South African National Grid :
A blueprint for e-Infrastructure in Sub-Saharan Africa ?

Bruce Becker, Meraka Institute, CSIR
Coordinator, South African National Grid
South African National Research Network

Overview

- … in my own words …
- The goal : e-Science in (South) Africa
- Starting with trust
- Existing e-Infrastructure in South Africa
- SAGrid organisational model and activities
- Interoperability / cooperation
- The future : A blueprint for Africa

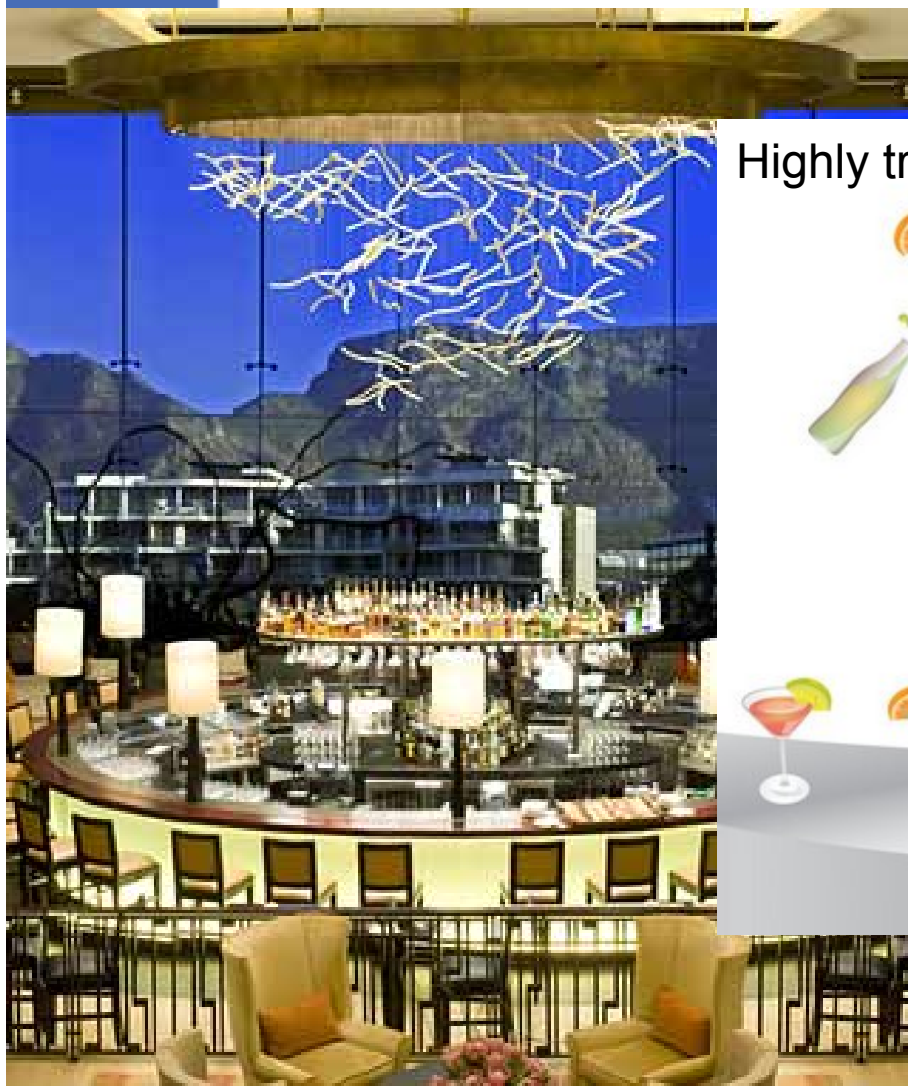# A scenario : this scientist walks into a bar...

What users want –
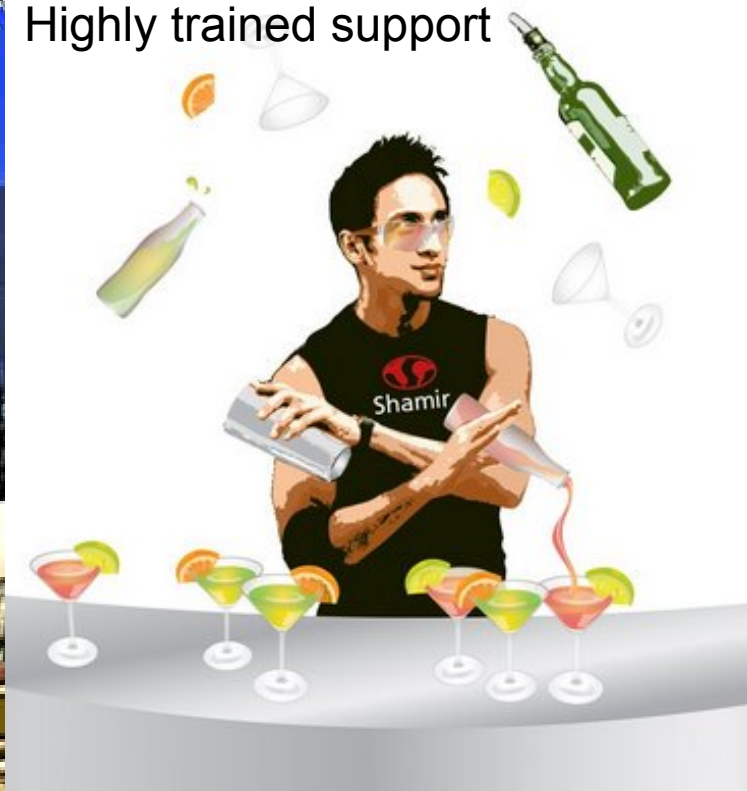# best of all worlds for no extra effort

Some users want highly varied, modular application selection
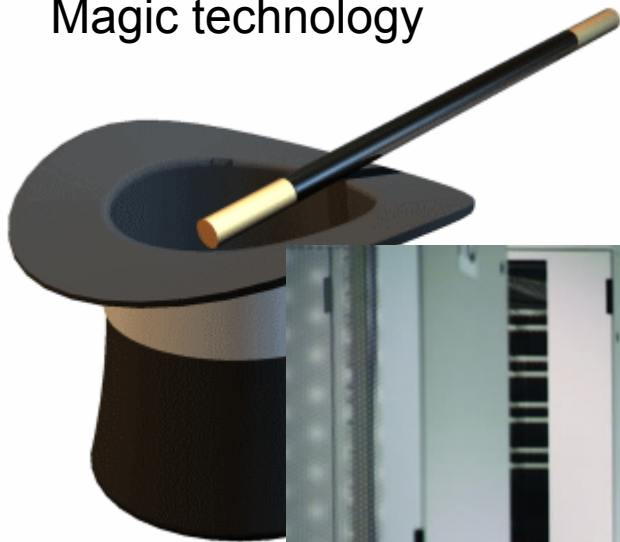
Amazing infrastructure

Highly trained support

Vertically integrated Highly specialised applications

Magic technology

PROFIT !!!
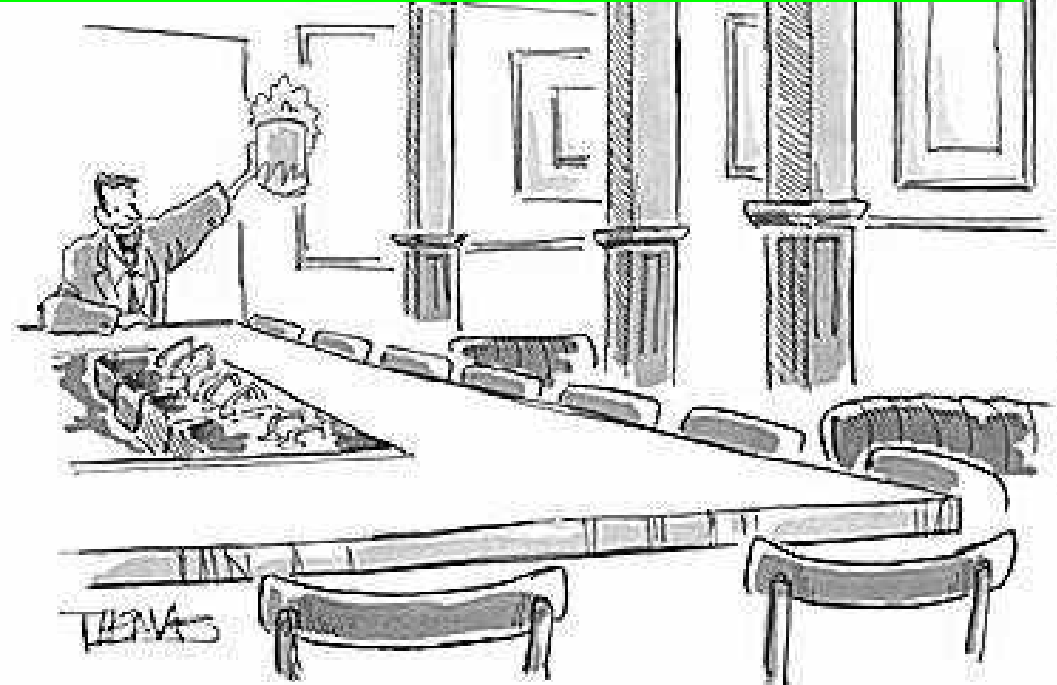
? ? ?

Generic HPC data centre with shared resources

5

Strange experiments on users

Over-ambition ending in disaster

Best laid plans...
potential alternate realities we want to avoid...

Build a great bar, but Nobody comes to the party

"Drinks are on me!"

search ID: tbrn93

6

We have to standardise on something...
not just anything : good, functional, palatable, interoperable
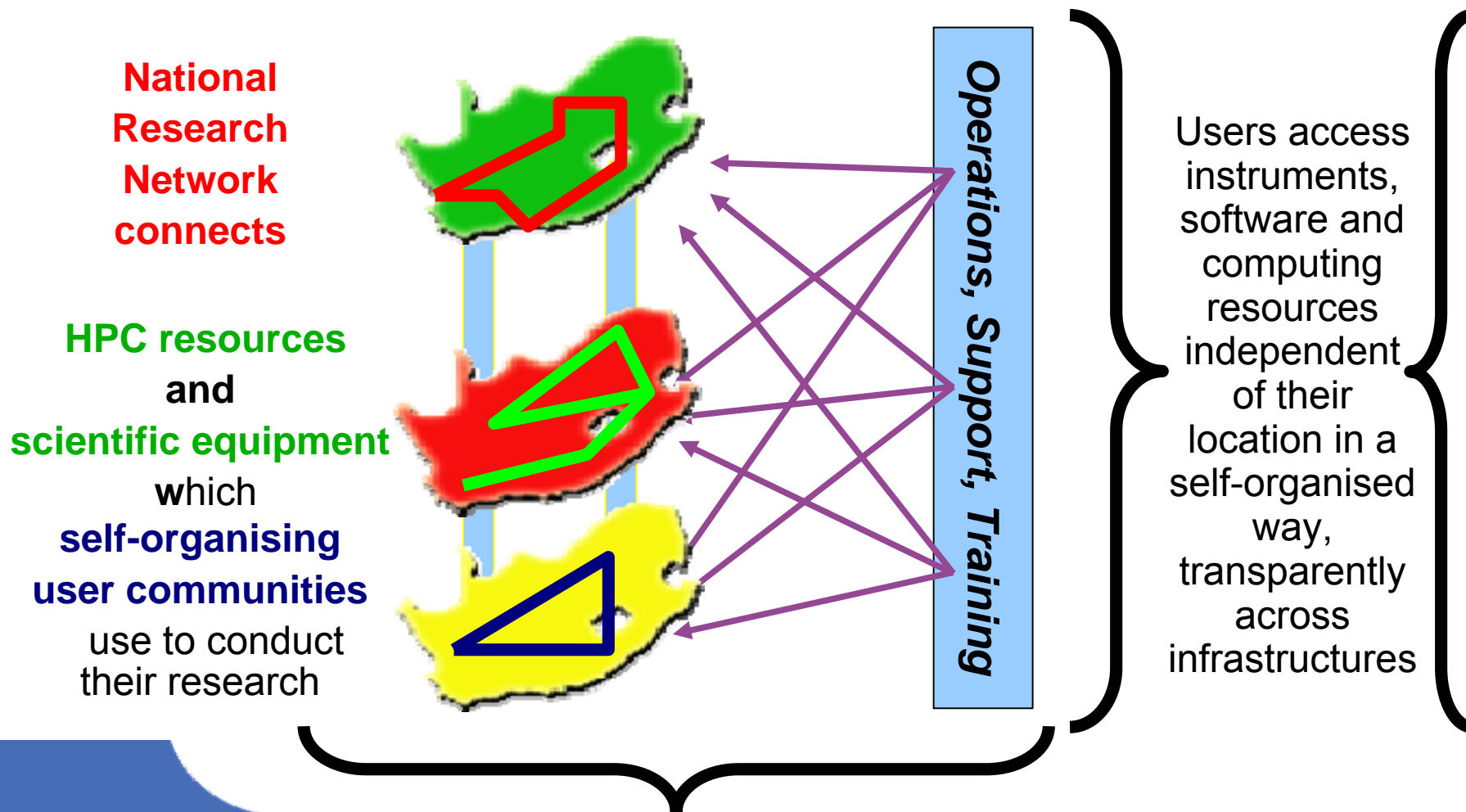## not necessarily perfect...

To talk about infrastructure, we have to standardise on something...
- Technologies
- **Methodologies**

An integrated federal infrastructure to the end-user's benefit - Requires interoperability and collaboration at all levels :

National Research Network connects

HPC resources and scientific equipment **w**hich **self-organising user communities** use to conduct their research

Operations, Support, Training

Users access instruments, software and computing resources independent of their location in a self-organised way, transparently across infrastructures

e-infrastructure

Federal/Joint Management, development, planning and research activities

meraka
INSTITUTE

Let's start from the very beginning : The SAGrid Certificate Authority

Enabling International Collaboration Through Trust : SAGrid Policy Management Authority

- Security and identity infrastructure is crucial to the entire undertaking, yet no PMA exists for Africa
- SAGrid started with trust anchor in INFN CA (similar process to participants in EUMedGrid)
  - http://security.infn.it/CA/en/RA
- SAGrid CA under development – first presentation to EUGridPMA October 2009 (Berlin)
- Can act as trust centre for the region, evolve into AfricaGridPMA

10
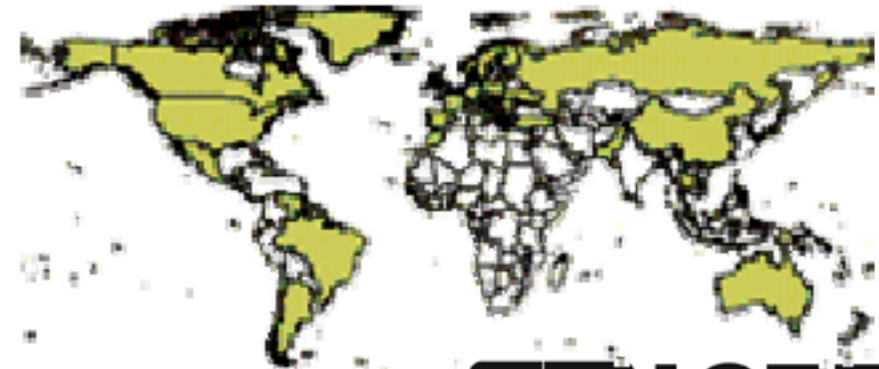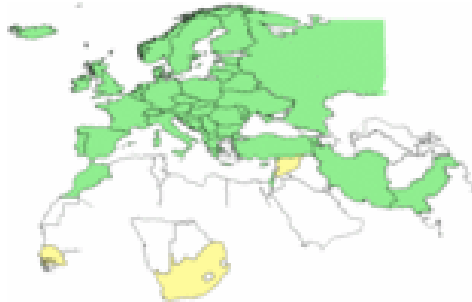
**The "global" network coverage means high-speed access...
to computing resources, data, instruments – and other researchers !**
See http://www.feast-project.org/ ; soon AfricaConnect project (Dante)
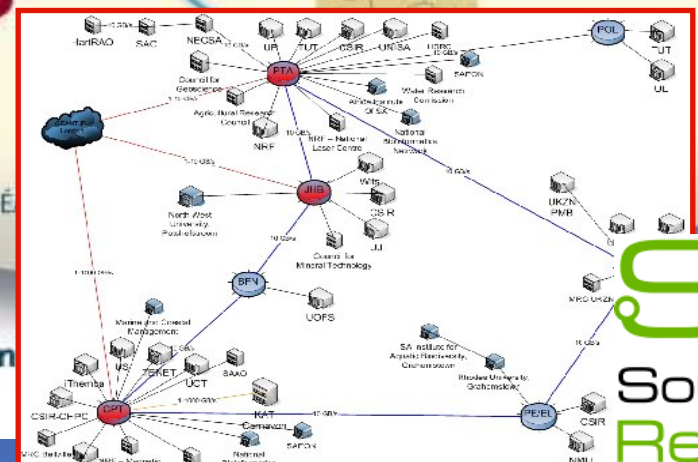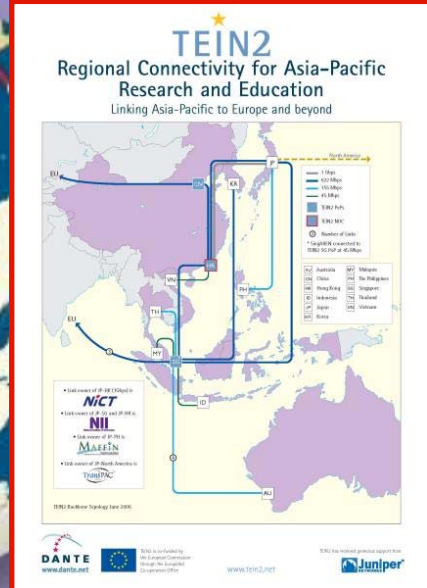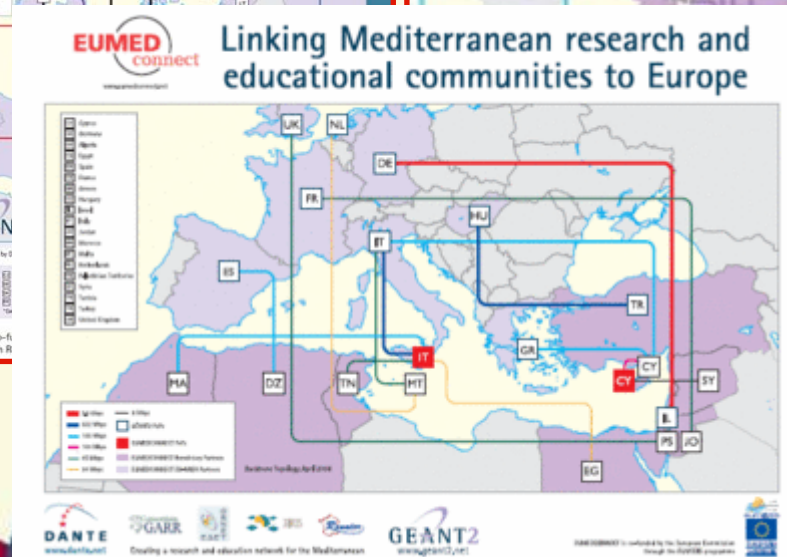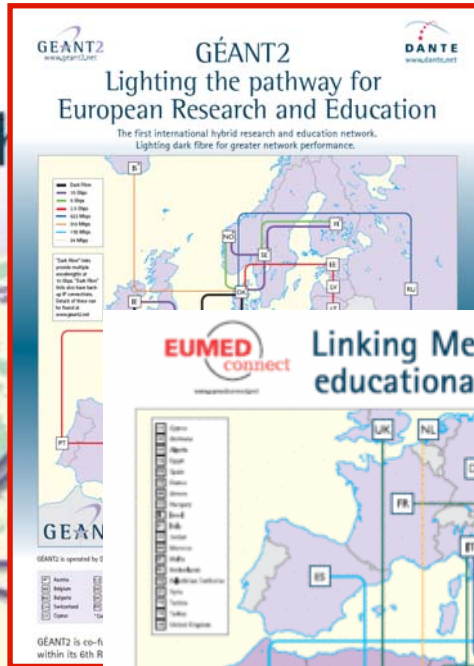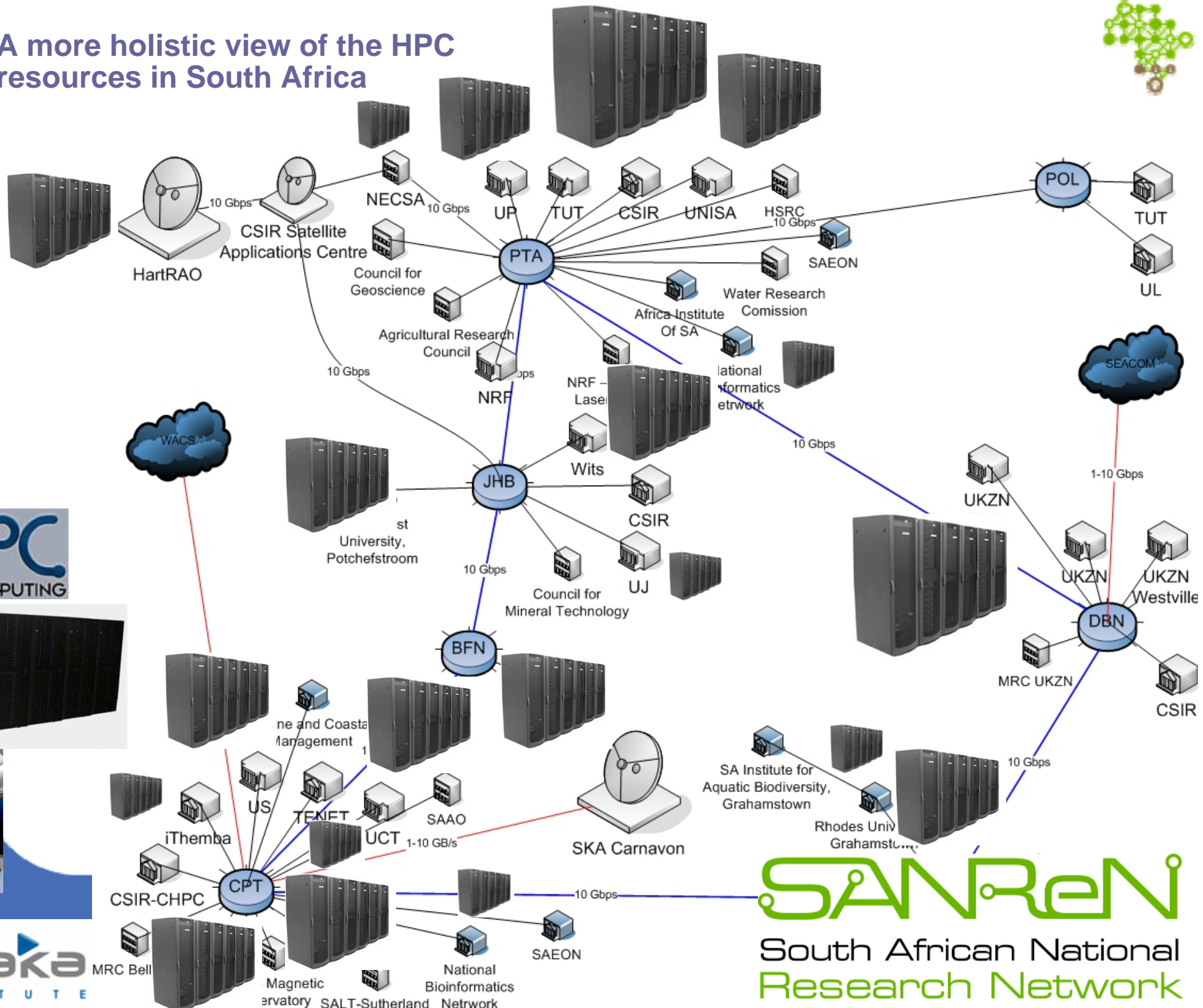
South Africa's investment in e-Infrastructure

- The Centre for High-Performance Computing :
  - Premier HPC site in the country, provides more resources than any given site
  - Situated in Cape Town, managed by Meraka
  - Designed and managed to tackle exceedingly large problems, perhaps requiring specialised hardware
  - Used on a project basis, through major "flagship" projects and collaborative "consortium" projects

BlueGene/p
PPC arch
~14 Tflops Peak

IBM e1350 cluster
AMD Opteron x86_64
~3.3 Tflops Peak

Not shown : CHPC phase 13

Figures from www.chpc.ac.za

A more holistic view of the HPC resources in South Africa

7 functional sites
O(1000) cores
O(100 TB) permanent storage

Phase 1 sites (JRU)

"interested/under discussion"

Core services

External infrastructure
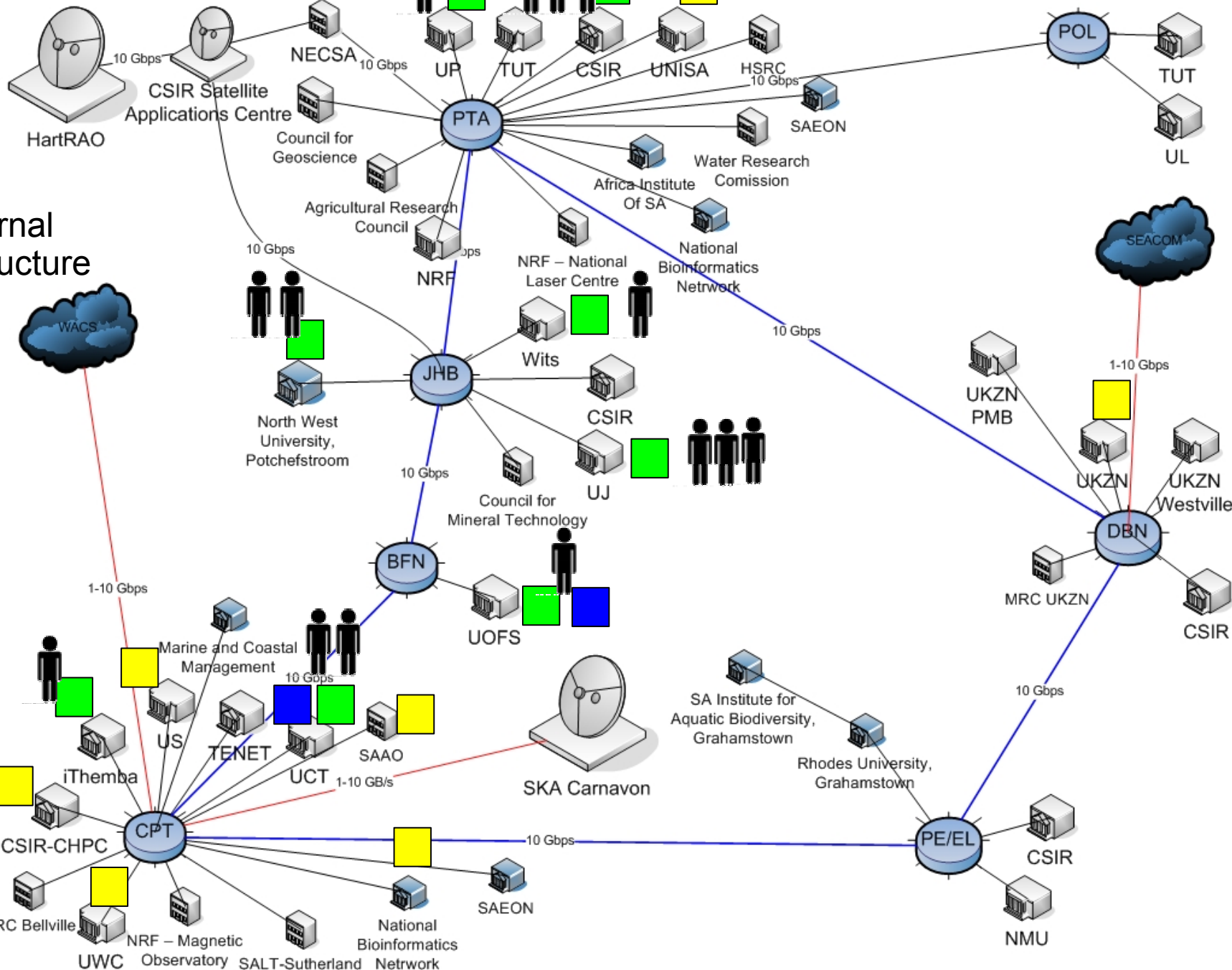
SAGrid : An example of an integrated infrastructure initiative

- A coordinated effort to address all aspects of e-Infrastructure deployment, management and usage, modelled on experiences of similar projects worldwide
- Services provided by SAGrid to **users and scientific collaborations :**
  - Identity and security infrastructure
  - User Training
  - Application and User Support
- Services undertaken by **SAGrid resource providers**
  - Compute resource deployment
  - Operations
- **Overall coordination and coordination** with other infrastructures
  - Interoperability (EGI and other international infrastructure)
  - Network coordination (TENET, SANReN)
  - VO-level collaboration and new collaborations on the grid

16

Frameworks

Management
Support
Training
Application Deployment

SAGrid Joint Research Unit

- SAGrid is a federation :
  - A technical collaboration to enable scientific collaboration, members of a Joint Research Unit
  - Adhere to principles and themes which they find self-evident
  - Agree to pool, share and jointly manage resources
  - Provides a framework for participation to collaborative projects aimed at deploying, maintaining and improving e-Infrastructure
  - Designates a "flagbearing institute" to represent the federation (currently Meraka)
  - An open document, inspired by very successful Italian Grid Initiative (IGI)
  - Precursor to a more formal National Grid Initiative (NGI) – along the lines of EGI model.
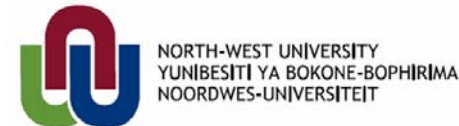
Support : South African National Grid Operations Portal

Discussions under way for integration of ops.sagrid into GGUS and GOCDB → new : AfricaRoc recently created !

Training

- User communities need to be provided with relevant and easy-to-use services
- Coherent deployment and operations requires that various communities are trained in parallel :
  - Technical specialists, site administrators
  - Users and application developers
- An entire infrastructure is required in order to allow this training : a **t-Infrastructure !**
- T-Infrastructure serves as :
  - Immediate access point for user communities in areas with no grid infrastructure
  - Model for creation of new SAGrid VO's
  - Incubation and testing ground for prototype deployment

20

The EPIKH Project : www.epikh.eu

"Exchange Programme to advance e-Infrastructure Know-How" (EPIKH)

EU FP7- Marie Curie Actions – People - International Research Staff Exchange Scheme (IRSES)

Consortium "numbers":

23 partners;

18 countries;

4 continents (**Africa**, Latin America, Asia, Europe);

115 persons involved;

>650 researcher-months;

>500 secondments;

Duration: March, 1, 2009 – February, 28, 2013 (48 months)

EC contribution: 1,188,000 €

The EPIKH Partners
(coordinator : Roberto Barbera. INFN (CT) and Consorzio COMETA)

No. 1	COMETA	No. 2	MTASZTAKI	No. 3	CIEMAT	No. 4	UTFMS
No. 5	ERI	No. 6	CSIR	No. 7	IHEP	No. 8	CESNET
No. 9	CEFET-RJ	No. 10	ISI	No. 11	JUNET	No. 12	CERIST
No. 13	UTIC	No. 14	PKU	No. 15	SCS-TAU	No. 16	CNRST
No. 17	BUAA	No. 18	CNRS	No. 19	UFRJ	No. 20	UNAM
No. 21	INFN	No. 22	GRNET	No. 23	UCT		

Partners in common with EUMEDGRID-Support

22

Interoperability

- Service-level infrastructure interoperability crucial to collaboration
- SAGrid provides a near clone of Italian Grid Infrastructure middleware release for all services
- Sites 100 % self-supported, maintained in concert with IG middleware releases
- Training material and events provided in joint events with GILDA

Bruce Becker, for SAGrid JRU and Ops | BBecker@csir.co.za | www.sagrid.ac.za

Interoperability means : EGEE VO's are more easily accessible to (South) Africans – research activities open up.

Virtual Organisations supported on South African infrastructure

Catch-all SAGrid VO for SA applications and individual users

GILDA VO for new users and development of new applications

https://voms.virt.sagrid.ac.za/voms/sagrid

A Large Ion Collider Experiment

European Organisation for Nuclear Research

CERN/LHC : HEP, experimental nuclear physics

Enable access to all EGI-supported VO's for African researchers, by participating directly as NGI

eGI

Bio-chemistry, materials

WISDOM
Initiative for grid-enabled drug discovery against neglected and emergent diseases

Bioinformatics/ Drug discovery

25

SAGrid : ported applications and development projects

- EGEE/EGI VO's and applications enabled per-site "on demand"
- Permanent call for applications – treated prior to porting workshops
- Several new applications ported during porting workshop :
    - Selected 8 from ~ 36 submissions ; Others will be addressed at other porting schools.
    - Ported :
    - Molecular Dynamics (NAMD, AutoDock)
    - Gene sequencing (WCD)
    - Genetic Programming (Comp Sci)
    - HTK (Human Languages Technologies)
    - Detector design, simulation (GEANT4)
    - Earth Observation and Disaster Modelling (GIS/GRASS)
- Developing the "client" model for cooperative projects which do not provide resources
    - Eye towards sustainability and wider use of infrastructure
    - Strong collaboration with Foreign Affairs Ministries of France and Italy
- Investigating Data Management system for SALT, SKA.

A blueprint for the future : exporting what we have learned

First Contact

dissemination of concept

demonstration of concept

formation of Joint Research Unit (resource providers)

identification or creation of National Certificate Authority

user training

Site admin training

assignment of interim Registration Authorities at local institutes, from accredited CA

new users join GILDA VO, using GILDA resources

prototype sites deployed with clones of GILDA services

accreditation of national CA to EUGridPMA

Formation of Foundation Services Joint training and development to bring new users and applications to the grid

incubation phase

stable operations

local t-Infrastructure created : extension/clone of GILDA

New Virtual Organisations supported on the grid, through local VOMS Services anchored in EUGridPMA-accredited CA Local Applications ported natively, managed by local software admin

National Grid Infrastructure / National Grid Initiative

mera
INSTIT

African Collaborations "on the grid"

• Brain Gain Initiative managed by UNESCO

- 20 institutes in Africa, Middle east,

- Several scientific domains, Operational and technical support from SAGrid

• Other projects :

 • ERINA4Africa

 • "Lighthouse" projects

• Regional scientific research projects (SKA, ASP2010, EGY)

Coordination &
Harmonisation of
Advanced
eINfrastructures
CHAIN

UbuntuNet Alliance
for Research and Education Networking

Summary

- SAGrid is coordinating distributed computing, focal point for e-Science and collaboration in South Africa and the region
- Good participation to FP7 programme, expected to continue
- Fully interoperable with EGEE/EGI "by design", strong effort to ensure that Sub-Saharan region will be able to follow
- Infrastructure level is in place, applications and VO-level work in 2010 to ensure wide use and relevance of services.

Where to find more information

- SAGrid websites :
  - http://www.sagrid.ac.za (main website)
  - http://ops.sagrid.ac.za   (user and site operations portal)
  - http://docs.sagrid.ac.za (digital repository)
  - http://gstat.sagrid.ac.za (monitoring)
  - http://www.facebook.com/pages/South-African-National-Grid/162132040926

Become a fan and follow "South African National Grid" on Facebook