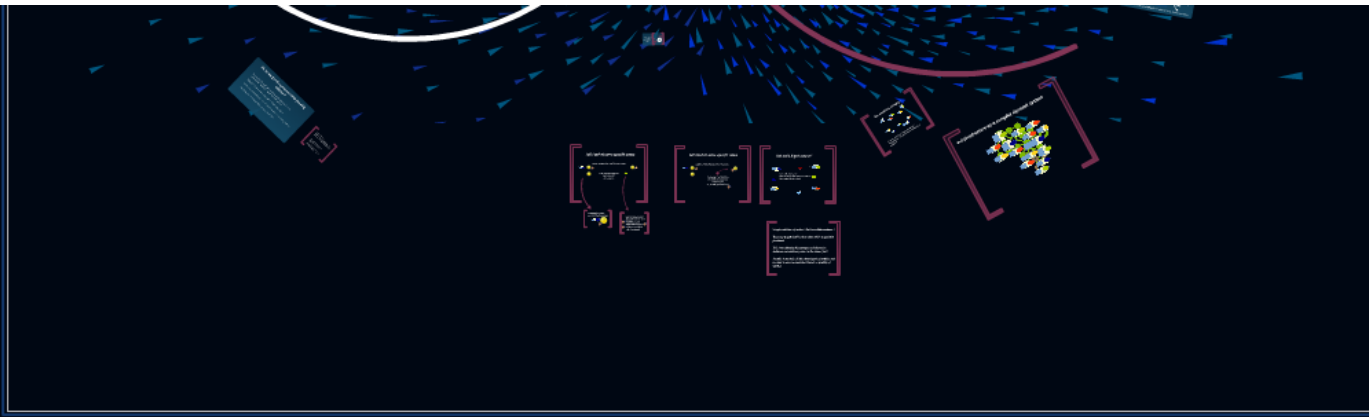"Out of turbulence"
or
"In principio era Chaos"

"Out of turbulence"

or

"In principio era Chaos"

The story I will tell...

At the risk of being called an optimist, it's likely that everything you need to do your research is already available...

This is perhaps not the story that you'll often hear.

You'll hear that pieces are missing from a big puzzle.

While actually, what's missing from the puzzle - and from the universe in general

is simply...

Order is not usually found in nature.

For example - you can't mess with the second law of thermodynamics.

Wrong ! That's only true on a global scale.

So, this story will take you back home.

This story comes with a warning

WARNING

Highly simplified scientific analogies will be made

Don't panic

There is a big friendly bird

and some cats

to get you through it.

Mr Miyagi's here too...

Let's start with the basics: *e-Science in Africa

There is much **potential** for productive e-Science in Africa. Several institutes moving to conduct research (as opposed to just teaching) and have appreciated the need for **access** to e-Infrastructures

Africa has a large diaspora to collaborate with.

E-Science (or eScience) is computationally intensive science that is carried out in highly distributed network environments, or science that uses immense data sets [...] ; the term sometimes includes technologies that enable distributed collaboration [...] E-science has been more broadly interpreted [...], as "the application of computer technology to the undertaking of modern scientific investigation, including the preparation, experimentation, data collection, results dissemination, and long-term storage and accessibility of all materials generated through the scientific process.

Let's start with the basics: e-Science in Africa

There is much **potential** for productive e-Science in Africa.
Several institutes moving to conduct research (as opposed to just teaching) and have appreciated the need for **access** to e-Infrastructures
Africa has a large diaspora to collaborate with.
There are new bandwidth opportunities connecting institutes and people.
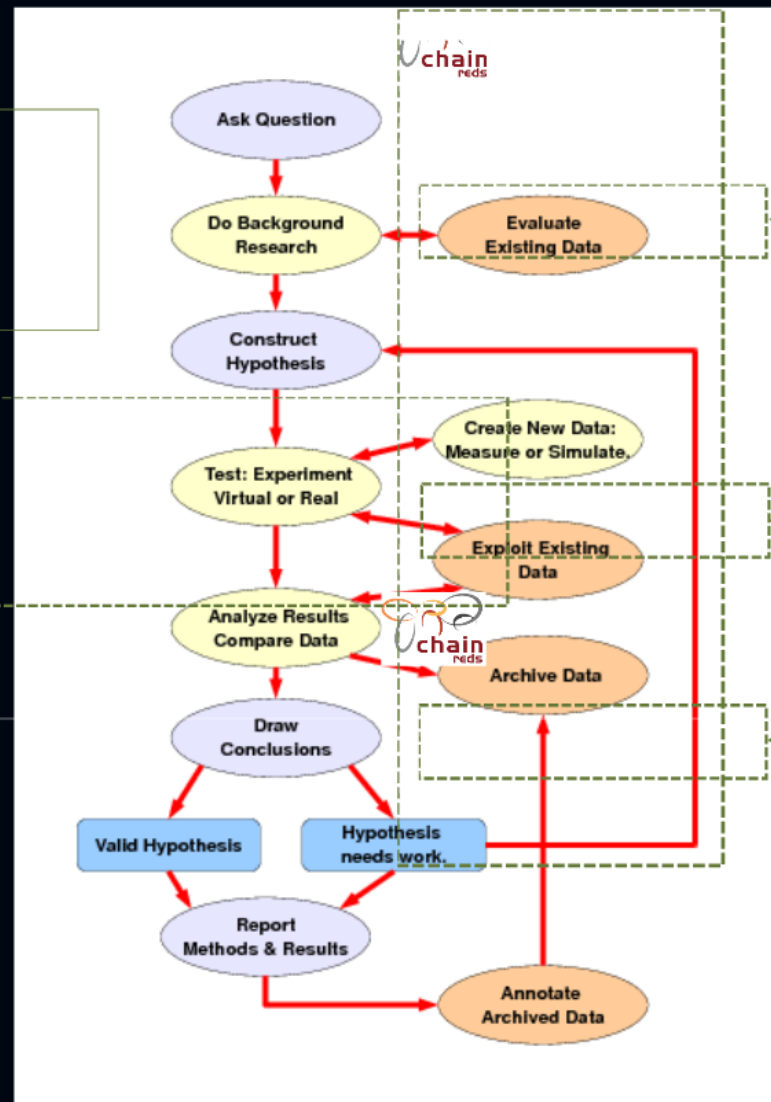There is a long list of important scientific problems to tackle.
...However...
Despite all this **potential**, the available **energy** is not all **transformed** into work.

e-Science
requires
e-Infrastructures

The Scientific Method
(worldwide,
cross-domain,
collaboration enabler)

Data Infrastructures
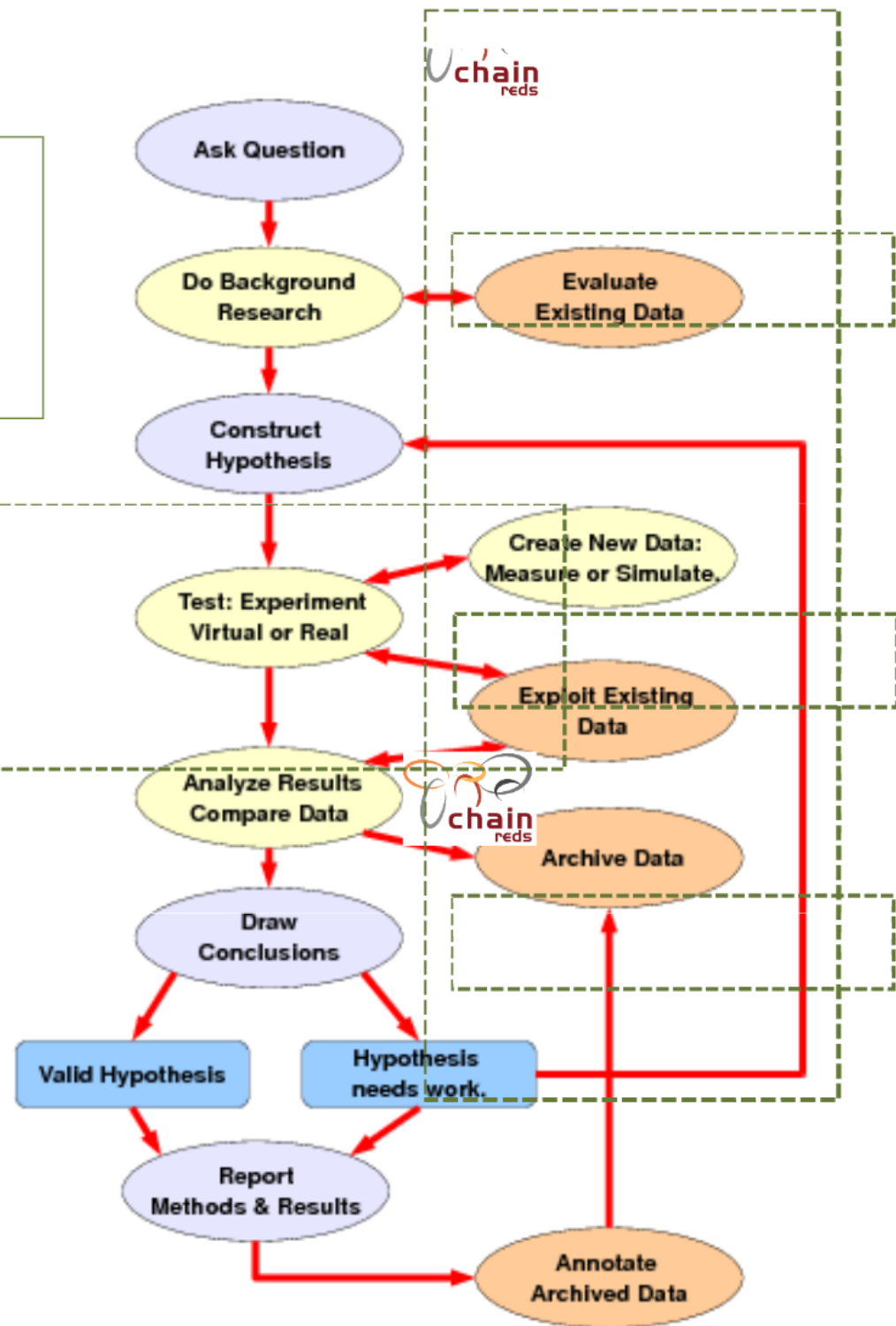Open Access Doc. Repos.
Data Repos.

HTC/HPC Clusters Grids, Clouds

Correllate papers among them and with data

Ask Question

Do Background Research

Evaluate Existing Data

Construct Hypothesis

Create New Data: Measure or Simulate,

Test: Experiment Virtual or Real

Exploit Existing Data

Analyze Results Compare Data

Archive Data

Draw Conclusions

Valid Hypothesis

Hypothesis needs work.

Report Methods & Results

Annotate Archived Data

PREZI

uctures

ific Method
e,
ain,
ion enabler)

HTC/HPC
Clusters
Grids, Clouds

Ask Question

Do Background Research

Evaluate Existing Data

Construct Hypothesis

Create New Data: Measure or Simulate.

Test: Experiment Virtual or Real

Exploit Existing Data

Analyze Results Compare Data

Archive Data

Draw Conclusions

Valid Hypothesis

Hypothesis needs work.

Report Methods & Results

Annotate Archived Data

Correllate papers among them and with data

South Africa as an example

South Africa has built several e-infrastructure components with a long-term focus (the SKA).
 1) Biggest supercomputer in Africa
 2) Fastest research network in Africa
 3) Biggest research data repository in Africa

From the outside, it appears that this was a top-down approach

Ok, so we get the government to fund big science ?

How can this be applicable to African countries where there is much smaller capacity to support such initiatives ?

What if there **is** no single driving force ? Deus ex machina ?

Utopia: top-down approach - "Here's a lot of money, go solve the problem..."
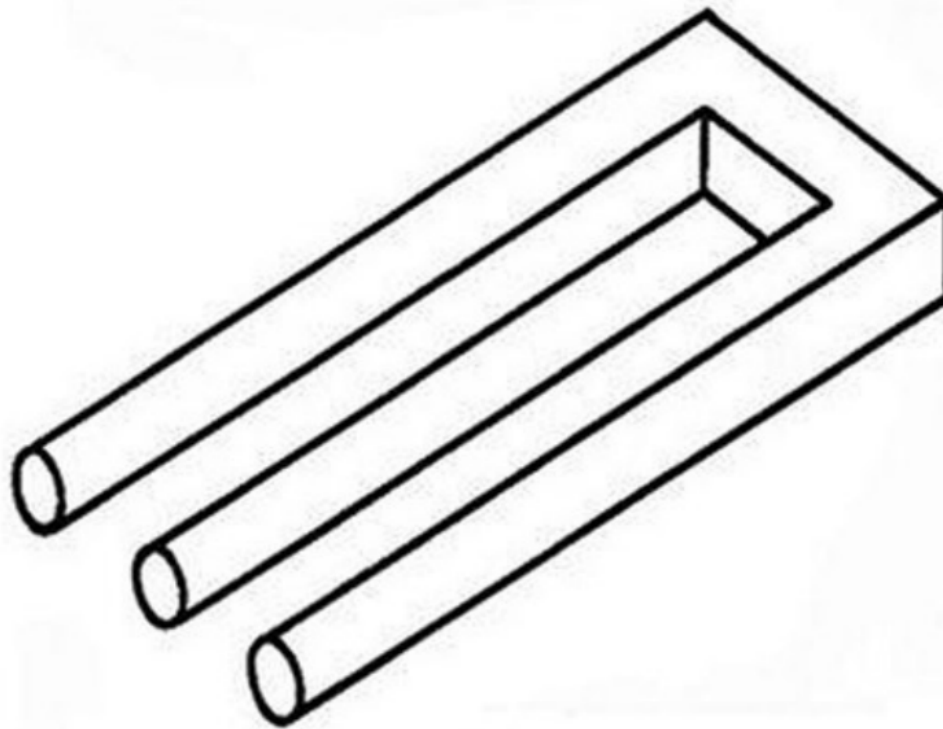
Take some time to think about the problems with that...

Eyes on the prize !
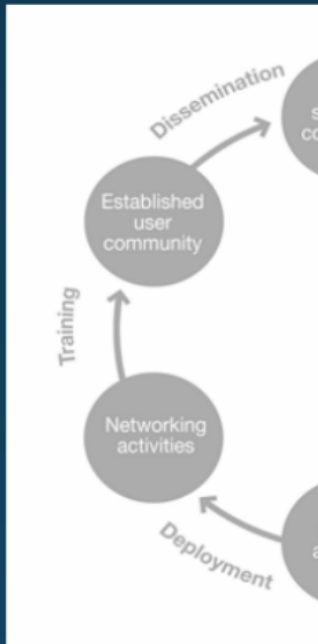*Scalable research in Africa needs an*
open, self-sustainable e-Infrastructure

There is also room for emergent
phenomena, based on local
interactions.

Many outcomes are permissible, but a
stable solution is not guaranteed.

WORKS ON PAPER

Let's go back to basic assumptions

In order to discover and curate research outputs, you first need to do some research.
Performing research almost always requires the interaction of **people, applications and data**. (It's already a 3-body problem)

Applications and data need computing and data resources;
The resources drive usage, which in turn drive the acquisition of new resources – looks like a driven harmonic oscillator...
Hypotheses:
   1) Decays in the absence of a use case
   2) Strongly damped in the case of weak input signals
   3) Requires an initial catalyst

Local User Community

enable research

Local User Community

Local Computational
Data resources

Funding
and
operational
environment

strong research
agenda

justify support

PREZI

Funding and operational environment

strong research agenda

User/Resource feedback loop:
In the case of a strong research agenda

Input:
User community drives computing centre

Two feedback loops

1) User activity drives research agenda, which drives user community
2) Interesting research agenda justifies computing resources;
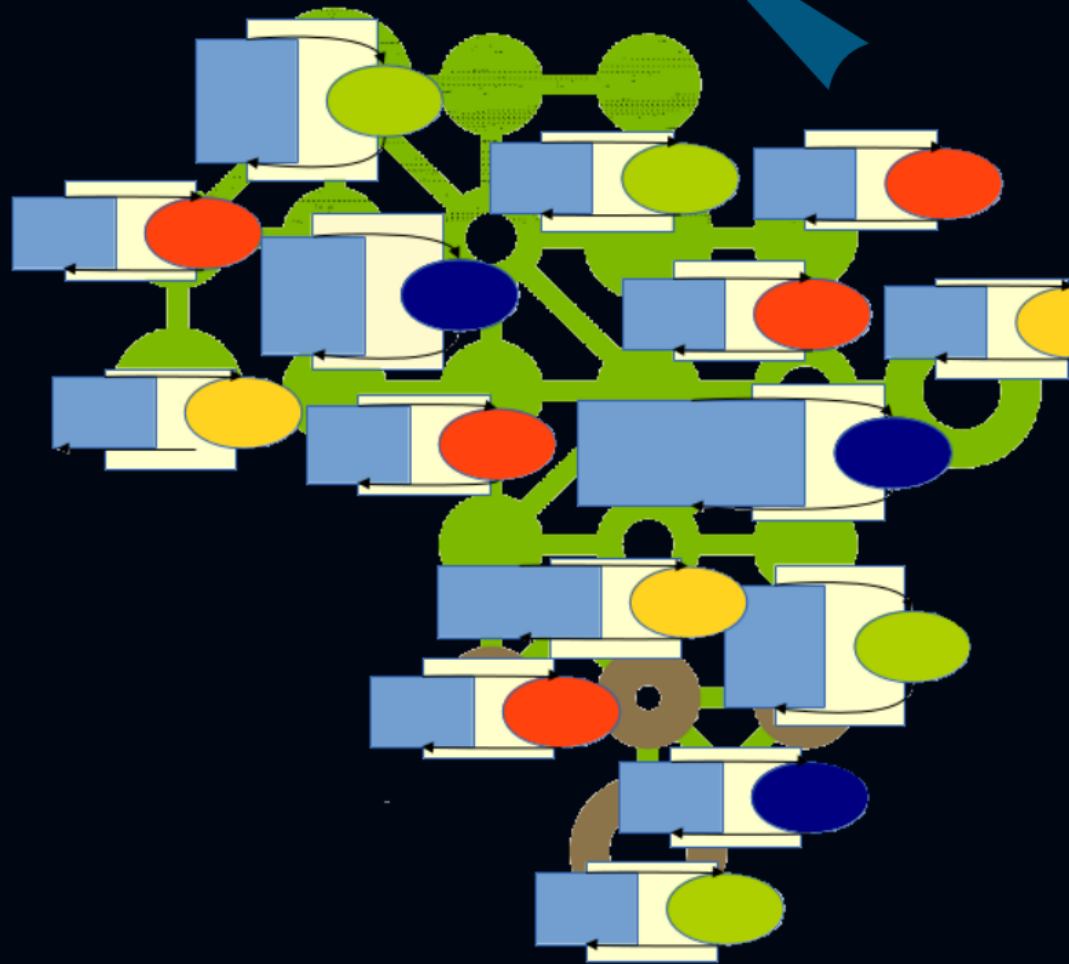Computing resources improve research agenda.

enable research

Local User Community

Local Computational
Data resources

justify support

**publications
data**

However, you need **output** to counterbalance to external forces
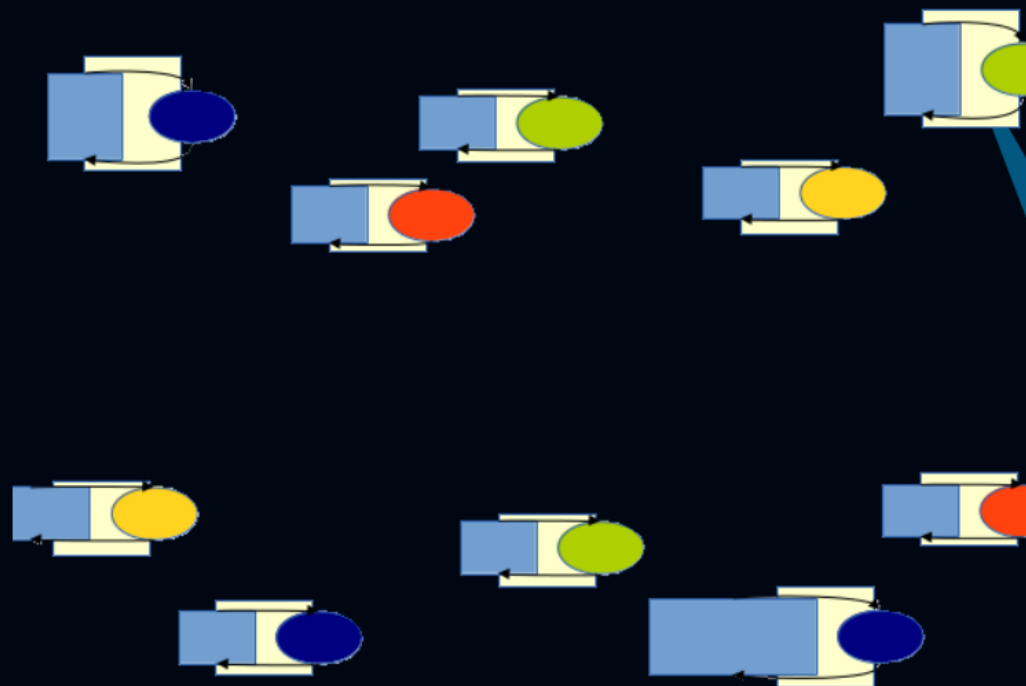
Works on paper...

but will it scale ?

e-Infrastructure as a complex dynamic system

Two questions, wiseguy :

How much work can these systems do ?
How much work does it take to sustain these systems ?
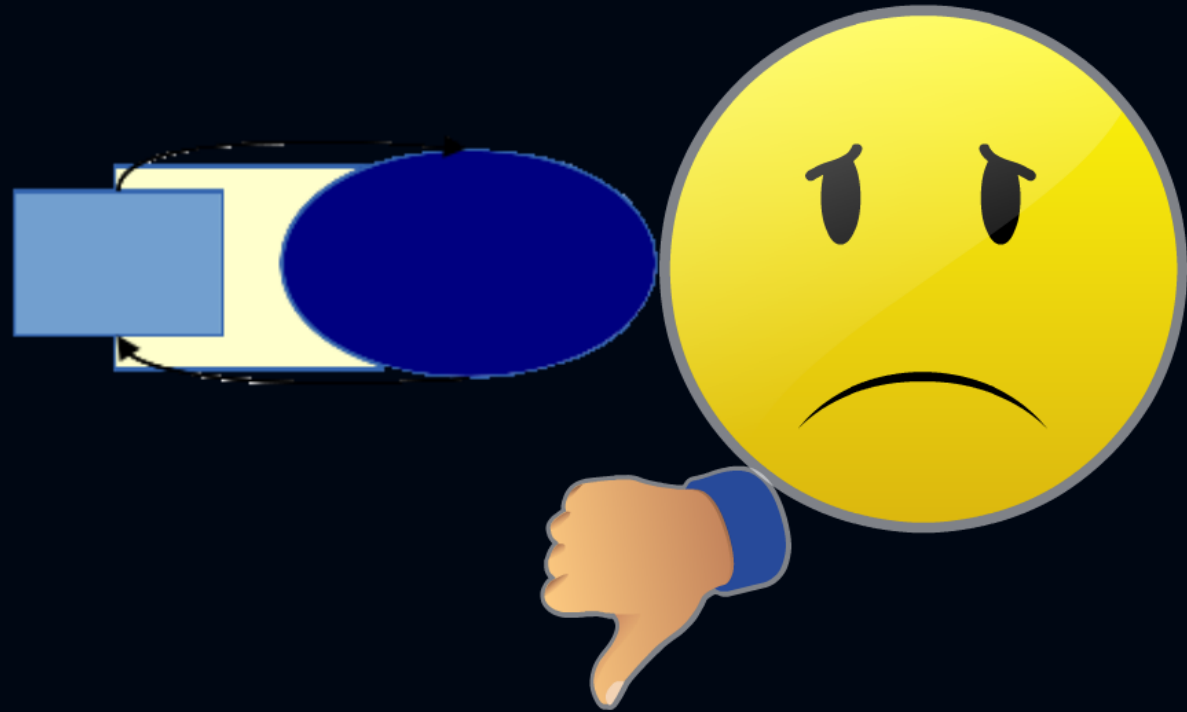
Let's look at some specific cases

single communities and their resources

some communities are
not connected
to resources

communities merge, overload their resources

Let's look at some specific cases

single communities and their resources

some communities are
not connected
to resources

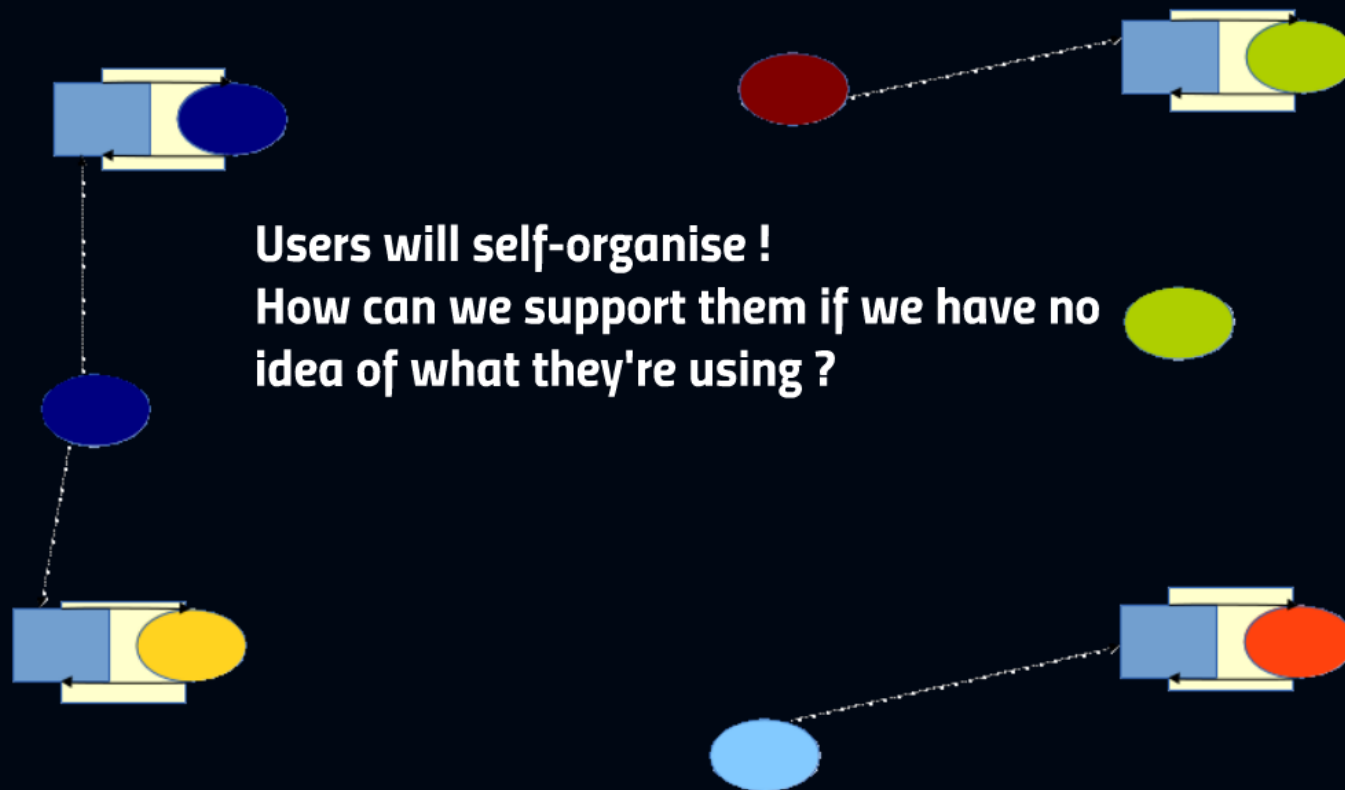# Let's look at some specific cases

single communities and their resources

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

But wait, it gets worse!

Users will self-organise!
How can we support them if we have no
idea of what they're using?

Simple addition of nodes ? Ad-hoc collaborations ?

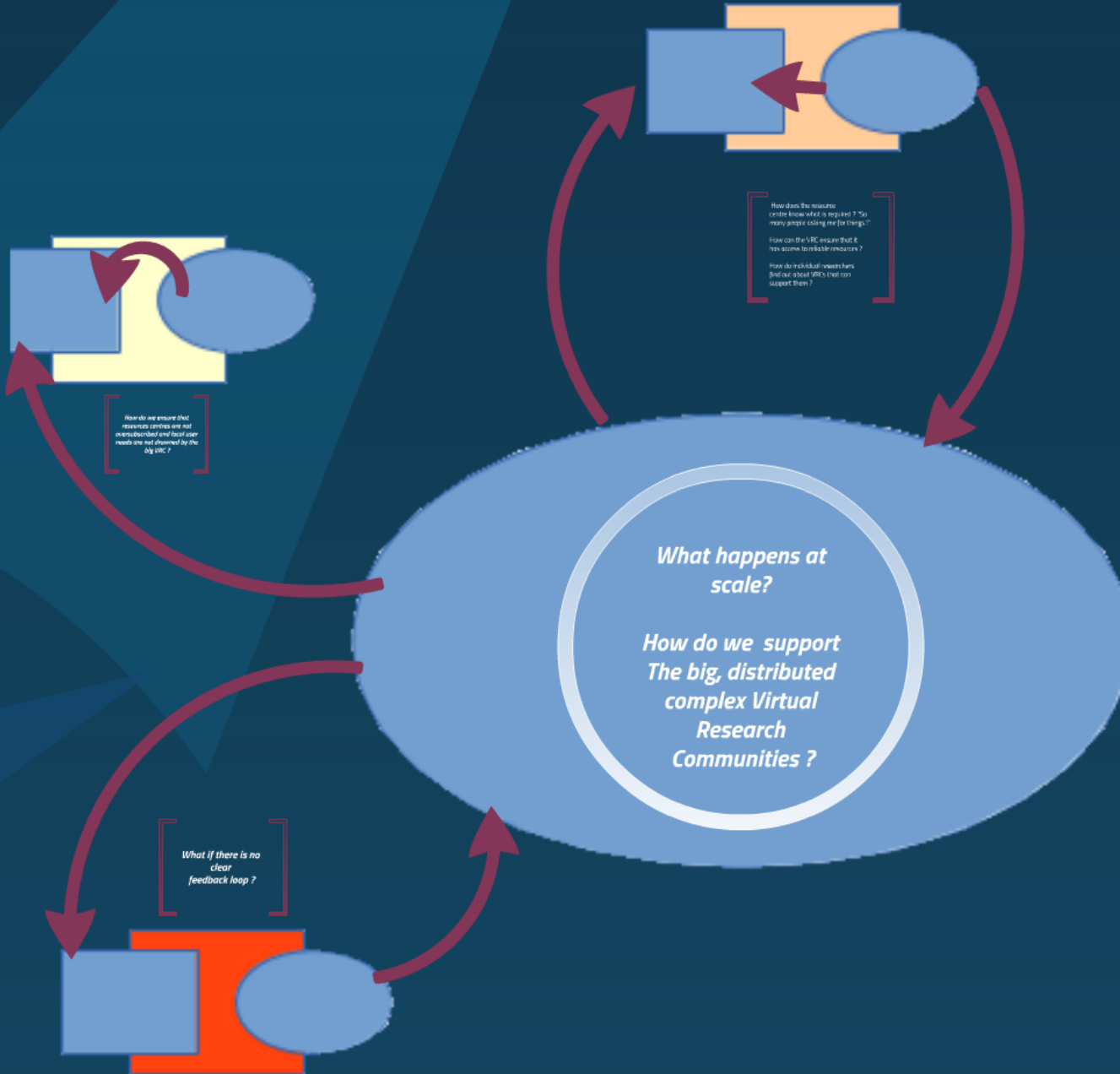Too easy to put load load on sites with no positive feedback.

This immediately discourages collaboration between researchers, even in the same field.

Results in pockets of disadvantaged scientists and no way to ensure consistent levels or quality of service

What happens at scale?

How do we support
The big, distributed
complex Virtual
Research
Communities ?

ant

ium

How does the resource
centre know what is required ? "So
many people asking me for things !"

How can the VRC ensure that it
has access to reliable resources ?

How do individual researchers
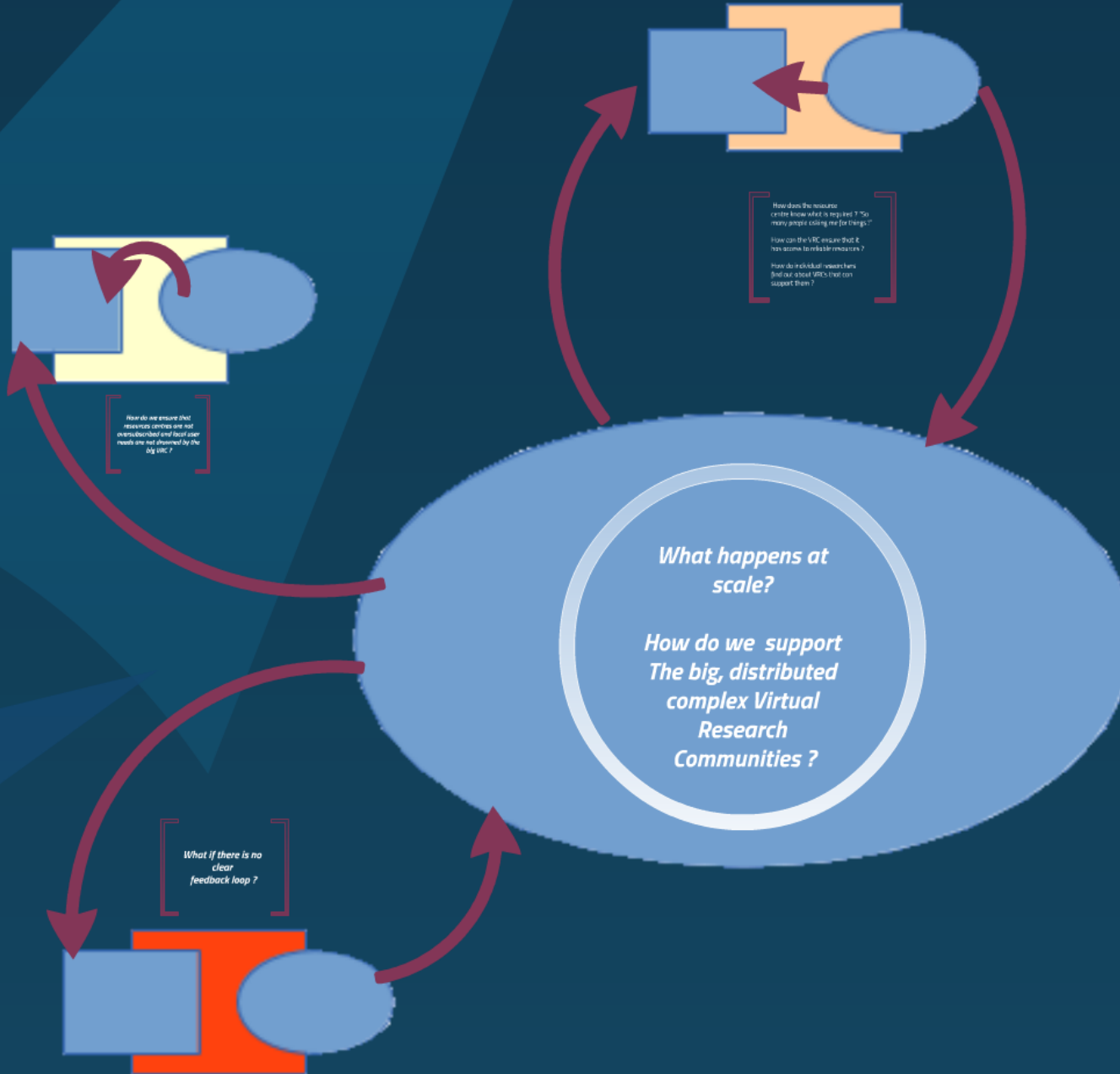find out about VRCs that can
support them ?

ant

ium

How does the resource
centre know what is required ? "So
many people asking me for things ?"

How can the VRC ensure that it
has access to reliable resources ?

How do individual researchers
find out about VRC's that can
support them ?

How do we ensure that
resources centres are not
oversubscribed and local user
needs are not drowned by the
big VRC ?

**What happens at
scale?**

**How do we support
The big, distributed
complex Virtual
Research
Communities ?**

What if there is no
clear
feedback loop ?

What if there is no clear feedback loop ?

ant

ium

How does the resource
centre know what is required ? "So
many people asking me for things?"

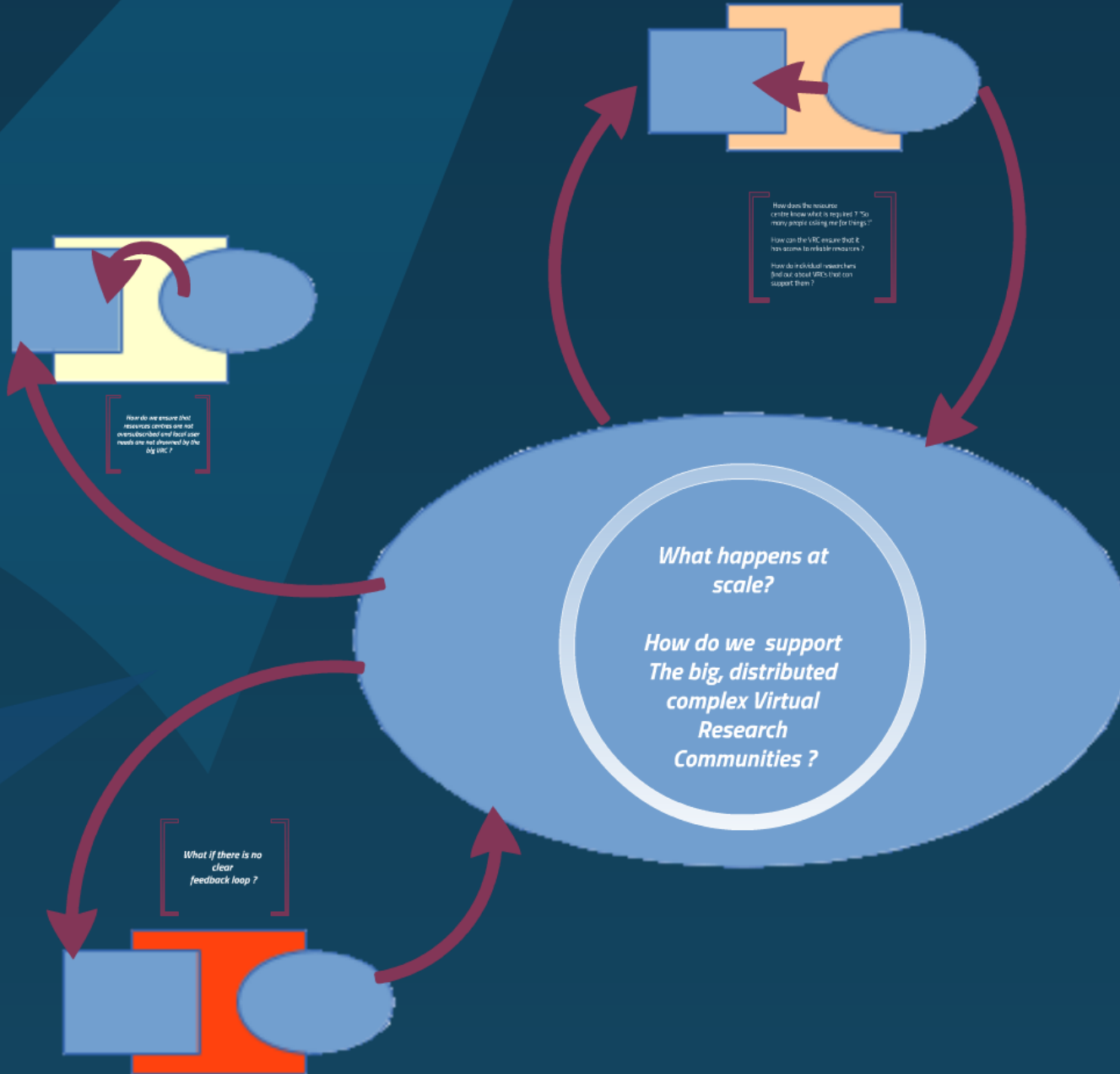How can the VRC ensure that it
has access to reliable resources ?
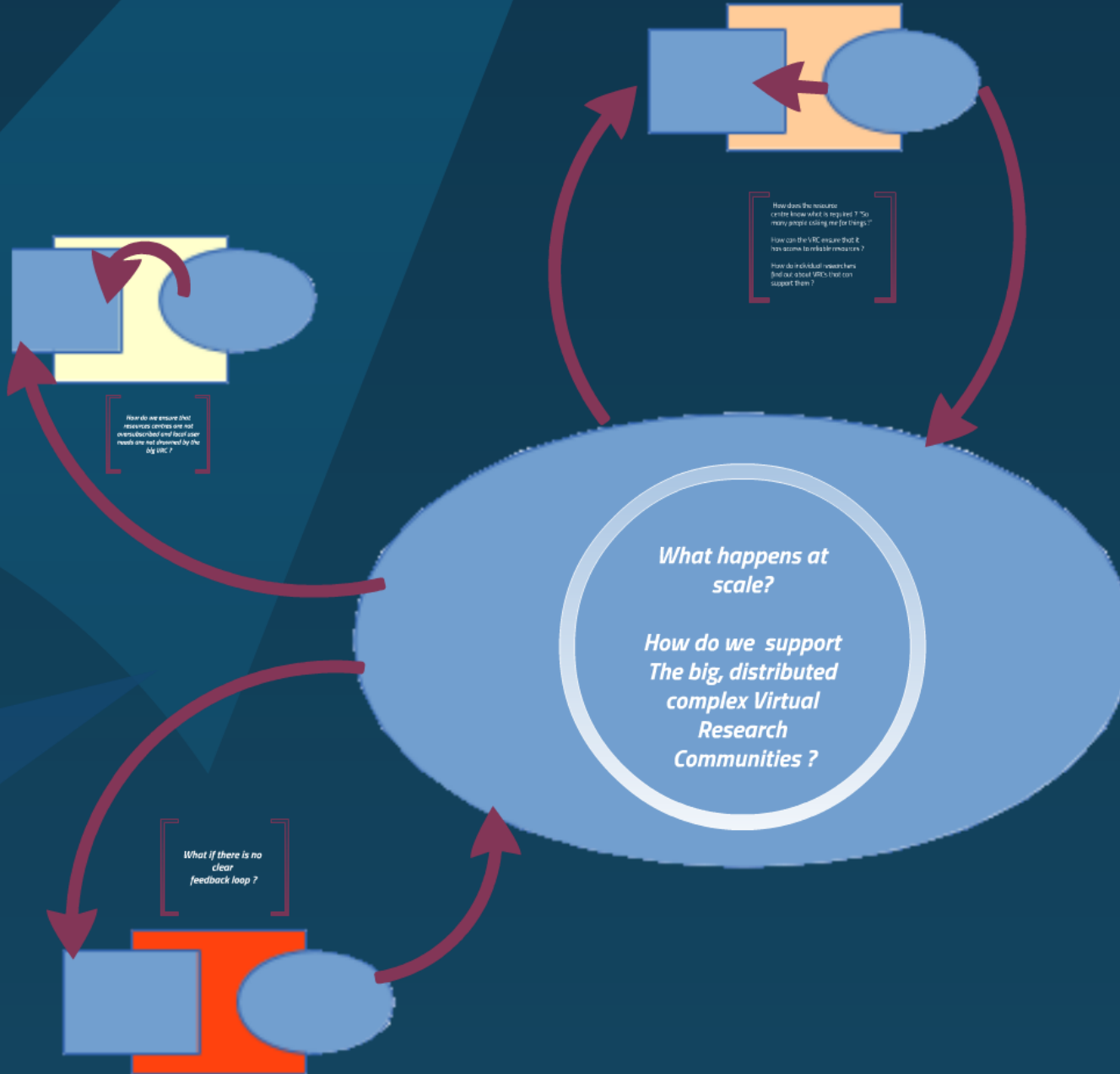
How do individual researchers
find out about VRCs that can
support them ?

How do we ensure that
resources centres are not
oversubscribed and local user
needs are not drowned by the
big VRC ?

What if there is no
clear
feedback loop ?

**What happens at
scale?**

**How do we support
The big, distributed
complex Virtual
Research
Communities ?**

How do we ensure that resources centres are not oversubscribed and local user needs are not drowned by the big VRC ?

ant

ium

How does the resource
centre know what is required ? "So
many people asking me for things ?"

How can the VRC ensure that it
has access to reliable resources ?

How do individual researchers
find out about VRCs that can
support them ?

How do we ensure that
resources centres are not
oversubscribed and local user
needs are not drowned by the
big VRC ?

What if there is no
clear
feedback loop ?

**What happens at
scale?**

**How do we support
The big, distributed
complex Virtual
Research
Communities ?**

PREZI

What do we learn from this ?

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**
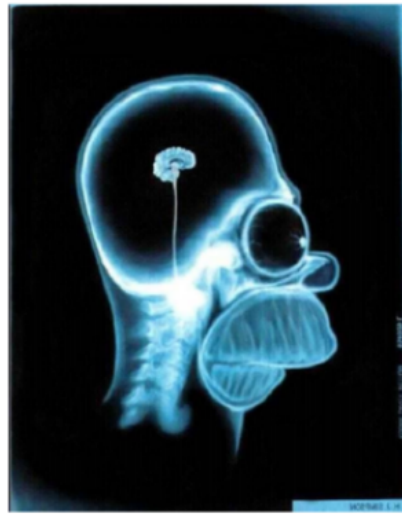
This is a luxury not all of our colleagues can **afford**

What should we do ?

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

Structural MRI vs. **Functional** MRI

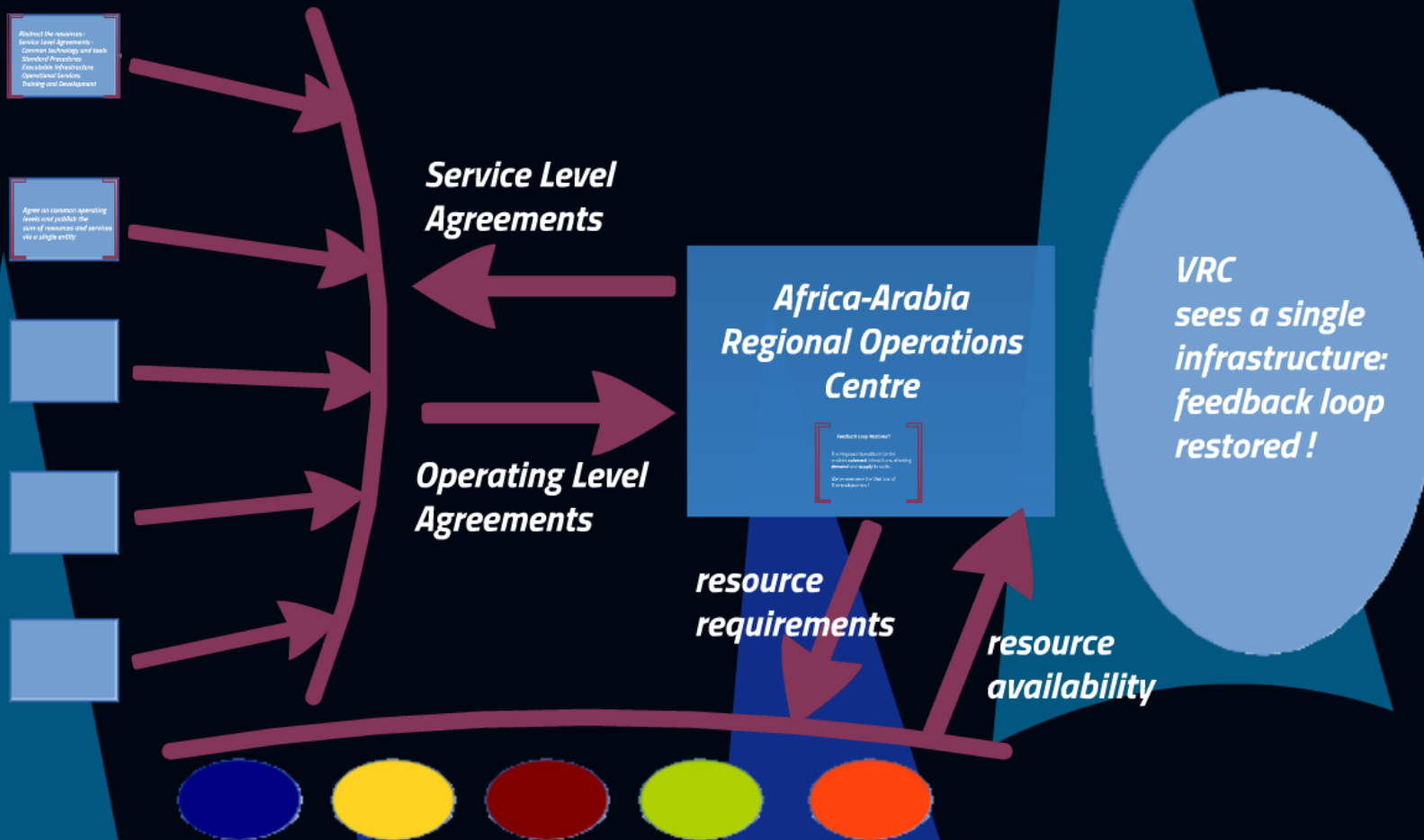**Structural** MRI reveals brain <u>anatomy</u>.

Functional MRI (fMRI) reveals brain <u>function</u>.

PREZI

Emergent Organisation

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

Service Level
Agreements

Operating Level
Agreements

Africa-Arabia
Regional Operations
Centre

VRC
sees a single
infrastructure:
feedback loop
restored !

resource
requirements

resource
availability

PREZI

Abstract the resources :
Service Level Agreements -
Common technology and tools
Standard Procedures
Executable Infrastructure
Operational Services
Training and Development

Agree on common operating levels and publish the sum of resources and services via a single entity

Feedback Loop Restored !

The Regional Operations Centre enables **coherent** interactions, allowing **demand** and **supply** to scale…

We've overcome the 2nd law of Thermodynamics !

"I imagine that right now, you're feeling a bit like Alice. Hmm? Tumbling down the rabbit hole?"

Luckily, everybody can be told what the ROC is...
and they can see it for themselves

The ROC is :
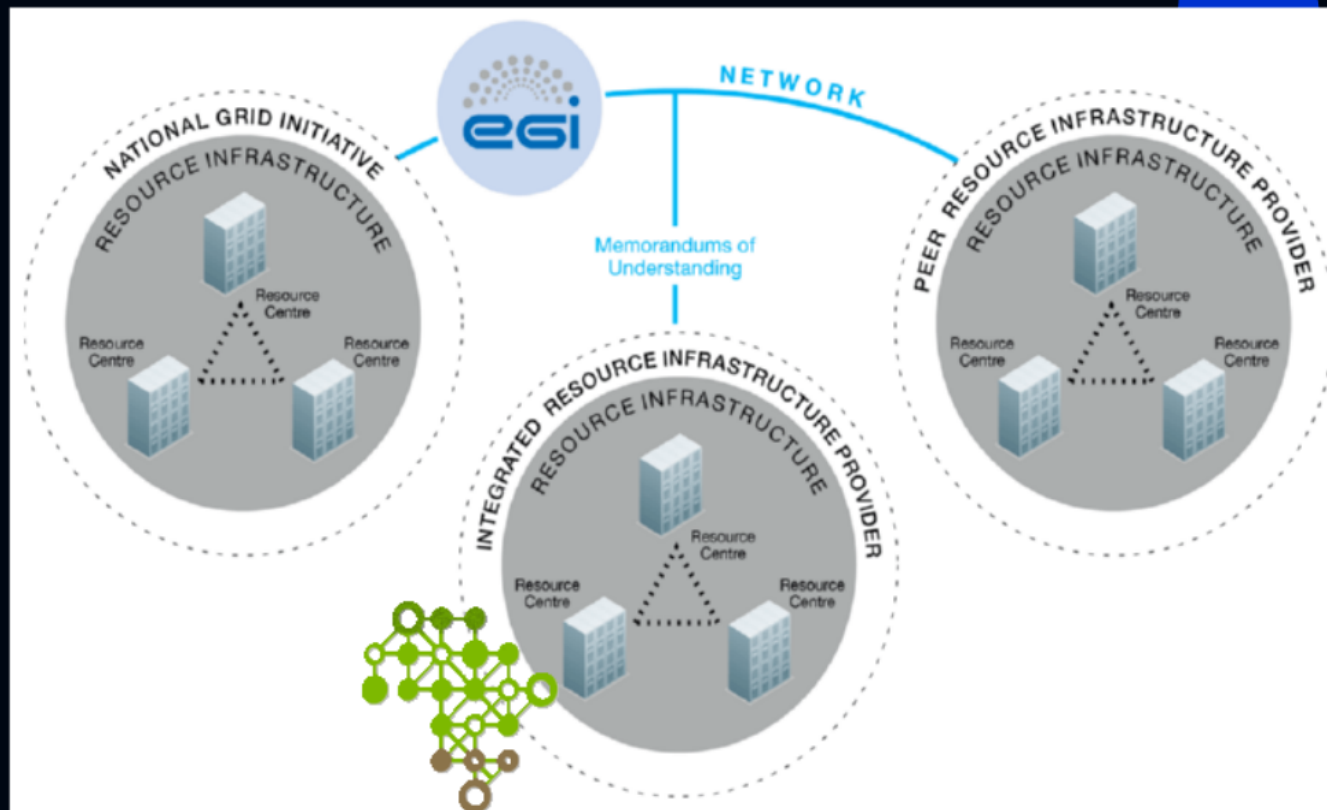A point of **coordination** for regional computational and data resources.

The **sum** of all **interoperable** computing resources

The **sum** of all **experts** and **support** staff at the sites

A **reference point** for best practice and documentation

A **support structure** for regional operations

An interface to **technology experts**

The ROC provides clean
interface to peer infrastructures

ROC Central Tools

The GOCDB is a central service where all sites, services, resources and personell are registered.

It helps you answe questions like :

Is there anything in Namibia ?
What is in Namibia ?
Who do we call in Namibia if someone from Cameroon wants to collaborate there ?

It's the starting point for every other service provided by the ROC

ROC Central tools

By interoperating with EGI.eu, we also provide other regionalised services for allowing cloud, grid and data infrastructures to transparently serve communities :

Operations Portal - http://operations-portal.egi.eu
Accounting Portal - http://accounting-portal.egi.eu
Monitoring and Alarms : https://nagios.c4.csir.co.za/nagios
Global Support System (GGUS) - https://ggus.eu

Lots more...

The ROC is not just about fancy tools ⭐

The power of a distributed computing infrastructure lies on its coherence
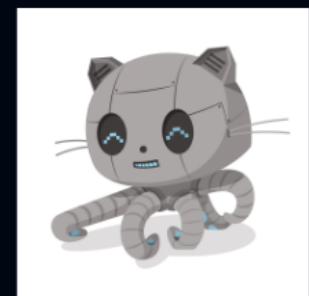
Standard Operating Procedures :
Clear, explicit procedures for performing common tasks

Automation:
A 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. Let experts code the implementation

I told you there would
be cats !

The ROC is not just about fancy tools ⭐

The power of a distributed computing infrastructure lies on its coherence
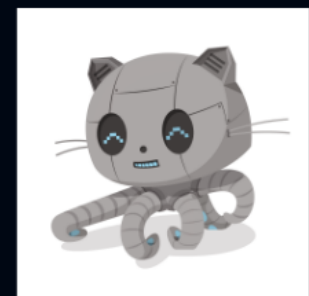
Standard Operating Procedures :
Clear, explicit procedures for performing common tasks

Automation:
A 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. Let experts code the implementation
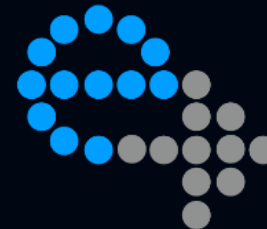
It doesn't end there

The mere existence of computationa and data infrastructures does not imply that it will be useful to you !

Let's talk about accessiblity and integration with other infrastructures

This is where the FP7-funded CHAIN-REDS project comes into play.

Objectives -

1) Extend and **consolidate** the international **cooperation** of Europe with other regions of the world in the domain of e-Infrastructures for R&E.

2) Promote, coordinate and support the effort of a critical mass of **non-European e-Infrastructures** for R&E to collaborate with Europe by addressing **interoperability** and **interoperation**.

3) Study the opportunities of **data sharing** across different e-Infrastructures and continents

4) Promote trust-building towards **Open Scientific Data** infrastructures across the world regions

5) Demonstrate the relevance of **intercontinental cooperation** in several scientific data fields addressing existing and emerging Virtual Research Communities (e.g. Earth Science, Climate Change, Genomics, etc.)

6) Provide guidance and recommendations for roadmaps for long-term global **collaboration** in e-Infrastructures and harmonisation of existing policies.
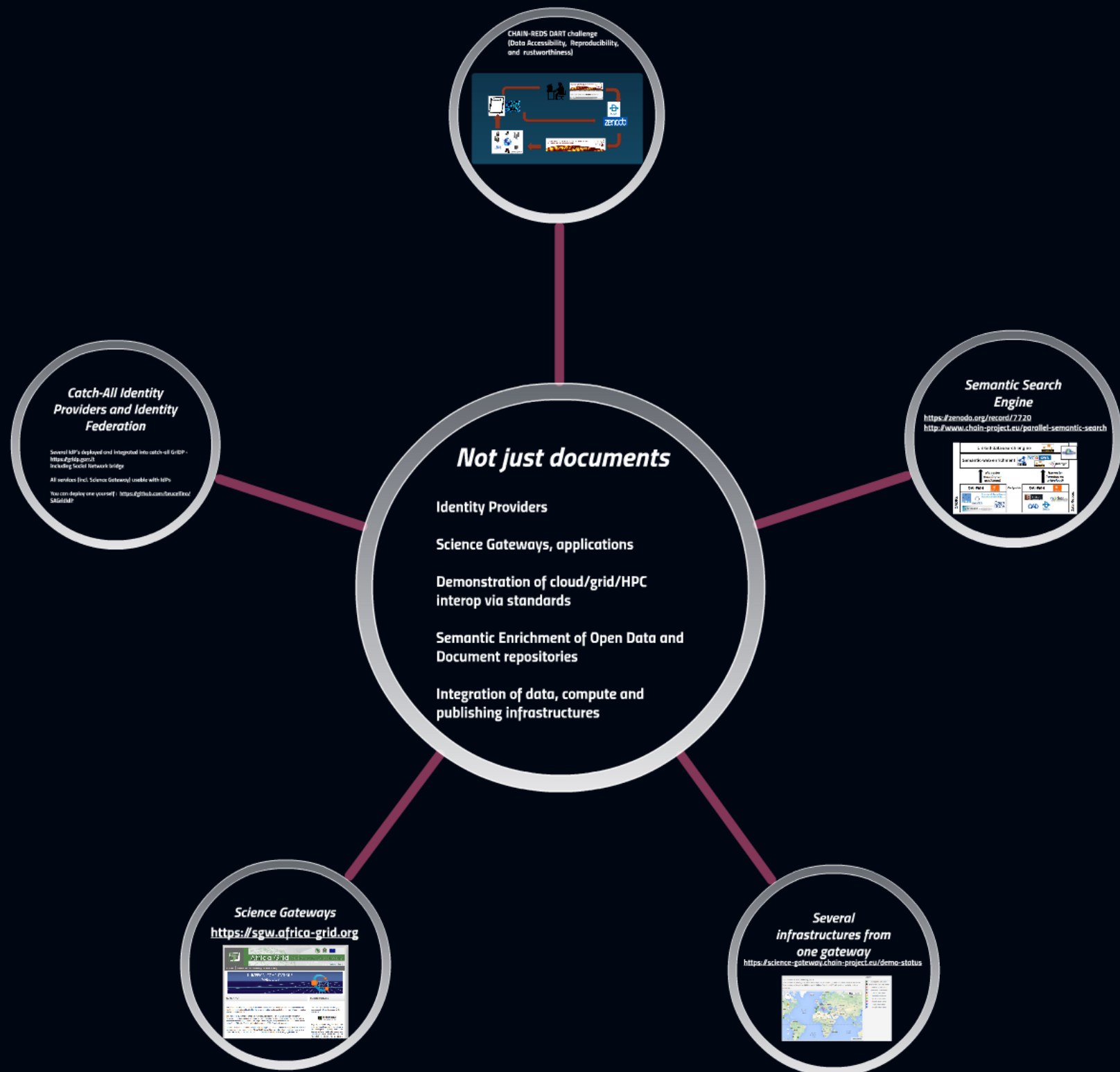
Some documents to read

http://www.chain-project.eu/deliverables

"Trans-continental Data Infrastructures and Data repositories" - http://documents.ct.infn.it/record/557/files/CHAIN-REDS-D4.1_v04.pdf

"Identity Federation and other alternate AA mechanisms" - http://documents.ct.infn.it/record/561/files/CHAIN-REDS-D5%201_v1_3%20Final.pdf

"Analysis of Data Infrastructures and Data repositories" - http://documents.ct.infn.it/record/560/files/CHAIN-REDS-D4.2_v05.pdf

CHAIN-REDS DART challenge
(Data Accessibility, Reproducibility,
and Trustworthiness)

Catch-All Identity Providers and Identity Federation

Several IdP's deployed and integrated into catch-all GriDP -
https://gridp.garr.it
Including Social Network bridge

All services (incl. Science Gateway) usable with IdPs

You can deploy one yourself : https://github.com/brucellino/SAGridIdP

Semantic Search Engine

https://zenodo.org/record/7720
http://www.chain-project.eu/parallel-semantic-search

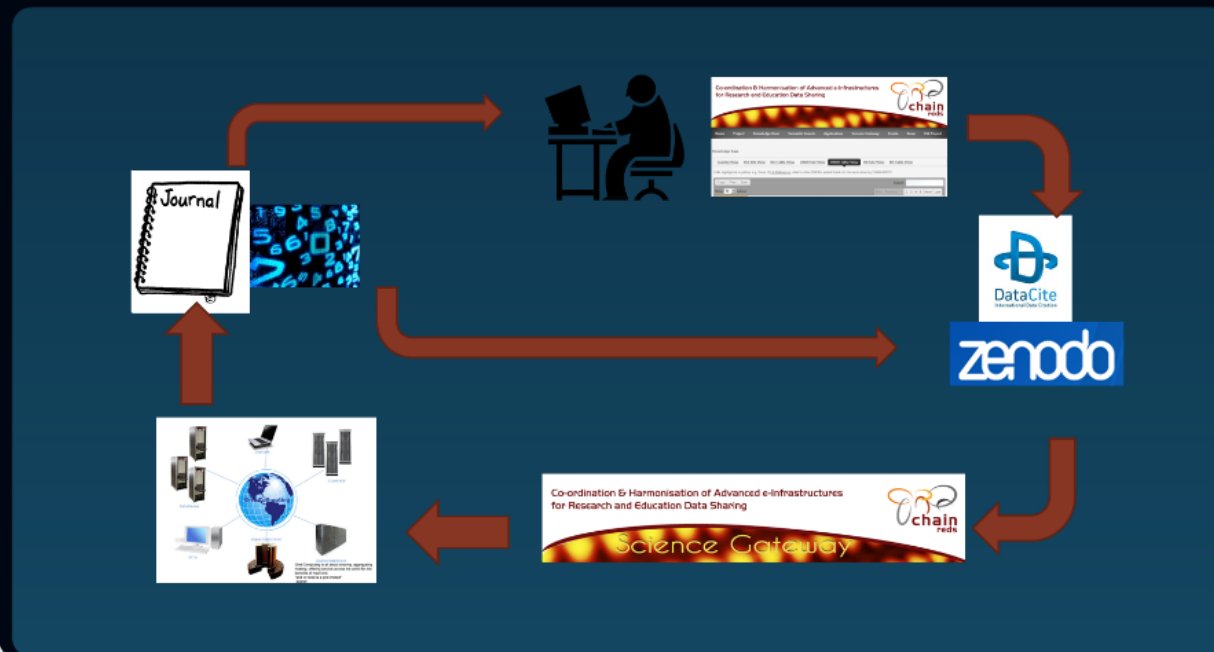## Not just documents

Identity Providers

Science Gateways, applications

Demonstration of cloud/grid/HPC interop via standards

Semantic Enrichment of Open Data and Document repositories

Integration of data, compute and publishing infrastructures

Science Gateways
https://sgw.africa-grid.org

Several infrastructures from one gateway
https://science-gateway.chain-project.eu/demo-status

PREZI

Not just documents

Identity Providers

Science Gateways, applications

Demonstration of cloud/grid/HPC interop via standards

Semantic Enrichment of Open Data and Document repositories

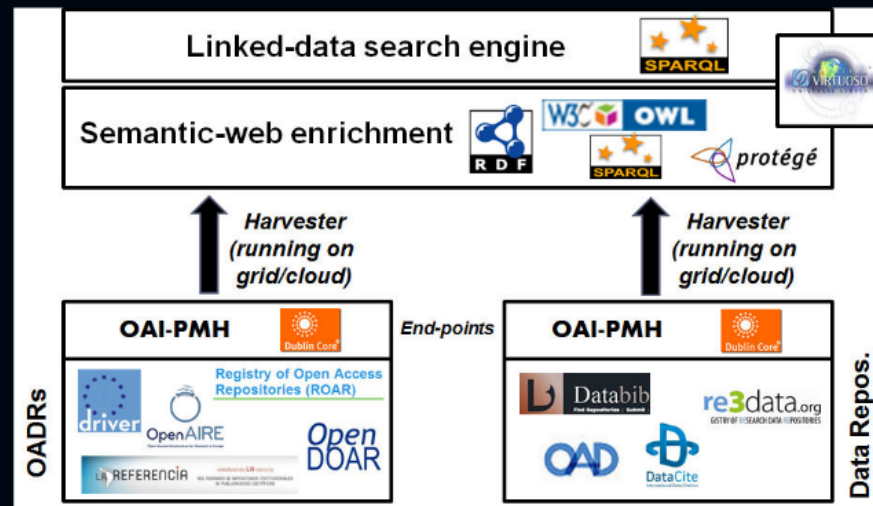**Integration of data, compute and publishing infrastructures**

CHAIN-REDS DART challenge
(Data Accessibility, Reproducibility,
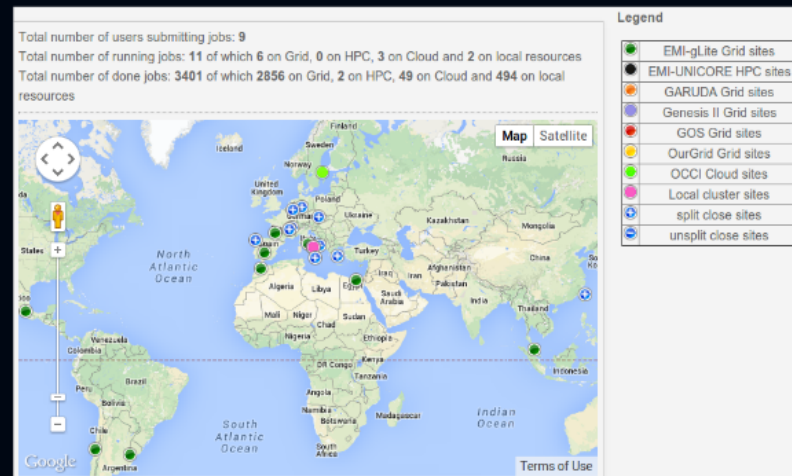and rustworthiness)
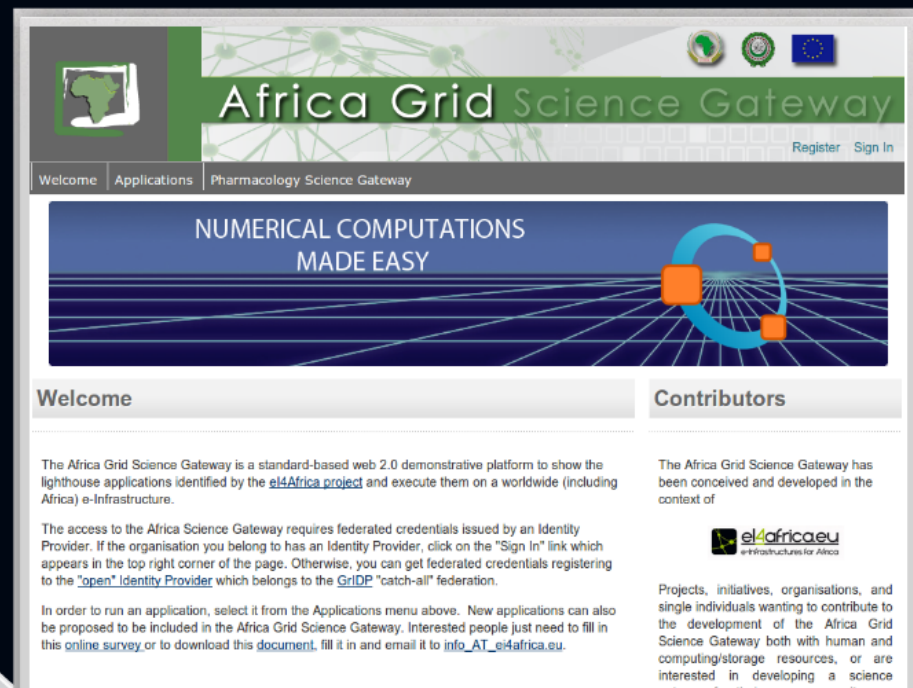
Semantic Search Engine

https://zenodo.org/record/7720
http://www.chain-project.eu/parallel-semantic-search

Several infrastructures from one gateway

https://science-gateway.chain-project.eu/demo-status

Science Gateways
https://sgw.africa-grid.org

Catch-All Identity Providers and Identity Federation

Several IdP's deployed and integrated into catch-all GrIDP -
https://gridp.garr.it
including Social Network bridge

All services (incl. Science Gateway) usable with IdPs

You can deploy one yourself : https://github.com/brucellino/
SAGridIdP

Let's talk...

Repository manager ?
Tell us about it and we'll put it on the map and work with you to get it OAI-PMH compliant if it isn't already

Researcher ?
Tell us about your research application and data needs, we'll liaise with the infrastructure providers

Developer ?
Work with us to implement relevant standards and demonstrate the power of a shared infrastructure

Thanks !
Get in touch

Bruce Becker :
bbecker@csir.co.za

@brucellino almost everywhere

http://www.sagrid.ac.za
http://roc.africa-grid.org

CHAIN-REDS project coordinator
Federico Ruggieri:
federico.ruggieri@roma3.infn.it
http://www.chain-project.org

Let's collaborate

Come hack with us
http://www.github.com/SAGridOps | AAROC

Come chat with us
http://www.facebook.com/SAGrid | AAROC
http://www.twitter.com/TheSAGrid

PREZI

Bruce Becker :
bbecker@csir.co.za

@brucellino almost everywhere

http://www.sagrid.ac.za
http://roc.africa-grid.org

CHAIN-REDS project coordinator
Federico Ruggieri:
federico.ruggieri@roma3.infn.it
http://www.chain-project.org

Let's

Come hack wi
http://www.g

Come chat wi
http://www.f
http://www.t

Let's collaborate

Come hack with us
http://www.github.com/SAGridOps | AAROC

Come chat with us
http://www.facebook.com/SAGrid | AAROC
http://www.twitter.com/TheSAGrid