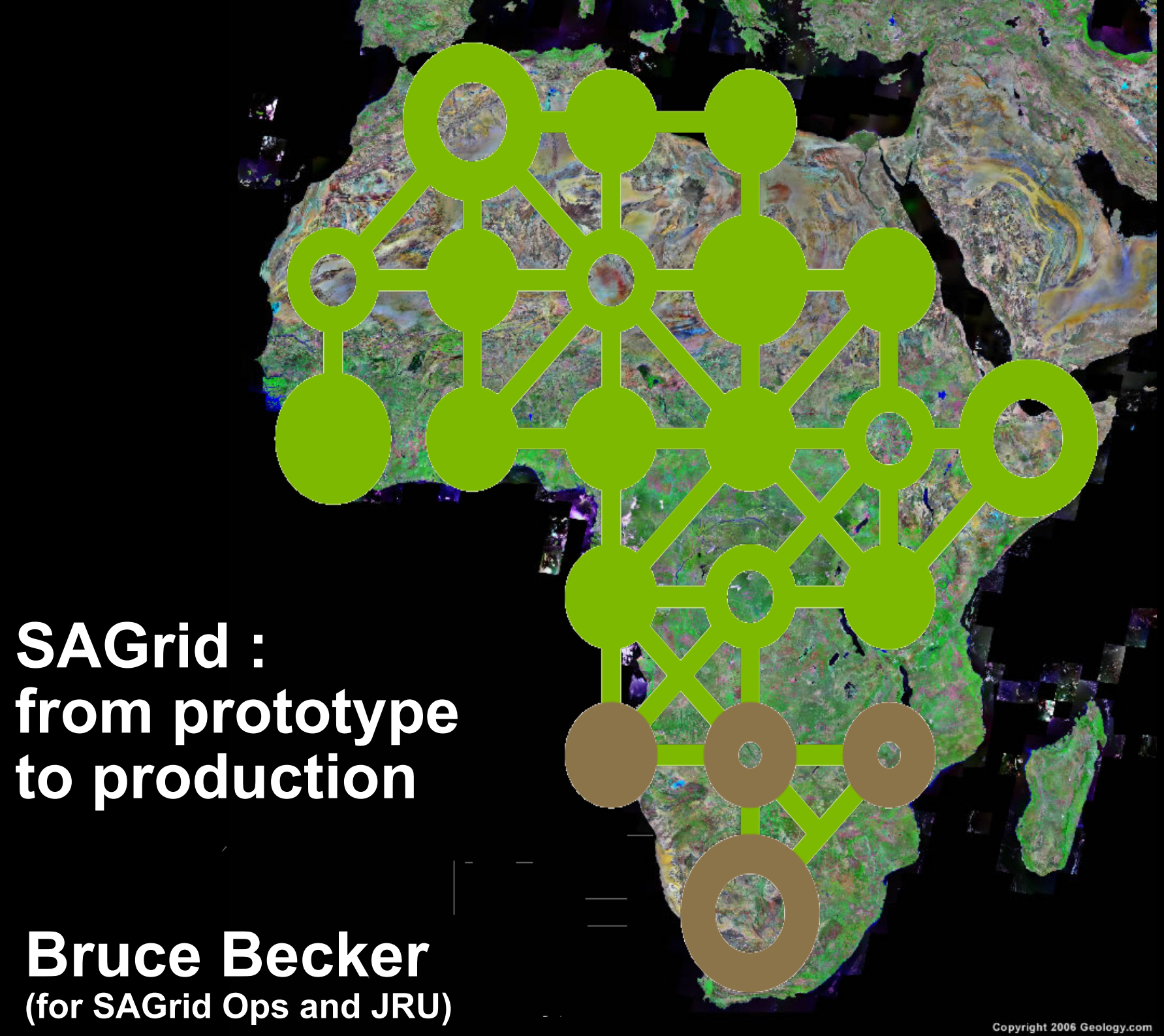


SAGrid : from prototype to production

Bruce Becker
(for SAGrid Ops and JRU)

Overview

- Starting at the end : imagining a pan-African e-Infrastructure
- Where are we at ? Existing infrastructure in South(ern) Africa
  - Networks
  - Computing resources
  - People
- What are we doing ? SAGrid projects under way
  - Certificate Authority
  - Deployment
  - Training
  - Virtual Research Communities
- Interoperability / cooperation
- The bigger picture : A call for collaboration

Effective research means effective collaboration

- Several aspects of scientific research imply that unilateral research programmes are disadvantaged :
  - Convergence and emergence of research domains : enablement of cross- and multi-disciplinary research
  - Massive investments needed for research infrastructure
  - Rapid pace of discovery, through usage of ever more advanced tools
- Research is increasingly (in some case exclusively) digital
  - **"Third mode"** of research – in-silico experimentation through algorithmic and Monte-Carlo models long recognised
  - **"Fourth mode**" - extraction of hidden information and emergence of features in massive data sets
- Geography is increasingly irrelevant
  - High-speed internet and communication services are killing the "distance" effect
- However, we need to take cognisance of the ***digital divide***
  - ***Exclusion*** from communication and information technologies has a crippling effect on the ability to conduct 21st century research.
  - Exclusion from cutting-edge research has a crippling effect on the ability to participate in a 21st century economy.
- **A massive problem with no clear single solution**

Starting at the end : Imagining a pan-African e-Infrastructure

- E-Infrastructure differs from “traditional” research infrastructure :
  - Rapid pace of developments
  - Enabling nature
  - Continuous possibility for “leapfrog” projects which provide massive advances skipping many intermediate steps.
  - Widely appealing, used by almost every research domain
- Collaboration / cooperation in Africa has generally seen know-how leave Africa :
  - Research mostly “south → north”
  - Some “north → south”, but not much in advanced research
  - Almost no “south → south”
- The “Digital Divide” is often the cause and the result of this situation
- E-Infrastructure is critical to addressing the Digital Divide
  - Heavily dependent of network infrastructure – death of distance
  - Transparent access and participation to global e-Scientific instruments and collaborations possible
  - Major technical and political obstacles to be address
- Given success of e-Infrastructure projects in other “under-developed” parts of the world (Latin America, South-East Europe, etc), why could the same not be done in Africa ?

What does infrastructure entail ? The difference between resources and services

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 which 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

The SAGrid Certificate Authority

Enabling International Collaboration Through Trust : SAGrid Policy Management Authority

- Certificate Authority in South Africa is essential to the entire infrastructure
  - Need for international collaboration means that “independent” CA is useless, must be to norm with EUGridPMA and eventually IGTF
  - Initial procedure started in June 2008, Catania, but **final accreditation can take several months**
  - To avoid delays, INFN assigned to RA's which can issue INFN certificates to South African hosts and individuals.
  - **https://security.fi.infn.it/CA/en/RA**

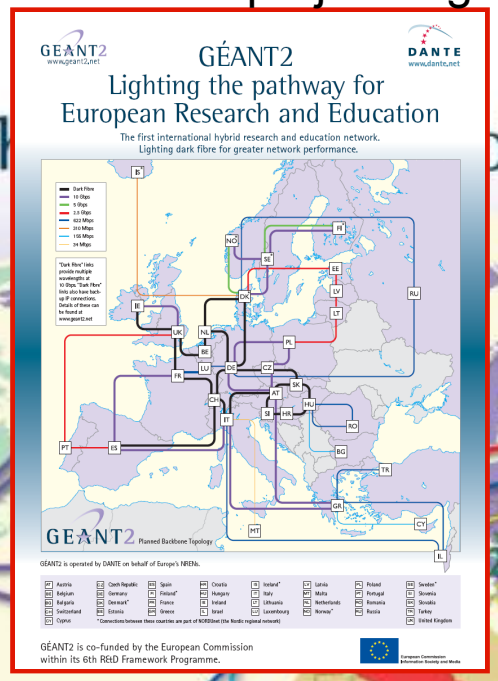
Does this look like a problem... ?
or an opportunity for SA ?

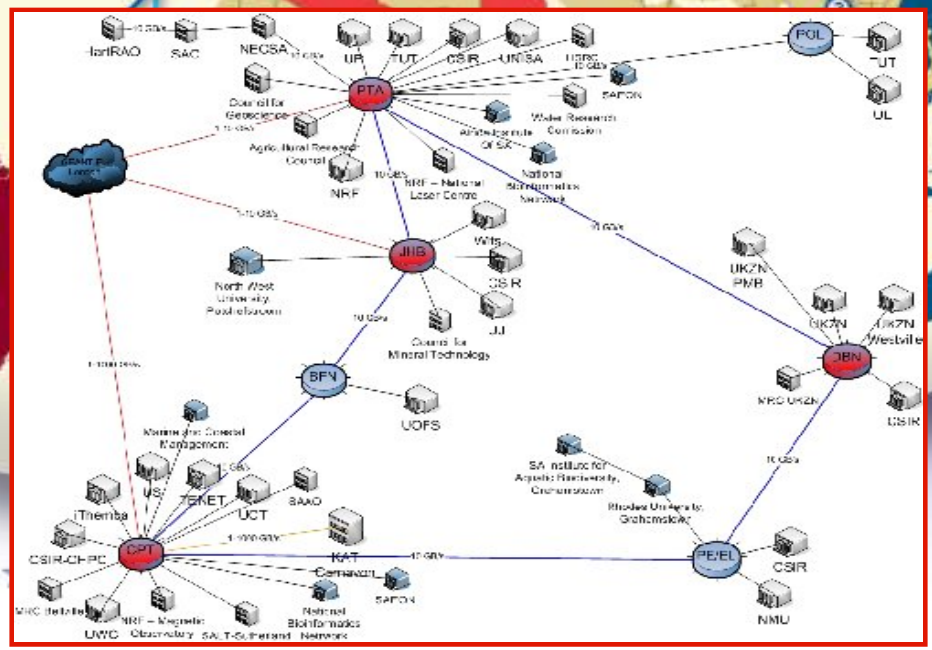
SA Certificate Authority will be hosted at Meraka
in collaboration with ISCoC
CP/CPS currently under 2nd review
Plans for accreditation presented to EUGridPMA October 2009
Prototype CA completed - Demonstrator CA installation under way

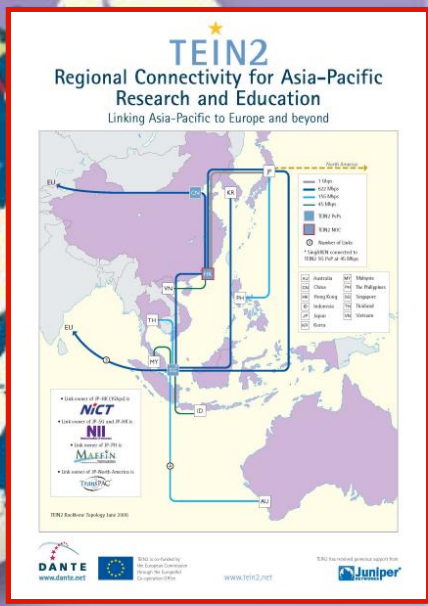
Continental networking and NRENs

The “global” network coverage means high-speed access... to computing resources, data, instruments – and other researchers !

See http://www.feast-project.org/



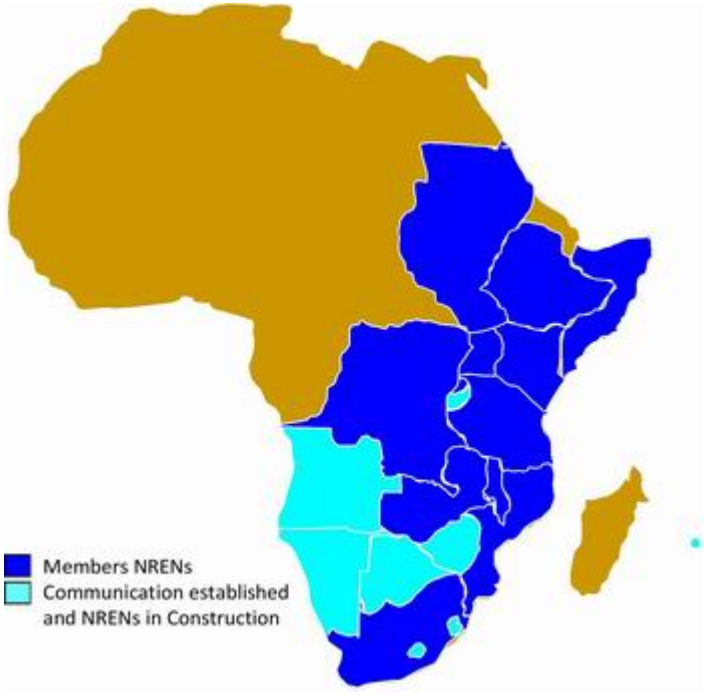




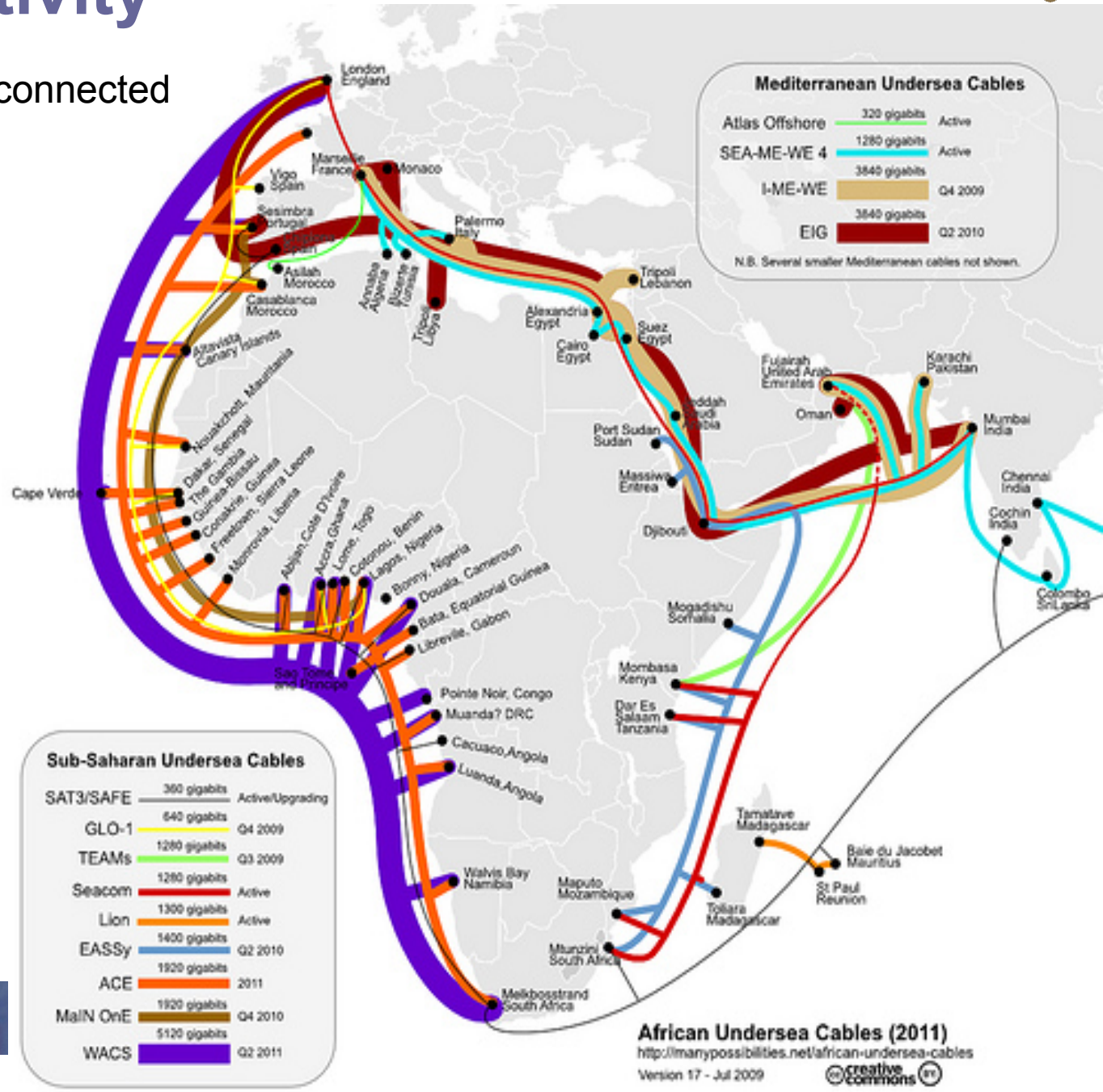
European Commission
Information Society and Media

African connectivity

- Africa will soon be very well connected



UbuntuNet Alliance
for Research and Education Networking

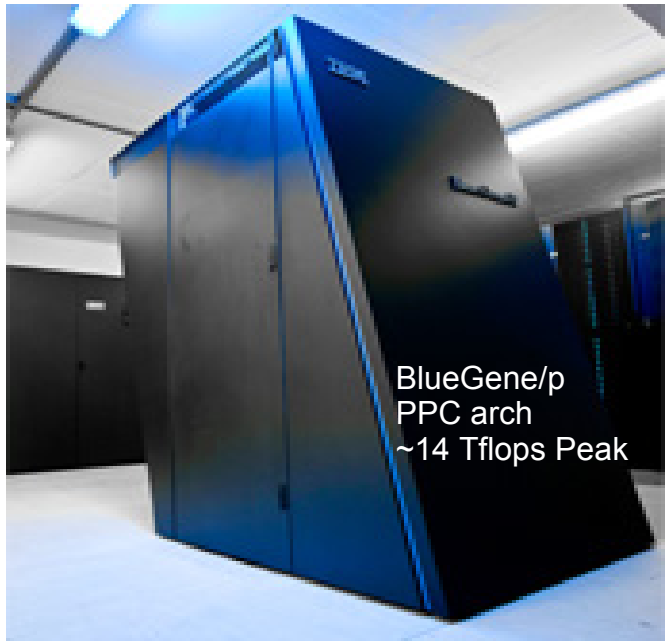


HPC resources around the country

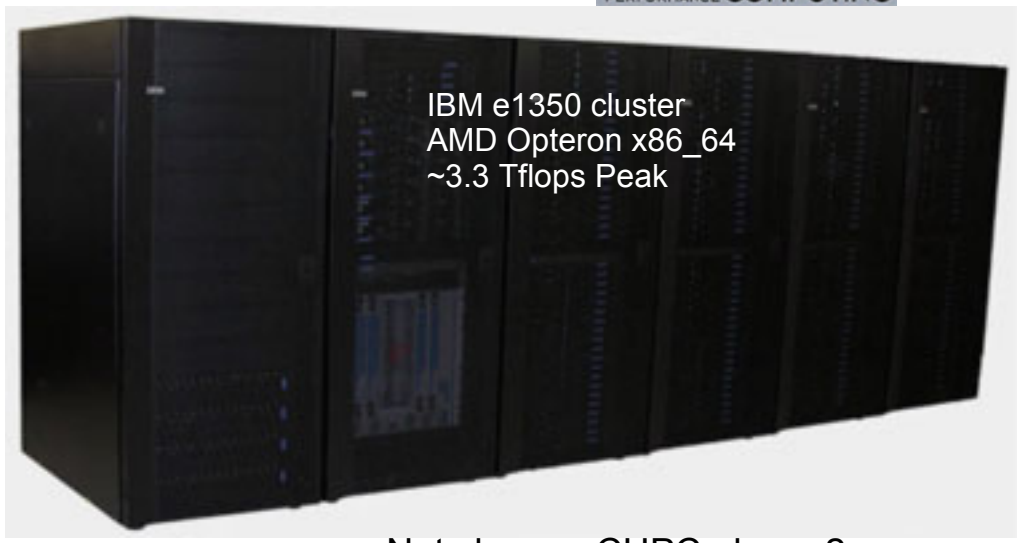
South African National CyberInfrastructure

http://www.chpc.ac.za

- 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



Figures from www.chpc.ac.za



Not shown : CHPC phase 2

A more holistic view of the HPC resources in South Africa

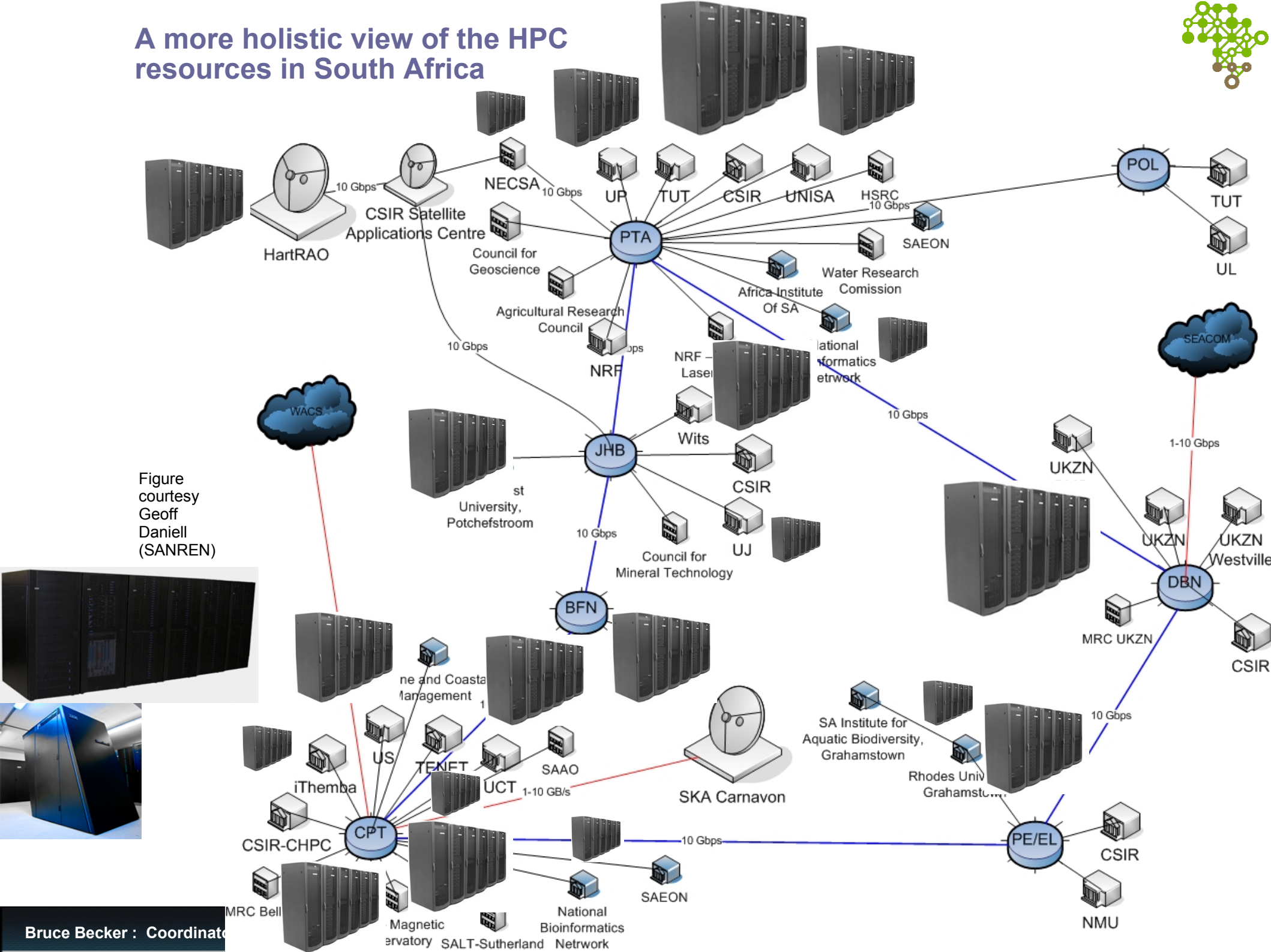


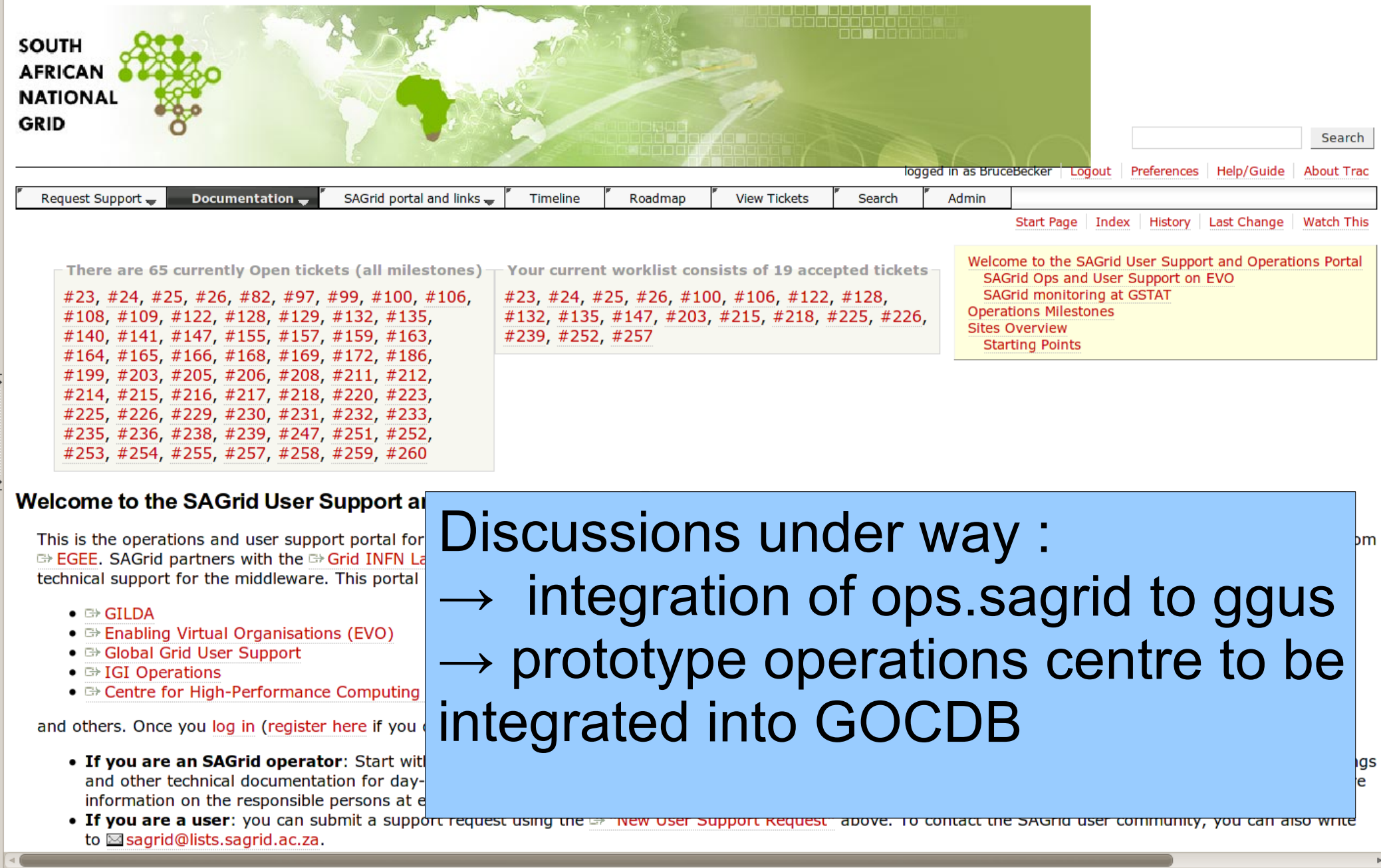
Figure courtesy Geoff Daniell (SANREN)

Management of Grid Infrastructure in South Africa – a technical collaboration to support scientific collaboration

- Joint Research Unit (JRU) of SAGrid :
  - Comprises IT department heads, heads of research groups, heads of laboratories
  - JRU document is a guideline to the kind of federation and organisation that SAGrid will have **– statement of intent**
  - Non-binding document, no legal status
  - Precursor to a ***National Grid Initiative (NGI)*** which is a binding document, has legal status.

JRU is an open document, there are few conditions for joining...

Support : South African National Grid Operations Portal



Discussions under way :

→ integration of ops.sagrid to ggus

→ prototype operations centre to be integrated into GOCDB

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

Interoperability with EGEE/EGI and related projects

- Interoperability is crucial to collaboration
- SAGrid provides a near clone of Italian Grid Infrastructure middleware release for all services – 100 % self-supported
- Training material and events provided in joint events with EGEE / GILDA

gLite Lightweight Middleware for Grid Computing

eGee Enabling Grids for E-sciencE

italian grid infrastructure

IGI

Cooperation

- Close collaboration with UbuntuNet
- MoU of training and support to HP/UNESCO Brain Gain project, co-investigation of emerging technologies
- Constant participation to FP7 proposals

UNESCO United Nations Educational, Scientific and Cultural Organization

LABS hp

Bruce Becker : Coordinator, SAGrid / bbecker@csir.co.za

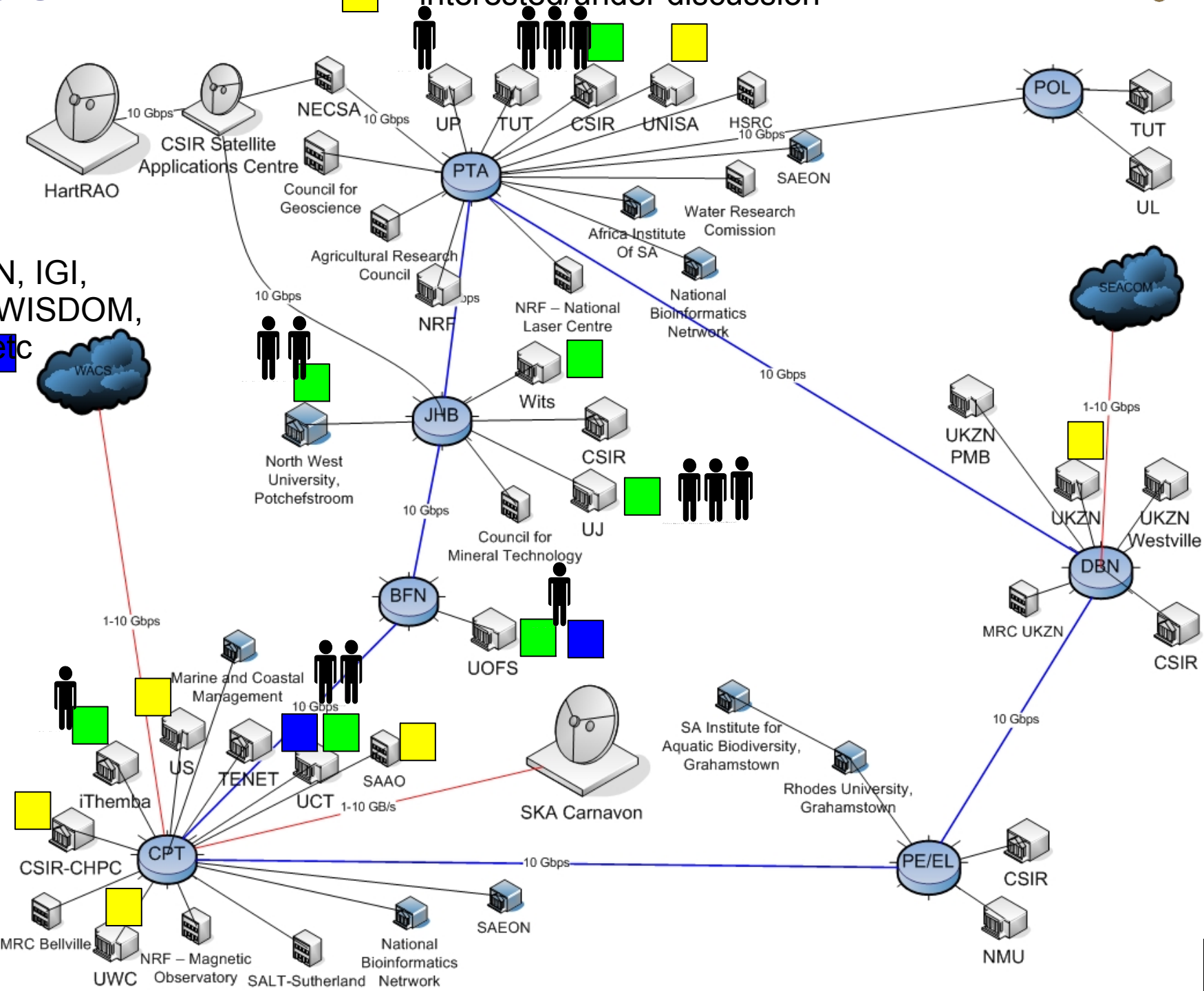
SAGrid through the eyes of SANREN

Phase 1 sites (JRU)

"interested/under discussion"

Core services

CERN, IGI, GILDA, WISDOM, etc



Virtual Organisations supported on South African infrastructure

e-nmr

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

Bio-chemistry, materials

Bioinformatics/ Drug discovery

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

EGI

ALICE
A Large Ion Collider Experiment
European Organisation for Nuclear Research
CERN

ATLAS

CERN/LHC : HEP, experimental nuclear physics

LCG

GENESI-DR

(under discussion)

SAGrid infrastructure ported applications and development projects

- 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)
    - Human Languages Technologies
    - Detector design, simulation (GEANT4)
    - Deployment model being tweaked for full deployment on the grid, deciding which VO's to support on sites.
- 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
  - Pilot project with Working for Water (DoWA), CESVI SpA, INFN on "Decision Support System"
- Investigating **Data Management System** for SALT, SKA.

Before we get ahead of ourselves.... where are we going with this ?

- 2009 was “the year of the operator / technician”
- 2010 is the “year of the expert user / big experiment”
  - Somewhat different activities
  - Much more complex to handle
  - Far more diverse requirements and community
  - A necessary step to convincing scientific communities en masse
- Success for infrastructure, and for all of the resource providers :
  - infrastructure is **USED** effectively, that users **TRUST** it.
- Need to communicate this goal in a transparent, responsible and honest way
  - SAGrid activities should undergo an external, public review, and call for suggestions – to be implemented as part of a coherent roadmap in the region.
- The region – and the amount of work – is **vast** : this can only be tackled by a big, loose collaboration
- A ***task force*** is needed combining the existing work being done by funded projects, defining the desired roadmap for an ***“AfricaGrid”*** and all related services, so that a coordinated, collaborative solution can be implemented.

A Sub-Saharan Task Force for Grid Deployment

- The enabling nature of e-Infrastructures and the death of distance means that researchers and technicians around the region can directly collaborate with South Africa (and others) immediately.
- **There is existing work which is under way in several areas :**
  - Internet networking (AfricaConnect, ERINA4Africa, etc)
  - HPC site deployment (Dar es Salaam, CHPC, HP/UNESCO, SA Universities, etc)
  - Training: site, user, application (EPIKH)
  - Scientific collaborations (many)
  - Dissemination, communication to decision makers (BELIEF-II, EGI_InSPIRE, CHAIN, etc)
- **Who can / should participate ?**
  - Existing funded FP-7 projects : CHAIN, ERINA4Africa and AfricaConnect, EUMedSupport, etc
  - Previous similar regional projects : EELA, EUIndia, EUChina (experience)
  - Regional contacts are crucial : UbuntuNet and UbuntuNet HPC SIG, WACREN, UNESCO Brain Gain project, SAGrid JRU, embassies of France and Italy in SA
- **What do we expect to obtain ?**
  - Coherent vision of the benefits of inter-African collaboration
  - Consensus on technology
  - Interoperable infrastructure
  - Possibilities for intermediate sharing of resources between nations/institutes
  - Greatly accelerated deployment and participation to global science

Projects of great relevance : EPIKH, HP/UNESCO

The EPIKH Project (www.epikh.eu)

"Exchange Programme to advance e-Infrastructure Know-How"

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

Coordinator : Consorzio COMETA (Italy)

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

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	

HP/UNESCO

20 institutes in Africa
Budget $1M

Several scientific domains

Operational and technical
support from SAGrid

United Nations
Educational, Scientific and
Cultural Organization

Bruce Becker : Coordinator, SAGrid / bbecker@csir.co.za

Summary

- South Africa has a well-established infrastructure, which includes SAGrid
  - http://www.sagrid.ac.za
- Federated / jointly managed resources well integrated with good networking provide unique point of access for users, collaborations
- SAGrid Operations are stable, user communities are beginning to make use of the infrastructure
- South Africa aims to act as blueprint for the region, and tight cooperation with EGI and NGI's is essential at all levels.
- A Task Force for e-Infrastructure Deployment in the region is being formed to ensure that the investment is made and exploited efficiently