGA

**PI, L. Verdes-Montenegro**
**Revise Her Talk for Main Results**
**(SpS3, Secular Evolution)**

**A**nalysis of the interstellar **M**edium of **I**solated **GA**laxies

- Multi-wavelength, multi-object study on isolated galaxies with strict isolation criteria
- Careful curation of data
- Very careful processing of new parameters from
  - Group's own observation programs and data reduction
  - Literature table scanning
  - Virtual Observatory table harvesting and parsing
- Emphasis on marrying astronomy and computer science, and buy-in of the VO

**E-Science Developers!**

# AMIGA

- Project goal: providing a baseline for galaxy properties to compare with other environments

- Interaction-free sample, ideal for tracing **HI infall**: we can use CIG galaxies to detect the **cosmic web**

- Need for very sensitive telescopes able to resolve faint HI ➡ Square Kilometre Array & pathfinders

**WE NEED TOOLS FOR OUR OWN SCIENCE ANALYSIS**

↳ **PARTICIPATING IN SKA.TEL.SDP PROTOCONSORTIUM**

# e-Science Tools & SKA

# e-Science Tools & SKA

- Distributed computing
  - Move computation to the data
  - Computing services ➡ Science-computing
- Collaborative environments
- Linked data

} **For Scientific Discussion & Science Extraction**

# Defining Computations

- Events & Processes
- Dependencies
  - Resources
  - Local & Remote Processes
  - Sequences
  - Concurrences
  - Triggers

**Formally, or at least Machine Readable**

**Workflow Definition Languages**

# AMIGA Contributions

- Wf4Ever
  - Workflows for **process** & **scientific methodology** specification
  - Web and command line **tools for data preservation**, **methodology preservation**, **reuse**, **repurposing**, & **collaboration**
  - Provide extra tools for **astronomical data processing** & **services**
- AMIGA4GAS
  - Use workflows as **process abstraction engines**
  - Use **federation** and **supercomputing models** for Taverna
  - Adapt Taverna (& workflows) to those computing models

Workflow 4Ever


## EU FUNDED FP7 STREP PROJECT
## DECEMBER 2010 – DECEMBER 2013

1. **Intelligent Software Components (iSOCO, Spain)**
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Centre (PSNC, Poland)
5. University of Oxford (OXF, UK)
6. Instituto de Astrofísica de Andalucía (IAA, Spain)
7. Leiden University Medical Centre (LUMC, NL)

Workflow 4Ever

Technological **infrastructure** for the **preservation** and **efficient retrieval** and **reuse** of scientific workflows in a **range of disciplines**

## Partners

- One SME
- Six public organisations

## Core Competencies (Tech)

- Digital Libraries
- Workflow Management
- Semantic Web
- Integrity & Authenticity
- Provenance
- Information Quality

## Case Studies

- Astronomy (IAA-CSIC)
- Genome-wide Analysis and Biobanking

## Goals

**Archival, classification, and indexing** of **scientific workflows** and their associated materials in scalable semantic repositories, providing advanced access and recommendation capabilities

**Creation of scientific communities** to collaboratively share, reuse, and evolve workflows and their parts, stimulating the development of new scientific knowledge

**Targeting Already Established Communities: myExperiment, Virtual Observatory**

Fork Me on GitHub

# AstroTaverna

Astronomy plugins for Taverna Workbench

## AstroTaverna

To install the AstroTaverna plugin to Taverna:

- Download and install Taverna 2.4
- Start Taverna
- Add a plugin site: `http://wf4ever.github.com/astrotaverna/`
- Restart Taverna
- The *VO services* perspective should now appear together with variouys local tools under Available Services

For more information, see the Astrotaverna wiki page.

Fork Me on GitHub

Taverna Workbench 2.4.0

Design | Results | myExperiment | VO services | Service Catalogue

Service panel

Filter: Clear

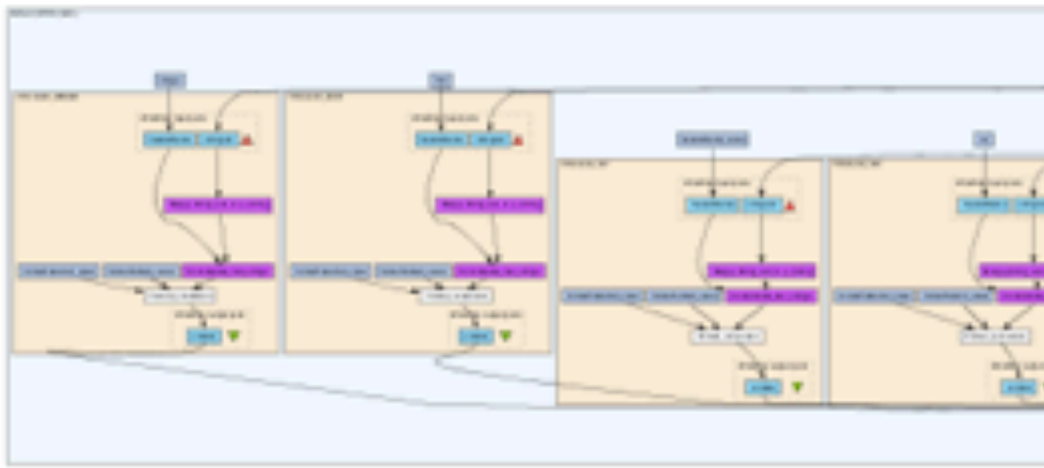
Import new services

- Local services
- Astro tools
  - Add Column – Add column using a expression
  - Add sky coordinates – Add sky coordinates
  - Cat n-tables – Cat a list of tables
  - Cat tables – Cat two tables
  - Check template filler – Check Template filler
  - Coordinates transformation – Coordenates transformation in a table
  - Format conversion – Table format conversion
  - List from column – Get list from column in a votable
  - Resolve coordinates – Resolve object coordinates
  - Select columns – Columns selection in a table
  - Select Rows – Rows selection in a table
  - Template filler – Template filler from a votable
  - tjoin – Join between tables
- Biomart @ http://www.biomart.org/biomart/martservice

Workflow explorer | Details | Validation report

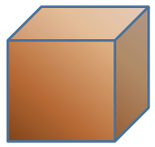
- Gathering_info_from_
  - Workflow input ports
    - init_table
  - Workflow output ports
    - VOtable
  - Services

Gatherin

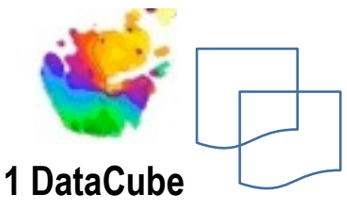


# AMIGA4GAS

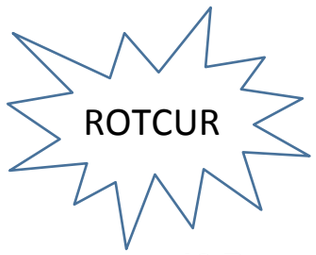
## 3D Kinematical modeling



Input Files



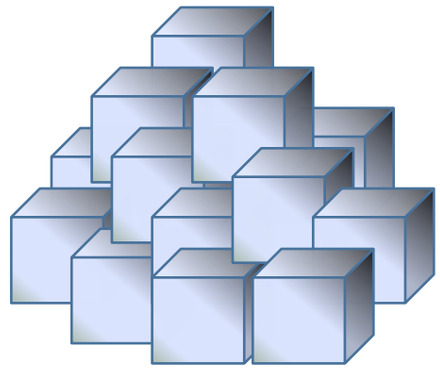
1 DataCube
1 Velocity Map
1 Config File Rotcur
1 Config File Galmod

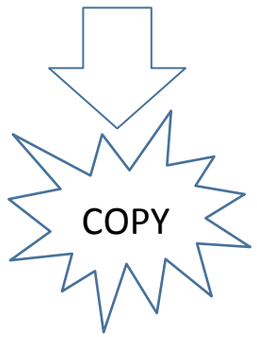


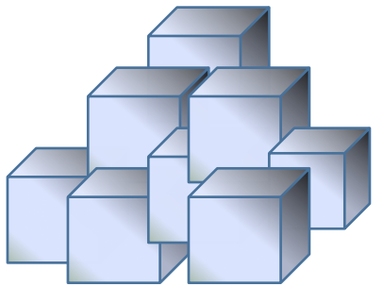

12 Runs
Possible combinations
in Input Parameters

12 ASCII Files

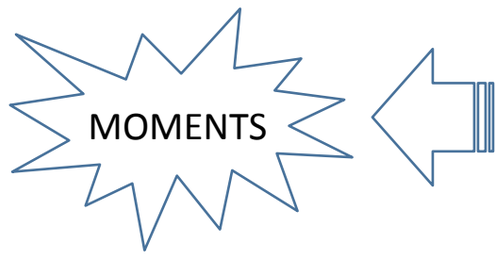

12 Cubes
• 4 Approaching
• 4 Receeding
• 4 Both

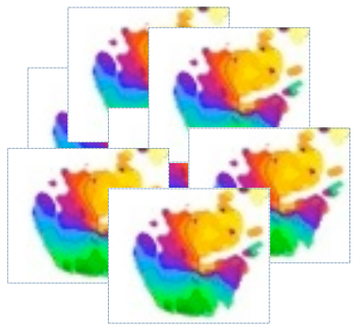




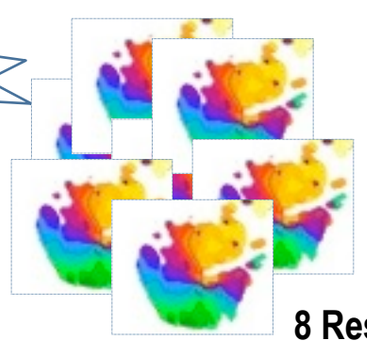



8 Cubes
• 4 Approaching + Receeding
• 4 Both

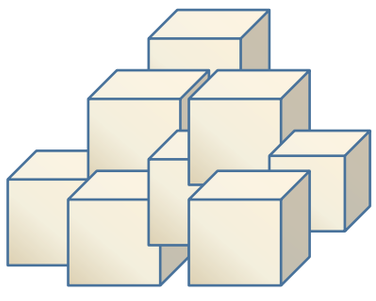


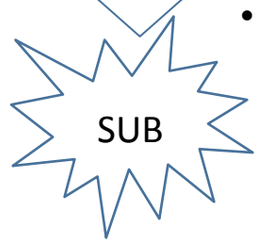



8 Velocity Maps

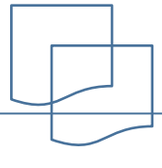
8 Residual Cubes
8 Residual Maps






## VARIABLE PARAMS

INSET
RADII, WIDTHS
WEIGHT
TOLERANCE
DENS
NV
Z0
VDISP




8 Values for Peaks in Cubes
8 Values for Peaks in Maps

GIPSY

# AMIGA4GAS

**AMIGA** for the **G**TC, **A**LMA, and **S**KA Pathfinders

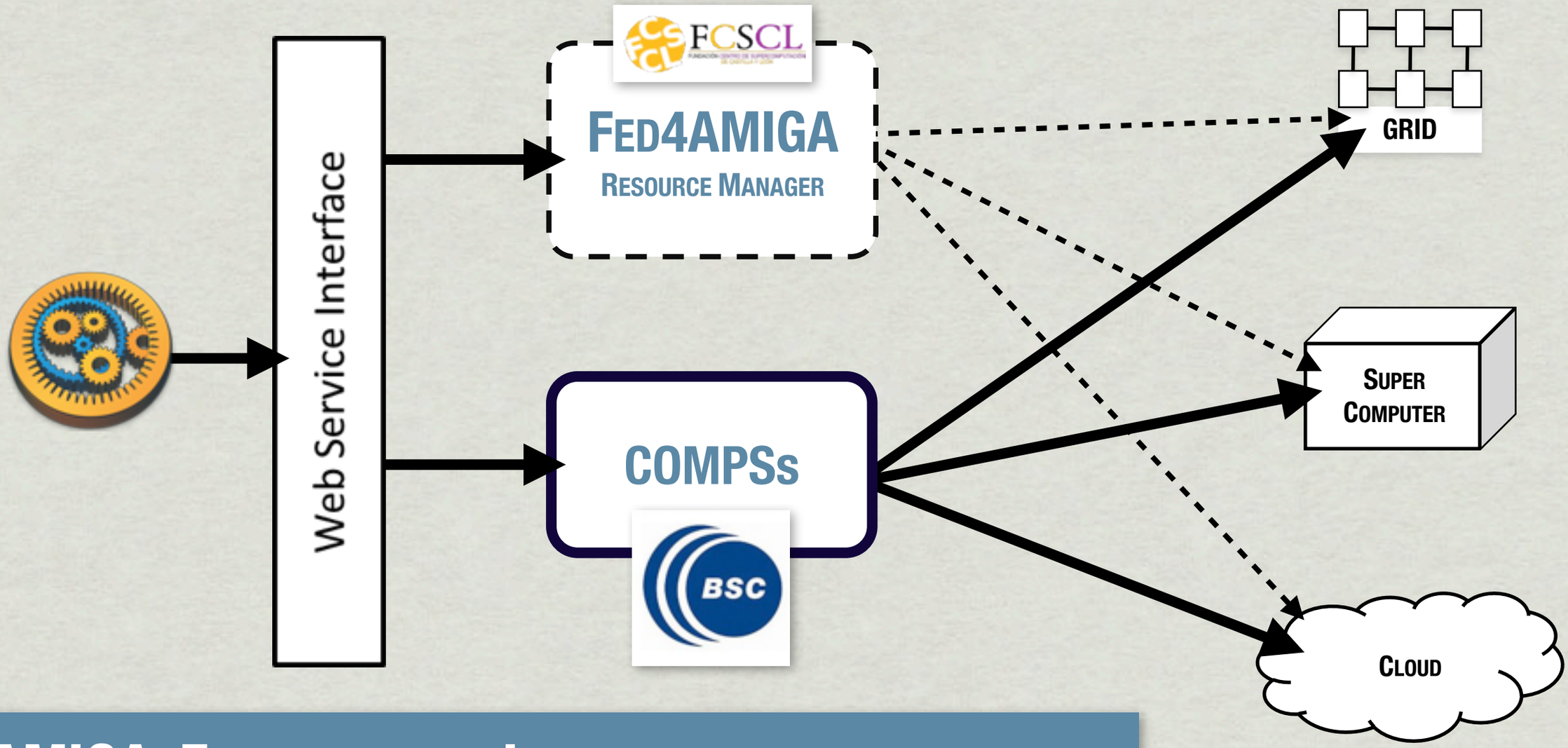
- Technical part, devoted to computing & data federation
  - Heterogeneous computing federation
    - Local computing cluster, grid, cloud computing

**In Partnership With**
**BSC, FCSCL**

- Main Goals
  - Porting the Taverna workflow engine to supercomputing environments
  - Development of an integration layer for automatic workflow deployment

**Direct Relevance to SKA Science Data Processor**

# Infrastructure Federation




## FED4AMIGA: FEDERATION OF INFRASTRUCTURES

- HOW TO INTEGRATE THE INFRASTRUCTURES IN A FEDERATED SYSTEM?
- HOW TO AUTHENTICATE THE USERS?
- HOW TO IMPLEMENT BUSINESS RULES TO DECIDE IN WHICH INFRASTRUCTURE THE TASK SHOULD RUN?

# Wf4Ever + CyberSKA

- Workflows can be used to formally specify processing tasks
  - Specially good when computing tasks exists as services
  - Even better if data can be referenced, instead of sent over the wire

# Wf4Ever + CyberSKA

- Complementary to CyberSKA (infrastructure)
- Wf4Ever places more emphasis on
  - End-user tools
    - process creation
    - data manipulation
    - annotation
    - interdisciplinary algorithm repurposing
  - Long-term Preservation, Quality Assurance

**VERY MUCH SCIENCE-ORIENTED**

# SKA Computing Synergies

# SKA Computing Synergies

- The SKA computing power amounts to being able to **sift through the entire Internet** more than **100 times per day**
  - Citizens can be empowered, through SKA-like tools, to process city wide, regional, or national data for insight
- Intelligent sensor networks can provide tools for better, instantaneous, resource planning

**In Line With H2020 Priorities**

# Conclusions

# Conclusions

- The SKA is the proverbial e-Science instrument
- Workflows can be used for both machine-readable, formal process description, and human-readable scientific tool development
- Federated, transparent workflow computing
- **Wf4Ever** & **AMIGA4GAS** are **complementary** to CyberSKA, SKA.TEL.SDP work

# Conclusions

- It is a **long road towards the SKA**, but we have to **get involved** in this problems **now**

# Thank you!

Julián Garrido
José Enrique Ruiz del Mazo
Susana Sánchez Expósito
Juan de Dios Santander-Vela <jdsant@iaa.es>
Lourdes Verdes-Montenegro