## SKA-Link kickoff

## Combining knowledge to pioneer Big-Data solutions for SKA Data Centres

Instituto de Astrofísica de Andalucía (CSIC) 3rd and 4th April 2017





## Cambrige team (SKA-LINK)







Paul Alexander (SDPc LEAD)

Rosie Bolton (SDP Project Scientist, AENEAS WP3 joint lead and SKAO Regional Centre Project Scientist)

Bojan Nikolic (SDP Project Engineer)







The University of Manchester

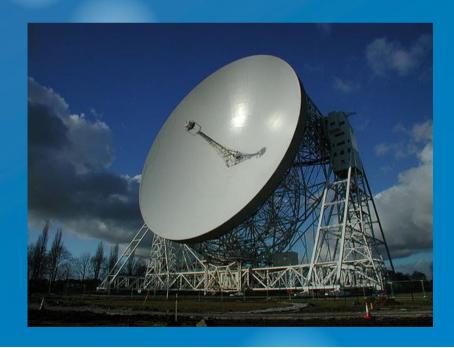


# SKA-link Meeting

University of Manchester/
Jodrell Bank Centre for Astrophysics

Chris Skipper 3rd April 2017







## People



## Anna Scaife

PIP.IMG consortium, based at Manchester

Interested in: radio imaging/interferometry, magnetism, galaxy clusters



## Chris Skipper

PIP.IMG consortium, based at Jodrell Bank

Interested in: Software, GPUs, algorithms, SQL, gridding, w-projection





The University

## Jodrell Bank Observatory

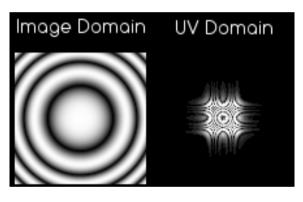




## SDP Imaging Pipeline (PIP.IMG)



Prototyping: Gridding, W- & A-projection, fast imaging, GPU code (CUDA), anti-aliasing, cleaning



w-Kernel Creation

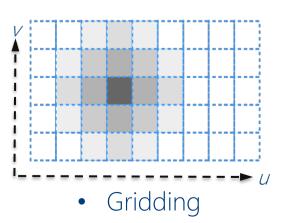
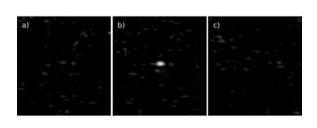


Image Domain UV Domain

• AA-Kernel Creation



Fast Imaging

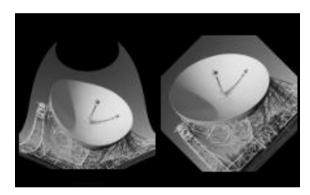
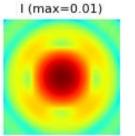


Image-plane Reprojection



I<-V (max=-15.43)

A-projection

## Thank You!

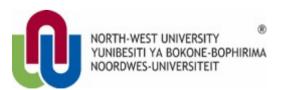






from big data to big ideas

SKA Link / SDP DELIV: University of Cape Town Rob Simmonds
IDIA Associate Director
Professor, Dept. Computer Science, UCT.











## UCT people involved in data delivery activities

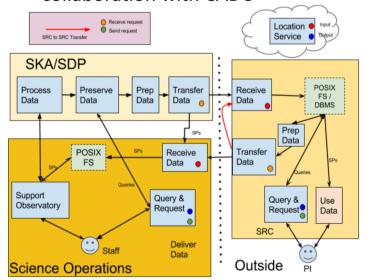
- Rob Simmonds (me)
  - Professor Computer Science UCT and Associate Director of IDIA
  - Lead of SDP DELIV (data delivery) work package
  - Previously CTO of WestGrid, Compute Canada
- David Aikema
  - IDIA Senior Data Scientist
- Adrianna Pinska
  - IDIA Visualization Specialist
- Stefan Coetzee
  - Cloud systems integrator / maintainer
- Joe Bochenek
  - IDIA Senior Data Scientist
- Russ Taylor Director of IDIA
- Brad Frank DOME collaboration Chief Scientist
- 4 more IDIA developer positions being advertised



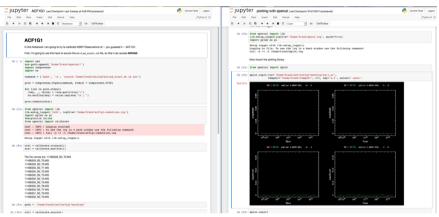


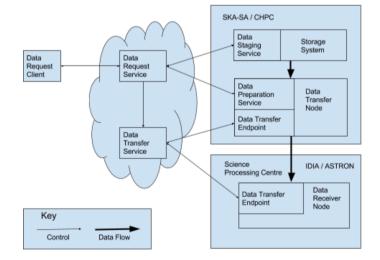


- SDP delivery architecture design with DELIV team
- Architecture prototyping includes:
  - MeerKAT data delivery from SKA-SA to IDIA and ASTRON (DOME collaboration)
  - MeerKAT query service in prototyping in collaboration with CADC





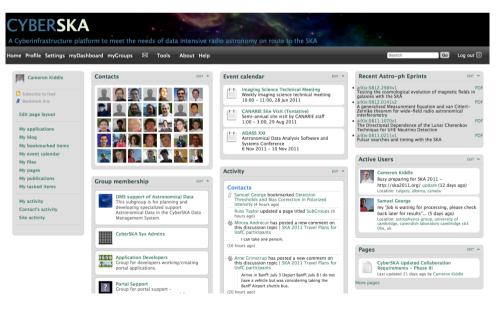


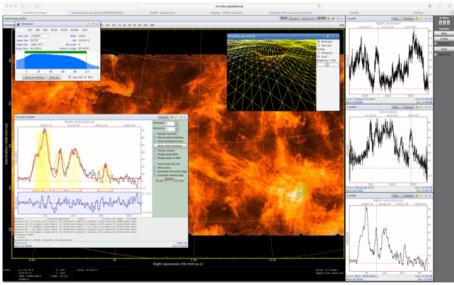




## CyberSKA portal

- CyberSKA portal
  - Radio Astronomy Gateway with over 700 users
  - Federated architecture enables globally distributed portal instances to share namespace information, find data, access data and transfer data between instances
  - Remote visualisation of image cubes from any federated site







## SKA-Link meeting objectives

- We offer:
  - SDP DELIV knowledge
  - MeerKAT focused collaboration
  - Existing radio astronomy portal and data delivery tools
- We would like:
  - Increased collaboration with European partners



Netherlands Institute for Radio Astronomy

# SKA-Link Activities at ASTRON Michael Wise Head, ASTRON Astronomy Group **SKA-Link Kickoff Meeting** Instituto de Astrofísica de Andalucía Granada, ES April 03, 2017 ASTRON is part of the Netherlands Organisation for Scientific Research (NWO)







#### • International LOFAR Telescope

- Operations for data collection, processing, and distribution to science community
- -Design and development of advanced processing algorithms and pipelines
- -Design and operation of the LOFAR Long-Term Archive (LTA) currently housing 28 PBytes

#### WSRT and APERTIF

- Operations of the Westerbork Synthesis Radio Telescope
- -Design and construction of APERTIF phased-array feed to upgrade survey capabilities
- -Design and operation of the APERTIF Long-Term Archive (ALTA)

### • SKA design consortia

- Contributing to LFAA, MFAA, and SDP consortia
- -Design and prototyping activities

#### ASTRON / IBM DOME collaboration

- Collaboration between ASTRON and IBM (20 M€ investment)
- -Development of new technologies and algorithms to enable SKA science



## NWO Related Projects



#### AENEAS

- -H2020 project to produce a design study for a distributed, European SKA Regional Centre
- Work packages on data transport, computing, data access, and user analysis

#### • SKA (Science, SDP, and DELIV)

- -ASTRON is contributing to both SDP and DELIV work packages consortium
- Significant participation by staff astronomers in SKA science teams

#### • ASTERICS

- -H2020 project to address common challenges shared by the various Astronomy ESFRI facilities
- -Facilities include SKA, CTA, KM3NeT, and the E-ELT
- Clear synergies with SKA computing, data management, and user access challenges

#### • DOME / SKA-SA / ASTRON data centre prototype

- Collaboration between ASTRON, IBM-DOME, SKA-SA, and IDIA
- Prototype distributed data management and processing
- -Prototype porting user processing strategies (MeerKAT, Apertif, LOFAR)
- -Support local science analysis with MeerKAT, Apertif, and LOFAR data



# Challenges of Interest



- Exascale data distribution, management, and processing
- Federated distributed data collections and processing
- Distributed user support models
- Algorithms, pipelines, and virtualization
- Standardization, interoperability, and the VO
- User data access and customized analysis
- Advanced analytics, machine learning, and deep learning



## NWO Points of Contact



#### Michael Wise

- -Head of ASTRON Astronomy Group
- -PI AENEAS H2020 project on European SKA Regional Centre

#### • Rob van der Meer

- -Program Officer European Collaboration at ASTRON
- Program Manager for AENEAS project

#### Hanno Holties

- -System engineer, ASTRON Radio Observatory
- Project lead for LOFAR LTA, Apertif LTA (ALTA)
- -Member of AENEAS WP3 on Computing and Storage

#### Yan Grange

- -Developer in the ASTRON Data Science Group
- -Background in astronomy, member of ASTRON computing group since 2012
- Member of SKA SDP consortium (DATA and DELIV work packages)
- Contributing to AENEAS WP5 on Data Access, Dome-SA-NL data centre prototyping project
- -Experience with LOFAR LTA, storage optimisation, workload characterisation (SKA and LOFAR)

## **Amy Krause**

Data Architect
The University of Edinburgh
Edinburgh Parallel Computing Centre (EPCC)

- Research software engineer and architect
- Data intensive architecture and design
- Cloud technologies
- User interfaces & mobile apps

- Many projects in the area of distributed dataintensive applications
- Collaboration with Informatics: Rosa Filgueira (now BGS), Malcolm Atkinson
  - Implemented dispel4py, a Python library for distributed stream-based workflows
- More recently focusing on cloud technologies
  - Commercial projects in large-scale data analytics: Genotype analysis, TV customer data analysis, letting agency data infrastructure
- Looking to support researchers by writing useful software and make their life easier

## Malcolm Atkinson

University of Edinburgh
Prof of e-Science
School of Informatics
Leading Data-Intensive Research Group

70's **Data** ⇔ Applications ⇔ People ⇔ Databases

CAMBRIDGE RANGOON NORWICH EDINBURGH

80's Data 

⇔ Applications 

⇔ Organisations 

⇔ Software Engineering

90's Data 

⇔ Applications 

⇔ Organisations 

⇔ Sun 

⇔ Java

00's Data 

⇔ Applications 

⇔ Nations 

⇔ eScience 

⇔ Distributed Systems

10's Data 

⇔ Applications 
⇔ eScience 
⇔ Federations and Data Diplomacy

EDINBURGH PHILADELPHIA GLASGOW INRIA 02

GLASGOW MOUNTAIN VIEW GLASGOW

> GLASGOW E-SCIENCE EDINBURGH

E-SCIENCE EDINBURGH

Persistent goal: Enable easy, effective and ethical use of your data to meet your challenges

## MALCOLM'S CONTRIBUTION

#### TO SKA-LINK THINKING

- Systems architecture based on experiencing change
- Organisational strategies based on many campaigns
- Recognition that people are key
  - Harnessing multiple viewpoints with conceptual frameworks
  - Improving their productivity by sustaining the value of their work
  - Accelerating induction by removing hurdles
  - Respecting identity issues
- Frameworks that facilitate collaboration

I AM NOT ADEPT
AT THE TECHNICAL
DETAILS OF
SPECIFIC SYSTEMS
AND VIEW THEM
AS PART OF THE
EVOLVING
DIGITAL
ECOSYSTEM IN
WHICH DATA-DRIVEN
RESEARCH MUST BE
NIMBLE TO THRIVE

# MALCOLM'S WANTS TO TAKE HOME FROM SKA-LINK MEETING

- A better understanding of the socio-technical challenges
- Opportunities to try to collaborate to meet such challenges
- Plans to muster or bid for resources to do the necessary R&D
- Expectation that this will advance our capacity to exploit data riches in many fields

## **Malcolm Atkinson**

Chat with me I'm interested in everything

E.G. Walking





Why are we so fortunate to live on a planet with such features?



# Introduction to SZTAKI and LPDS projects

Prof. Peter Kacsuk

Head of Laboratory of Parallel and Distributed Systems





## **About SZTAKI**

Hungarian Academy of Sciences Institute for Computer Science and Control

- Established in 1964
- EU Centre of Excellence in IT, Computer Science and Control
- Basic and applied research
- Contract-based R&Đ&I activity mainly on complex systems, turnkey realizations
- Transferring up-to-date results to industry and universities

- Basic research
  - Computer science
  - Systems- and control theory
  - Engineering and business intelligence
  - Machine perception and humancomputer interaction
- Applied research and innovation
  - Vehicles and transportation systems
  - Production informatics and logistics
  - Energy and sustainable development
  - Security and surveillance
  - Networking systems and services, cloud / distributed computing

#### **Key figures**

- Budget
  - 12-13 MEUR/year
  - ~30% basic funding
- Staff
  - 280
  - 67% scientific
- Fraunhofer ProjectCenter
- W3C member
- ERCIM member
- Contributor to the European Grid Infrastructure



## About the research lab LPDS

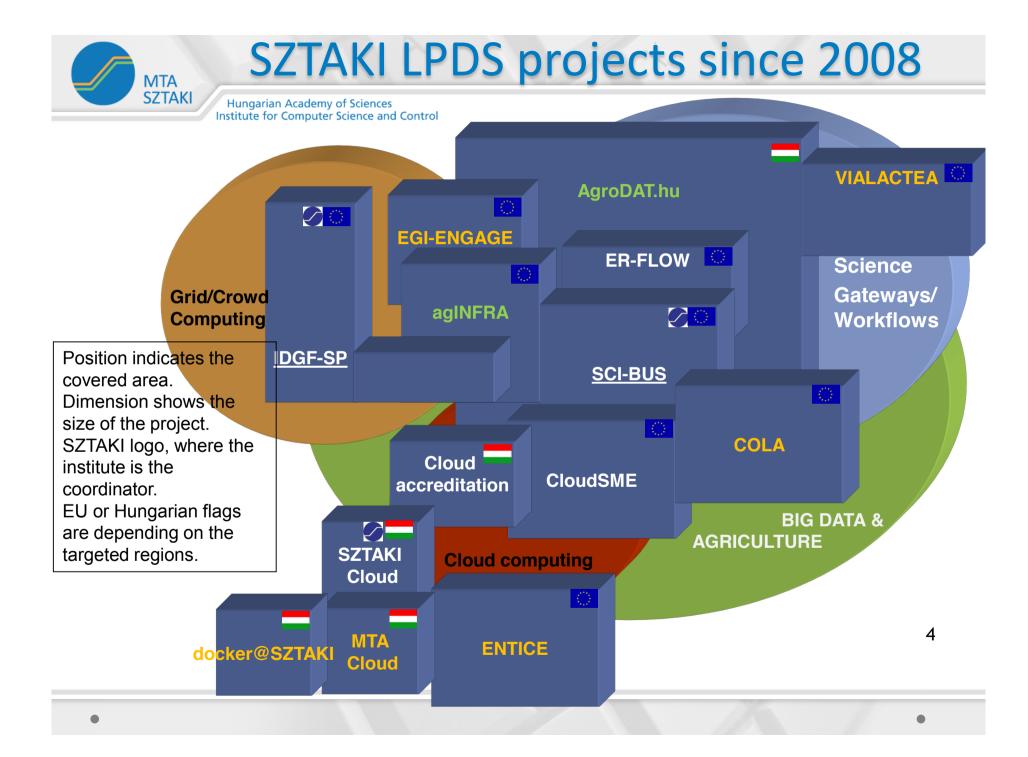
Hungarian Academy of Sciences
Institute for Computer Science and Control

- Research division of MTA SZTAKI from 1998
- Head: Peter Kacsuk, Prof.
  - editor in chief: Journal of Grid Computing (Springer)
  - coordinator of four e-Infrastructure projects in the EU 7th Framework Programme (FP7)
- Deputy head: Robert Lovas, PhD
  - coordinator of two projects in EU FP7
  - international liaison (European Grid Infrastructure)
  - secretary of International Desktop Grid Federation (Dutch foundation)
- 12 research fellows (full/part time)
- Foundation member
  - Hungarian National Grid Initiative (NGI\_HU)
  - International Desktop Grid Federation (IDGF)
  - OpenNebula (ONE) User Group, Hungary
- Participant and/or coordinator in European and national Grid & Cloud research, e-infrastructure, and educational projects (from 2000)
- By far the most successful research lab in the region in the field ICT / e-Infrastructures in terms of the number of coordinated FP7 EU projects









## MTA SZTAKI

## Our interest in SKA-Link

Hungarian Academy of Sciences Institute for Computer Science and Control

- Since the SCI-BUS project SZTAKI is engaged in supporting scientific user communities by gateway technology:
  - VIALACTEA science gateway
  - o EGI-ENGAGE:
    - DARIAH science gateway
    - Long Tail of Science gateway
  - MTA Cloud project
- Extending the gateway technology to support the easy use of cloud and docker technology
- During this meeting I will present our gateway technology and its support for cloud and docker
- Colleagues who will participate:
  - Zoltan Farkas and Istvan Márton (Gateway), Jozsef Kovács and Enikő Nagy (Occopus/cloud), Attila Farkas (docker)

## Rosa Filgueira

Senior Data Scientist British Geological Survey (BGS)



#### Background:

- PhD Computer Science HPC research Madrid/Carlos III
- 5 years as a Postdoc Data Intensive Research -University of Edinburgh/DIR group

#### Currently:

- Data sciences activities across Geoscience domains:
  - Data gathering, cleaning, filtering, analysis
  - Parallelization/optimization of applications
  - Promoting scientific workflows, data-frameworks, containers and reproducibility tools, etc.
- Writing research proposals
- Involved in many research projects H2020, NERC, BGS
- Collaborations with EPCC, University of Edinburgh, etc.
- Research activities in BGS-Informatics

# Contributions to Ska-Link thinking

Delivering easy-to-use frameworks to empower data-driven scientists

- Develop methods in their familiar environment
- Deploy methods in production systems without changing them
- Improve methods without introducing distractions by new technologies
- Enable sharing of methods and data

#### How ?

- New advanced abstract frameworks that adapt automatically
  - to different data scales and generation rates
  - to different computational platforms

need to be
easy to understood
platformindependent

# To take home from Ska-link meeting

- Opportunities to collaborate in the future
- Set us on a path to pioneer globally distributed use of data on an unprecedented scale.
- Plans for future grants and research proposals



## **SKA-link** kick-off

Jens Krüger

High Performance and Cloud Computing Group
Zentrum für Datenverarbeitung Eberhard Karls Universität Tübingen

> 03.04.17 Granada



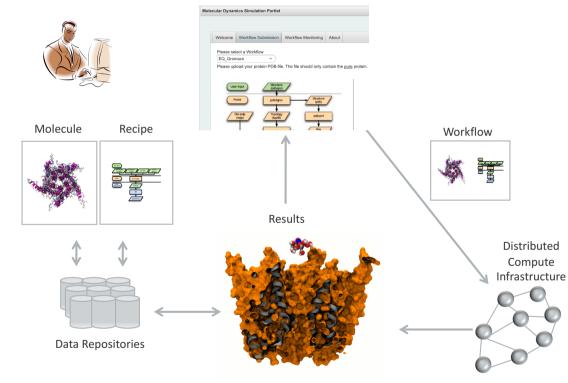
- Jens Krüger
  - Group Leader HPC and Cloud Computing
  - bwHPC competence center for bioinformatics and astronomy
  - Chemist, Bioinformatician, Simulant
- Felix Bartusch
  - PhD Student
  - CiTAR: Reproducible science, containerization, workflows
- Max Hanussek
  - PhD Student
  - de.NBI Cloud: Virtualization environment for bioinformatics
- Fabian Wannenmacher
  - Master Student
  - Networking among OpenStack services
- Volker Lutz
  - Technician
  - HPC infrastructure, storage, software
- Werner Dilling
  - Head of the 'Zentrale Systeme' division





- Core developer and maintainer of MoSGrid
- Operating a science gateway for 6+ years on production level gives some insight with respect to:
  - User management
  - Authentication
  - Community management
  - Keeping software in sync
  - Workflows
  - Data staging
  - Metadata annotation
  - Distributed storage
  - Middlewares

- ...







#### **Networking**

Personal contact Discussion New trends

. . .

#### **Technologies**

Science gateways Micro services Workflows Containerization

. . .

#### **Sharing expertise**

Portal operation
Community management
Storage concept
Identity Management

. . .

#### Reproducible science

Annotation Standards Semantic search

. . .

#### **Outreach**

Conferences Workshops Dissemination

. . .

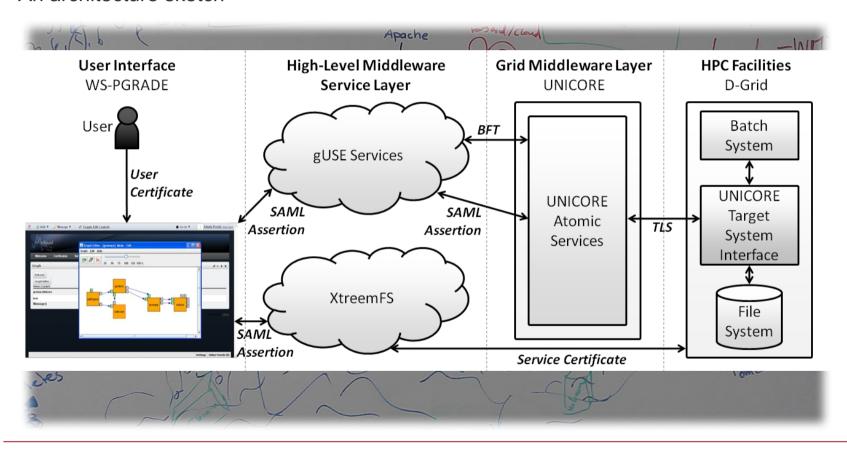
#### **Future projects**

EU calls
Research visits
Exchange programs

. . .



An architecture sketch



## KNMI Team @ RDWD



## R&D Observation and Data Technologies Prof. Albert Klein Tank

albert.klein.tank@knmi.nl



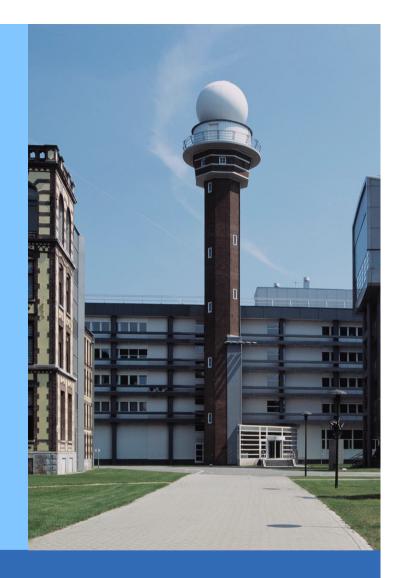
<u>Alessandro Spinuso</u> - R&D Data-intensive VREs & Provenance spinuso@knmi.nl

Wim Som de Cerff - R&D Climate Services, Agile Methods Expert wim.som.de.cerff@knmi.nl

Andre Pagani - R&D Data Scientist - DataLab andrea.pagani@knmi.nl

Maarten Plieger - GIS and OGC Engineer maarten.plieger@knmi.nl

Andrej Mihajlovski - Metadata Engineer andrej.mihajlovski@knmi.nl



### **Group Mission:**

"Developing an optimal infrastructure for the collection, processing, accessing and visualisation of geophysical information"

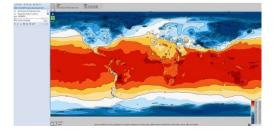


### Relevant Projects and R&D

- Climate4Impact Climate data access, online processing and visualisation <a href="http://climate4impact.knmi.nl">http://climate4impact.knmi.nl</a> (OGC standards, Metadata Vocabularies, Federated Archives, VREs)
- **EPOS-VERCE** Earthquake simulations and Misfit analysis workflows <a href="http://porta.verece.eu">http://porta.verece.eu</a> (HPC & data-intensive workflows, provenance, VREs,)
- R&D S-ProvFlow Data-Intensive provenance. Capturing and exploitation <a href="https://github.com/KNMI/s-provenance">https://github.com/KNMI/s-provenance</a> (Reproducibility as a Service)
- R&D DataLab KNMI "hub" for data-driven innovation (Facilities Data Science Infrastructure and knowledge uptake)



C41



**EPOS-VERCE** 



R&D S-ProvFlow

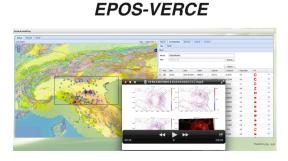


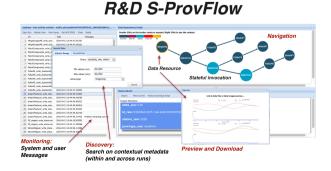


#### Role in The SKA-Link

- Share personal experience on the process for the realisation of a usable community-driven Virtual Research Environment (VERCE)
- Challenges of implementing HPC and data-intensive Applications as a Service
- Role and Integration of Reproducibility and Provenance Services in Modern Online Scientific Tools (RaaS).

# C41





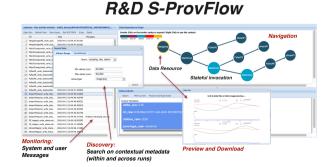


## Take Home.. Knowledge and Network

- Inspiring group of experts from academy and public institutions (creating bridges)
- Discuss the role of users in the automation of methods, especially when scale surpass cognitive and technical possibilities. (finding the right balance)
- KNMI is starting evaluating a program for Space Weather Services (new projects) bert.van.den.oord@knmi.nl
- People I want to work with. Great recent experience. Result oriented. Extended network (recipe for innovation)

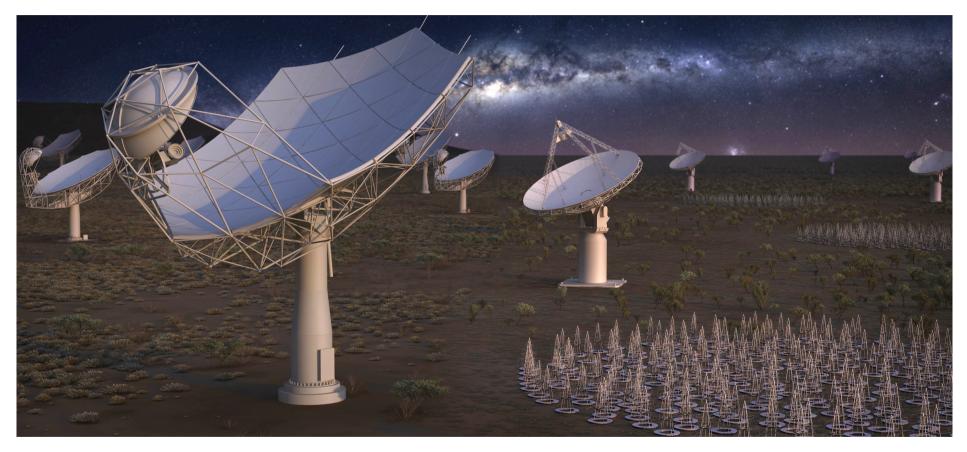
And the first





## **SKAO Introductions**Who are we - what do we do?





#### SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

**Dr Antonio Chrysostomou**Head of Science Operations Planning







## **Dr Antonio Chrysostomou**

#### Head of Science Operations Planning at SKA

planning for the operation of the SKA observatory across three sites

#### Chair of the SKA Regional Centres Coordination Group

 ensuring that the SKA has a viable platform for its users to be able to access and analyse their SKA data products

#### Travel and Hiking

documenting those activities with

my camera!







Exploring the Universe with the world's largest radio telescope



## **Dr Rosie Bolton**

SKAO Regional Centre Project Scientist

Member of the SKA Regional Centre Coordination

Group

Leading task to develop the high level requirements for SRCs and the Alliance of SRCs

- what must each SRC be able to pledge in order to be accredited and how will this be monitored?
- how will the alliance of SRCs function as a unified body, in spite of the diverse SRC landscape?
- to where do the responsibilities of the SKA Observatory extend?
- how will the picture evolve over the lifetime of the SKA?



CERN and the WLCG, LSST, etc



## SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope



www.skatelescope.org



Russ Taylor

SKA Research Chair University of Cape Town & University of the Western Cape

Director

Inter-University Institute for Data Intensive Astronomy













**Prof. Russ Taylor** Director



**Prof. Rob Simmonds Associate Director** 



Dr. Sarah Blyth



Asoc. Prof. Michelle Kuttel



Dr. Michelle Cluver UWC



**Prof. Roy Maartens** UWC



**Prof. Claude Carignan** UCT



Prof. Romeel Davé UWC



Dr. Bradley Frank UCT



Prof. Renée Kraan-Korteweg UCT



**Prof. Mario Santos** UWC



**Prof. Patrick Woudt** UCT



## Developments relevant to SKA-link

- SKA Precursor Regional Science and Data Centre
- User-centric, scalable and extensible solutions for data-to-science for MeerKAT Large Survey Projects
- MeerKAT Tier 2 processing facility
- Multi-wavelength data fusion initiatives
- Cloud-based Visualization of remote big data sets
- Cyber-infrastructure platforms for distributed teams working on data intensive radio astronomy research
- African Data Intensive Research Cloud



## Take-aways from SKA-link meeting

- Advance a joint vision of a distributed network of centres to enable science in the data-intensive paradigm with a user-facing philosophy
- Discussion of use cases and priorities
- Engagement of stakeholders
- Communication toward alignment of development plans and programs
- Identification of potential areas of collaboration



from big data to big ideas

www.idia.ac.za

















## IAA Asteroseismology team

Prof. Rafael Garrido







## **Activities**

#### Economic Resources from two research programs

#### **Regional Funds (CoRot)**

CoRot data exploitation.

Data Mining and Databases.

Support from the ground from Sierra Nevada Observatory (SVO).

#### **National plan Funds (PLATO)**

Is our technical responsibility to develop the Main Electronic Unit (MEU). Co-PI: Julio Rodriguez.

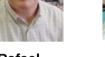
Is our scientific responsibility to develop tools to transform the raw data into physical information.





















Rafael Garrido

Javier Pascual-Granado

Sebastiano **DeFrancisis** 

**Mariel Lares** 

José Ramón Rodón

Guillermo Cortés



**Juan Carlos** Suárez



Antonio García Hernández



**Andrés Moya** 



**Maria Angeles** Mendoza







Julio Rodriguez Beatriz Aparicio Rosario Sanz

Teams from IAA and University of Granada

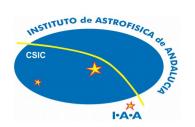






## Impact on the community

- Non standard time series analysis.
- Numerical stellar interior models including rotation.



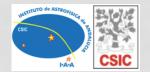




## What can we do for SKA-link What SKA-link can do for us

- Learning new e-Science techniques for a reliable and transparent work that can be applied to the study of stellar interiors through asteroseismology techniques.
  - Knowing first hand what the SKA Regional Centres will be, and how this can be of interest beyond radioastronomers
    - Exploring whether we can, on our side, contribute or support an SRC at the IAA (see talk tomorrow)

## **AMIGA** team in SKA-Link



Analysis of the interstellar Medium of Isolated GAlaxies. <a href="http://amiga.iaa.es">http://amiga.iaa.es</a>



#### Lourdes Verdes-Montenegro (lourdes@iaa.es)

- PI of the group
- Radio-astronomer, expert on HI studies of isolated galaxies and compact groups
- Coordinator of the Spanish participation in SKA
- Science Data Processor (SDP) consortium Board member
- SKA Regional Centre (SRC) coordination group member



#### Julian Garrido (jgs@jaa.es)

- Software engineer. Expertise on IVOA standards and astrophysical workflows development.
- · Project manager of the Spanish participation in SKA
- SDP Member: DATA group, designing the platform for preserving SKA data



#### Susana Sánchez (sse@iaa)

- Software engineer. Expertise on e-Science tools, distributed computing and astrophysical workflows development.
- SDP Member: DELIV group, designing the platform for delivering SKA data



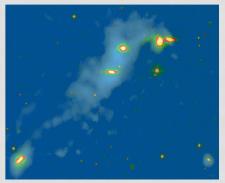
#### Marina Fernández (marina@iaa.es)

- Biochemist and communication specialist
- · Coordination assistant, supporting the Spanish participation in SKA
- LOC of the SKA-Link kick-off meeting

## **AMIGA** team background

## Specialists in galaxy environment

- Multi-wavelength catalogue of ~1000 isolated galaxies
- Study of the Atomic gas (HI) as tracer of the galaxy formation and evolution (radio-astronomical techniques)



Hickson Compact Group 16 Verdes-Montenegro+. 2001

## Contributors to IVOA standards

Multi-wavelength studies require access to heterogeneous archives

 Needs on data interoperability standards (IVOA), but: IVOA standards for radio data missing

## e-Science (tools) developers

Radio data volumes are increasing exponentially

- Tools for extracting scientific information from 3D data
- Migration to Grid and Cloud in preparation for huge data volumes
- Development of AstroTaverna, a plugin for integrating VO services in Taverna workflow management system.



RADAMS: First VO-compliant Radio Astronomical Data Model. It was used to implement the 30m antenna IRAM archive

## AMIGA's role in SKA-Link and its interests

## As members of the SKA community, we are

- Worried about the <u>current crisis of reproducibility</u> in science
- Interested on innovative ways of producing reproducible science
- Decided to trigger the idea that <u>SKA could be the first scientific</u> infrastructure taking the lead of trustable, reproducible science

## As members of the groups designing the

## Science Data Processor

 Contributing to the design of the platforms for preserving and delivering SKA data and metadata -We are interested on <u>technologies</u> for supporting both <u>data-intensive &</u> <u>reproducible science</u>

## SKA Regional Centres

- H2020 AENEAS project participants
- LVM external advisor of the SRC committee group

## **AMIGA's expectations from SKA-Link**

Together with SKA-Link participants, we would like to:

- Share ideas about what means "good" science:
  - There should be just Science, as the one following the Scientific Method
  - Reproducible / Trustable / Verifiable
- Suggest innovative metrics to assess the SKA success
  - Metrics that promote and support the Scientific Method
- Identify innovative technologies to support SKA users:
  - in exploiting huge data volumes
  - in producing Science
- Evaluate the feasibility of adopting these innovative metrics and technologies by SKA and the SRCs

## SKA-Link kickoff

## Combining knowledge to pioneer Big-Data solutions for SKA Data Centres

Instituto de Astrofísica de Andalucía (CSIC) 3rd and 4th April 2017



