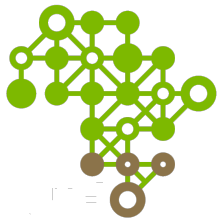


Africa-Arabia Regional Operations Centre

Bruce Becker | bbecker@csir.co.za
Coordinator, SAGrid
Meraka Institute, CSIR


WARNING
Highly simplified scientific analogies will be made

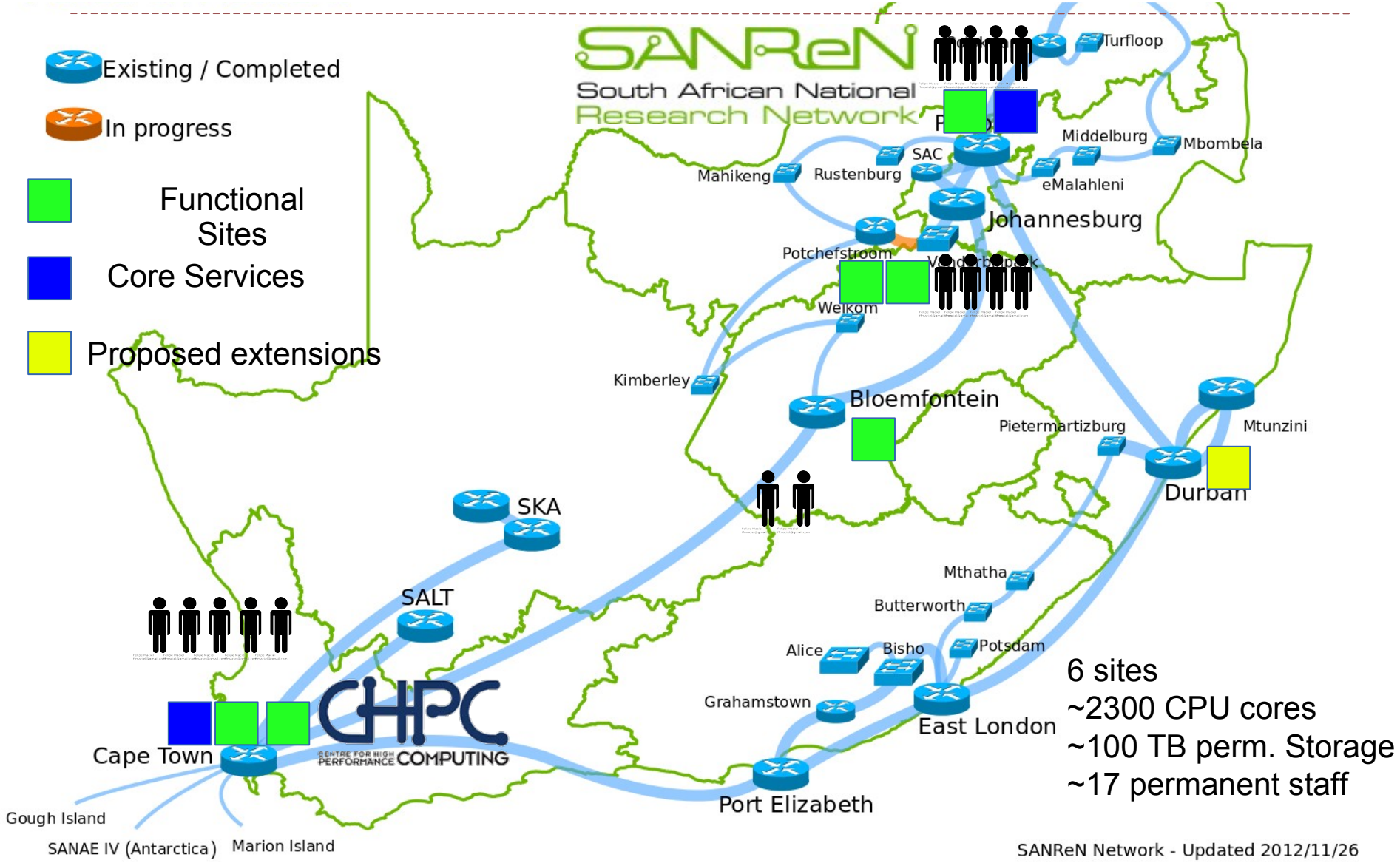
Outline

- What have we learned
- Status of computational and data infrastructure in Africa
- How to grow these resources sustainably
  - User-resource-support interactions
  - Characteristic features at various scales
- What is the Africa-Arabia Regional Operations Centre ?
  - Components vs procedures
  - Automation
  - Services and Resources
- Outlook

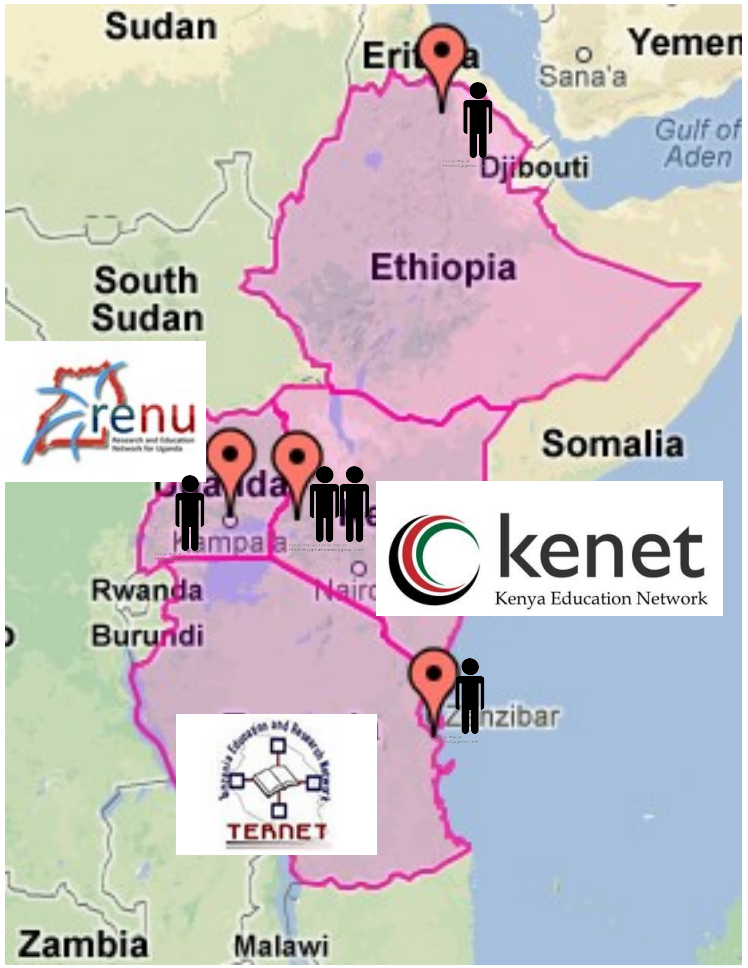
What have we learned

- There is much potential for productive e-Science in Africa
  - Institutes moving to conduct research, appreciated the need for access to e-Infrastructures
  - Large diaspora to collaborate with
  - New bandwidth opportunities connecting institutes and people
  - Long list of important scientific problems to tackle
- Despite the ***potential***, the available ***energy*** is not all transformed into ***work***

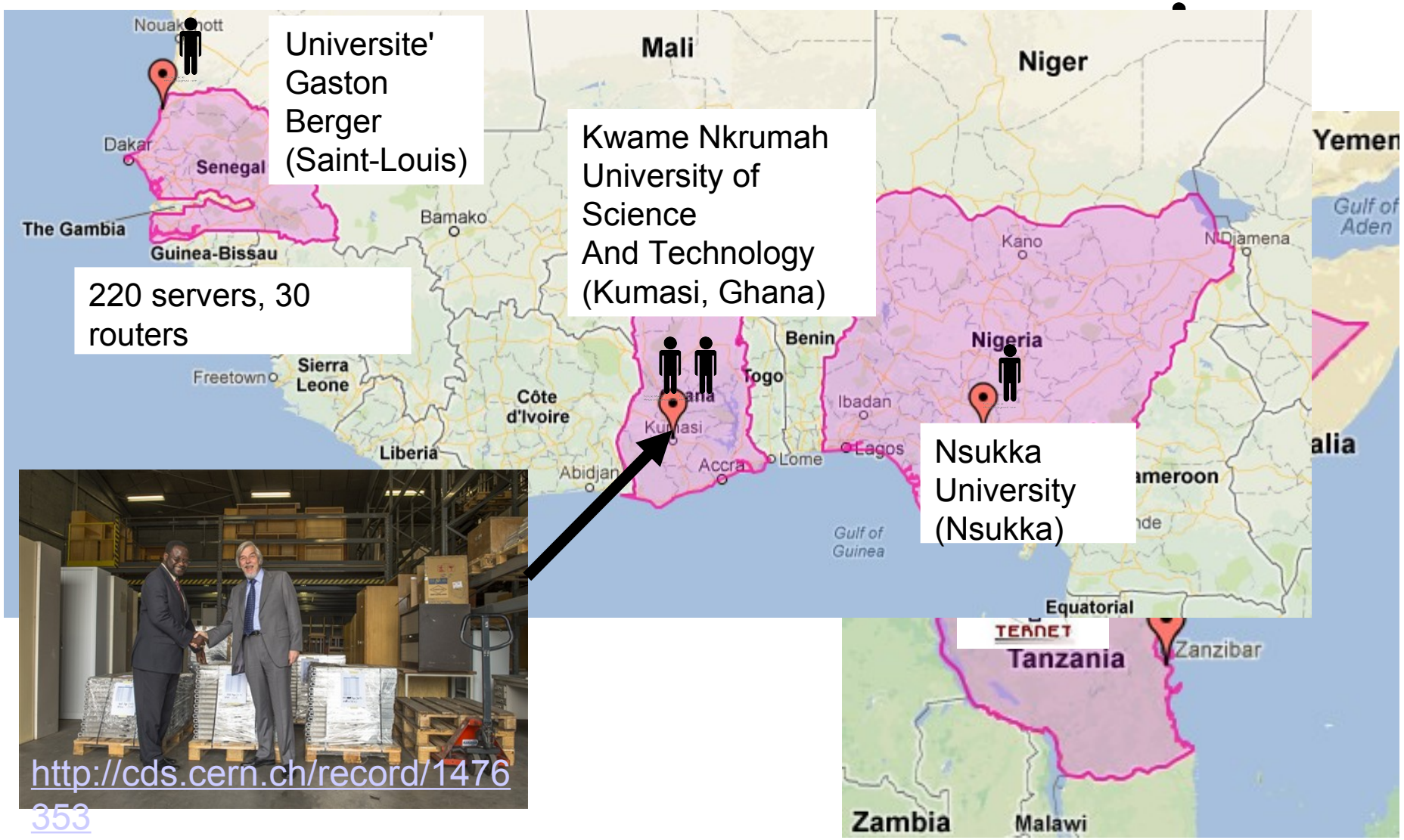
South African Integrated e-Infrastructure Initiative



Other African sites



African sites



http://cds.cern.ch/record/1476353

Plenty of resources and activity -
how do we ensure that we can **sustainably** support this infrastructure ?

Lessons from South Africa

- South Africa has built several e-infrastructure components with a long-term focus (the SKA).
  - Biggest **supercomputer** in Africa
  - Fastest **research network** in Africa
  - Biggest **research data** repository in Africa
- From the outside, it appears that this was a top-down approach
  - How can this be applicable to African countries where there is
    - much smaller capacity to support such initiatives ?
    - No single driving force ?
- There is also room for **emergent phenomena,** based on local interactions

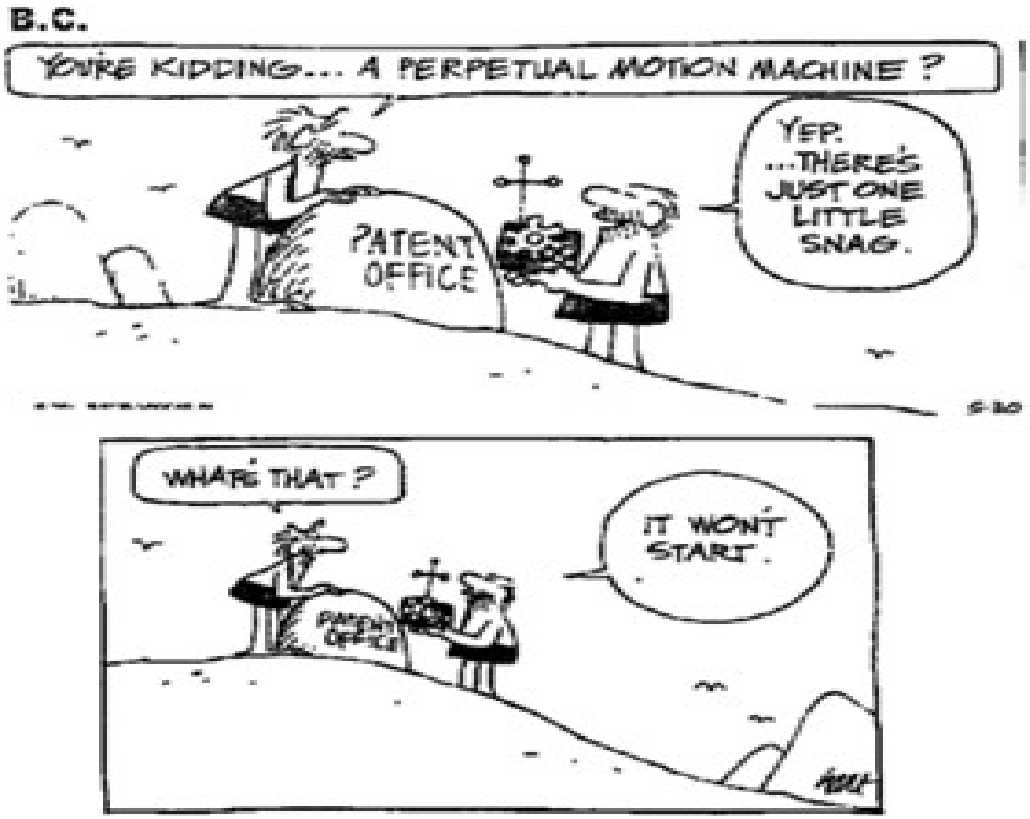
Lessons from Africa (CHAIN)

- Sites, resources and people in Africa are in general hard to find -
  - Who do you talk to ?
  - Which funding agency do you approach ?
  - Does researcher X speak on behalf of their community ?
- General strategy: use the network
  - Several projects to develop a contact database, curated by Ubuntunet
- Great, we found lots of eager scientists !
  - **NOW WHAT ?!**

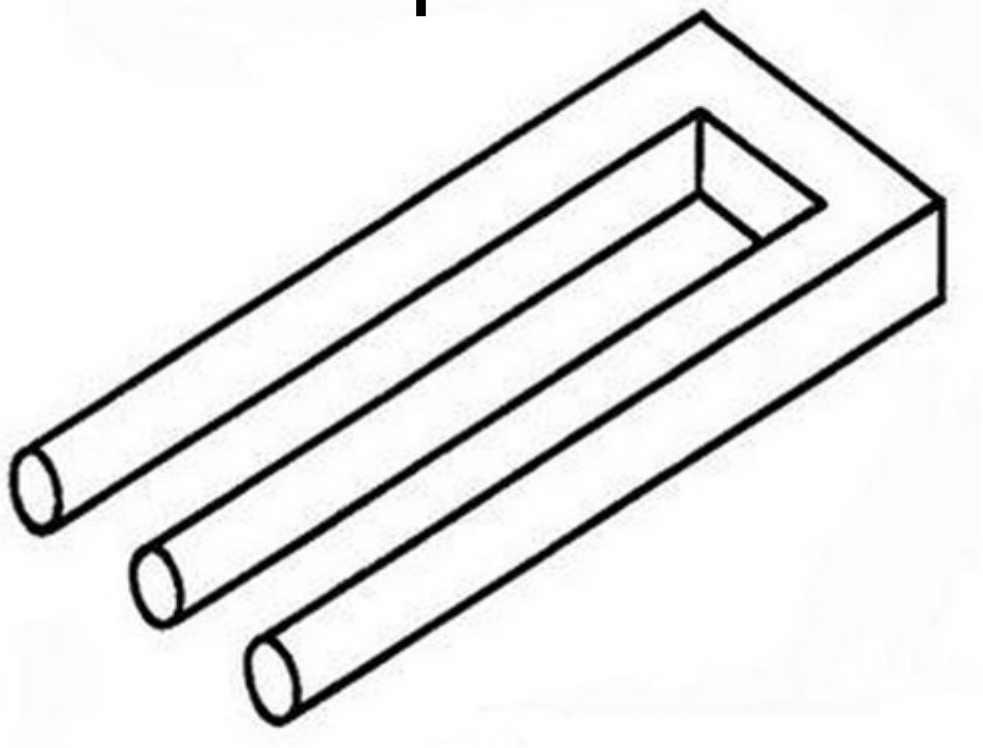
Real-world complexities

- Utopia: top-down approach - “Here's a lot of money, go solve the problem...”
  - Take some time to think about the problems with that...
- There is also room for **emergent phenomena,** based on local interactions
- Many outcomes are permissible, but a stable solution is not guaranteed.
- **Eyes on the prize** : Scalable research in Africa → needs open, self-sustainable e-Infrastructure

Perhaps our model is too simplistic... or optimistic



Perhaps our model is too simplistic... or optimistic



WORKS ON PAPER

User/Resource feedback

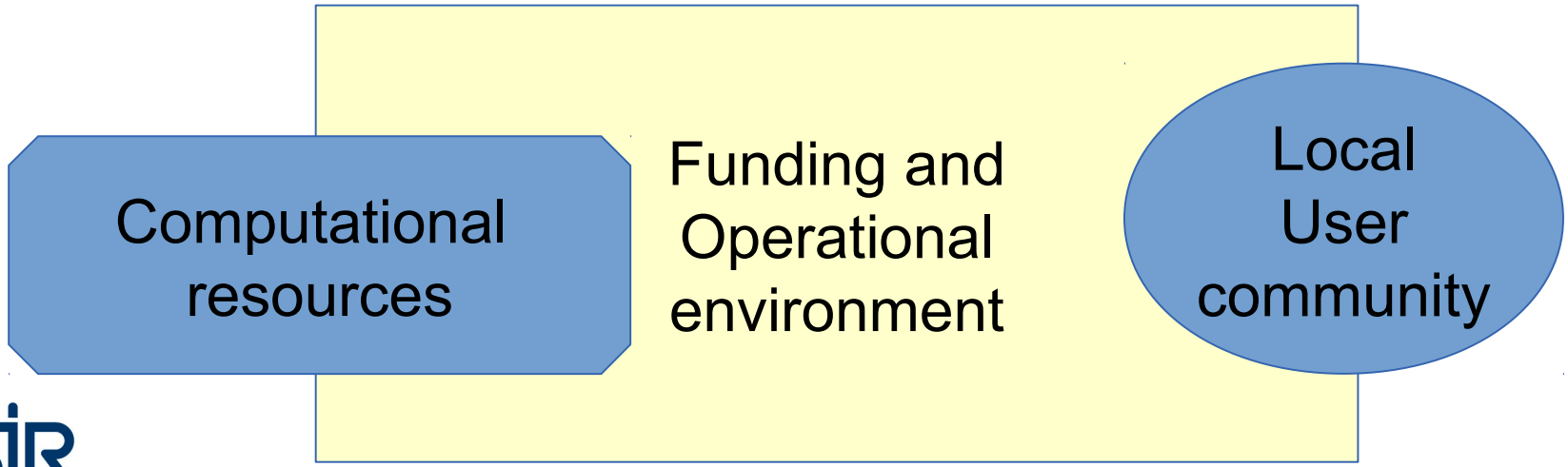
- The **resources** drive **usage**, which drive the acquisition of new **resources** – a *driven harmonic oscillator ?*
- Hypotheses:
  - Decays in the absence of a use case
  - Strongly damped in the case of weak input signals
  - Requires an initial catalyst

User/Resource feedback

- The **resources** drive **usage**, which drive the acquisition of new **resources** – a *driven harmonic oscillator ?*
- Hypotheses:
  - Decays in the absence of a use case
  - Strongly damped in the case of weak input signals
  - Requires an initial catalyst

User/Resource feedback

- The **resources** drive **usage**, which drive the acquisition of new **resources** – a *driven harmonic oscillator ?*
- Hypotheses:
  - Decays in the absence of a use case
  - Strongly damped in the case of weak input signals
  - Requires an initial catalyst



User/Resource feedback

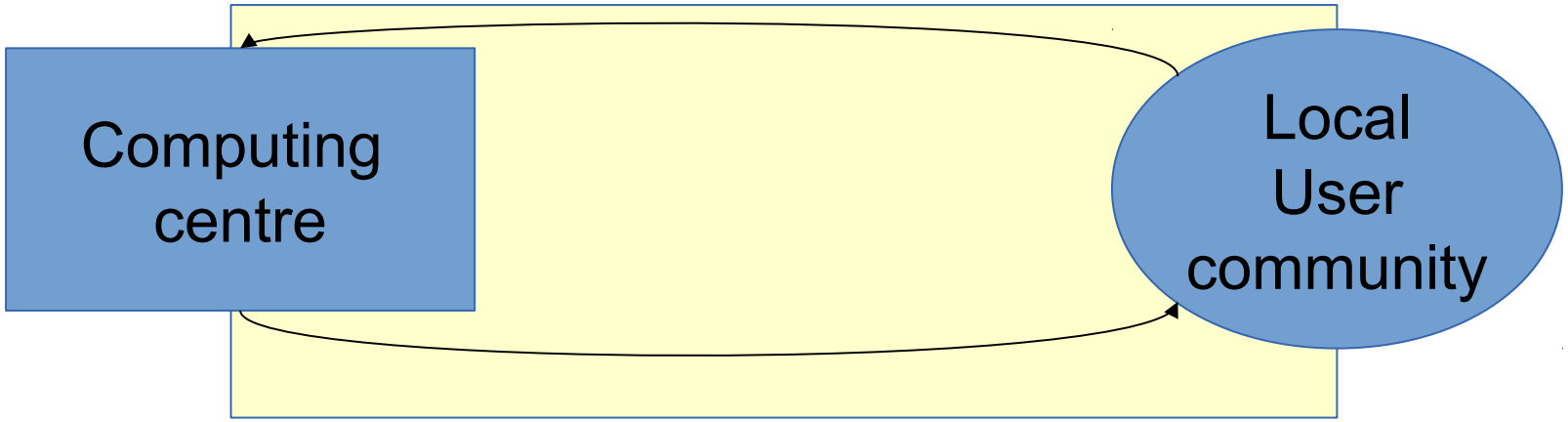
- The **resources** drive **usage**, which drive the acquisition of new **resources** – a *driven harmonic oscillator ?*
- Hypotheses:
  - Decays in the absence of a use case
  - Strongly damped in the case of weak input signals
  - Requires an initial catalyst

User/Resource feedback

- Hypothesis - the case of a strong **research agenda**:
  - Input: User community drives computing centre
  - Two feedback loops
    - User activity drives research agenda, which drives user community
    - Interesting research agenda justifies computing resources
    - Computing resources improve research agenda.
  - Output to counterbalance to external forces

User/Resource feedback

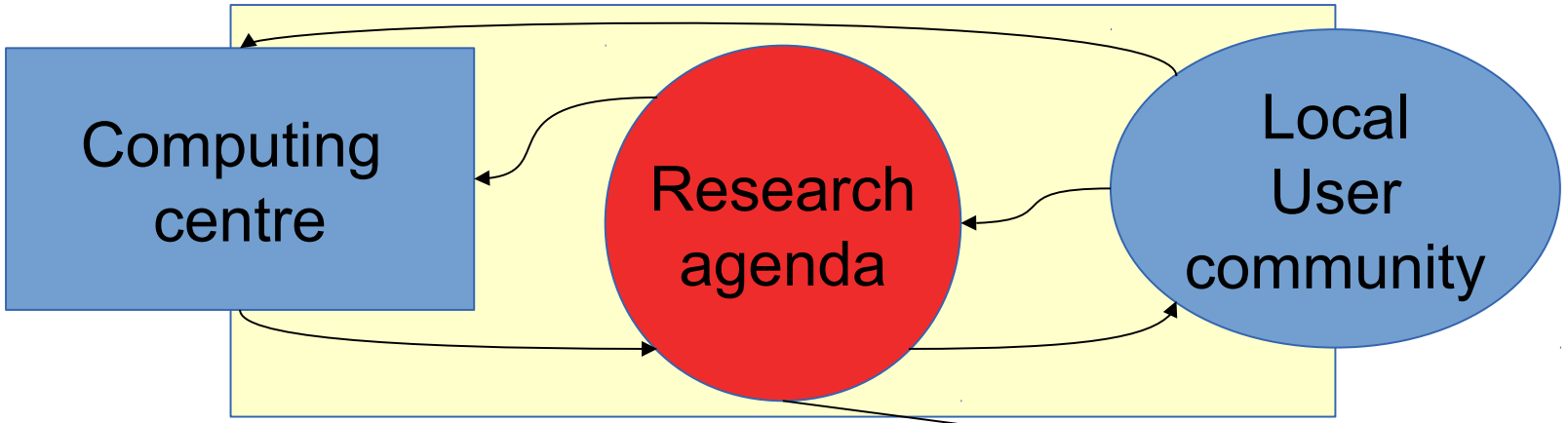
- Hypothesis - the case of a strong **research agenda**:
  - Input: User community drives computing centre
  - Two feedback loops
    - User activity drives research agenda, which drives user community
    - Interesting research agenda justifies computing resources
    - Computing resources improve research agenda.
  - Output to counterbalance to external forces



Publications/data

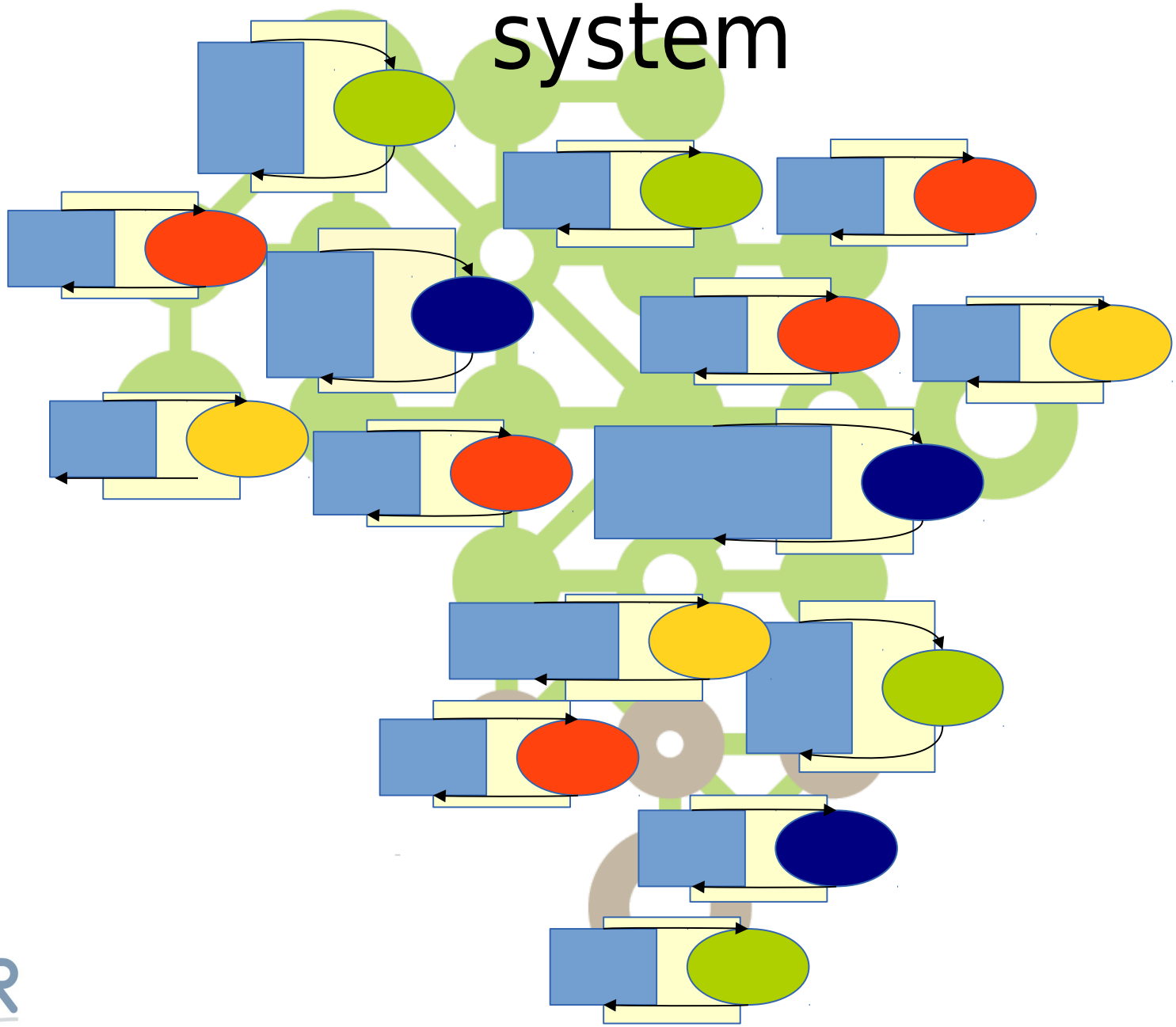
User/Resource feedback

- Hypothesis - the case of a strong **research agenda**:
  - Input: User community drives computing centre
  - Two feedback loops
    - User activity drives research agenda, which drives user community
    - Interesting research agenda justifies computing resources
    - Computing resources improve research agenda.
  - Output to counterbalance to external forces

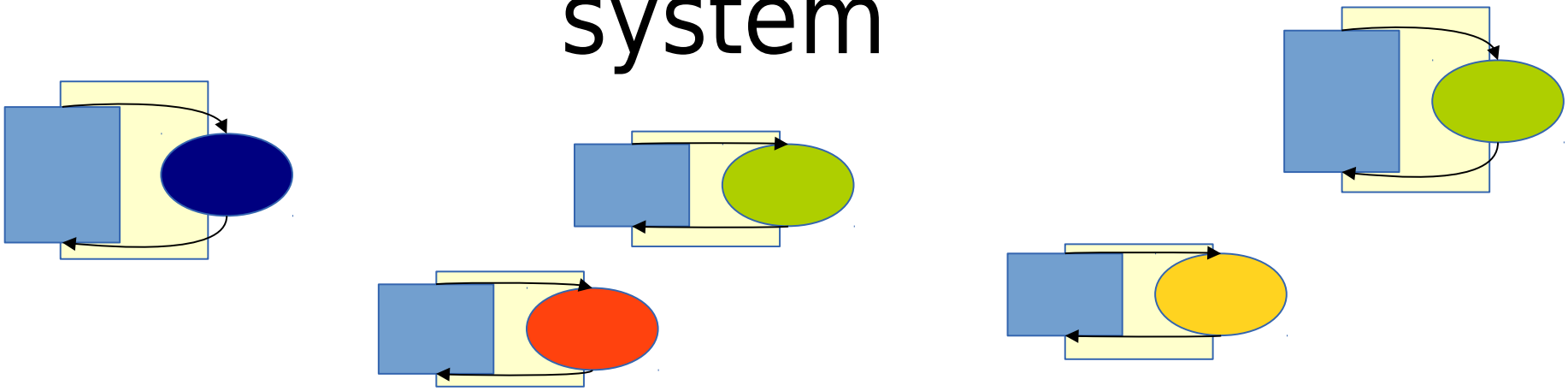


Works on paper... how do we scale it ?

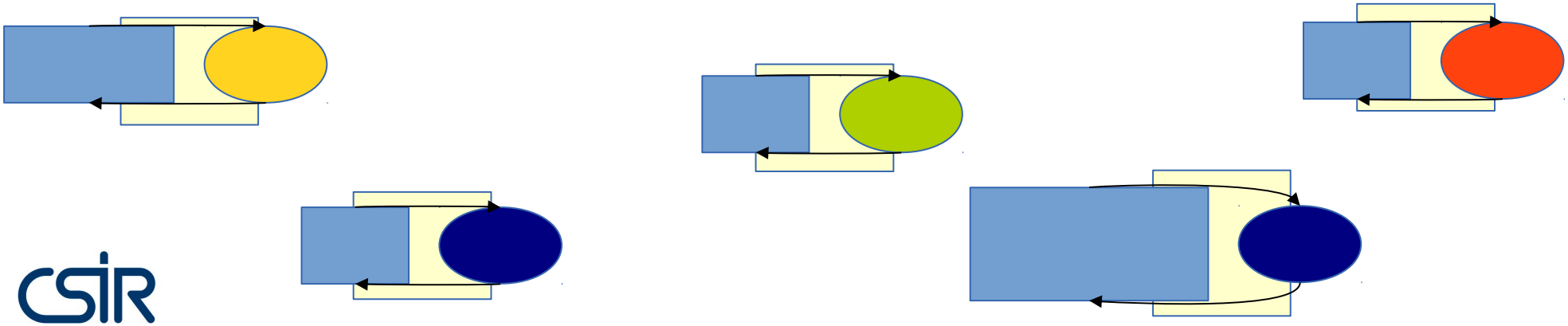
E-Infrastructure as a complex system



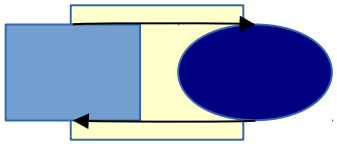
E-Infrastructure as a complex system

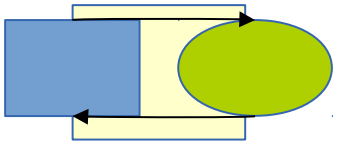


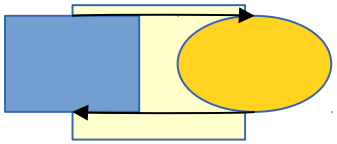
1) How much work can these systems do ?
2) How much effort does it take to sustain these systems ?

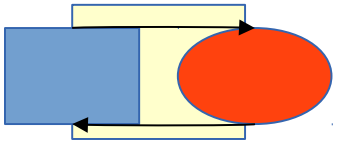


E-Infrastructure as a complex system

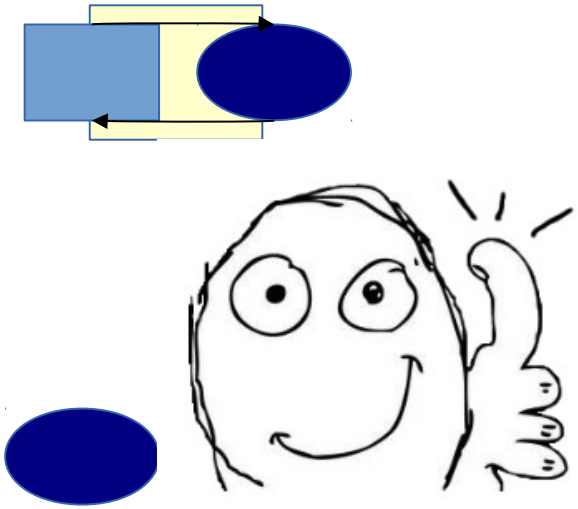


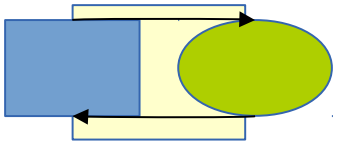


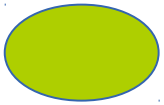


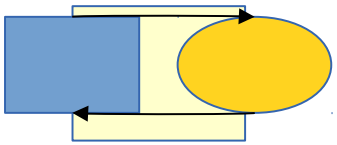


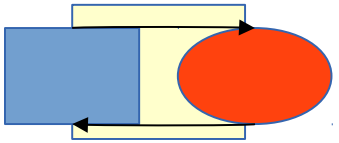
E-Infrastructure as a complex system



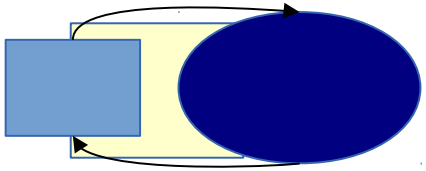


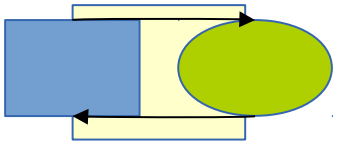




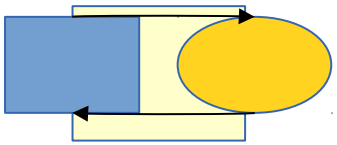


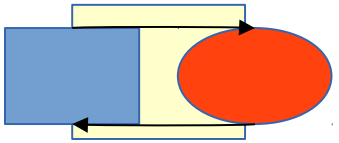
E-Infrastructure as a complex system



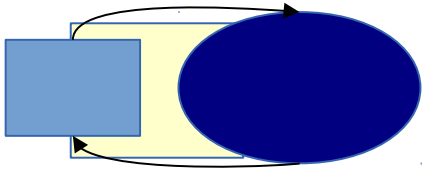


Communities merge
→ load on resources exceeds site capacity

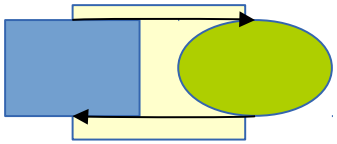




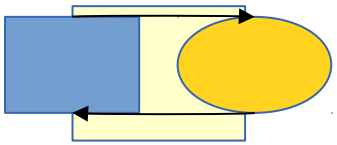
E-Infrastructure as a complex system

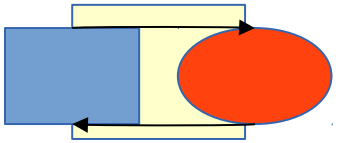


Communities merge
→ load on resources exceeds site capacity

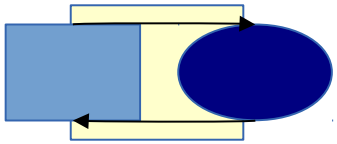
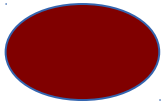
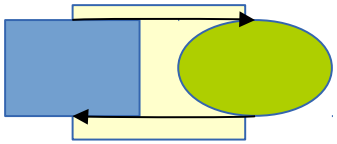
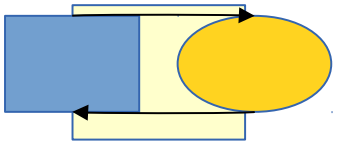
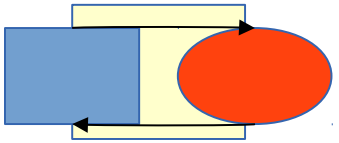


Communities remain ignorant of each other
→ inability to scale
→ duplication of effort
→ sections (countries) disadvantaged

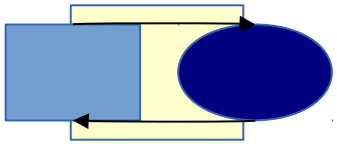


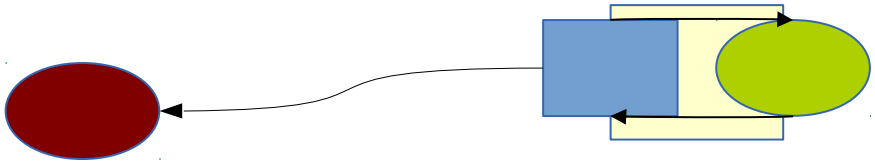


E-Infrastructure as a complex system

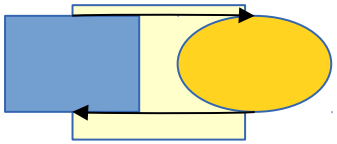






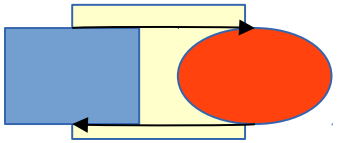
E-Infrastructure as a complex system



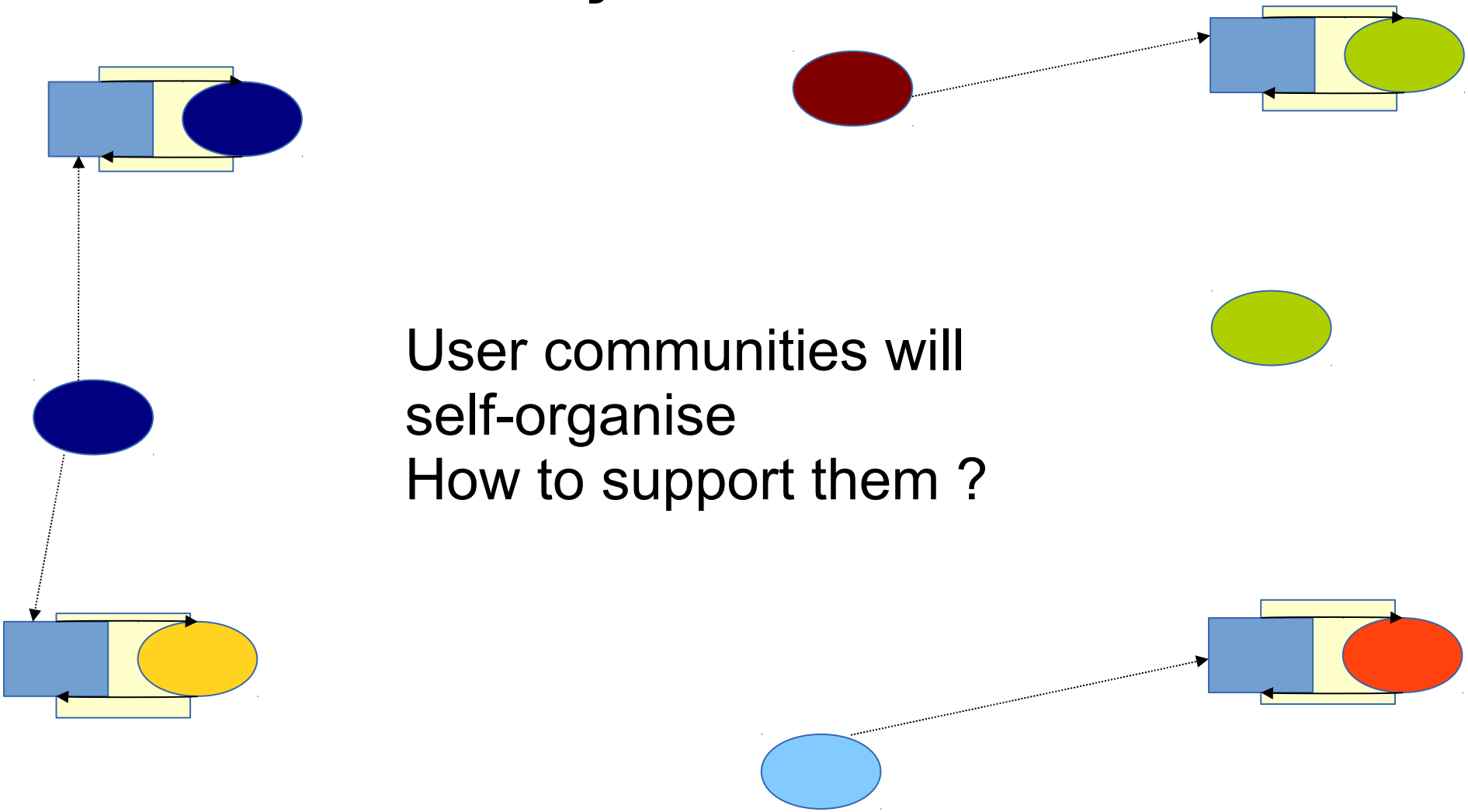


”Scavenger” communities use whatever resources they can find
→ no positive feedback loop



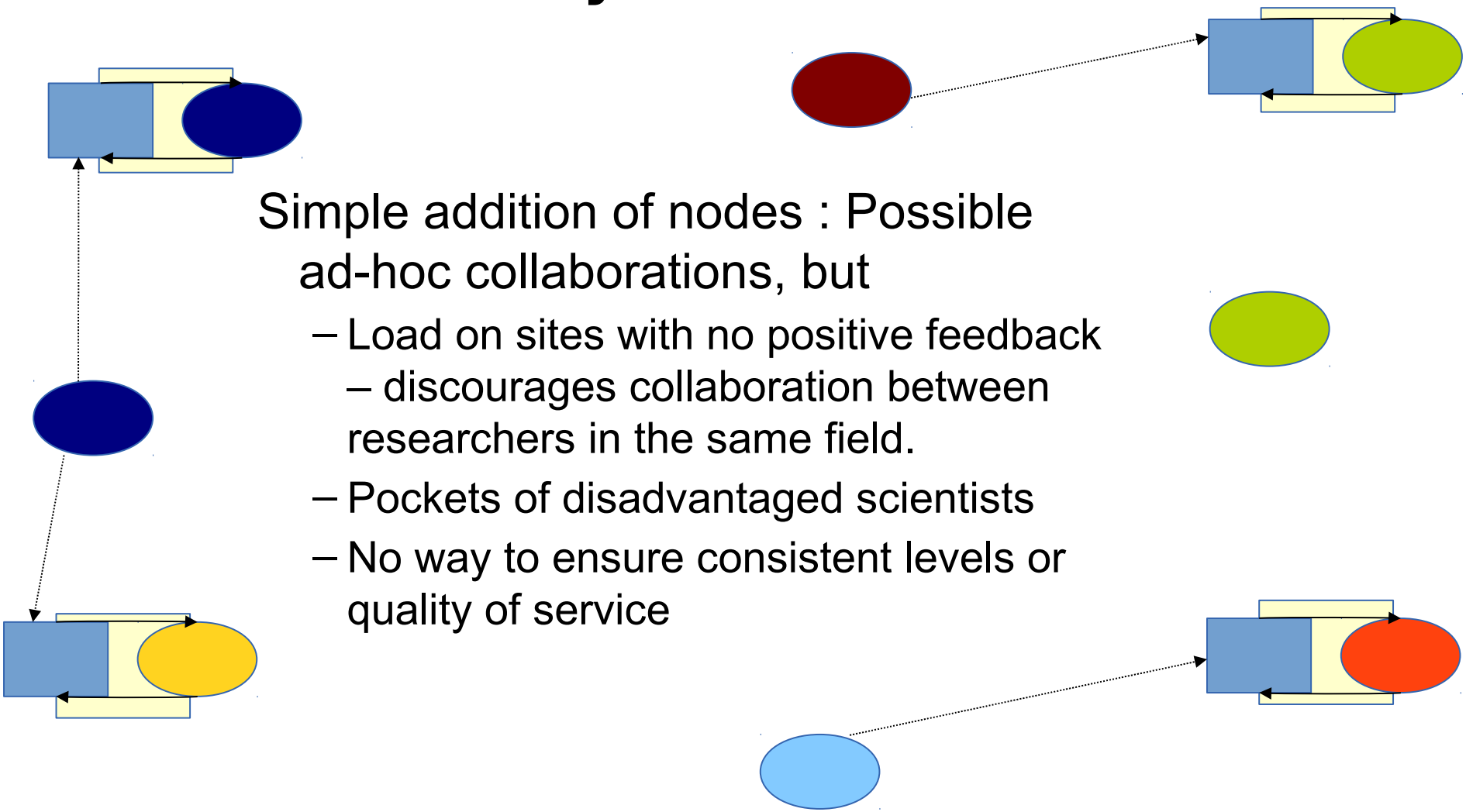


E-Infrastructure as a complex system



User communities will self-organise
How to support them ?

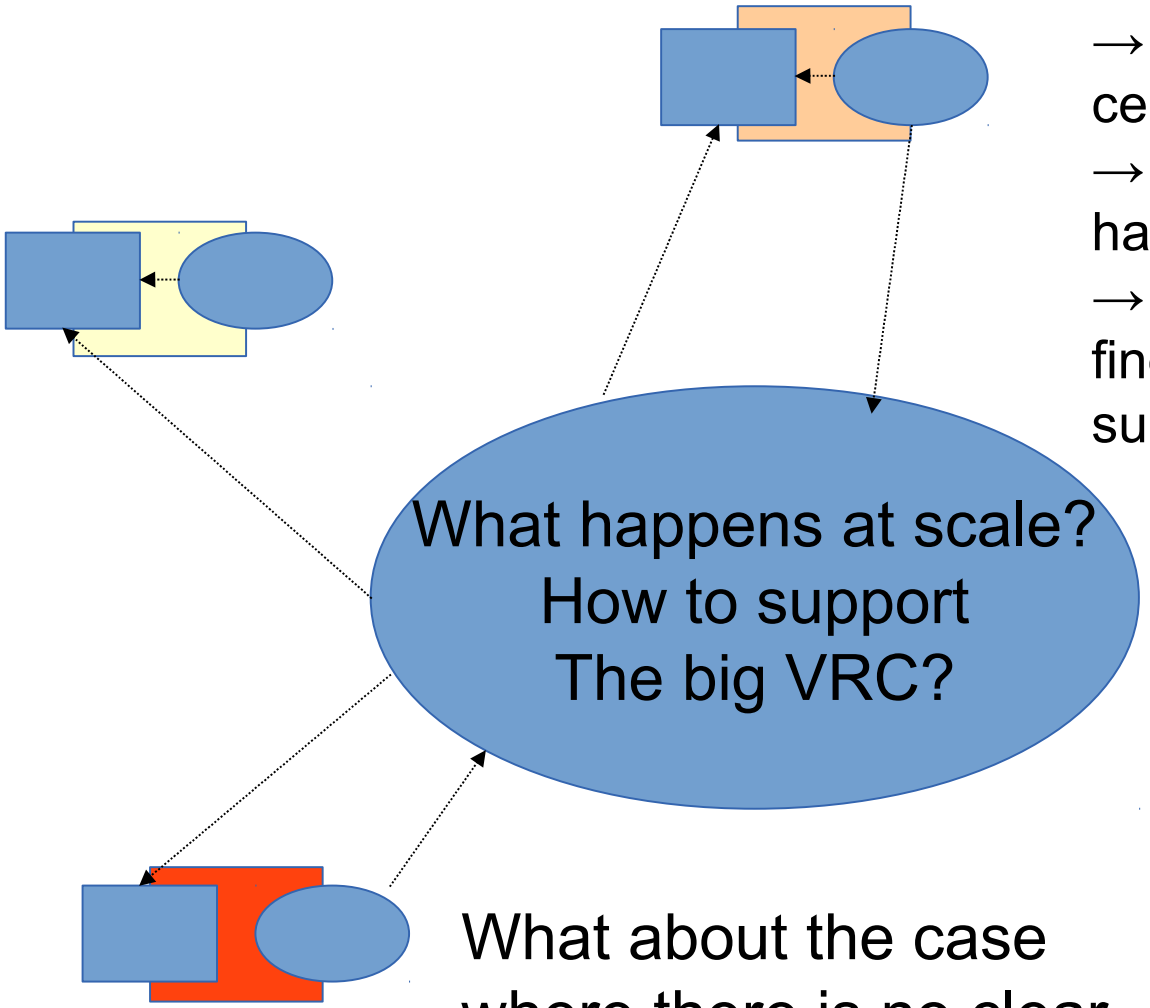
E-Infrastructure as a complex system



Simple addition of nodes : Possible ad-hoc collaborations, but

- Load on sites with no positive feedback – discourages collaboration between researchers in the same field.
- Pockets of disadvantaged scientists
- No way to ensure consistent levels or quality of service

Far-from equilibrium states



→ How does the resource centre know what is required ?
→ How can the VRC ensure that it has access to reliable resources ?
→ how do individual researchers find out about VRCs that can support them ?

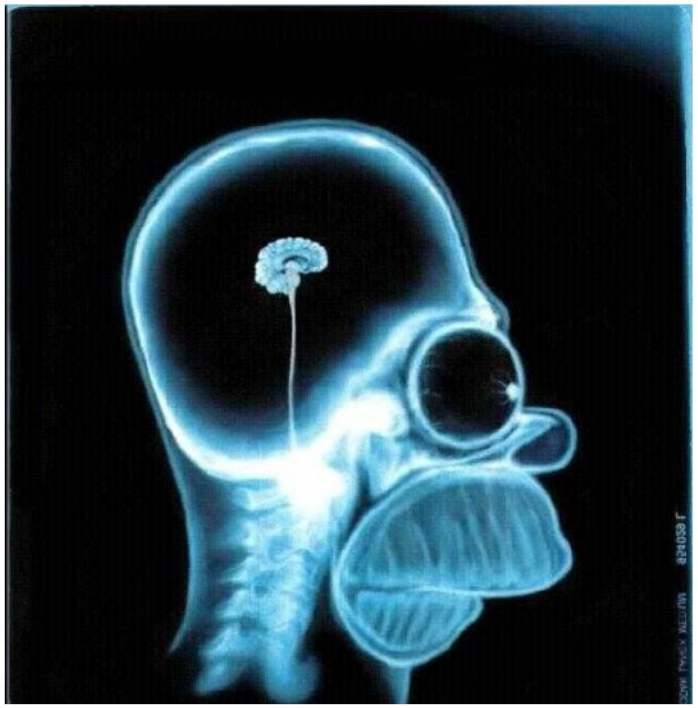
What about the case where there is no clear feedback loop ?

What does this mean ?

- In fairly closed systems, investment in resources and support of feedback mechanisms can lead to self-supporting e-Science systems
  - However, it cannot scale
- Usually an intervention is required to initially drive the system far from its rest state of doing nothing
  - This is a luxury not all of our colleagues can afford
- What should we do ?

Structural MRI vs. *Functional* MRI

Structural **MRI reveals brain anatomy.**



Functional **MRI (fMRI) reveals brain function.**

A complex system is better described by its **processes** than its **parts**

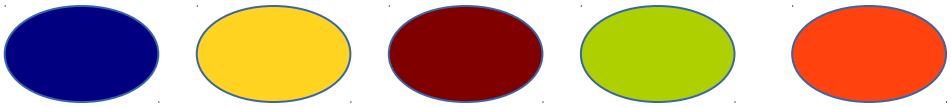
Perhaps we can re-organise ?

- Many outcomes are permissible, but a stable solution is not guaranteed.
- There is also room for **emergent phenomena,** based on local interactions, but in a different operating environment
  - Can a local equilibrium drive similar resonances in a more complex system ?
  - What kind of **interactions** will result in a higher stable state
  - Or result in overload and damp the resonance ?

Emergent organisation

How do we restore the feedback loop between resources providers and resource consumers ?

VRC sees a single infrastructure, feedback loop restored



Emergent organisation

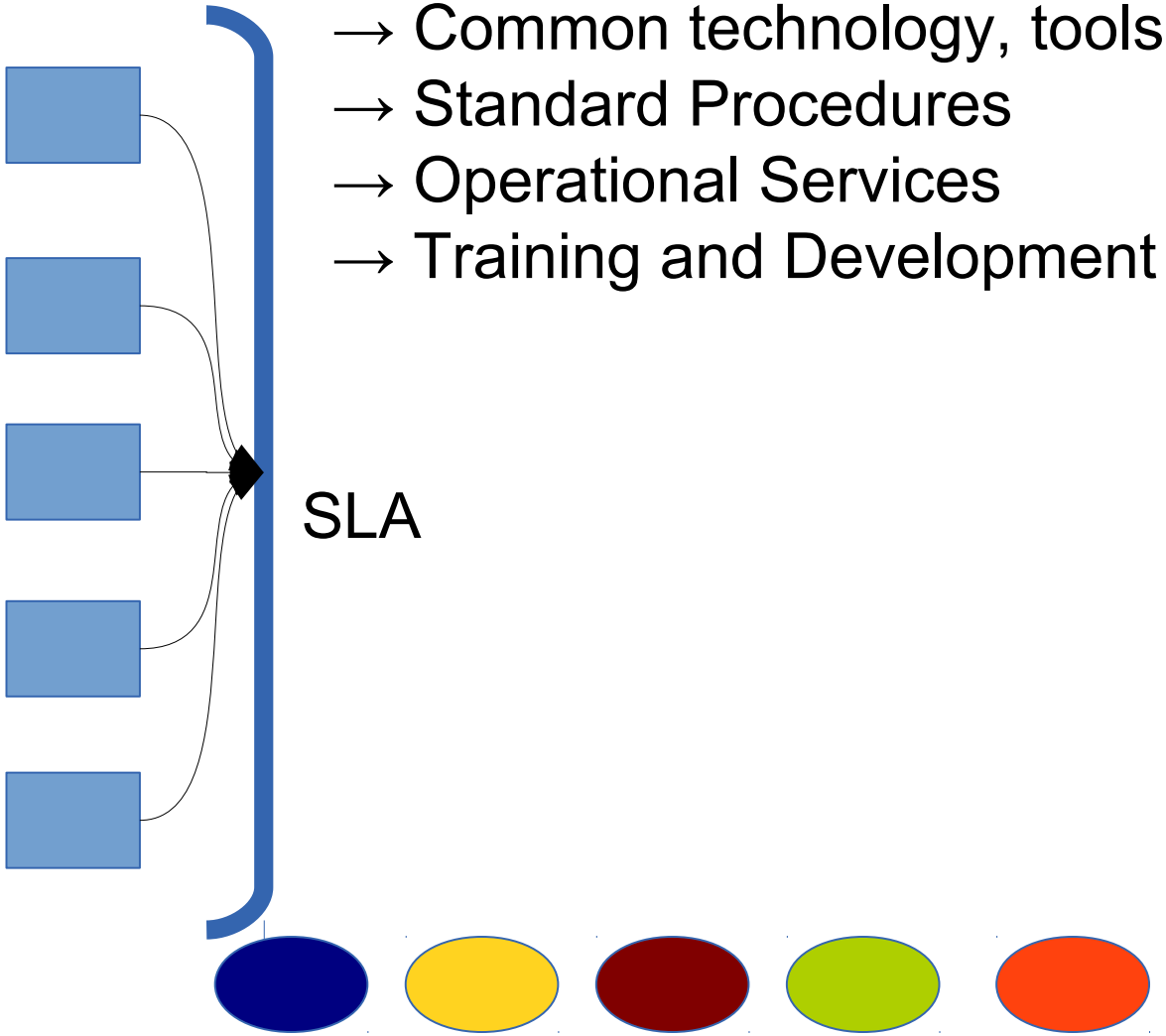
Abstract the resources :

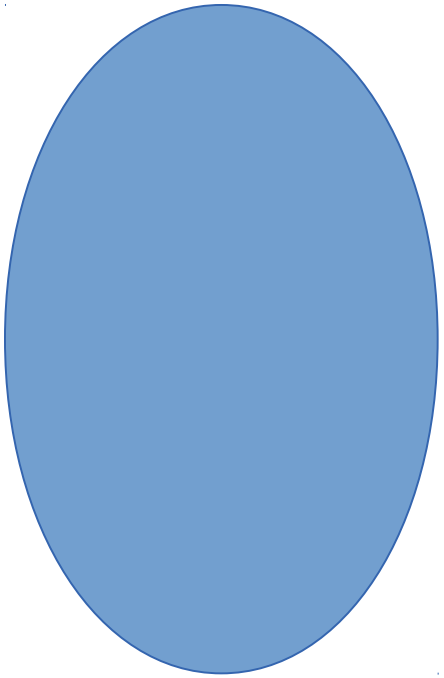
→ Common technology, tools

→ Standard Procedures

→ Operational Services

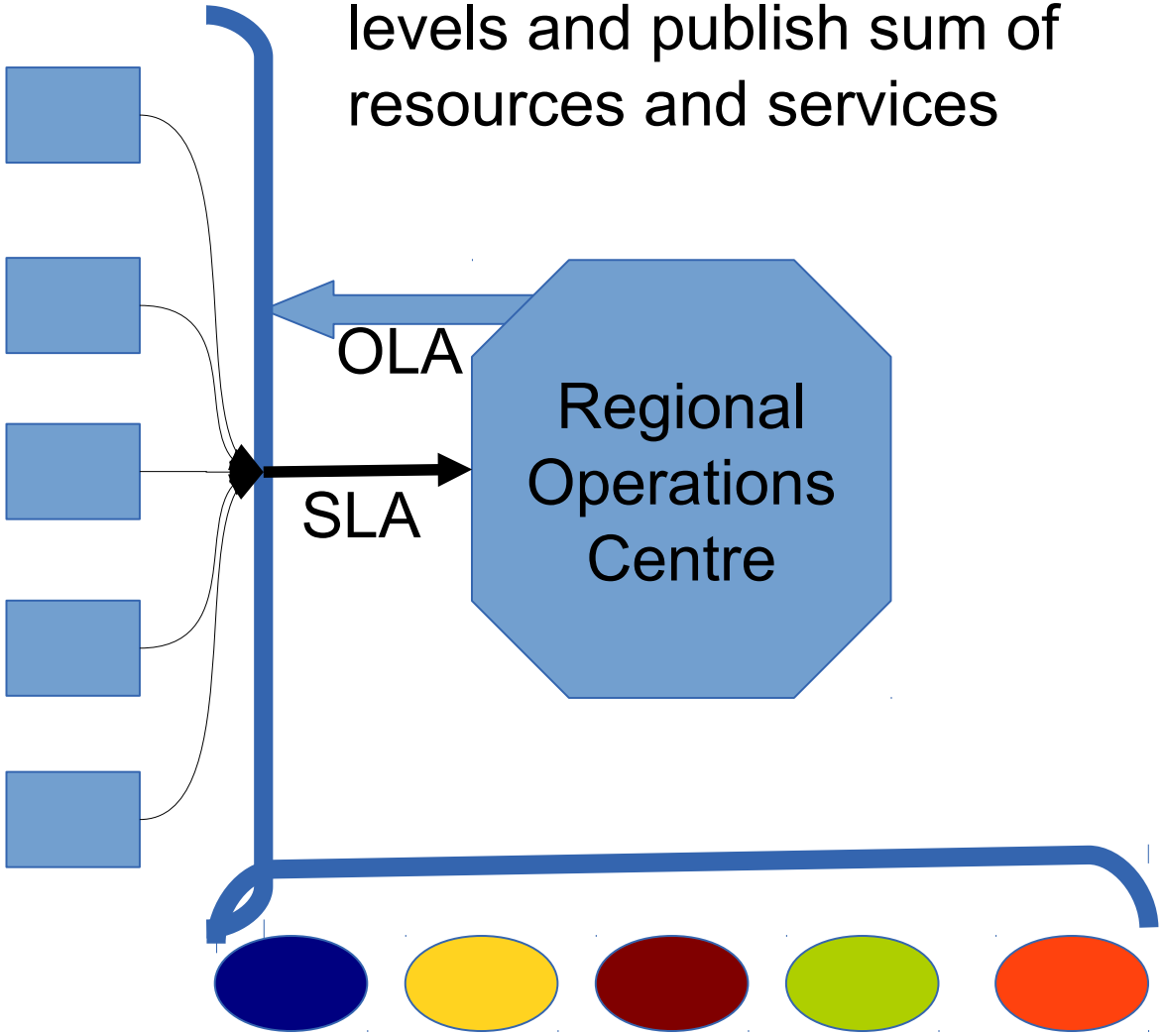
→ Training and Development





Emergent organisation

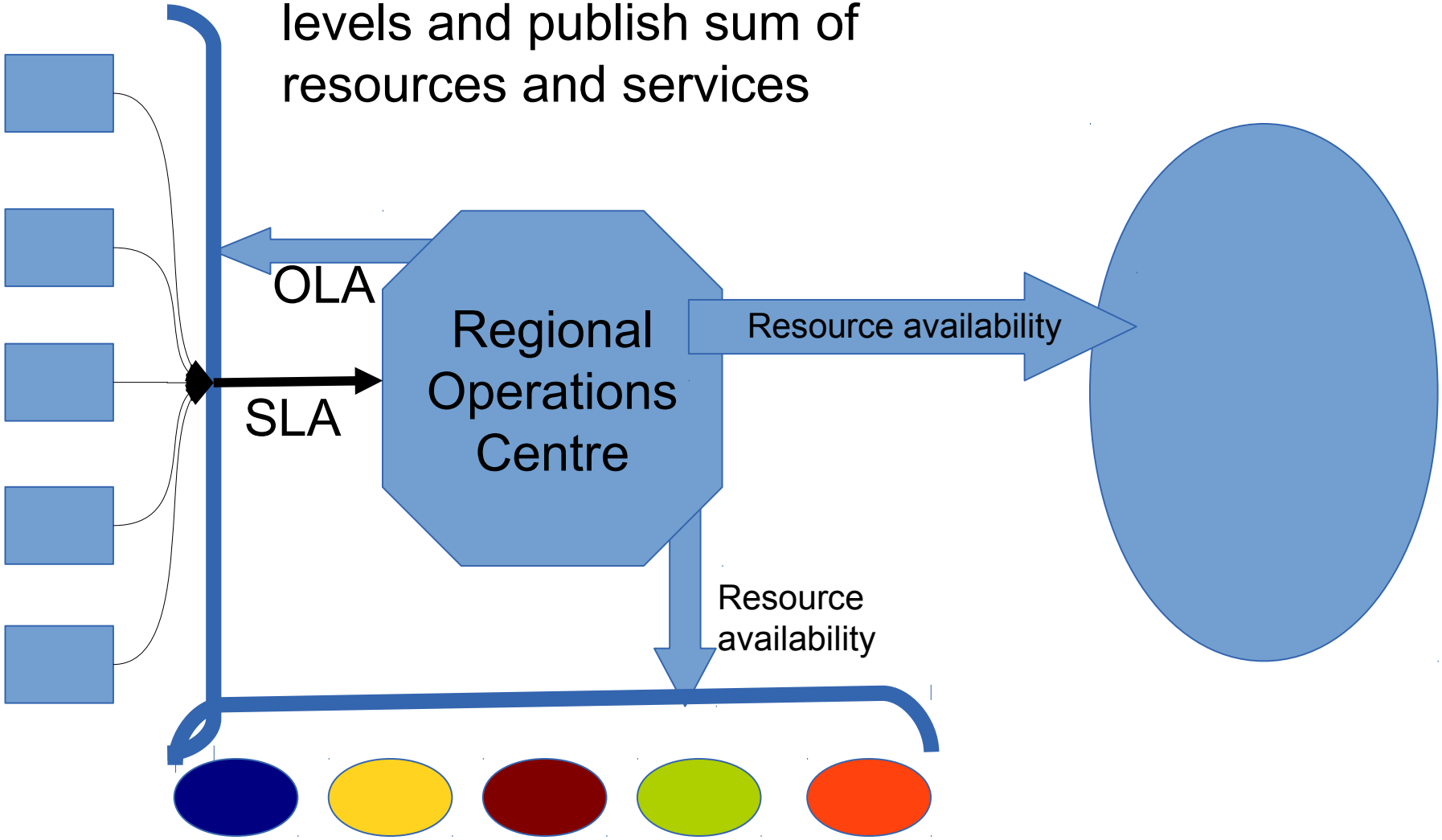
Agree on common operating levels and publish sum of resources and services



VRC sees a single infrastructure, feedback loop restored

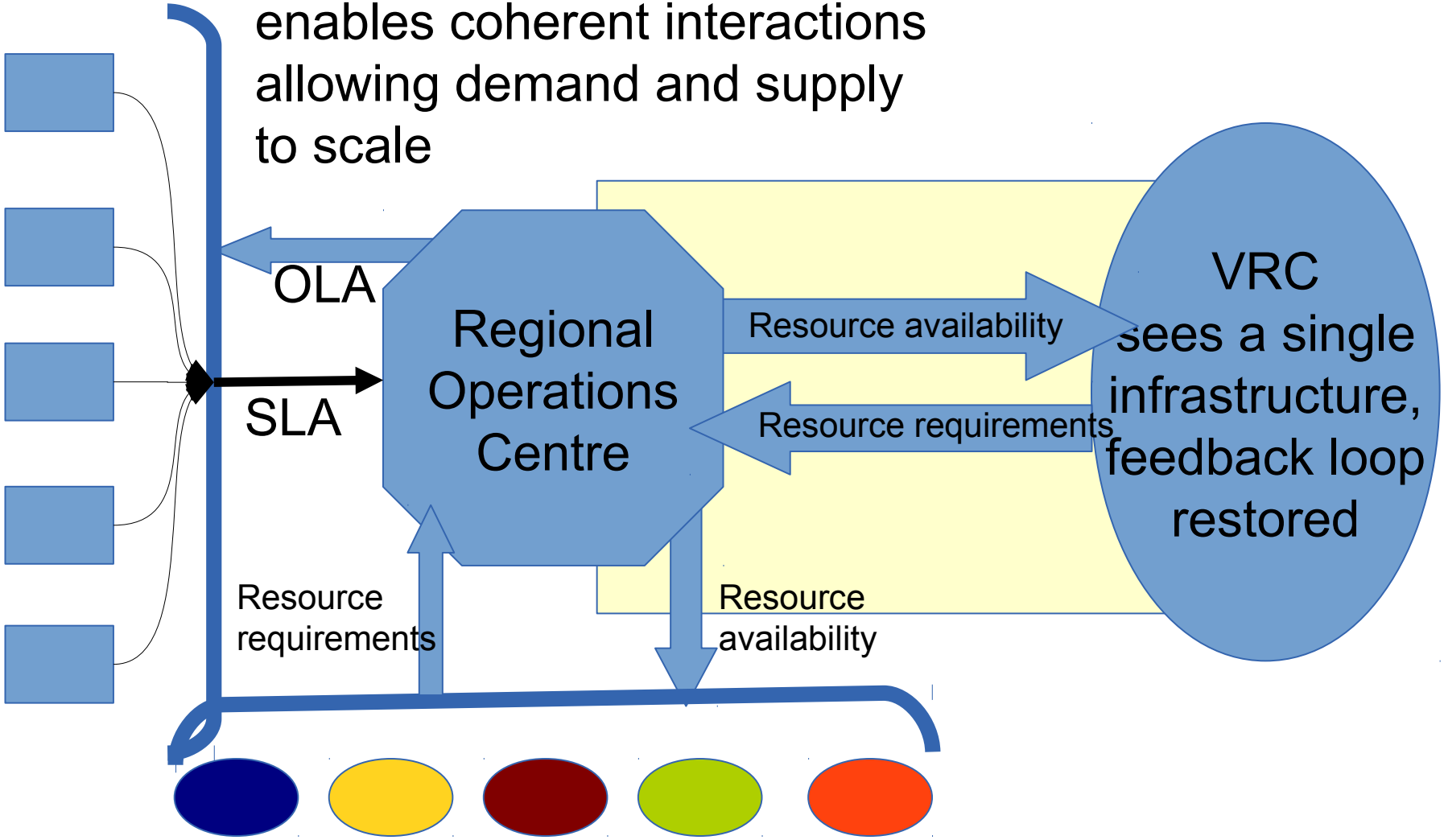
Emergent organisation

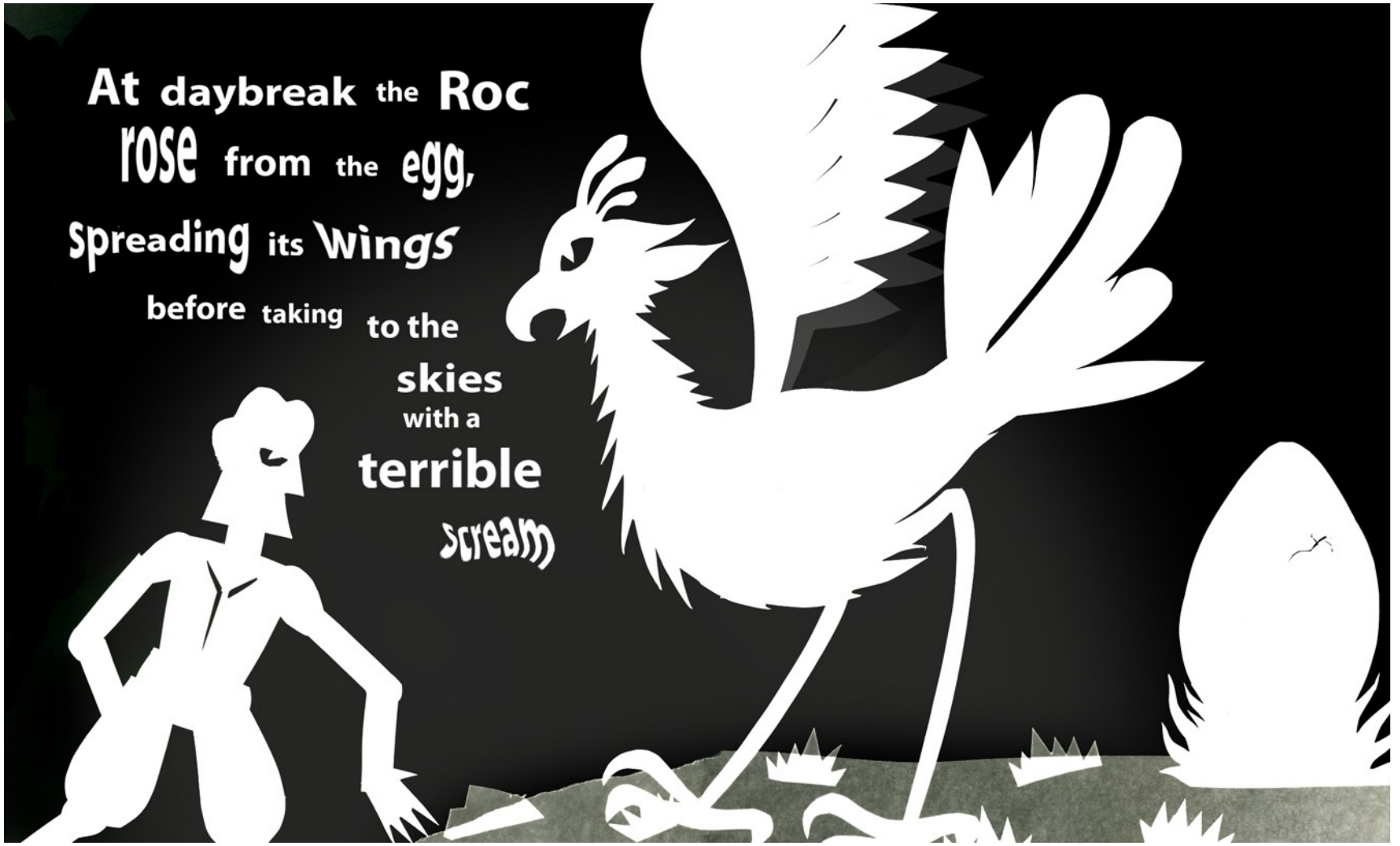
Agree on common operating levels and publish sum of resources and services



Restore feedback loop !

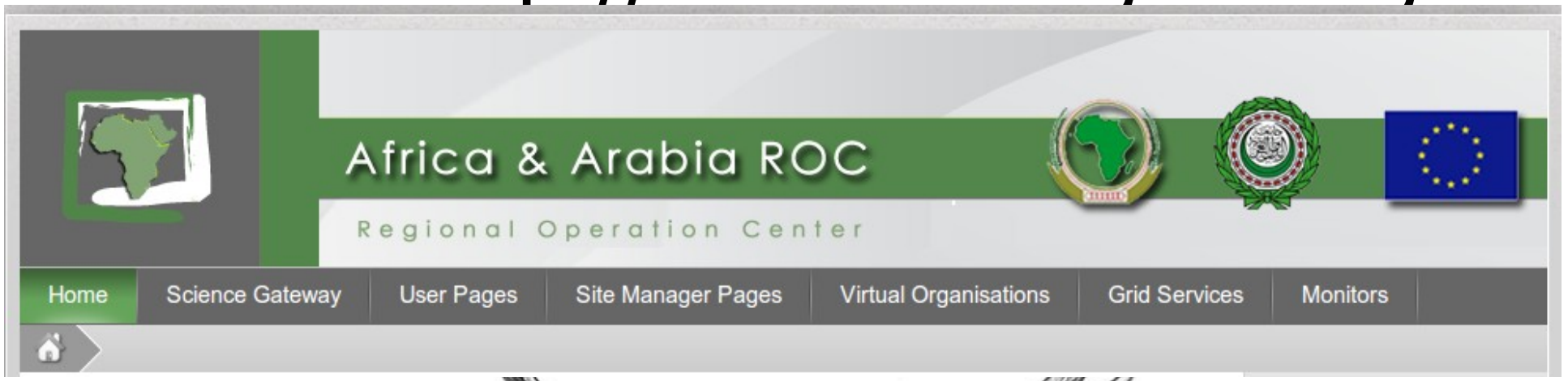
→ Regional Operations Centre enables coherent interactions allowing demand and supply to scale



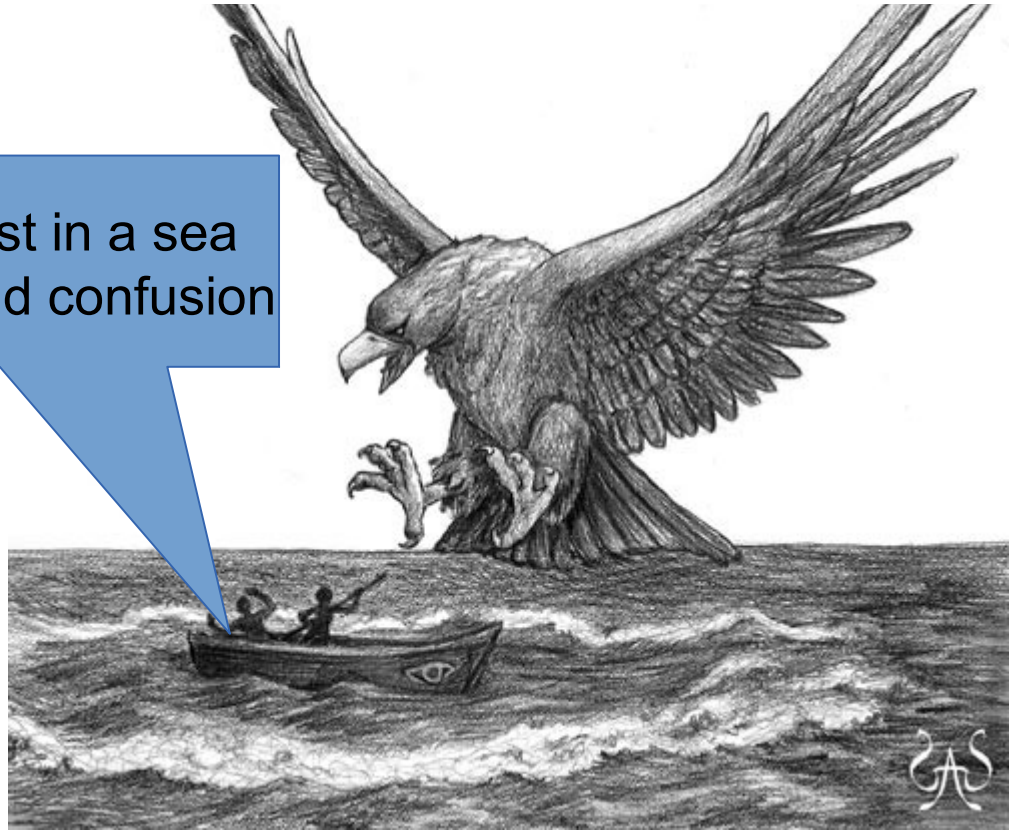

At daybreak the Roc
rose from the egg,
spreading its wings
before taking to the
skies
with a
terrible
scream

http://roc.africa-grid.org

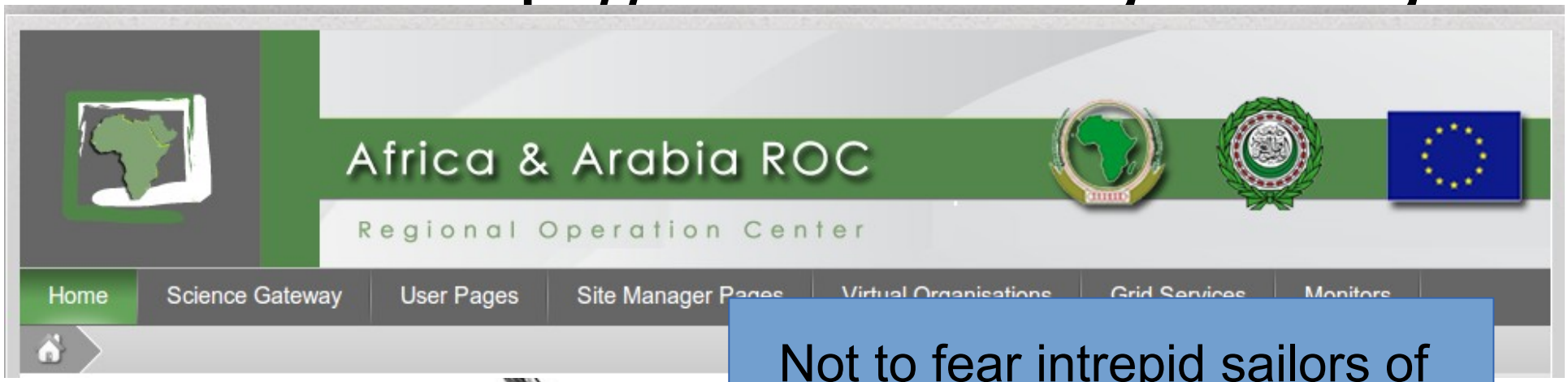
http://roc.africa-grid.org



Help ! We're lost in a sea of uncertainty and confusion

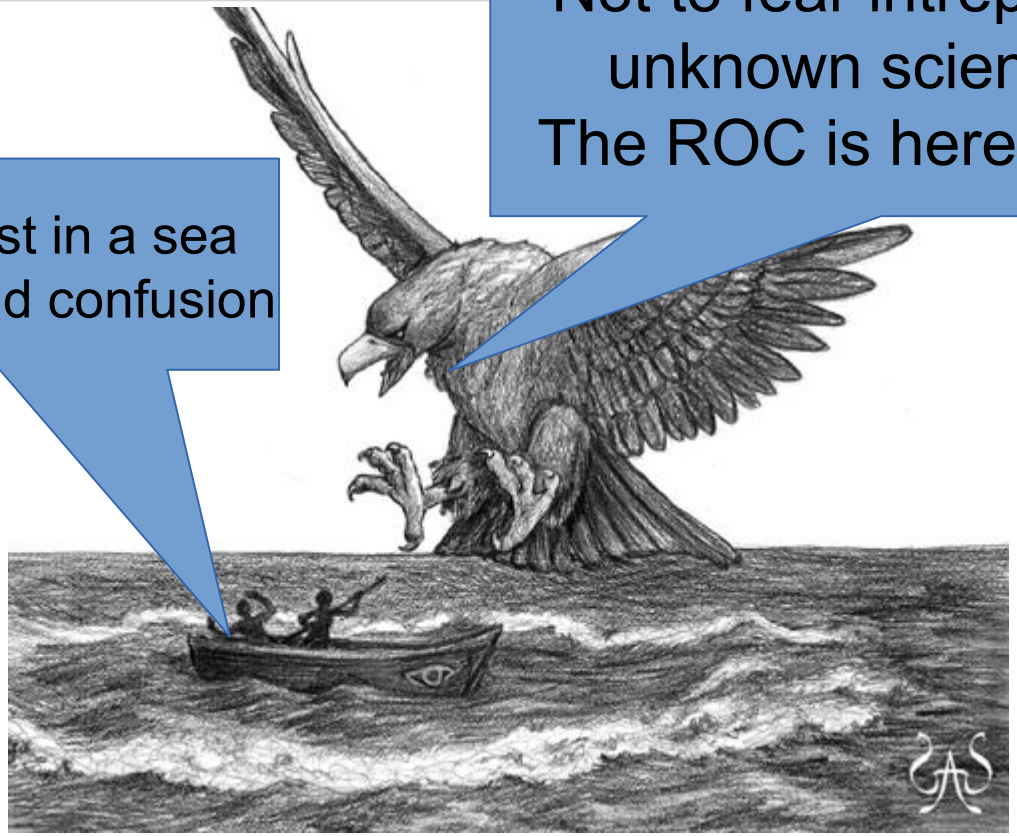


http://roc.africa-grid.org



Not to fear intrepid sailors of unknown scientific seas !
The ROC is here to help you !

Help ! We're lost in a sea of uncertainty and confusion



What is the ROC ? ***Components***

- Point of **coordination** for regional resources.
  - The sum of all **interoperable** computing resources
  - The sum of all **experts** and **support staff** at the sites
  - reference point for best practice and documentation
- **Support** structure for regional operations
  - Interface to **middleware** or **technology** experts via support system
  - Insight into **performance** and **state** of services via monitors
  - Insight into usage via accounting portals

What is the ROC ?
Global Operations Database

- Central service where all sites, services, resources and personell are registered.
  - Is there anything in Namibia ?
  - What is in Namibia ?
  - Who do we call in Namibia if someone from Cameroon wants to collaborate there ?
- Starting point for every other service provided by the ROC

What is the ROC ? ***Procedures***

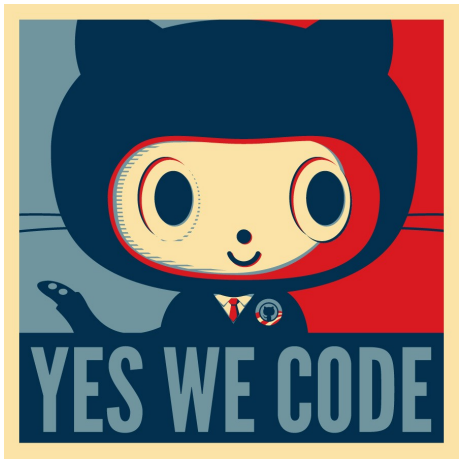
- The power of a distributed computing infrastructure lies on its ***coherence***
  - Sites may employ different **technology** (as long as it's interoperable), but should emply standard **procedures**
- Standard procedures improve ease of collaboration
  - common operational vocabulary
  - Predictable changes in state of services (upgrades, removal, etc)
  - Reduce the need for everyone to be an expert. If you don't know how to do something technical, just follow the recipe
  - Reduce impact of external factors (middleware providers, technology changes)

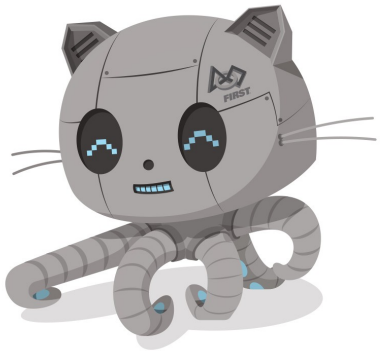
What is the ROC ?
Standard Operating Procedures

- Clear, explicit procedures for performing common tasks
  - Who should I inform when a site is underperforming ?
  - How do I escalate an issue ?
  - How are applications to be deployed ?
  - When should I announce a downtime ?
- Focus on **how** – reduce the need for expert knowledge
- Provide templates for the ”what”

What is the ROC ? ***Automation***

- Distributed computing infrastructure is by its nature a complex environment.
  - Reduce the need to know ***how*** to perform tasks
  - Limit to the need to know ***what*** services you want to deploy
- Rely on a core team of experts to maintain the ”golden standard” for configuration of services





What is the ROC ? ***Services***

- Several services in order to collaborate technically on the infrastructure level:
  - GOCDB
  - Monitoring and alarms
  - Support system, integrated with technology providers and other infrastructures
    - Middleware, network, etc.
  - Accounting portal
- Provides user communities with a coherent view of the infrastructure

ROC on : http://roc.africa-grid.org

- The Africa-Arabia Regional Operations Centre can be a catalyst for **emergent** e-Infrastructure in our region
- **Central** point of contact,
  - direct access to applications via SG
  - Direct access to support and operations team.
- **Reliable** resources are needed
  - Common set of SLA's and OLA's, as well as SOP's
  - Training and documentation updated
- **Scalable** resources are needed
  - Adopt an open, interoperable infrastructure which can easily accommodate both new resources and new users, at scale

ASREN Arab States Research and Education Network

EUMEDGrid Support

chain

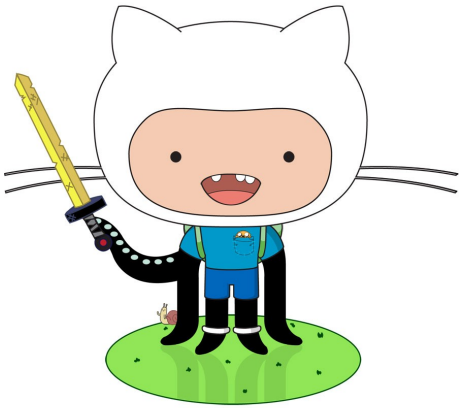
chain reds

el4africa.eu e-Infrastructures for Africa

Prospects for extending the infrastructure

- The ROC is a **catalyst** for new sites, acting as the initial stimulus for development in the absence of sufficient local energy
- However :
  - Care should be taken to not overload the ROC
  - Careful balance between extending the infrastructure and maintaining operational capability
- Computing and data sites in Africa are encouraged to be included in the ROC to publish their resources
- Sites encouraged to **share** their manpower.
- Currently, 1LS/TPM is an Africa-Arabia wide activity

How do we achieve this ?



- Maintain a **predictable feedback loop**
  - Entice big VRCs to use the infrastructure
  - Communicate success
  - develop open, interoperable ecosystem
- Ensure the system remains in **dynamic equilibrium**
  - Ensure that the operations level is in good working order – consistent policies and procedures at all sites
  - Evaluate new applications and use cases in terms of their impact on the infrastructure and vice versa
- Ensure access to a **large "energy bath"** to absorb large fluctuations in load
  - MoU to interoperate with external infrastructures
  - Explore ad-hoc access to external resource providers
- **Ensure that the entire infrastructure is visibile**

Removing barriers

- How does an African infrastructure interoperate with other global infrastructures in order to allow scientists to easily collaborate ?
- N*M MoU's is unfeasible -
  - ROC represents all sites, negotiates peer activities with other regional infrastructures
- Expose a single interface -
  - Single coordination effort required – currently undertaken by Meraka for European interoperation

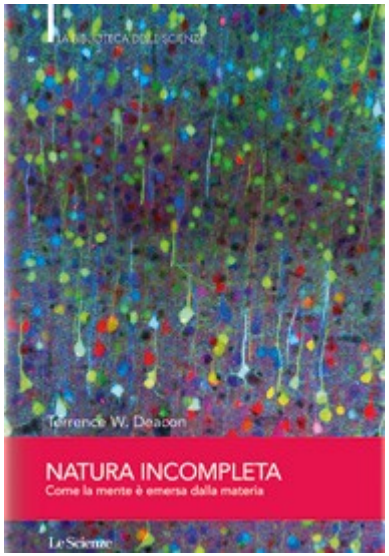
Conclusion

- Things are (thankfully) not as simple as we may think
- A feedback loop between resources and researchers is essential
- A re-organisation of interactions can allow components of a given system to reach higher-lying **states** and **functions**
- Coordination of services and resources is necessary to achieve self-sustainability

Thank you !

- It's a community, come and join us !
  - Africa-Arabia ROC on Facebook
  - On github:
    - https://github.com/AAROC
  - SAGrid on Facebook and Twitter
    - https://www.facebook.com/SAGrid
    - https://www.twitter.com/TheSAGrid

ISBN 978-0-393-04991-6

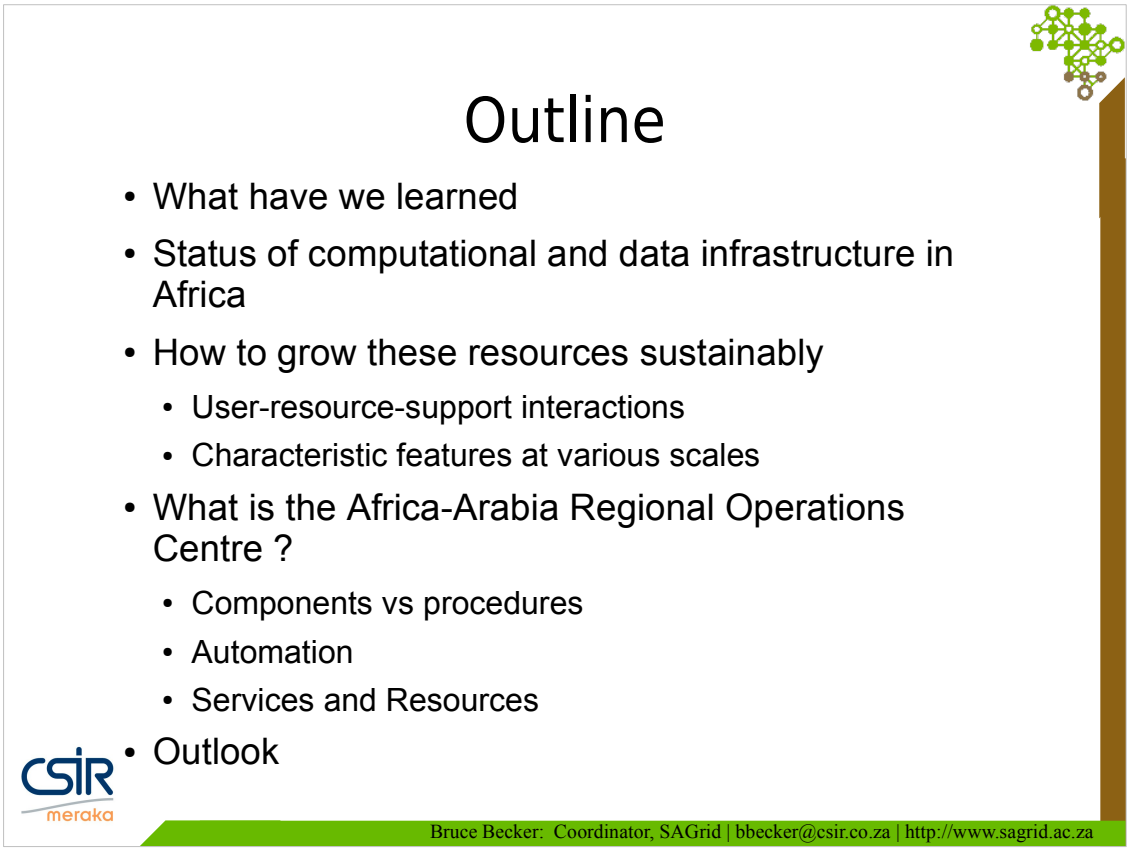


Many of the ideas presented here were inspired by the book
”Natura Incompleta” - Terrence Deacon

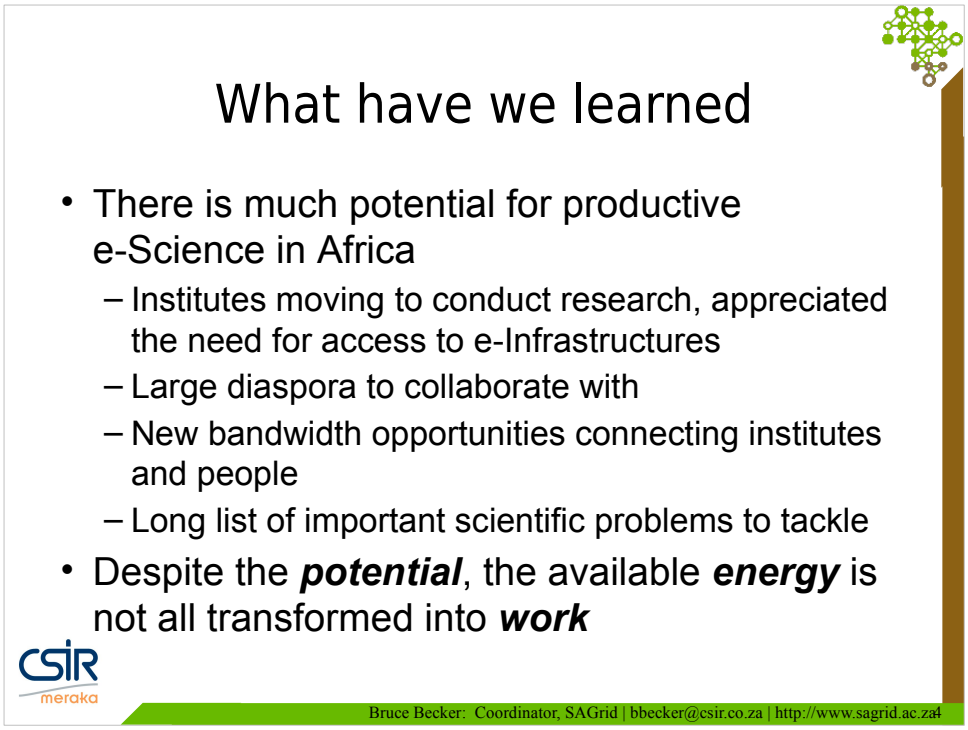
Africa-Arabia Regional Operations Centre

Bruce Becker | bbecker@csir.co.za
Coordinator, SAGrid
Meraka Institute, CSIR

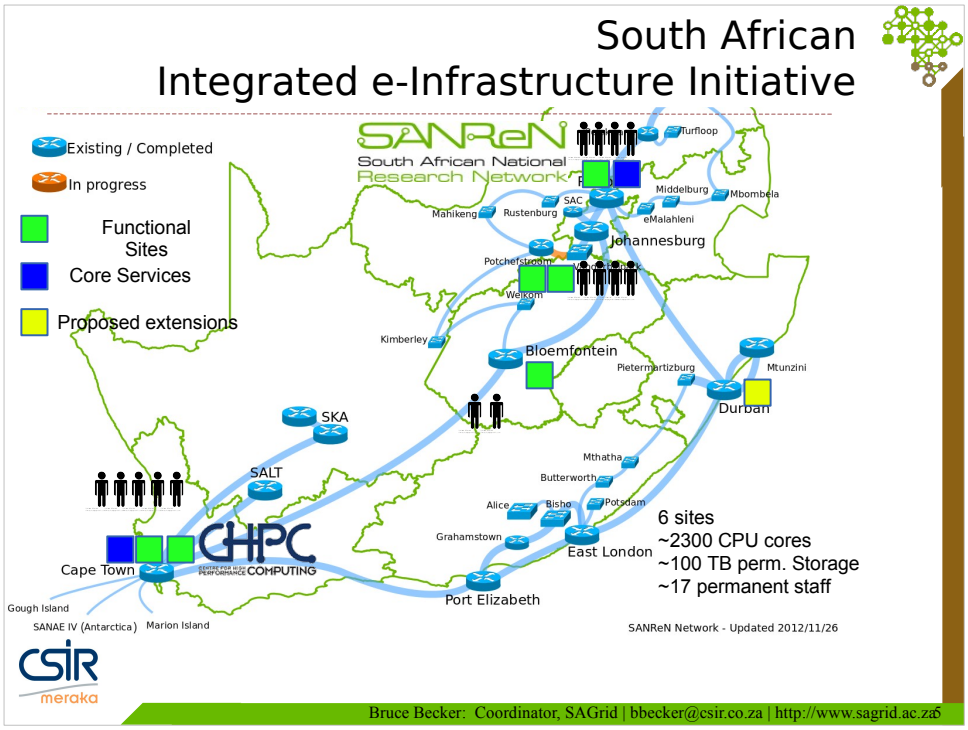
Bruce Becker: Coordinator, SAGrid | bbecker@csir.co.za | http://www.sagrid.ac.za



This is the outline of my presentation. I will describe some of the philosophy, interventions and lessons we have learned in South Arica regarding the development of an integrated e-Infrastructure; then I will briefly give a snapshot of the status of such activities in Africa.
Thereafter, I will go into some consideration on how to sustainably support and grow these, leading into the core of my presentation – the Africa and Arabia Regional Operations Centre. Finally, a summary and outlook will be given.

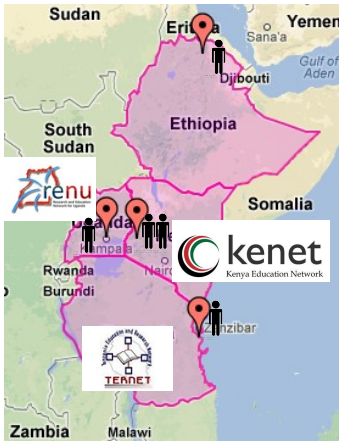


So, what have we learned in the last few years ? A few general trends, being followed worldwide, such as the adoption of the digital paradigm of scientific research, and the role that e-Infrastructures play in it... but we've documented and to some extent liberated the latent potential for collaboration in Africa. The digital divide still separates many, but the increase in access to bandwidt and other digital platforms has been almost revolutionary. It has connected us with a large diaspora with which to collaborate and allowed us to meaningfully contribute to international research project. Importantly, a great many neglected problems have become visible and thus piqued the interest of research communities locally, which are more optimistic of their ability to address these challenges.
However, despitethes positive tones, one must remain realistic: there is indeed vast potential, but the available energy is not all transformed into actual work. (more thermodynamic analogies later)

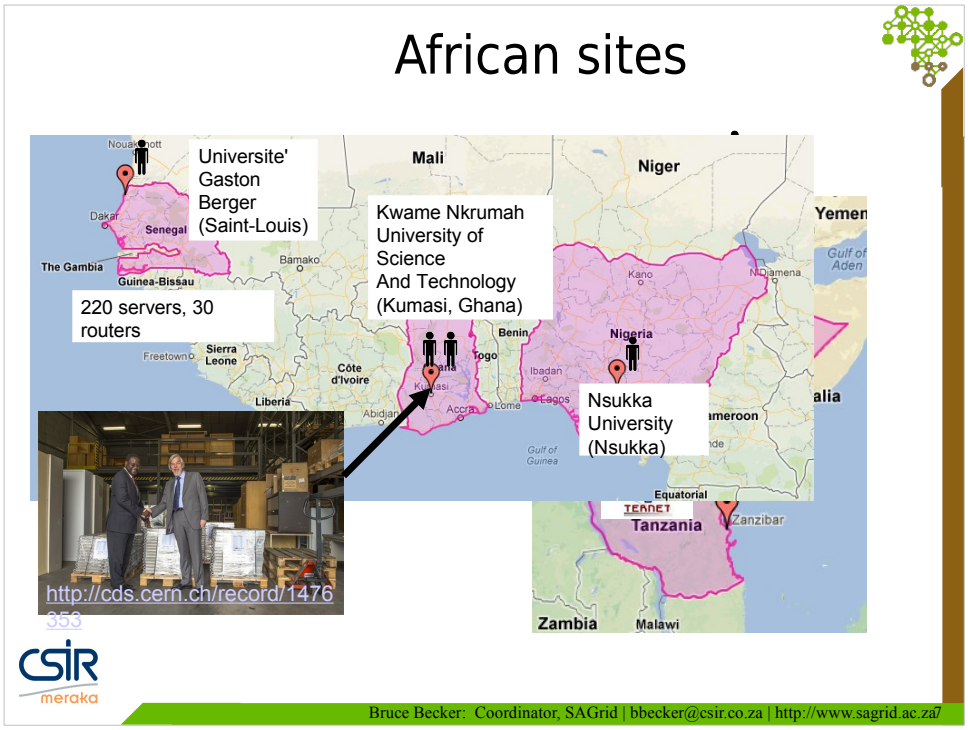


So, what does one of these ”e-Infrastructures” look like ? In South Africa, we have one of the youngest. A significant investment in computing and data resources, both centrally managed and funded by the government,as well as at universities and research laboratories. A complex ecosystem... but the vast majority of the funding so far has gone into the South African Research Network. The South African NREN connects more people and resources than was even conceivable just 5 years ago, at 10 GB/s. The impact of this single investment is difficult to estimate, since it is an ***enabling*** infrastructure and indeed some projects in science and education can now be undertaken thanks to it.
A large portion of the computing resources have been integrated into a distributed computing infrastructure based on the grid paradigm which you are all familiar with – the South African National Grid. You see at the sites the boxes and figures representing the people and services which operate as a single entity in a federation.

Other African sites



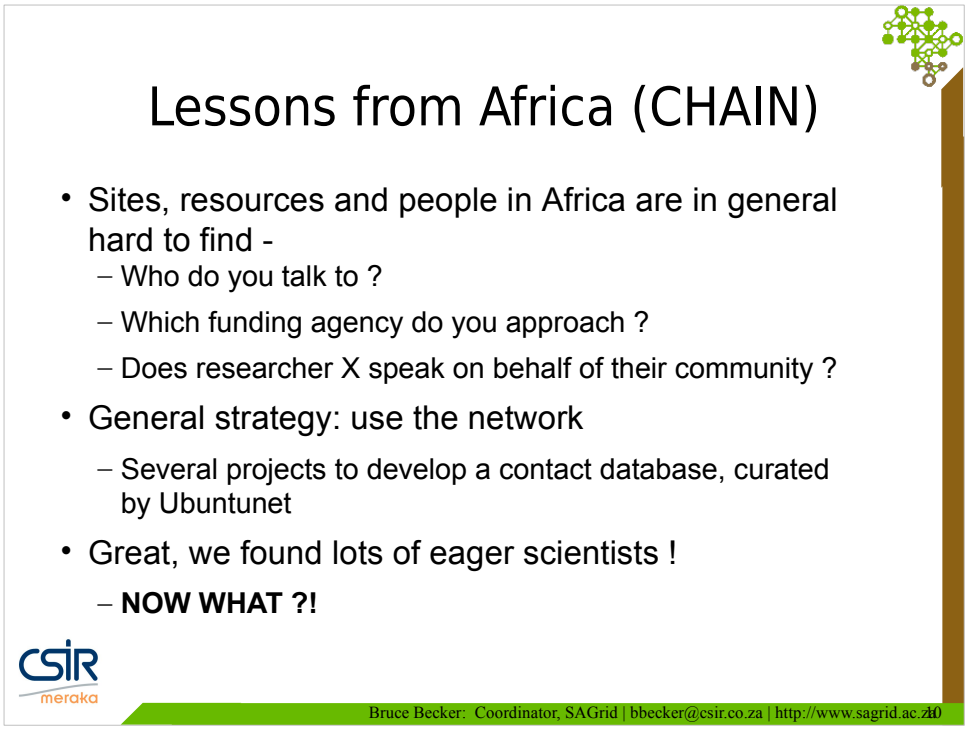
Now, South Africa is of course not alone in these developments. The BGI project itself has pushed such developments on both sides of the continent, not to mention those on the northern coast, supported by EUMedGrid, Support and Connect. There are strong NRENs in Kenya and Tanzania, which, together with Uganda, form good regional capacity in Eastern Africa.



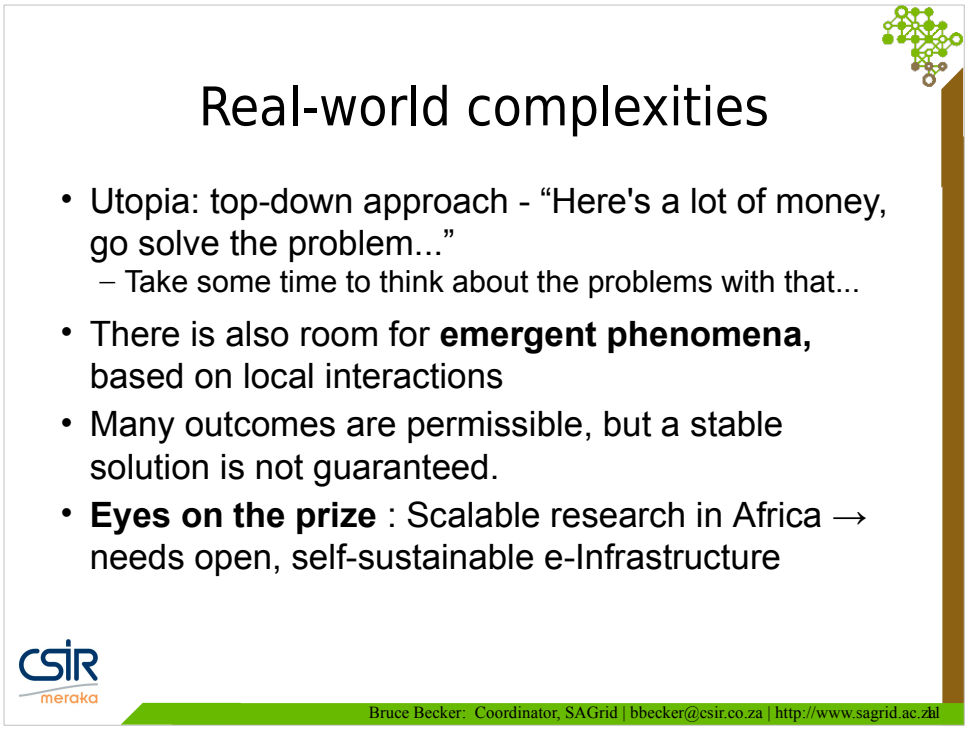
West Africa is of course a bit more complex than other parts of our continent, but here we see a lot of activity. In order to cut down on time, I've shown only a few, such as the site at UCGB in Seneal, the site at Nsukka University in Nigeria and the site under development at Kwame Nkrumah University of Science and Technology, which recently was the recipient of a large transfer of materiel from CERN.
As in the case of Eastern Africa, these form the nucleus the model I will describe in a moment. But let us go back to South Africa for a moment and take stock.

Let us start off with a matter of fact. There are plenty of resources and plenty of activity. The question we have to pose ourselves then is:
How can we ensure that this
a) transitions from being a bunch of disconnected resources to a true infrastructure ?
And
b) how can we sustainably support this infrastructure ?

So, South Africa has some serious e-Infrastructure with a long-term focus (and the SKA bid was of course central to this): the biggest supercomputer in Africa, at the Centre for High Performance Computing in Cape Town; The fastest NREN in Africa as I just mentioned, and the biggest academic data repository in Africa, currently under development in the DIRISA project. To the outside observer, it may appear that in order to achieve success one must adopt a top-down approac and throw money at the problem until it goes away !
But how applicable is this in other parts of Africa where there is a smaller capacity to support such initiatives and perhaps no single driving force, as we were lucky to have in the case of the SKA ? Should we give up … or give ourselves to politics to attempt to change ”the system” ?
There is another way, and I want to convince you all that there is plenty of room for ***emergent*** organisation, without having to rely on a heavy top-down intervention.



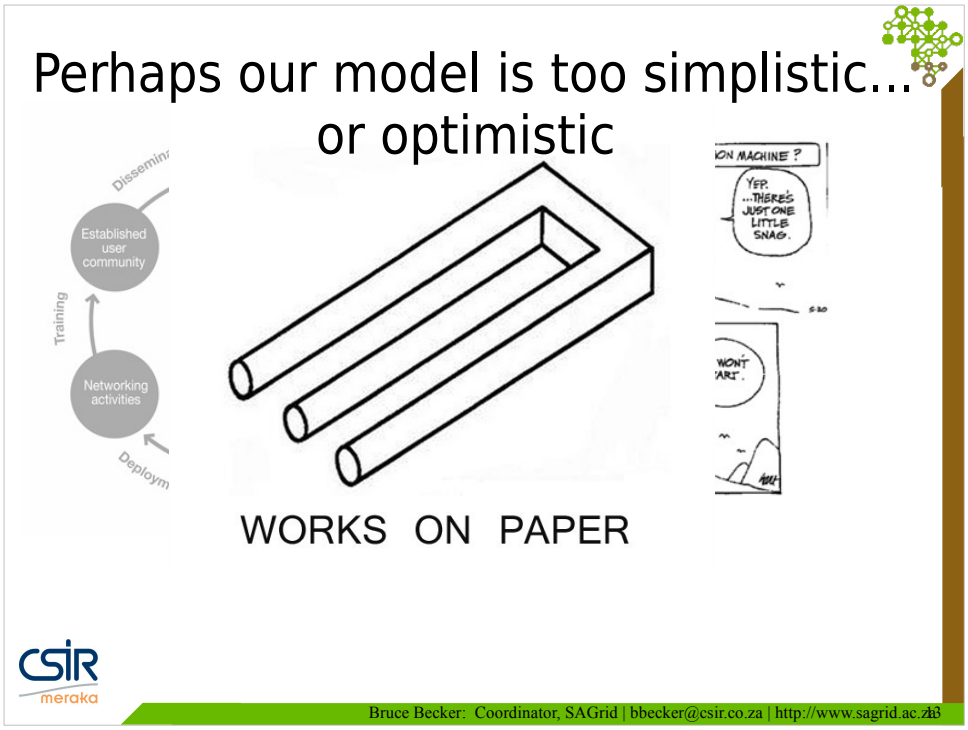
So, South Africa has some serious e-Infrastructure with a long-term focus (and the SKA bid was of course central to this): the biggest supercomputer in Africa, at the Centre for High Performance Computing in Cape Town; The fastest NREN in Africa as I just mentioned, and the biggest academic data repository in Africa, currently under development in the DIRISA project. To the outside observer, it may appear that in order to achieve success one must adopt a top-down approac and throw money at the problem until it goes away !
But how applicable is this in other parts of Africa where there is a smaller capacity to support such initiatives and perhaps no single driving force, as we were lucky to have in the case of the SKA ? Should we give up … or give ourselves to politics to attempt to change ”the system” ?
There is another way, and I want to convince you all that there is plenty of room for ***emergent*** organisation, without having to rely on a heavy top-down intervention.



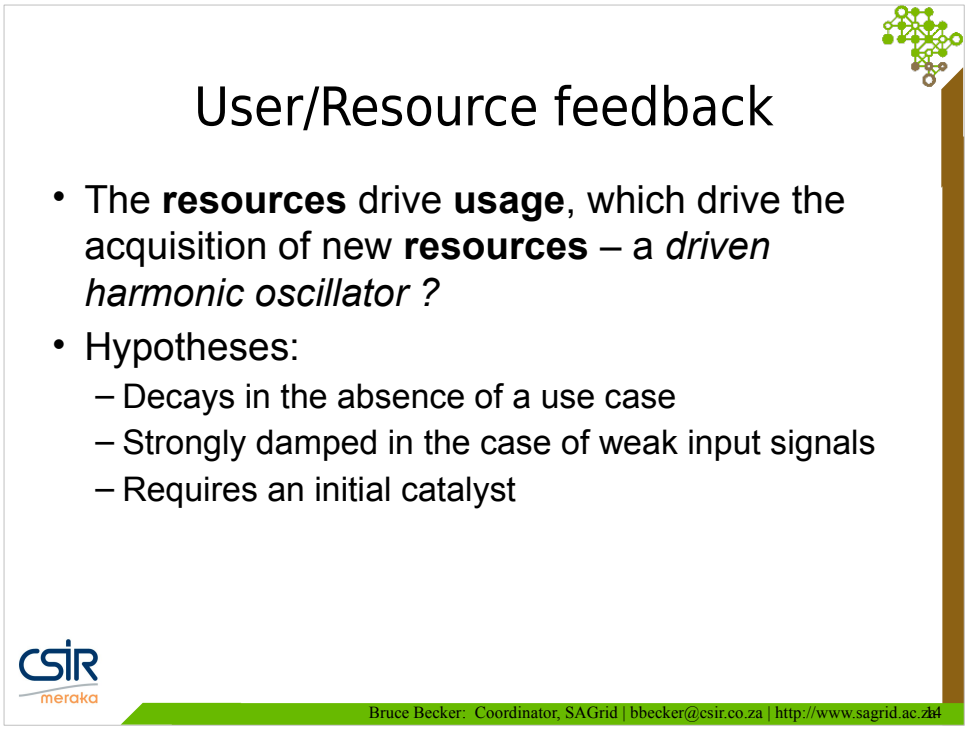
So, South Africa has some serious e-Infrastructure with a long-term focus (and the SKA bid was of course central to this): the biggest supercomputer in Africa, at the Centre for High Performance Computing in Cape Town; The fastest NREN in Africa as I just mentioned, and the biggest academic data repository in Africa, currently under development in the DIRISA project. To the outside observer, it may appear that in order to achieve success one must adopt a top-down approac and throw money at the problem until it goes away !
But how applicable is this in other parts of Africa where there is a smaller capacity to support such initiatives and perhaps no single driving force, as we were lucky to have in the case of the SKA ? Should we give up … or give ourselves to politics to attempt to change ”the system” ?
There is another way, and I want to convince you all that there is plenty of room for ***emergent*** organisation, without having to rely on a heavy top-down intervention.

Perhaps our model is too simplistic... or optimistic

Of course, this way is a bit more complicated... so, perhaps we should look again at the models we've been using to generate success. Starting with the linear interpretation of the so-called ”virtuous cycle”, where it is assumed that success is at worst self-sustaining. This is like referring to the perpetual motion machine you see on the right... and the question remains ”how do we get it started”!?

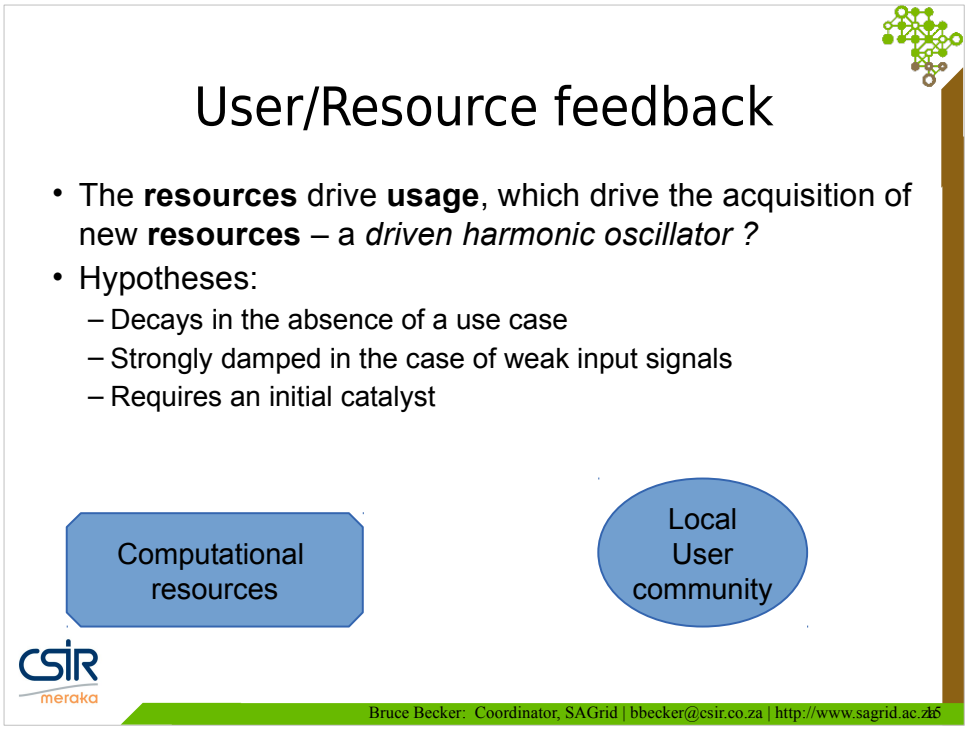


These ideas work in theory, but now it has come time to apply our experience in our wonderfully complex environment, which is Africa.



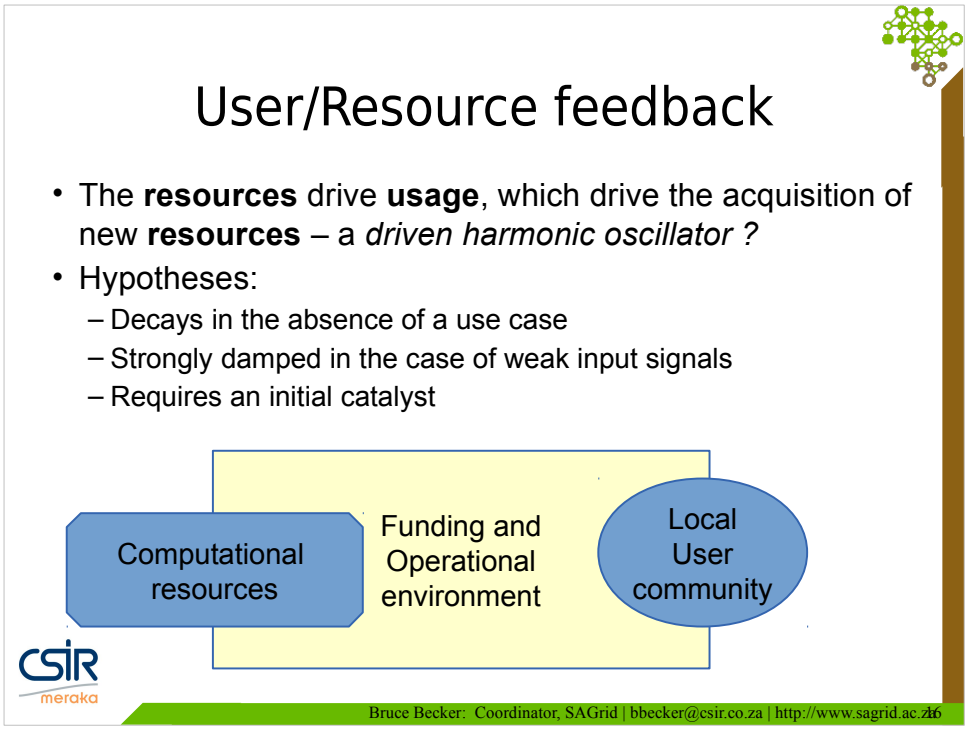
So,let us make a perturbation on our previous assumptions which were so linear in nature, by introducing a feedback loop between resources and their user community. Let us make the reasonable proposal that the existence of resources can drive their usage, which can drive the increase in the number of resources. To a physicist,this sounds like a driven harmonic oscillator, which I've illustrated below. In our model the existence of resources enables a user community to undertake research, which justifies support for the same or more resources. If this sounds like a simplified version of our our ”virtuous cycle” from before, it is because we have not yet placed our model in the real world...
So, let's make the additional hypotheses that :
1) The entire system decays in the absence of a ***use case***
2) the system is strongly damped in the case of ***weak input signals***
3) In order to start the oscillator, an initial catalyst is needed.

You can see now that our model is a bit more realistic and mirrors, I'm sure a lot of our personal experiences : an exciting new project comes along, but after a few months is dead or never gets started due to the factors I just mentioned.



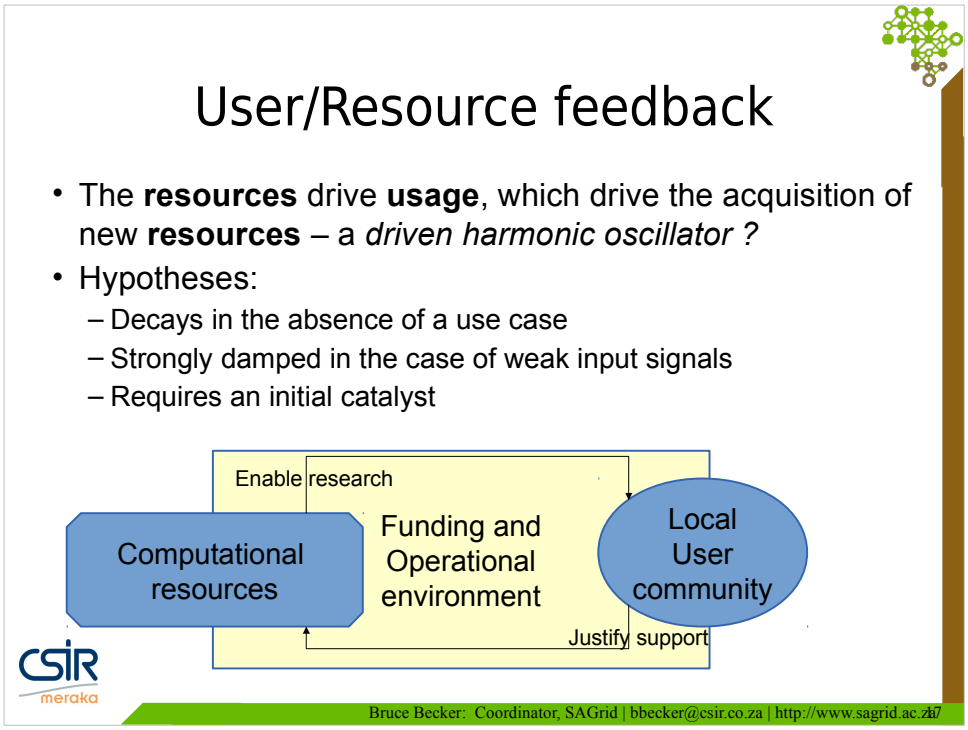
So,let us make a perturbation on our previous assumptions which were so linear in nature, by introducing a feedback loop between resources and their user community. Let us make the reasonable proposal that the existence of resources can drive their usage, which can drive the increase in the number of resources. To a physicist,this sounds like a driven harmonic oscillator, which I've illustrated below. In our model the existence of resources enables a user community to undertake research, which justifies support for the same or more resources. If this sounds like a simplified version of our our ”virtuous cycle” from before, it is because we have not yet placed our model in the real world...
So, let's make the additional hypotheses that :
1) The entire system decays in the absence of a ***use case***
2) the system is strongly damped in the case of ***weak input signals***
3) In order to start the oscillator, an initial catalyst is needed.

You can see now that our model is a bit more realistic and mirrors, I'm sure a lot of our personal experiences : an exciting new project comes along, but after a few months is dead or never gets started due to the factors I just mentioned.



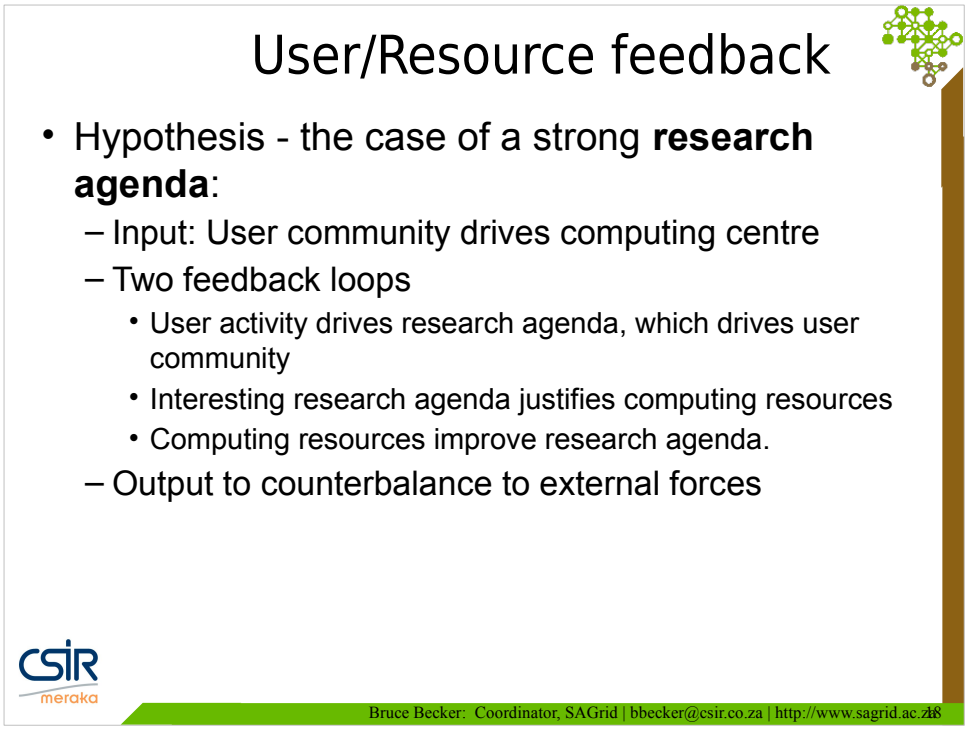
So,let us make a perturbation on our previous assumptions which were so linear in nature, by introducing a feedback loop between resources and their user community. Let us make the reasonable proposal that the existence of resources can drive their usage, which can drive the increase in the number of resources. To a physicist,this sounds like a driven harmonic oscillator, which I've illustrated below. In our model the existence of resources enables a user community to undertake research, which justifies support for the same or more resources. If this sounds like a simplified version of our our ”virtuous cycle” from before, it is because we have not yet placed our model in the real world...
So, let's make the additional hypotheses that :
1) The entire system decays in the absence of a ***use case***
2) the system is strongly damped in the case of ***weak input signals***
3) In order to start the oscillator, an initial catalyst is needed.

You can see now that our model is a bit more realistic and mirrors, I'm sure a lot of our personal experiences : an exciting new project comes along, but after a few months is dead or never gets started due to the factors I just mentioned.

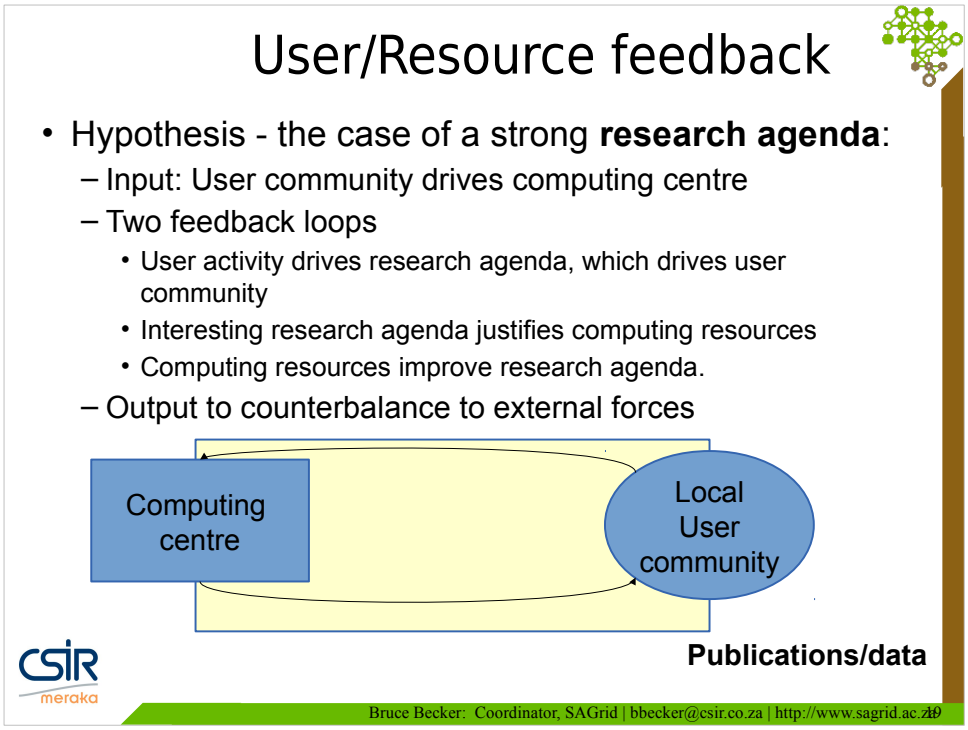


So,let us make a perturbation on our previous assumptions which were so linear in nature, by introducing a feedback loop between resources and their user community. Let us make the reasonable proposal that the existence of resources can drive their usage, which can drive the increase in the number of resources. To a physicist,this sounds like a driven harmonic oscillator, which I've illustrated below. In our model the existence of resources enables a user community to undertake research, which justifies support for the same or more resources. If this sounds like a simplified version of our our ”virtuous cycle” from before, it is because we have not yet placed our model in the real world...
So, let's make the additional hypotheses that :
1) The entire system decays in the absence of a ***use case***
2) the system is strongly damped in the case of ***weak input signals***
3) In order to start the oscillator, an initial catalyst is needed.

You can see now that our model is a bit more realistic and mirrors, I'm sure a lot of our personal experiences : an exciting new project comes along, but after a few months is dead or never gets started due to the factors I just mentioned.



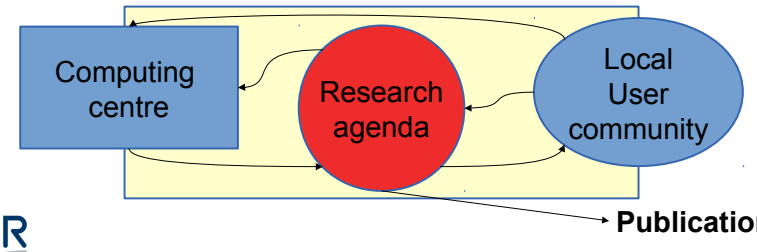
Of course, this can be resovled with an external force to artificially drive the system, but it can also be self-driven, in the case of an internal ”energy”. In our model, this is the existence of a strong research agenda. We now can have two more internal feedback loops:
1) the user activity drives the research agenda,
2) which justifies existence and support of the computing centre as before, but also
3) produces WORK, in the form of publications and/or data, which counterbalance the forces of decay we had in our simpler model..
4) The better-justified computing resources are therefore better supported and in turn improve the capacity of the user community to conduct their research agenda...
Note that a fundamental aspect of this slightly more complicated model is the fact that there is **output** from the system. This output has to be valid and visible !



Of course, this can be resovled with an external force to artificially drive the system, but it can also be self-driven, in the case of an internal ”energy”. In our model, this is the existence of a strong research agenda. We now can have two more internal feedback loops:
1) the user activity drives the research agenda,
2) which justifies existence and support of the computing centre as before, but also
3) produces WORK, in the form of publications and/or data, which counterbalance the forces of decay we had in our simpler model..
4) The better-justified computing resources are therefore better supported and in turn improve the capacity of the user community to conduct their research agenda...
Note that a fundamental aspect of this slightly more complicated model is the fact that there is **output** from the system. This output has to be valid and visible !

User/Resource feedback

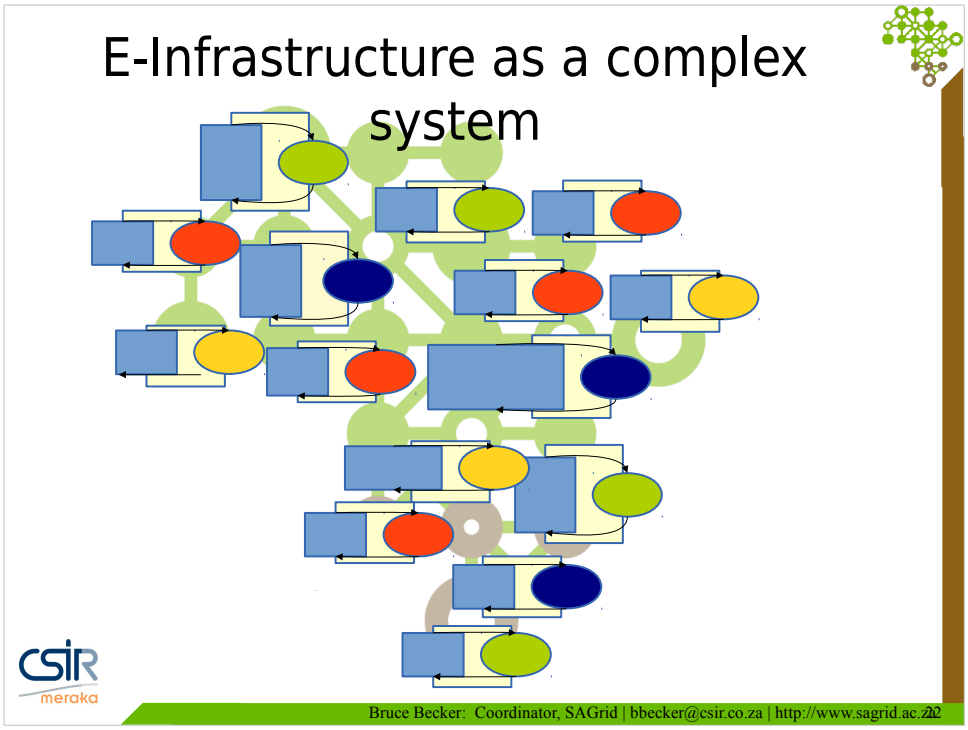
- Hypothesis - the case of a strong **research agenda**:
  - Input: User community drives computing centre
  - Two feedback loops
    - User activity drives research agenda, which drives user community
    - Interesting research agenda justifies computing resources
    - Computing resources improve research agenda.
  - Output to counterbalance to external forces



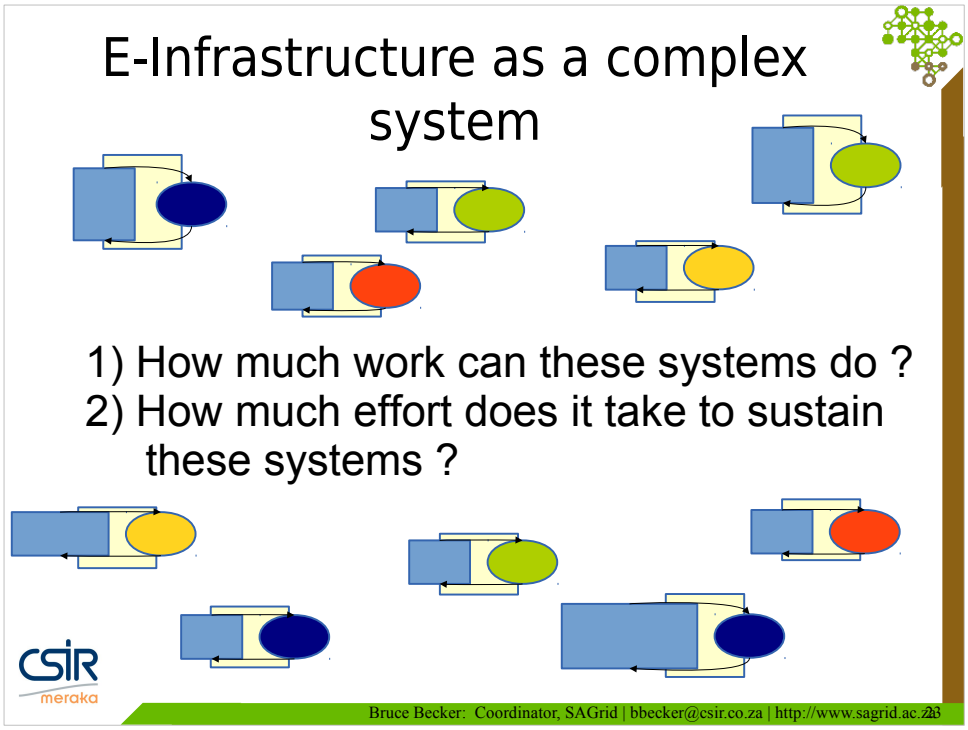
Publications/data

Of course, this can be resovled with an external force to artificially drive the system, but it can also be self-driven, in the case of an internal ”energy”. In our model, this is the existence of a strong research agenda. We now can have two more internal feedback loops:
1) the user activity drives the research agenda,
2) which justifies existence and support of the computing centre as before, but also
3) produces WORK, in the form of publications and/or data, which counterbalance the forces of decay we had in our simpler model..
4) The better-justified computing resources are therefore better supported and in turn improve the capacity of the user community to conduct their research agenda...
Note that a fundamental aspect of this slightly more complicated model is the fact that there is **output** from the system. This output has to be valid and visible !

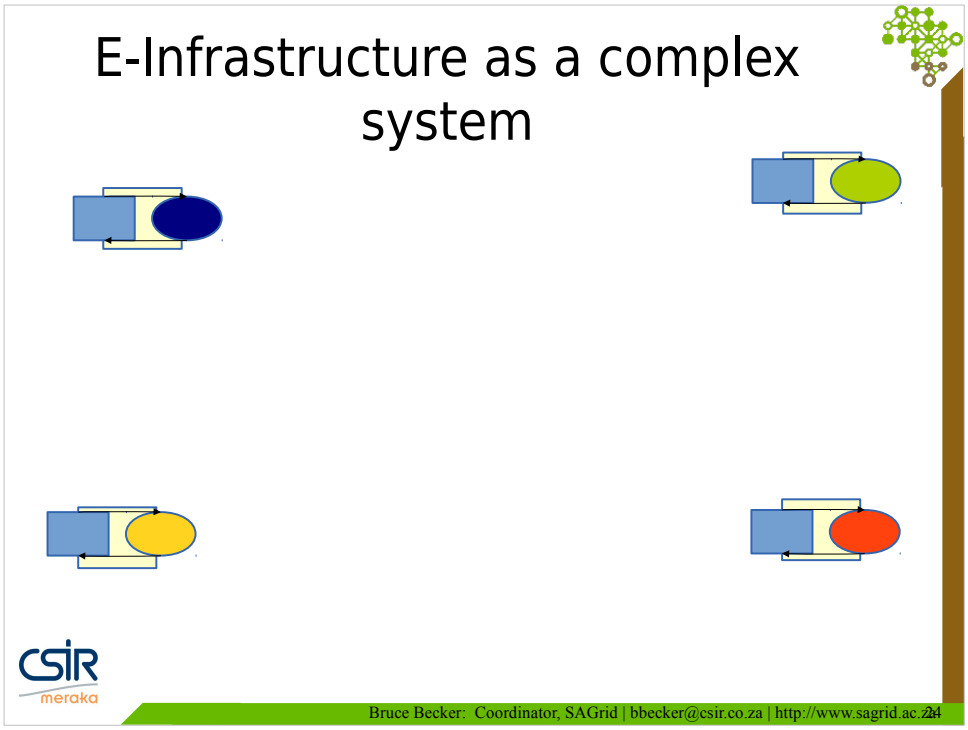
Works on paper...
how do we scale it ?



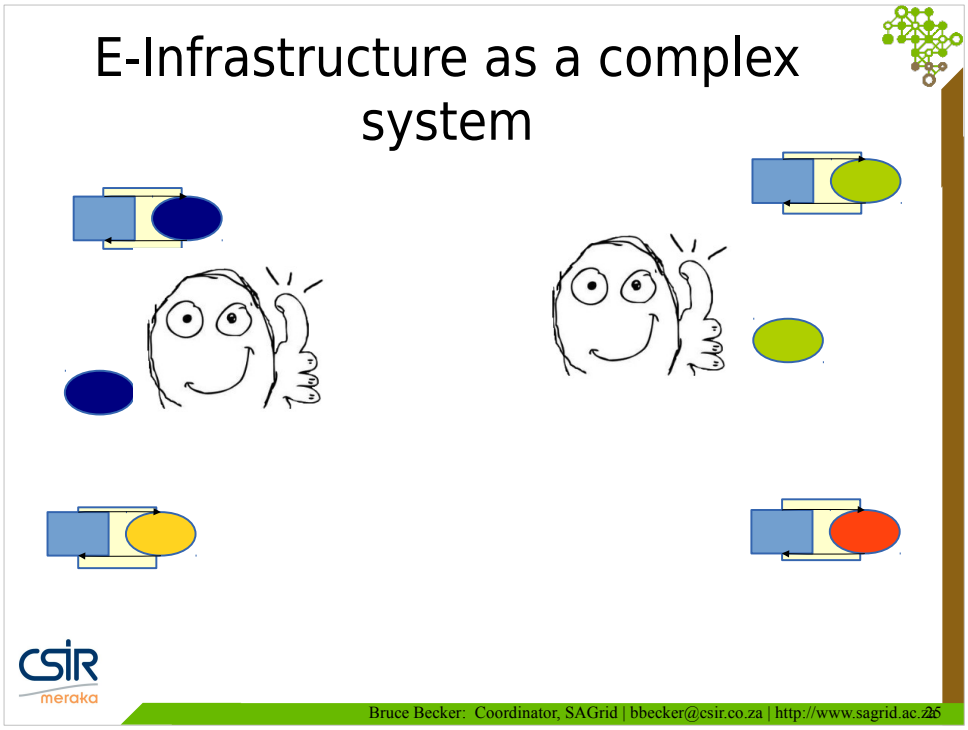
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



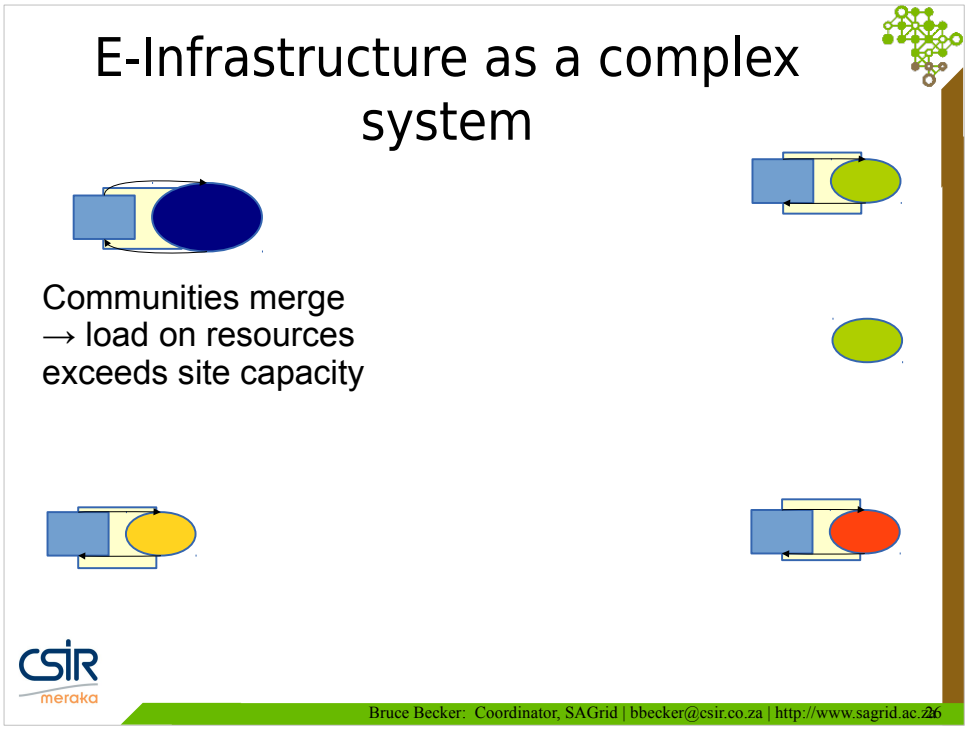
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



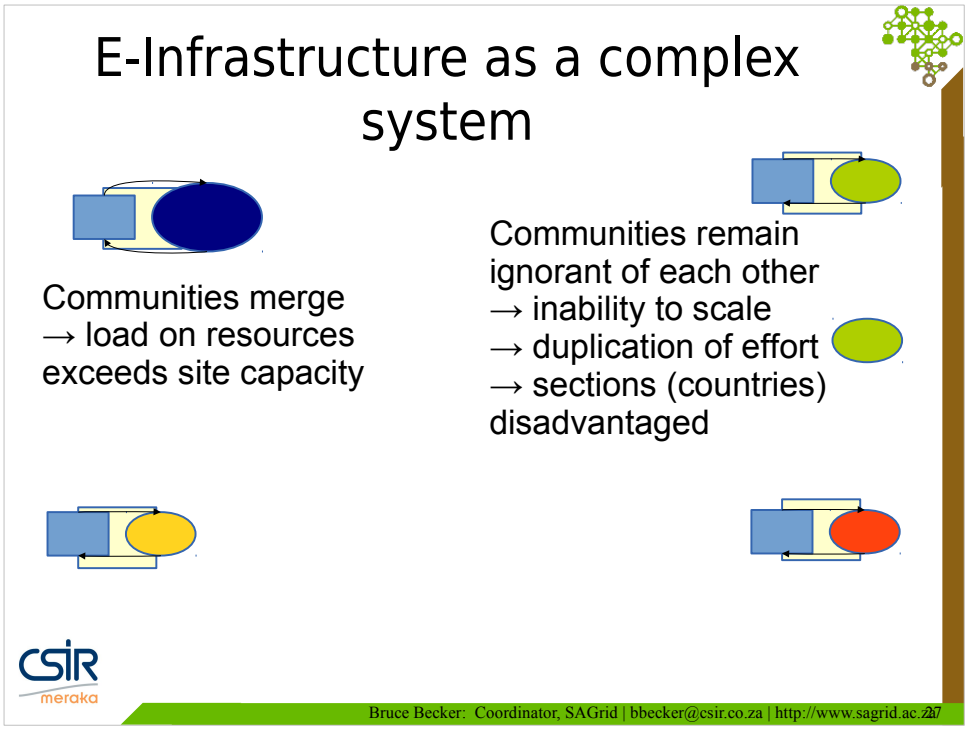
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



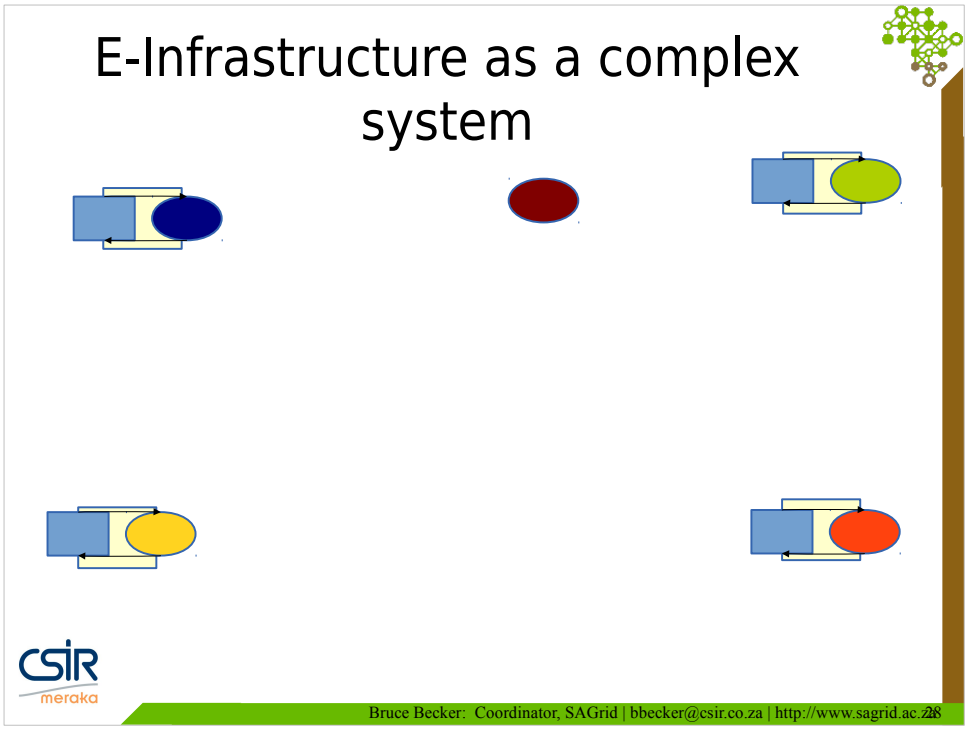
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



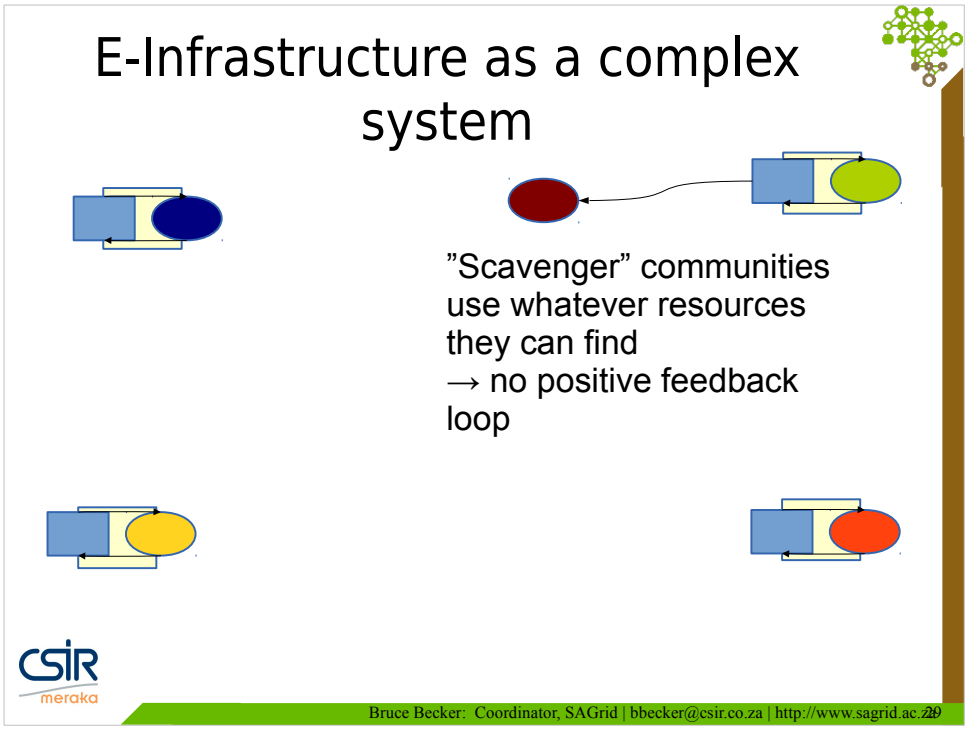
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



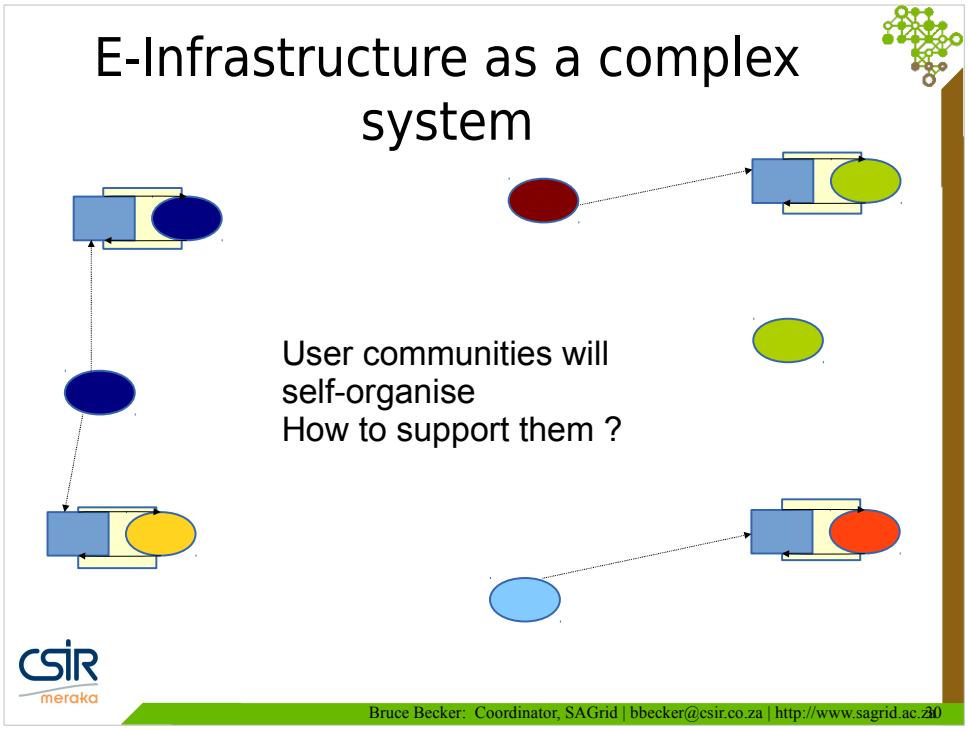
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



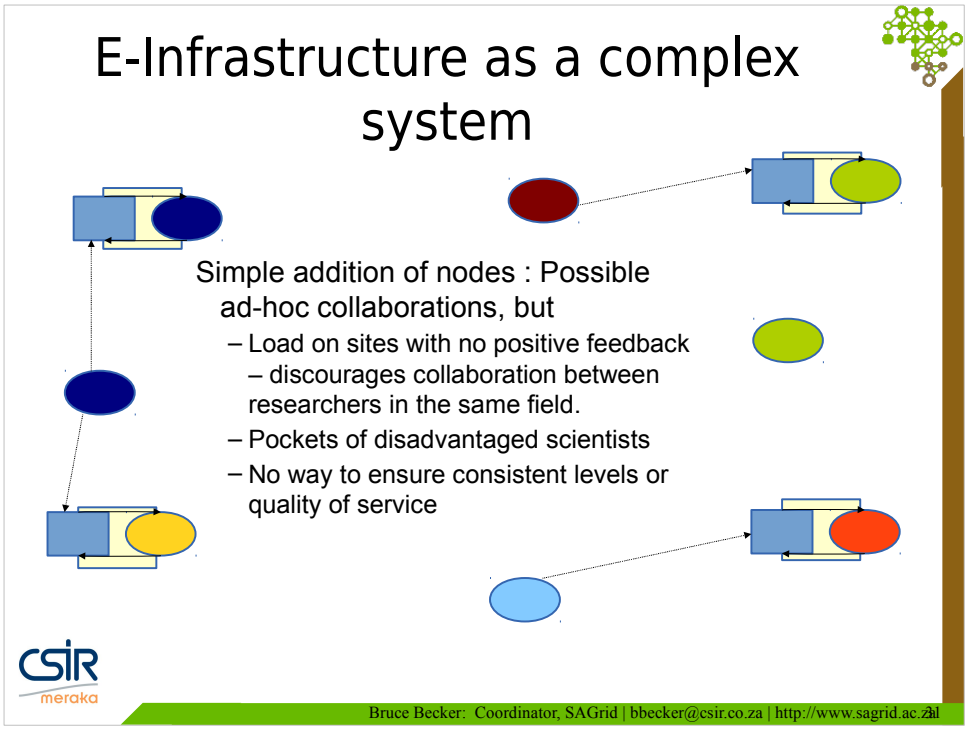
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



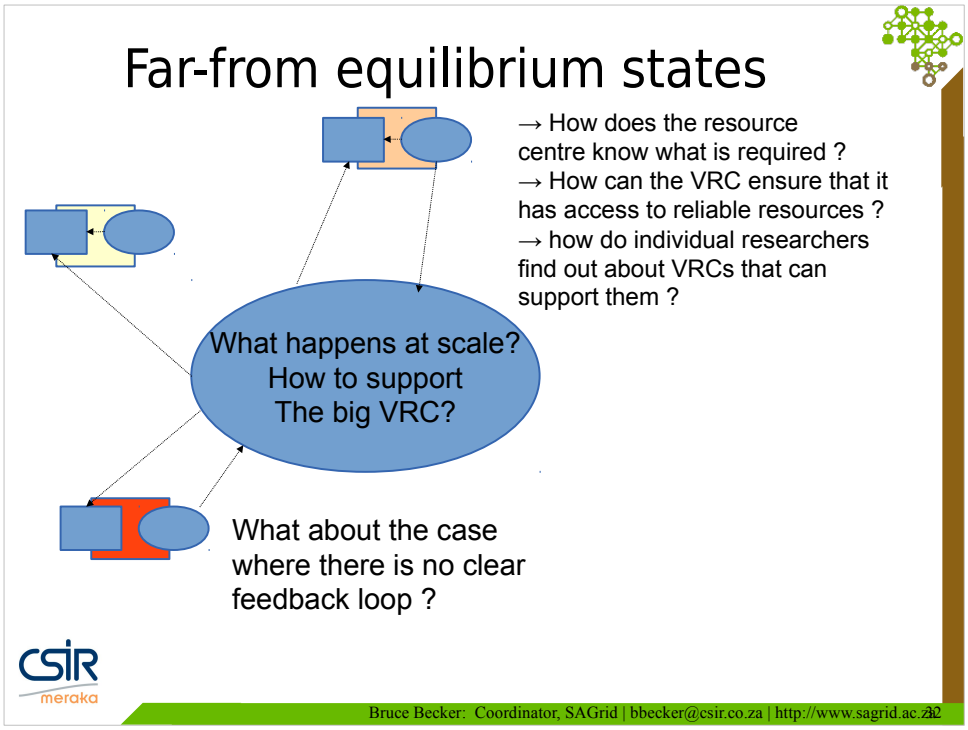
Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



Apart from this simple, quite self-contained interaction, what other kinds of interactions can we forsee ? If we multiply our little ”engine” , we can consider ad-hoc combinations of them, but they are likely to create extra load on the resources, without ensuring that there is the positive feedback loop we mentioned before. We can consider other disagreable configruations, where there are pockets of disadvantaged scientists who have no access to resources. Futhermore, as you may imagine, this incoherent state has no ability to scale.



Now, let us consider the case where there is a ”grand challenge” - an important research questions around which a large virtual research communty forms. How are they to be supported in the incoherent scenarios ? This is a far-from-equilibrium state – and clearly a desirable one for African research. But the support of a large VRC in such an environment is far more complex than the VRC itself ! Issues such as how many and what kind of resources are needed, and where, and when have to be dealt with by the VRC itself, which needs to have sevaral MoU's in place, each taking into account the next. Furthermore, this large VRC runs the risk of overpowering local user communities – who may even be performing the same or similar research ! - at the sites, by being in competition for the limited resources.
Note as well that we've ”extinguished” the generator – the positive feedback between research and resources, which was the key to sustainability.

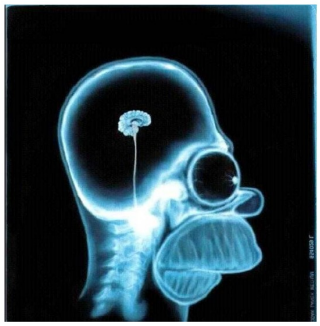
What does this mean ?

- In fairly closed systems, investment in resources and support of feedback mechanisms can lead to self-supporting e-Science systems
  - However, it cannot scale
- Usually an intervention is required to initially drive the system far from its rest state of doing nothing
  - This is a luxury not all of our colleagues can afford
- What should we do ?

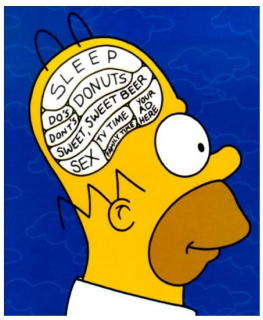
So, what message do we take away from this ?
In fairly closed systems, the investment in resources and the explicit creation and support of feedback mechanisms is fundamental to self-sustainability. This is fine on an ”atomic” (research or institute) level, but it cannot scale.
In a simplistic analysis, an external force is necessary to drive the system far from its rest state (of doing nothing), but this is a luxury that we can't always afford.
Furthermore, there are many permissible outcomes, but a stable solution is not guaranteed – the system is undeniably complex.
However, it is this same complexity that provides room for emergent phenomena, based on local interactions, but in a different operating environment – not the closed system of the research group or even the institute.
Now, the questions we need to address are :
1) can a local equilibrium drive similar resonances in a more complex system which it forms part of ? ?
2) what kind of interactions will result in a higher stable state ?
3) and which will, on the contrary, damp the resonance ?

Structural MRI vs. *Functional* MRI

Structural **MRI reveals brain anatomy.**



Functional **MRI (fMRI) reveals brain function.**

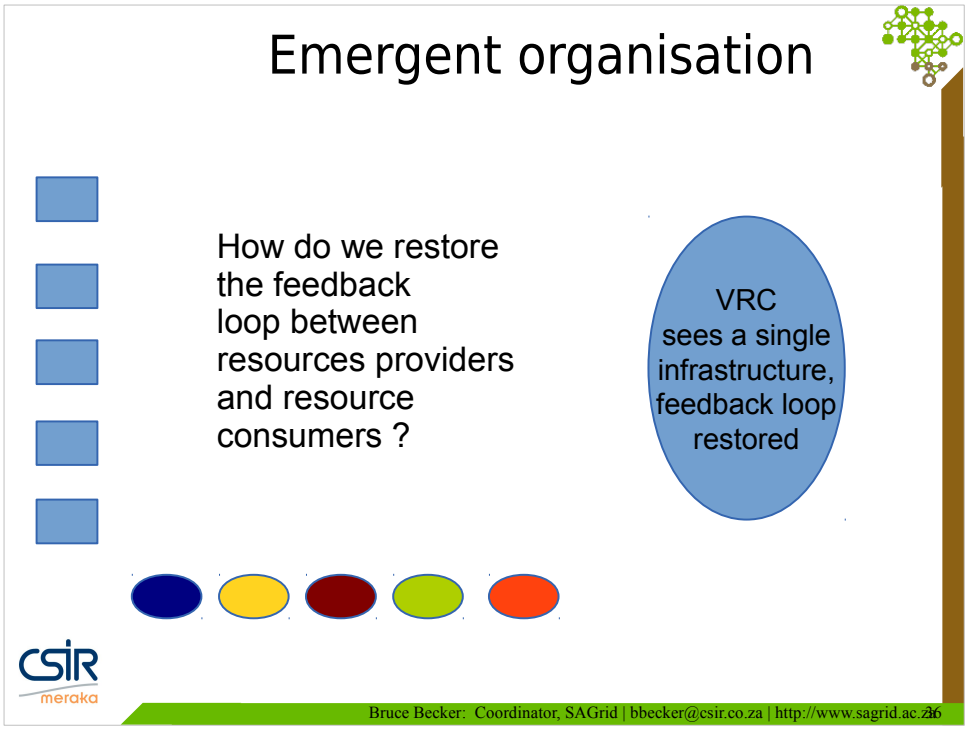


A complex system is better described by its **processes** than its **parts**

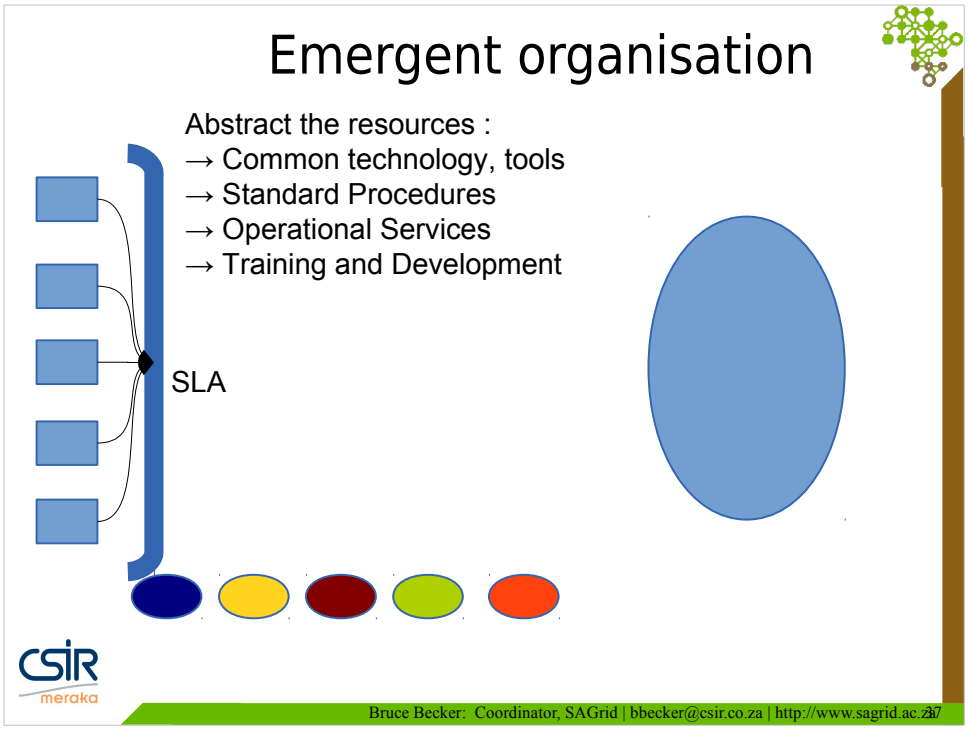
So, is it impossible to sustainably conduct large-scale research reliant on computing infrastructure in Africa ? Or are we being too simplistic in our model ?
Instead of the ***components*** of this system, let us try instead to think about the ***functions***. I show here a cartoon trying to draw an analogy between the field of neuro-imagery and our case. Depending on the way you organise something, given the same components, the functionality can change. So, it's better to describe a complex system by its ***processes*** than by its ***parts.***

Perhaps we can re-organise ?

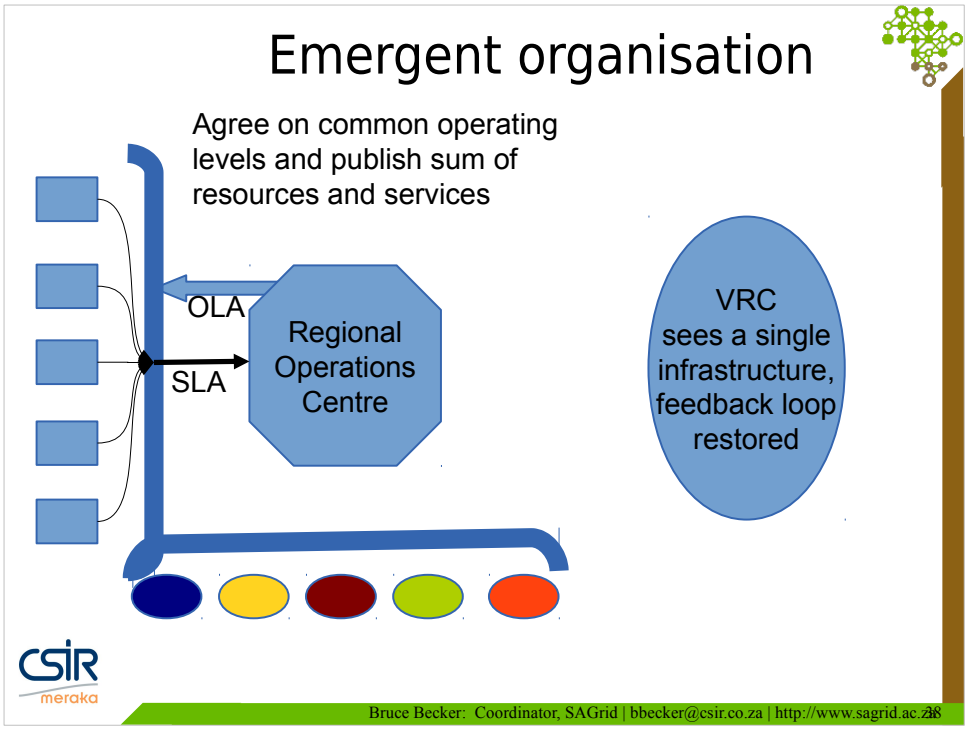
- Many outcomes are permissible, but a stable solution is not guaranteed.
- There is also room for **emergent phenomena,** based on local interactions, but in a different operating environment
  - Can a local equilibrium drive similar resonances in a more complex system ?
  - What kind of **interactions** will result in a higher stable state
  - Or result in overload and damp the resonance ?



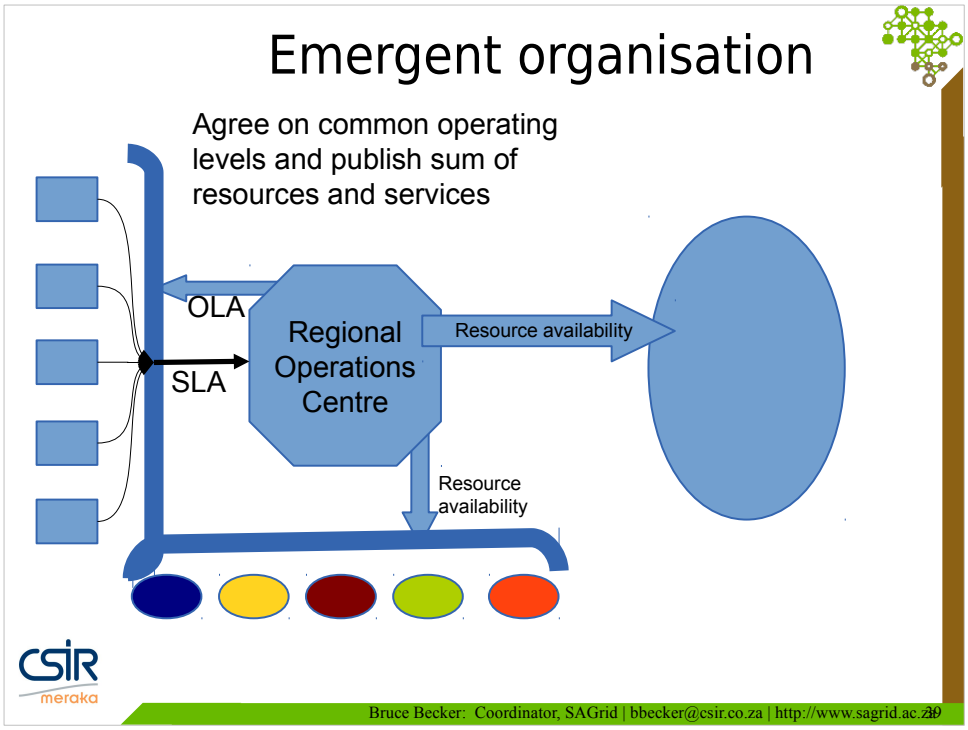
Let us finally then consider a reorganisation of the components of African e-Infrastructure and their interaction with research communities and vice-versa. Here, we introduce the concept of a Regional Operations Centre, or ROC. The ROC encourages coherent interactions between researchers and resources, by introducing a set of Service Level Agreements and Operating Level Agreements to the sites. This in turn presents the research communities – large and small – with a coherent view of what is potentially available, and how much work it can do. The ROC communicates, by publishing the state of the infrastructure (which is now no longer a simple ad-hoc collection of resources), allowing a meaningful negotiation even with large VRCs. Additionally, the VRCs need only to consider a single MoU, brokered by the ROC with the resource providers.
But the ROC is not just a re-organisation of the resource providers, there is the benefit of ***emergent*** functionality thanks to the ***scale*** of the system. We can now consider the adoption of common procedures, tools and technology, along with the capacity to train new personnell and therefore introduce further resources to the system, in a self-contained way. By scaling the system coherently, we actually reduce the load on the operational staff.



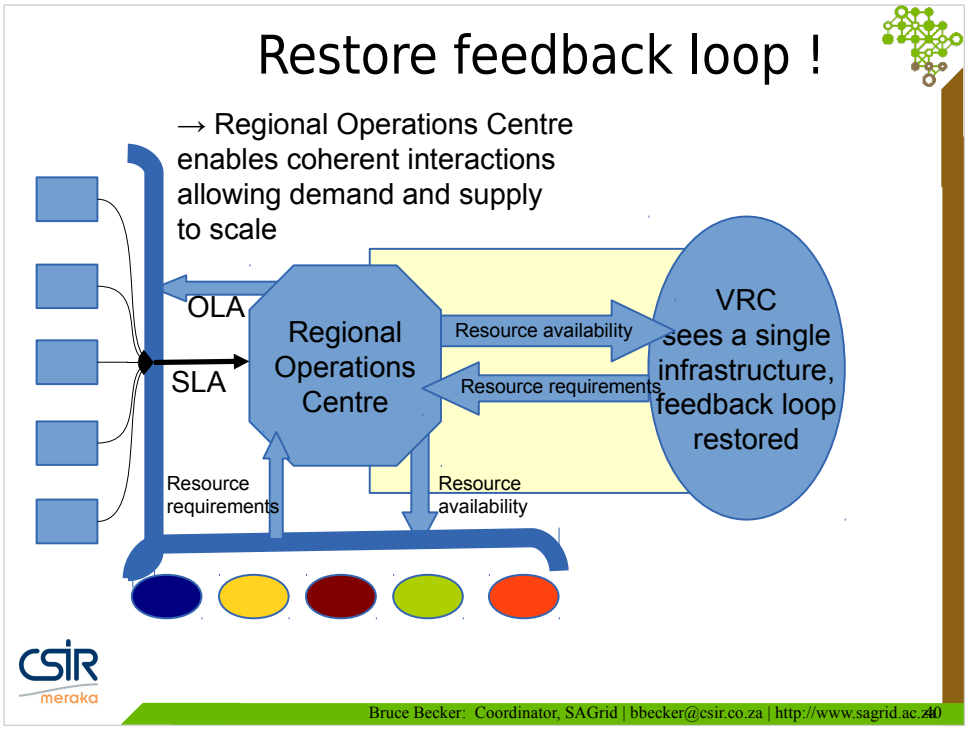
Let us finally then consider a reorganisation of the components of African e-Infrastructure and their interaction with research communities and vice-versa. Here, we introduce the concept of a Regional Operations Centre, or ROC. The ROC encourages coherent interactions between researchers and resources, by introducing a set of Service Level Agreements and Operating Level Agreements to the sites. This in turn presents the research communities – large and small – with a coherent view of what is potentially available, and how much work it can do. The ROC communicates, by publishing the state of the infrastructure (which is now no longer a simple ad-hoc collection of resources), allowing a meaningful negotiation even with large VRCs. Additionally, the VRCs need only to consider a single MoU, brokered by the ROC with the resource providers.
But the ROC is not just a re-organisation of the resource providers, there is the benefit of ***emergent*** functionality thanks to the ***scale*** of the system. We can now consider the adoption of common procedures, tools and technology, along with the capacity to train new personnell and therefore introduce further resources to the system, in a self-contained way. By scaling the system coherently, we actually reduce the load on the operational staff.



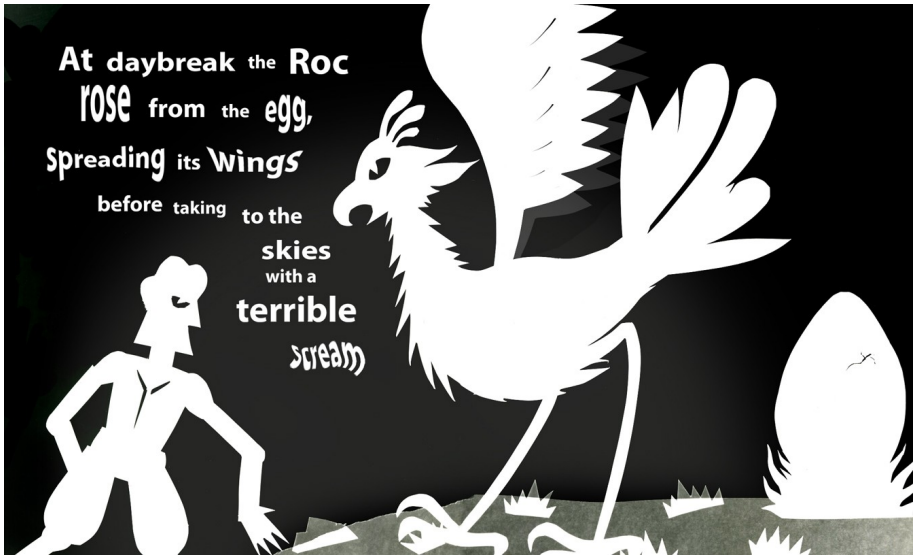
Let us finally then consider a reorganisation of the components of African e-Infrastructure and their interaction with research communities and vice-versa. Here, we introduce the concept of a Regional Operations Centre, or ROC. The ROC encourages coherent interactions between researchers and resources, by introducing a set of Service Level Agreements and Operating Level Agreements to the sites. This in turn presents the research communities – large and small – with a coherent view of what is potentially available, and how much work it can do. The ROC communicates, by publishing the state of the infrastructure (which is now no longer a simple ad-hoc collection of resources), allowing a meaningful negotiation even with large VRCs. Additionally, the VRCs need only to consider a single MoU, brokered by the ROC with the resource providers.
But the ROC is not just a re-organisation of the resource providers, there is the benefit of ***emergent*** functionality thanks to the ***scale*** of the system. We can now consider the adoption of common procedures, tools and technology, along with the capacity to train new personnell and therefore introduce further resources to the system, in a self-contained way. By scaling the system coherently, we actually reduce the load on the operational staff.



Let us finally then consider a reorganisation of the components of African e-Infrastructure and their interaction with research communities and vice-versa. Here, we introduce the concept of a Regional Operations Centre, or ROC. The ROC encourages coherent interactions between researchers and resources, by introducing a set of Service Level Agreements and Operating Level Agreements to the sites. This in turn presents the research communities – large and small – with a coherent view of what is potentially available, and how much work it can do. The ROC communicates, by publishing the state of the infrastructure (which is now no longer a simple ad-hoc collection of resources), allowing a meaningful negotiation even with large VRCs. Additionally, the VRCs need only to consider a single MoU, brokered by the ROC with the resource providers.
But the ROC is not just a re-organisation of the resource providers, there is the benefit of ***emergent*** functionality thanks to the ***scale*** of the system. We can now consider the adoption of common procedures, tools and technology, along with the capacity to train new personnell and therefore introduce further resources to the system, in a self-contained way. By scaling the system coherently, we actually reduce the load on the operational staff.

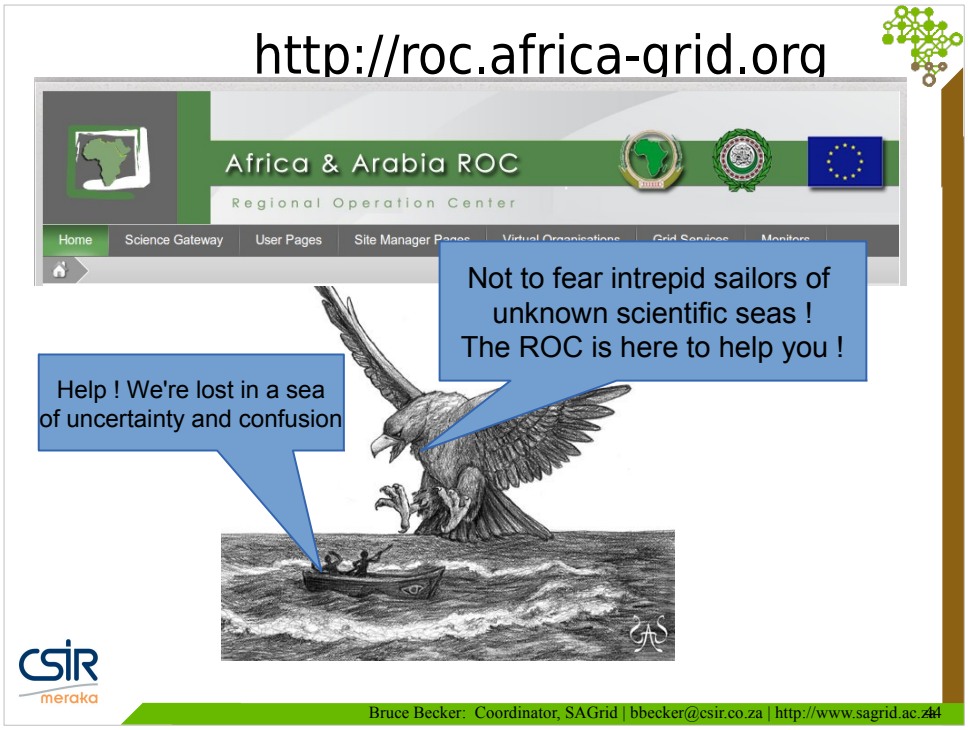


Let us finally then consider a reorganisation of the components of African e-Infrastructure and their interaction with research communities and vice-versa. Here, we introduce the concept of a Regional Operations Centre, or ROC. The ROC encourages coherent interactions between researchers and resources, by introducing a set of Service Level Agreements and Operating Level Agreements to the sites. This in turn presents the research communities – large and small – with a coherent view of what is potentially available, and how much work it can do. The ROC communicates, by publishing the state of the infrastructure (which is now no longer a simple ad-hoc collection of resources), allowing a meaningful negotiation even with large VRCs. Additionally, the VRCs need only to consider a single MoU, brokered by the ROC with the resource providers.
But the ROC is not just a re-organisation of the resource providers, there is the benefit of ***emergent*** functionality thanks to the ***scale*** of the system. We can now consider the adoption of common procedures, tools and technology, along with the capacity to train new personnell and therefore introduce further resources to the system, in a self-contained way. By scaling the system coherently, we actually reduce the load on the operational staff.

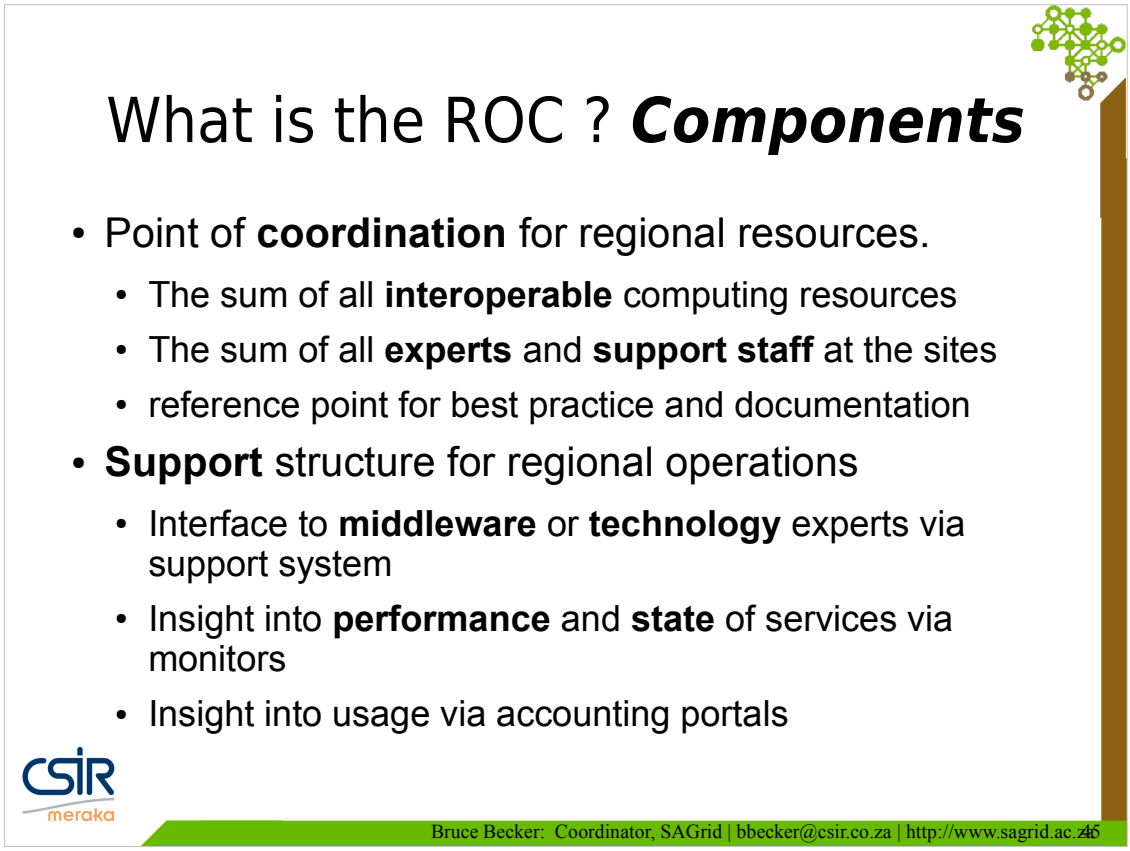

At daybreak the Roc
rose from the egg,
spreading its wings
before taking to the
skies
with a
terrible
scream

I didn't know this, but I found out durig the course of my work, that the ROC was actually a mythical bird, described in several legends, including the tales of Sinbad the Sailor. Descriptions of it vary, but in some stories, the ROC save the poor sailors from a terrible storm. I'd like to leave you with this impression...

I didn't know this, but I found out durig the course of my work, that the ROC was actually a mythical bird, described in several legends, including the tales of Sinbad the Sailor. Descriptions of it vary, but in some stories, the ROC save the poor sailors from a terrible storm. I'd like to leave you with this impression...



I didn't know this, but I found out durig the course of my work, that the ROC was actually a mythical bird, described in several legends, including the tales of Sinbad the Sailor. Descriptions of it vary, but in some stories, the ROC save the poor sailors from a terrible storm. I'd like to leave you with this impression...



I'll close off this presentation by going a bit more into detail about the ROC.

The ROC is a set of components:
1) a point of coordination for regional resources, resulting in the sum of ***interoperable*** resources. It is also a multiplying factor of the human capacity in the region.
2) The ROC is a support structure for regional operations, with an interface to experts for middleware and other technology support in the region.
3) it provides insight to the performance and state of services available via monitors and
4) it is a source of reference for best practices and documentation.

What is the ROC ?
Global Operations Database

- Central service where all sites, services, resources and personell are registered.
  - Is there anything in Namibia ?
  - What is in Namibia ?
  - Who do we call in Namibia if someone from Cameroon wants to collaborate there ?
- Starting point for every other service provided by the ROC

What is the ROC ? ***Procedures***

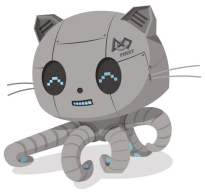
- The power of a distributed computing infrastructure lies on its ***coherence***
  - Sites may employ different **technology** (as long as it's interoperable), but should emply standard **procedures**
- Standard procedures improve ease of collaboration
  - common operational vocabulary
  - Predictable changes in state of services (upgrades, removal, etc)
  - Reduce the need for everyone to be an expert. If you don't know how to do something technical, just follow the recipe
  - Reduce impact of external factors (middleware providers, technology changes)

What is the ROC ?
Standard Operating Procedures

- Clear, explicit procedures for performing common tasks
  - Who should I inform when a site is underperforming ?
  - How do I escalate an issue ?
  - How are applications to be deployed ?
  - When should I announce a downtime ?
- Focus on **how** – reduce the need for expert knowledge
- Provide templates for the ”what”

What is the ROC ? ***Automation***

- Distributed computing infrastructure is by its nature a complex environment.
  - Reduce the need to know ***how*** to perform tasks
  - Limit to the need to know ***what*** services you want to deploy
- Rely on a core team of experts to maintain the ”golden standard” for configuration of services



What is the ROC ? ***Services***

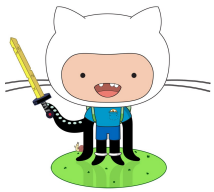
- Several services in order to collaborate technically on the infrastructure level:
  - GOCDB
  - Monitoring and alarms
  - Support system, integrated with technology providers and other infrastructures
    - Middleware, network, etc.
  - Accounting portal
- Provides user communities with a coherent view of the infrastructure

In summary, the ROC can be a catalyst for emergent e-Infrastructure in our region, acting as a central point of contact for technical and user communities alike. We have direct access via the ROC to the AfricaGrid Science Gateway, which I haven't discussed here, and direct access to the support and operations team.
The ROC needs ***reliable*** resources, and these are provided via Service and Operating Level agreements, as well as Standard Operating Procedures.
Apart from their reliability, we also need these resources to be ***scalable*** and the answer to this requirement is to adopt an open, interoperable infrastructure, which can easily accomodate both new resources and new users, at scale.

Prospects for extending the infrastructure

- The ROC is a **catalyst** for new sites, acting as the initial stimulus for development in the absence of sufficient local energy
- However :
  - Care should be taken to not overload the ROC
  - Careful balance between extending the infrastructure and maintaining operational capability
- Computing and data sites in Africa are encouraged to be included in the ROC to publish their resources
- Sites encouraged to **share** their manpower.
- Currently, 1LS/TPM is an Africa-Arabia wide activity

How do we achieve this ?



- Maintain a **predictable feedback loop**
  - Entice big VRCs to use the infrastructure
  - Communicate success
  - develop open, interoperable ecosystem
- Ensure the system remains in **dynamic equilibrium**
  - Ensure that the operations level is in good working order – consistent policies and procedures at all sites
  - Evaluate new applications and use cases in terms of their impact on the infrastructure and vice versa
- Ensure access to a **large "energy bath"** to absorb large fluctuations in load
  - MoU to interoperate with external infrastructures
  - Explore ad-hoc access to external resource providers
- **Ensure that the entire infrastructure is visibile**

Removing barriers

- How does an African infrastructure interoperate with other global infrastructures in order to allow scientists to easily collaborate ?
- N*M MoU's is unfeasible -
  - ROC represents all sites, negotiates peer activities with other regional infrastructures
- Expose a single interface -
  - Single coordination effort required – currently undertaken by Meraka for European interoperation

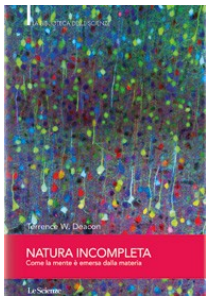
Conclusion

- Things are (thankfully) not as simple as we may think
- A feedback loop between resources and researchers is essential
- A re-organisation of interactions can allow components of a given system to reach higher-lying **states** and **functions**
- Coordination of services and resources is necessary to achieve self-sustainability

Thank you !

- It's a community, come and join us !
  - Africa-Arabia ROC on Facebook
  - On github:
    - https://github.com/AAROC
  - SAGrid on Facebook and Twitter
    - https://www.facebook.com/SAGrid
    - https://www.twitter.com/TheSAGrid

ISBN 978-0-393-04991-6



Many of the ideas presented here were inspired by the book
”Natura Incompleta” - Terrence Deacon