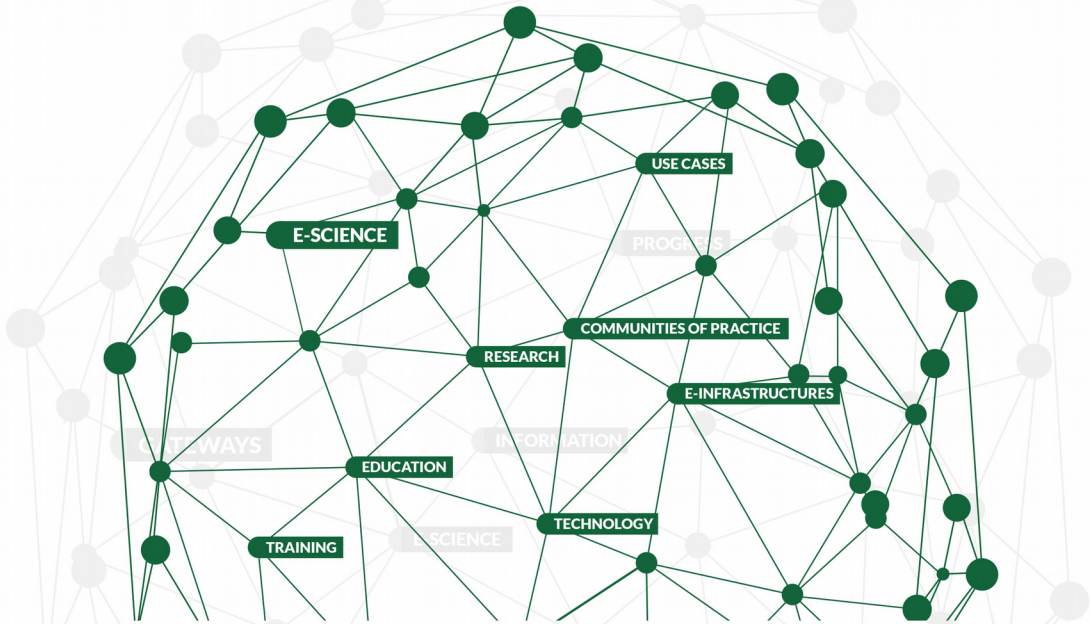
The Sci-GaIA Platform for Open Science



Sci-GaIA
www.sci-gaia.eu

The Sci-GaIA Project (www.sci-gaia.eu)

Bruce Becker, CSIR for Roberto Barbera and Sci-GaIA consortium.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement n°

Outline

- Introductory concepts, definitions and driving considerations
- How e-Infrastructures enable research
- Open Science – The Commons
- Summary and conclusions

The Sci-GaIA Project
(www.sci-gaia.eu)

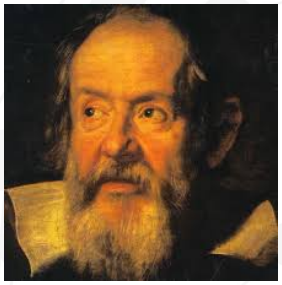
- **Energising Scientific Endeavour through Science Gateways and e-Infrastructures in Africa**
- Research Infrastructures – Coordination & Support Action
- Grant Agreement no. 654237
- EC contribution: ~1.4 MEuro
- Start date: 1 May 2015
- Duration: 24 Months

Bruce Becker

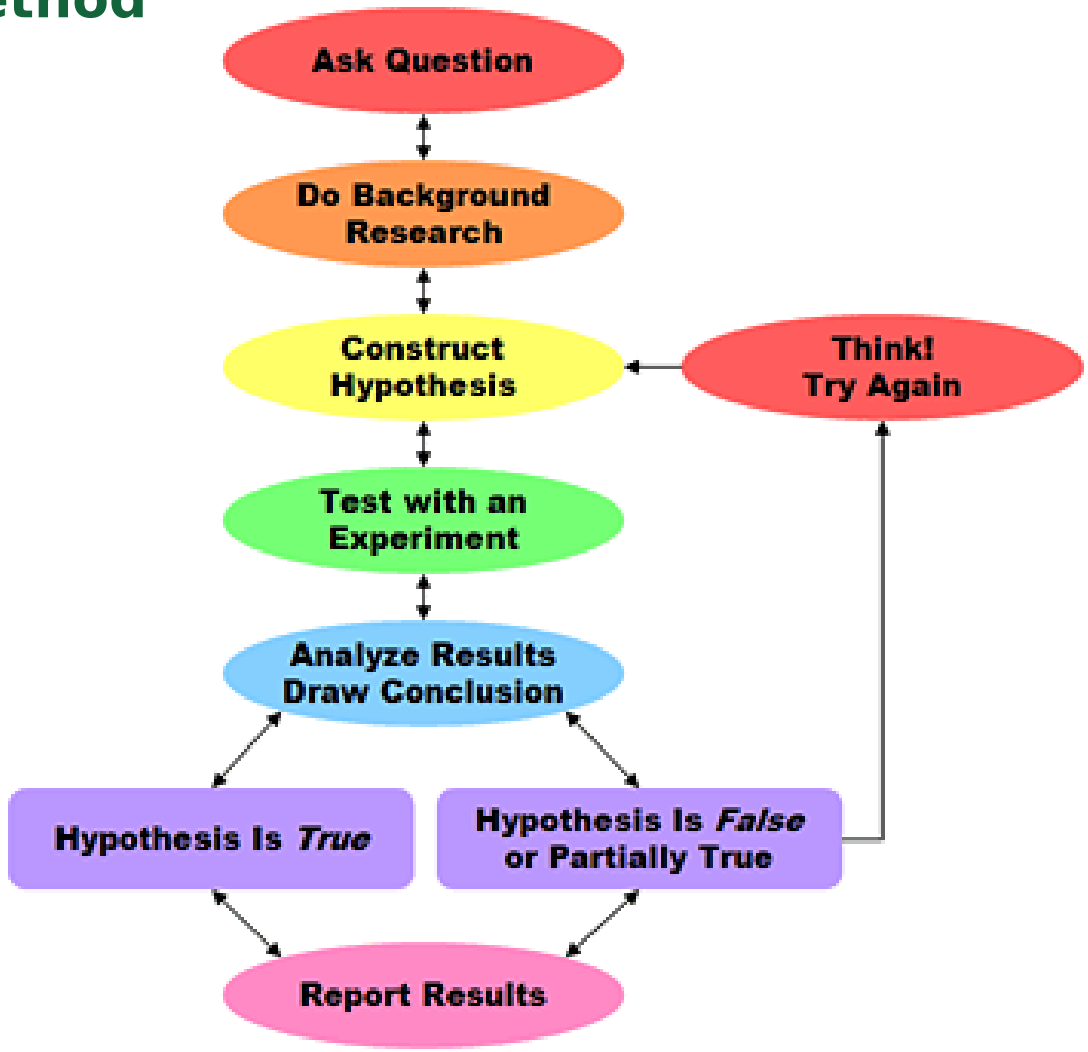
- Ph.D. 2007 from University of Cape Town
  - Study of the quark-gluon plasma at the ALICE experiment at CERN's Large Hadron Collider
  - Data acquisition and offline computing for ALICE
  - Postdocs at CEA Saclay, Paris and INFN Cagliari
- 2009 CSIR Meraka Institute
  - Coordinator of South African National Grid (9 universities)
  - Migrated to SANREN Competency Area in Meraka – Advanced Services for the network
- 2013 – Coordinator : Africa-Arabia Regional Operations Centre http://www.africa-grid.org
  - FP7 projects : CHAIN, CHAIN-REDS, ei4Africa
  - HP/UNESCO Brain Gain Initiative

Sci-GaIA
www.sci-gaia.eu

The Scientific Method

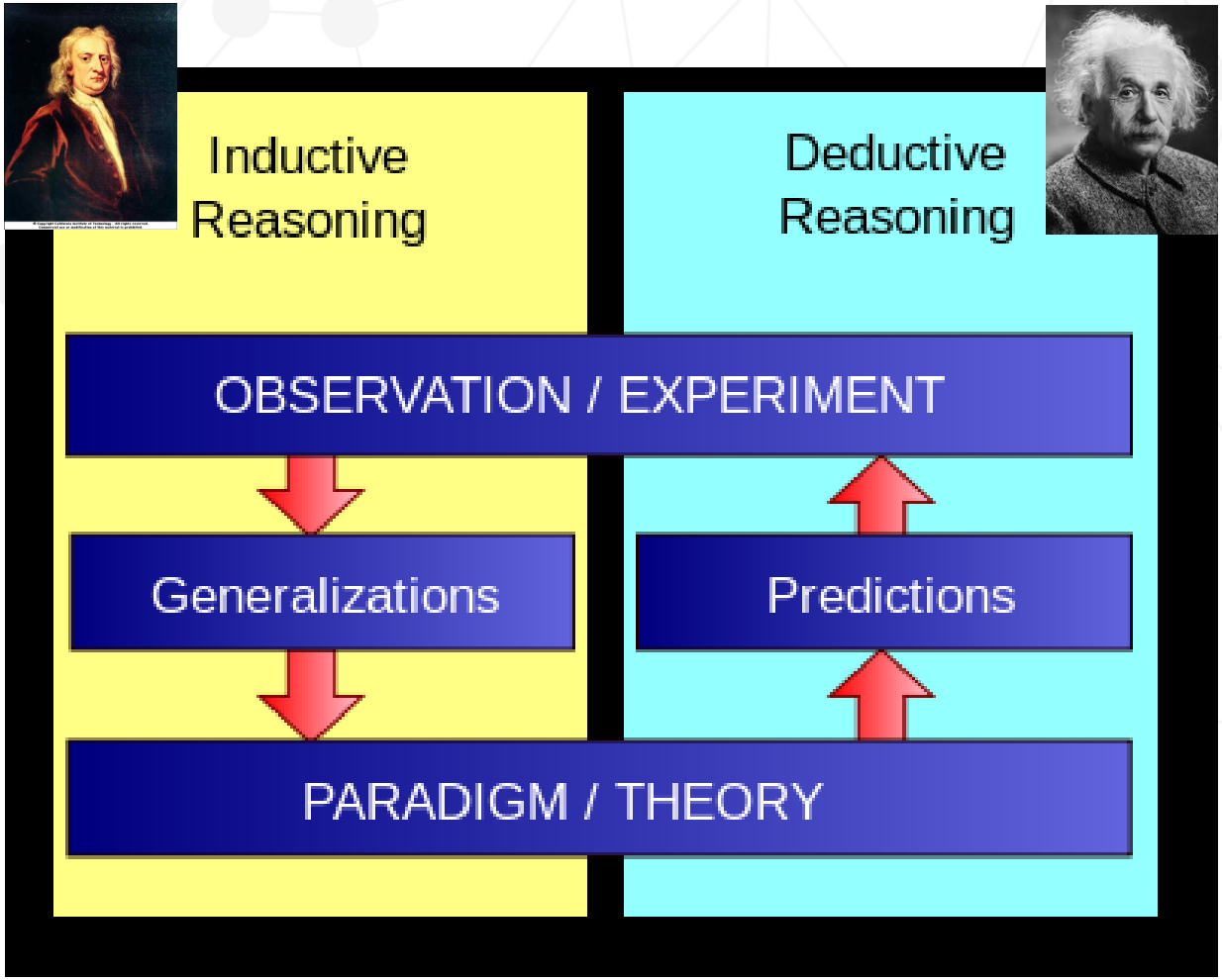


G. Galilei



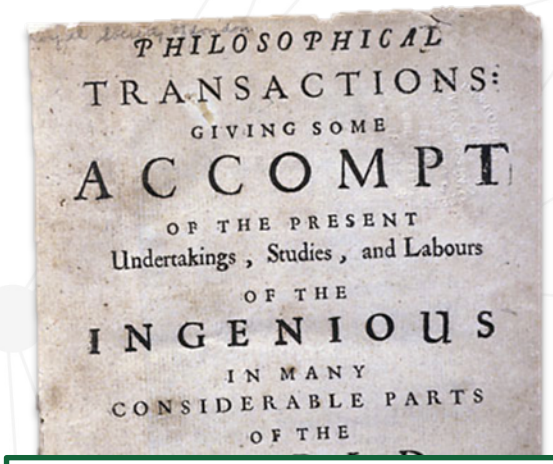
Sci-GaIA
www.sci-gaia.eu

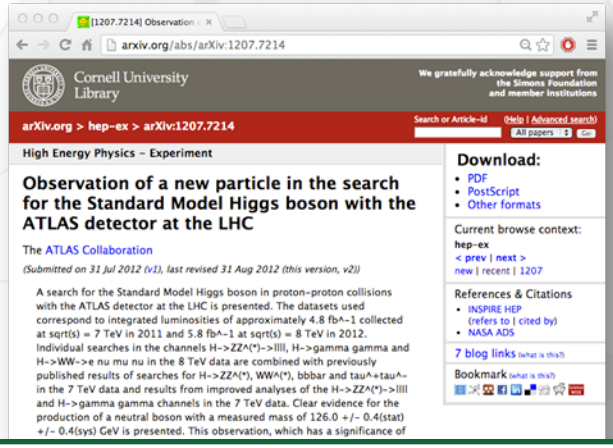
The Scientific Method

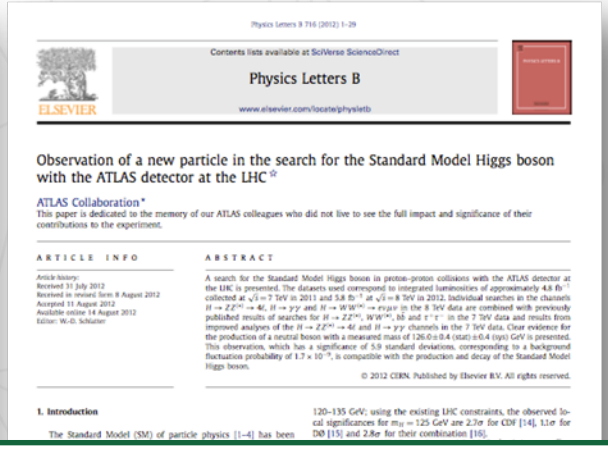


- **Examples of IR:**
  - Classical Mechanics
  - Newton's Gravitation Theory
- **Examples of DR:**
  - General Relativity
  - Standard Model

The “output” of the Scientific Method


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE


[1207.7214] Observation
arxiv.org/abs/arXiv:1207.7214
Cornell University Library
We gratefully acknowledge support from the Simons Foundation and member institutions
arXiv.org > hep-ex > arXiv:1207.7214
Search or Article-id (Help | Advanced search) All papers Go
High Energy Physics – Experiment
Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC
The ATLAS Collaboration
(Submitted on 31 Jul 2012 (v1), last revised 31 Aug 2012 (this version, v2))
A search for the Standard Model Higgs boson in proton-proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^-1 collected at sqrt(s) = 7 TeV in 2011 and 5.8 fb^-1 at sqrt(s) = 8 TeV in 2012. Individual searches in the channels H->ZZ^(*)->llll, H->gamma gamma and H->WW->e nu mu nu in the 8 TeV data are combined with previously published results of searches for H->ZZ^(*), WW^(*), bbbar and tau^+tau^- in the 7 TeV data and results from improved analyses of the H->ZZ^(*)->llll and H->gamma gamma channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 +/- 0.4(stat) +/- 0.4(sys) GeV is presented. This observation, which has a significance of
Download:
PDF
PostScript
Other formats
Current browse context: hep-ex
< prev | next >
new | recent | 1207
References & Citations
INSPIRE HEP (refers to | cited by)
NASA ADS
7 blog links (what is this?)
Bookmark (what is this?)


Physics Letters B
Contents lists available at SciVerse ScienceDirect
www.elsevier.com/locate/physletb
ELSEVIER
Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC
ATLAS Collaboration*
ARTICLE INFO
ABSTRACT
1. Introduction

Marked a real Scientific Revolution but...
it is the same since almost 4 centuries!

Sci-GaIA
www.sci-gaia.eu

The Pillars of the Scientific Method

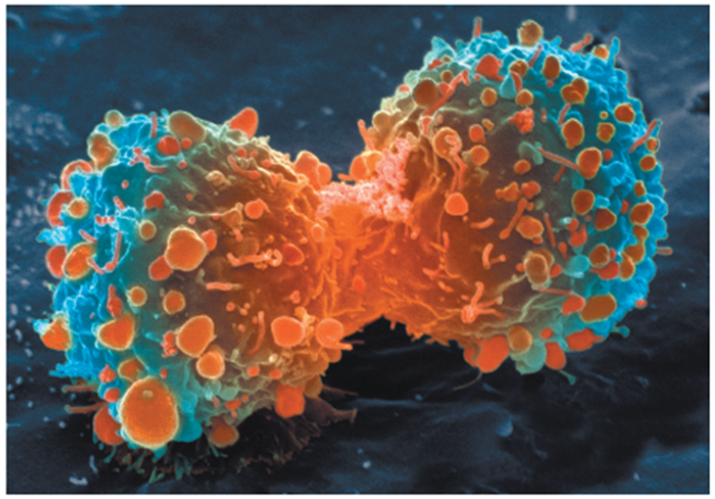
- **Repeatability**
  - The closeness of agreement between independent results obtained with the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory and after short intervals of time)
  - Affected by *random errors*

- **Reproducibility**
  - The closeness of agreement between independent results obtained with the same method on identical test material but under different conditions (different operators, different apparatus, different laboratories and/or after different intervals of time)
  - Affected by *systematic errors*

Reproducibility crisis (1/2)

COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge p.534 | EARTH SYSTEMS Past climates give valuable clues to future warming p.537 | HISTORY OF SCIENCE Descartes' lost letter tracked using Google p.540 | OBITUARY Wylie Vale and an elusive stress hormone p.542



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will

investigators must reassess their approach translating discovery research into gre clinical success and impact.

Many factors are responsible for the h failure rate, notwithstanding the inh ently difficult nature of this disease. C tainly, the limitations of preclinical to

47/53 "landmark" publications could not be replicated

[Begley, Ellis Nature, 483, 2012]

Must try harder

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorou at the data — and at themselves.

Error prone

Biologists must realize the pitfalls of work on massive amounts of data.

If a job is worth doing, it is worth doing twice

Researchers and funding agencies need to put a premium on ensuring that results are reproducible, argues Jonathan F. Russell.

The case for open computer programs

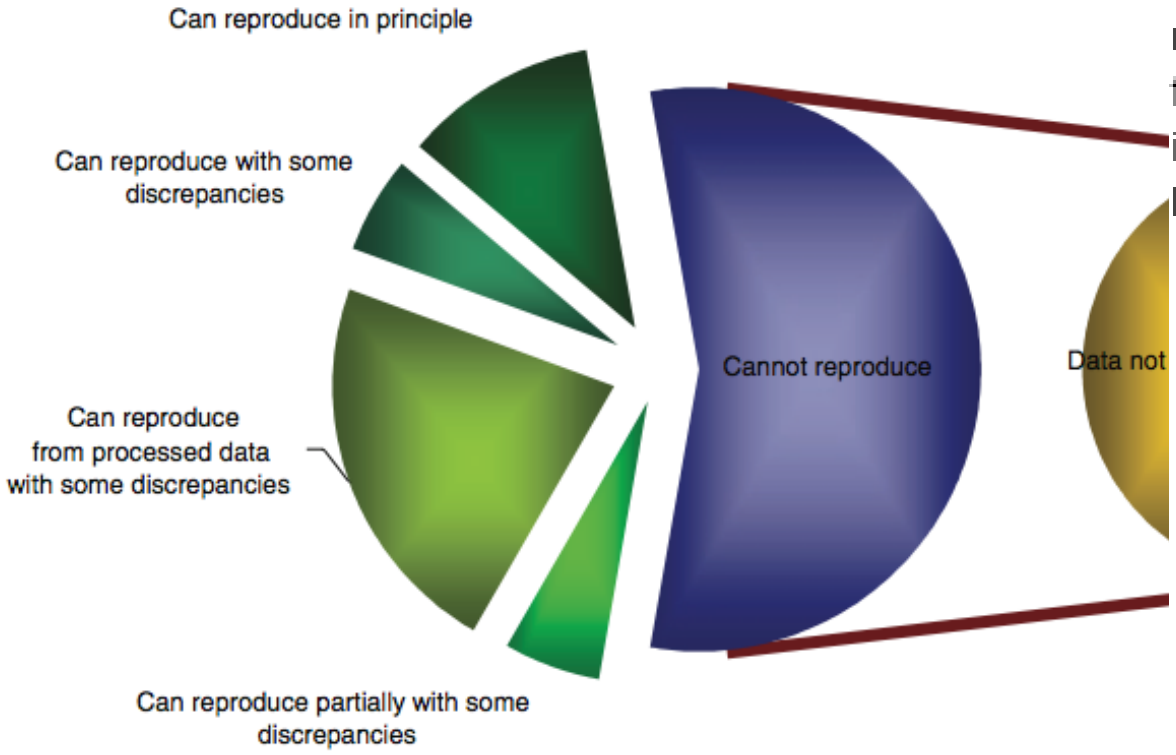
Six red flags for suspect work

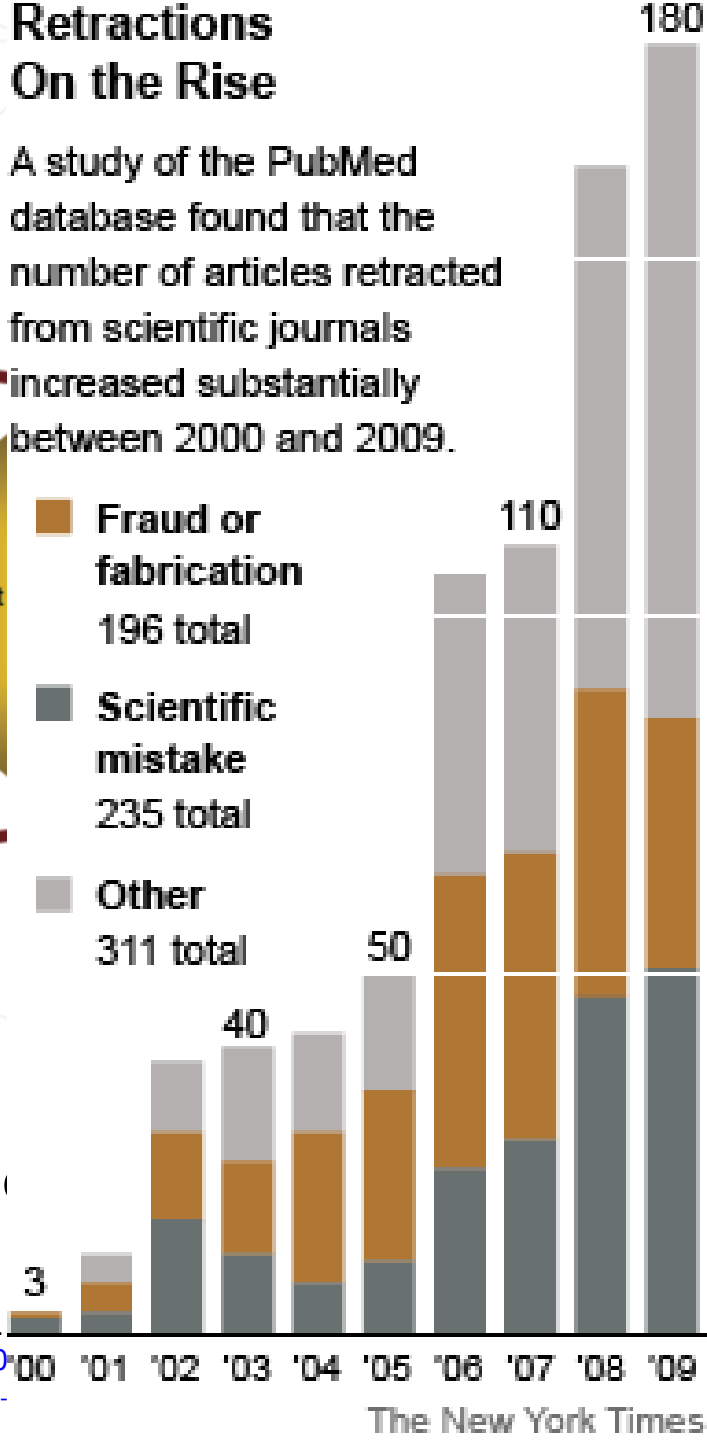
C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

Know when your numbers are significant

Sci-GaIA
www.sci-gaia.eu

Reproducibility crisis (2/2)





More retractions:

- >15x increase in the last 10 years
- At current rate, by 2045 as many pap published as retracted

1. Ioannidis et al., 2009. Repeatability of published microarray gene expression analyses.
2. Science publishing: The trouble with retractions http://www.nature.com/news/2011/1110
3. Bjorn Brembs: Open Access and the looming crisis in science https://theconversation.com/open-

Challenges in irreproducible research
(http://www.nature.com/nature/focus/reproducibility/index.html)

nature International weekly journal of science

Search Go ▸ Advanced search

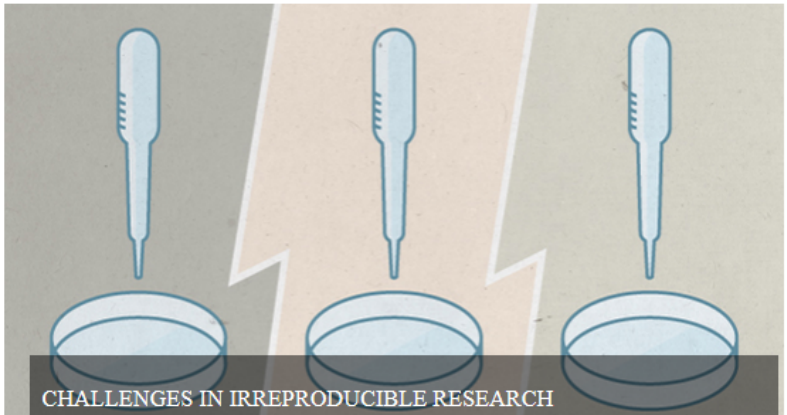
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive › Specials & supplements archive › Challenges in irreproducible research

Take part in Nature Publishing Group's annual reader survey here for the chance to win a Macbook Air. Find out more ▼ ✖

SPECIAL

▸ See all specials



CHALLENGES IN IRREPRODUCIBLE RESEARCH

No research paper can ever be considered to be the final word, and the replication and corroboration of research results is key to the scientific process. In studying complex entities, especially animals and human beings, the complexity of the system and of the techniques can all too easily lead to results that seem robust in the lab, and valid to editors and referees of journals, but which do not stand the test of further studies. *Nature* has published a series of articles about the worrying extent to which research results have been found wanting in this respect. The editors of *Nature* and the *Nature* life sciences research journals have also taken substantive steps to put our own houses in order, in improving the transparency and robustness of what we publish. Journals, research laboratories and institutions and funders all have an interest in tackling issues of irreproducibility. We hope that the articles contained in this collection will help.

Free full access

▾ Editorial ▾ Features ▾ News and analysis ▾ Comment ▾ Perspectives and reviews

EDITORIAL

Journals unite for reproducibility
Consensus on reporting principles aims to improve biomedical research.
Nature **515**, 7 (6 November 2014)

Code share
Papers in Nature journals should make computer code accessible where possible.
Nature **514**, 536 (29 October 2014)

Reducing our irreproducibility
Nature **496**, 398 (25 April 2013)

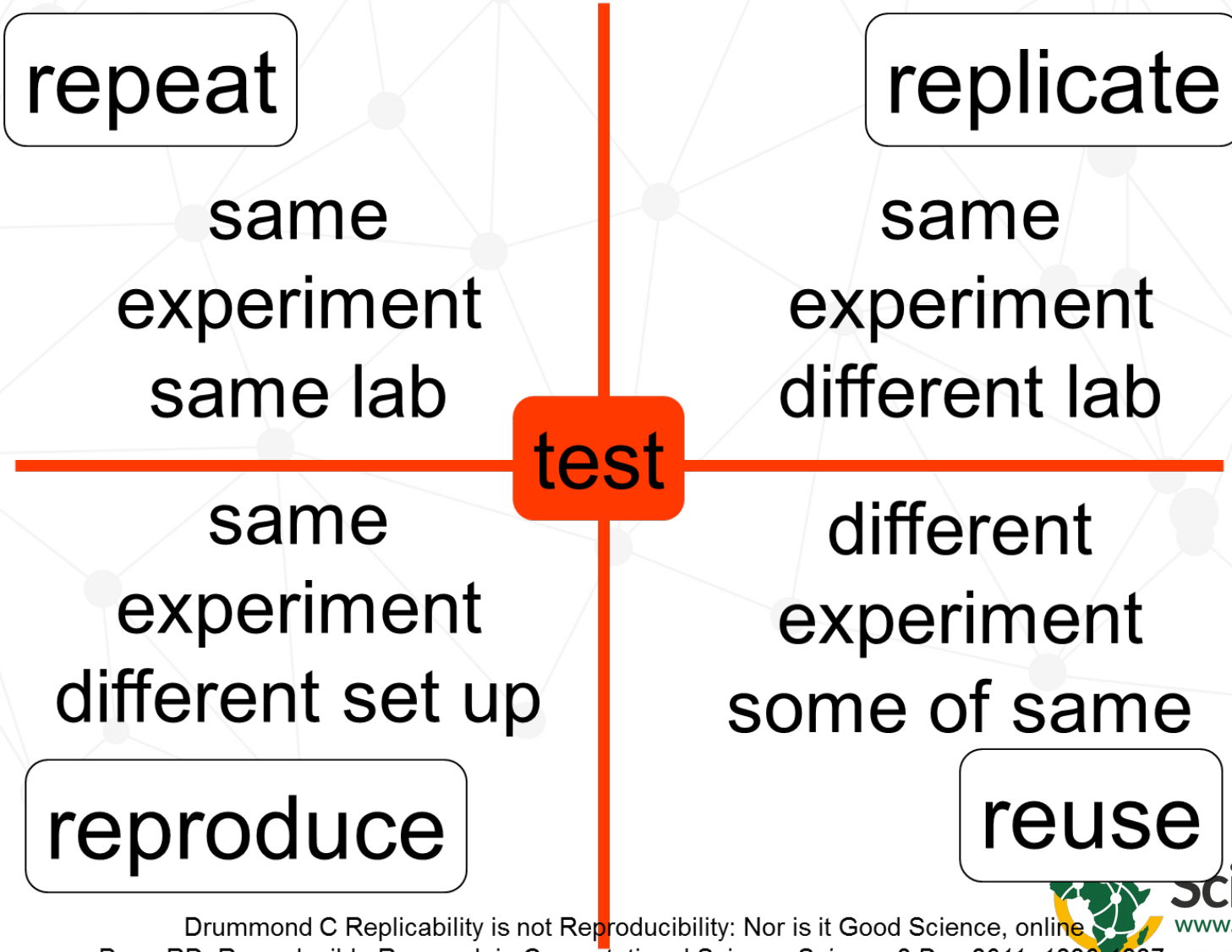
Further confirmation needed
A new mechanism for independently replicating research findings is one of several changes required to improve the quality of the biomedical literature.
Nature Biotechnology **30**, 806 (10 September 2012)

Error prone
Biologists must realize the pitfalls of work on massive amounts of data.
Nature **487**, 406 (26 July 2012)

Must try harder
Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.
Nature **483**, 509 (29 March 2012)

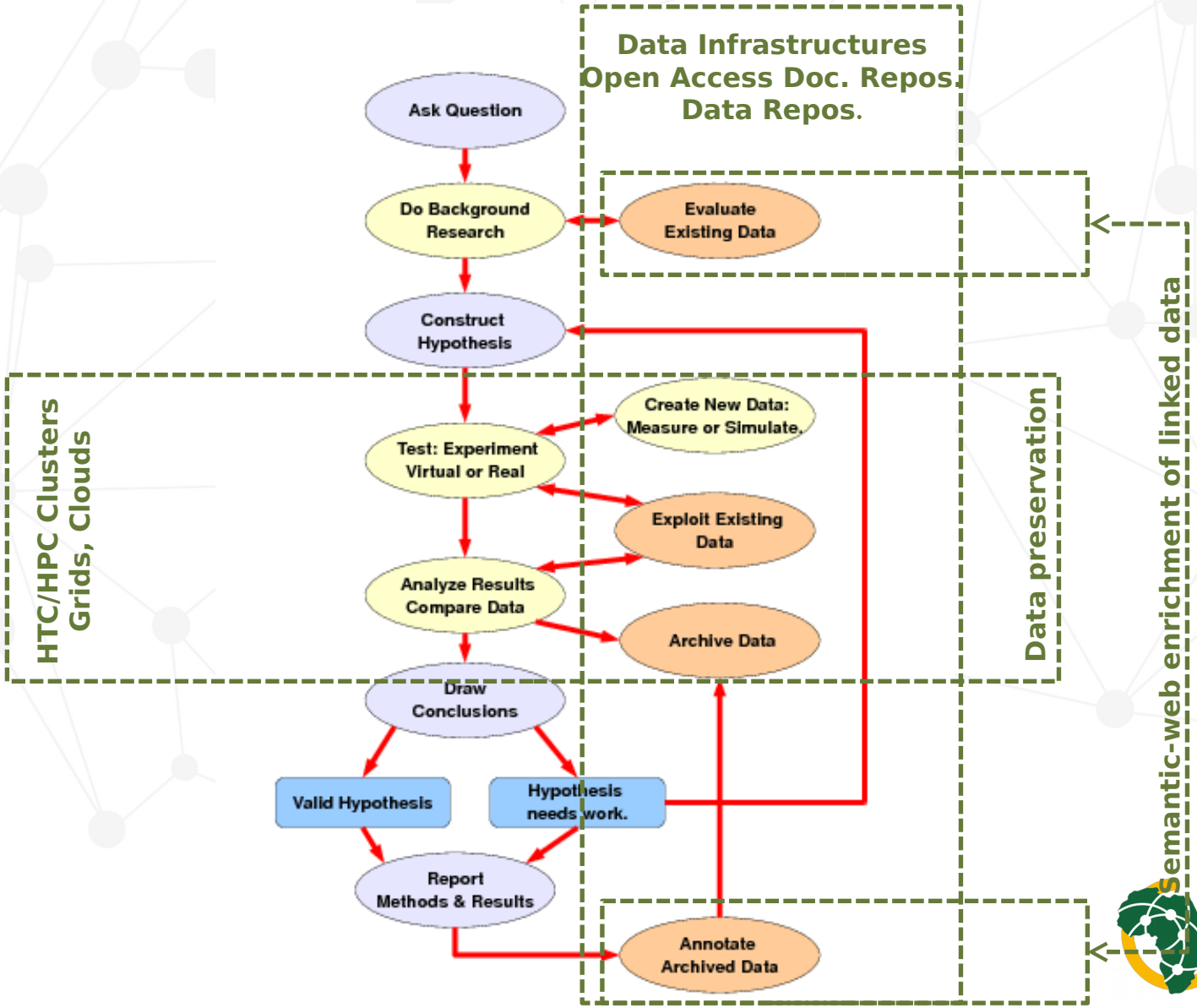
Sci-GaIA
www.sci-gaia.eu

Repeatability and Reproducibility are not all…



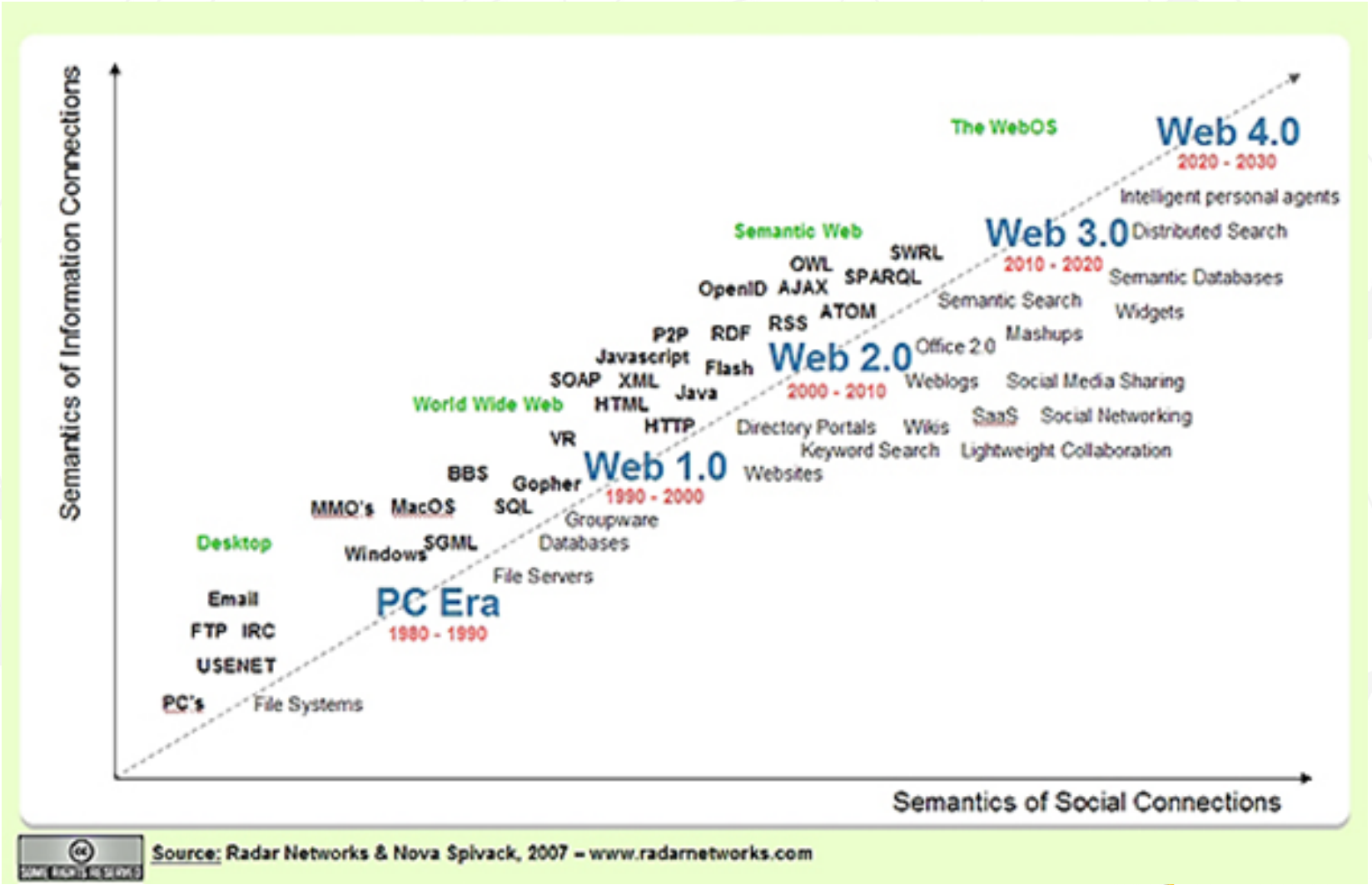
Drummond C Replicability is not Reproducibility: Nor is it Good Science, online
Peng RD, Reproducible Research in Computational Science *Science 2 Dec 2011: 1226-1227.*

How do e-Infrastructures support the Scientific Method?



Sci-GaIA
www.sci-gaia.eu

Evolution of the Web



Source: Radar Networks & Nova Spivack, 2007 – www.radarnetworks.com

Sci-GaIA
www.sci-gaia.eu

Web 2.0/3.0 technologies
(http://lifeboat.com/ex/web.3.0)

- **Ubiquitous Connectivity**:
  - Broadband adoption;
  - Mobile Internet access;
  - Mobile devices;
- **Network Computing**:
  - Software-as-a-service business models;
  - Web services interoperability;
  - Distributed computing (P2P, grid computing, cloud computing, […]);
- **Open Technologies**:
  - Open APIs and protocols;
  - Open data formats;
  - Open-source software platforms;
  - Open data (Creative Commons, Open Data License, etc.);
- **Open Identity:**
  - Open identity (OpenID);
  - Open reputation;
  - Portable identity and personal data (for example, the ability to port your user account and search history from one service to another);
- **The Intelligent Web:**
  - Semantic Web technologies (Semantic application platforms, and statement-based datastores such as triplestores, tuplestores and associative databases);
  - Distributed databases — […]
  - Intelligent applications (natural language processing, machine learning, machine reasoning, autonomous agents).”

Definitions
(http://book.openingscience.org)

- ***"Science 2.0** refers to all scientific culture, including scientific communication, which employs features enabled by Web 2.0* [and Web 3.0] *and the Internet"*
- ***"Open Science** refers to a scientific culture that is characterized by its openness. Scientist share results almost immediately and with a very wide audience"*

Opening Science

ISBN: 331900025X

Opening Science

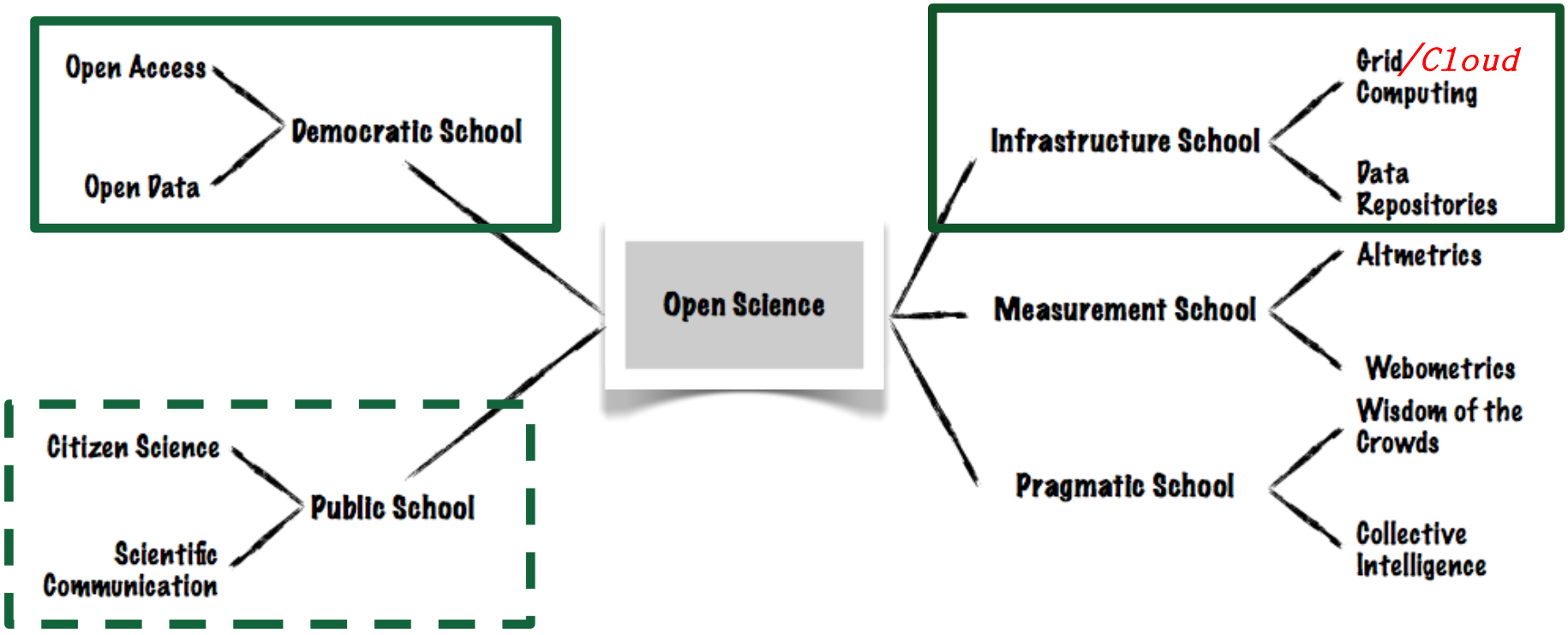
The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing

Figures
book.openingscience.org

Sci-GaIA
www.sci-gaia.eu

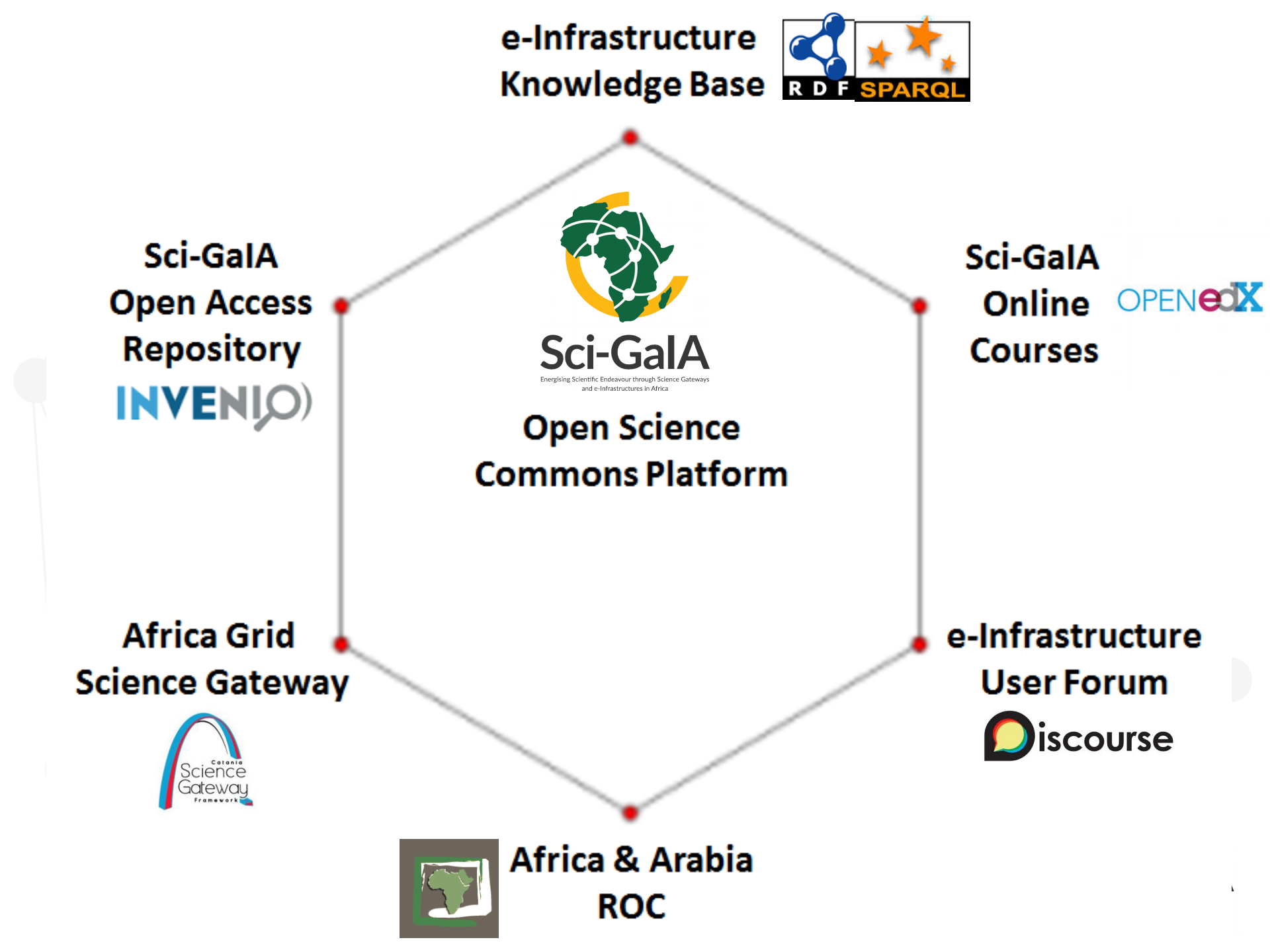
Open Science Schools of Thought

(http://book.openingscience.org)



Open Science Commons pillars

- Data:
  - *“Data is the subject matter for research. It should be dealt with according to the principles of open access and open science, while maintaining trust and privacy for researchers”*
- e-Infrastructures:
  - *“The technology and technical services supporting researchers, building towards integrated services and interoperable infrastructures across Europe and the world”*
- Scientific instruments:
  - *“The equipment and collaborations which generate scientific data, from small-scale lab machines to global collaborations around massive facilities”*
- Knowledge:
  - *“The human networks, understanding and material capturing skills and experience required to carry out open science using the three other pillars”*


e-Infrastructure
Knowledge Base
RDF SPARQL
Sci-GaIA
Open Access
Repository
INVENIO
Sci-GaIA
Energising Scientific Endeavour through Science Gateways
and e-Infrastructures in Africa
Open Science
Commons Platform
Sci-GaIA
Online
Courses
OPENedX
Africa Grid
Science Gateway
Catania
Science
Gateway
Framework
e-Infrastructure
User Forum
Discourse
Africa & Arabia
ROC

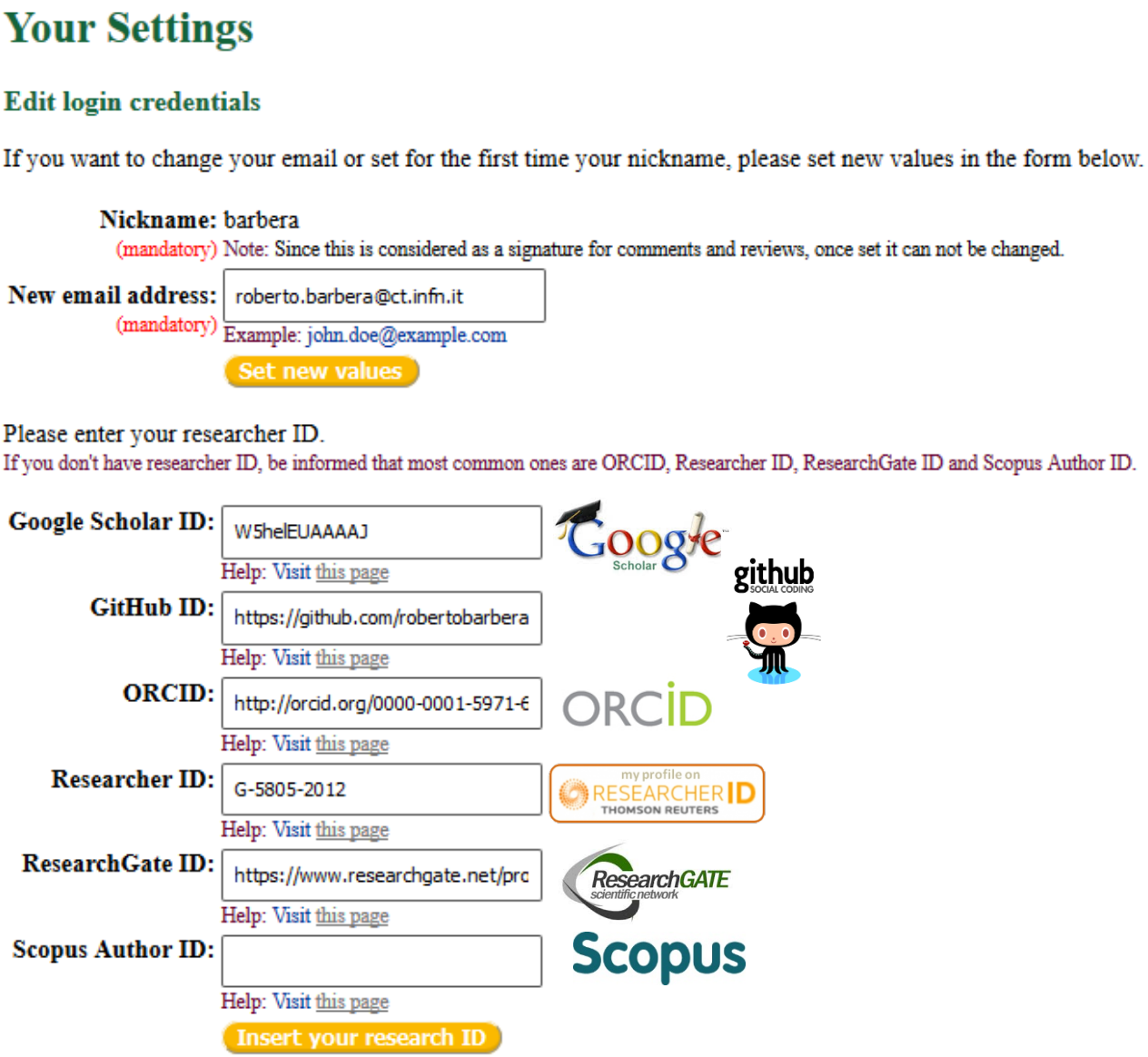
The Sci-GaIA requirements for an Open Access Repository

- Requirements:
  - Open source
  - Standard compliant
  - Well supported
  - Scalable up to O(10^6)-O(10^7) resources (to begin with)
- Choice:
  - **Invenio** (www.invenio-software.org) – actual version: 1.2.1 + our «add-ons»
- Motivations:
  - Fully compliant with OAI-PMH and Marc21 standards
  - Co-developed by an international collaboration comprising institutes such as CERN, DESY, EPFL, FNAL, SLAC and used by about 30 scientific institutions worldwide
  - ZENODO (OpenAIRE flagship repository), SCOAP^3 and inSPIRE HEP repositories are based on Invenio
  - The CERN Document Server (http://cds.cern.ch/), based on Invenio, contains more than 1.3 million documents
  - UNESCO and UEMOA are leading an initiative to create a virtual library based on Invenio in 8 West African countries

The Sci-GaIA Open Access Repository

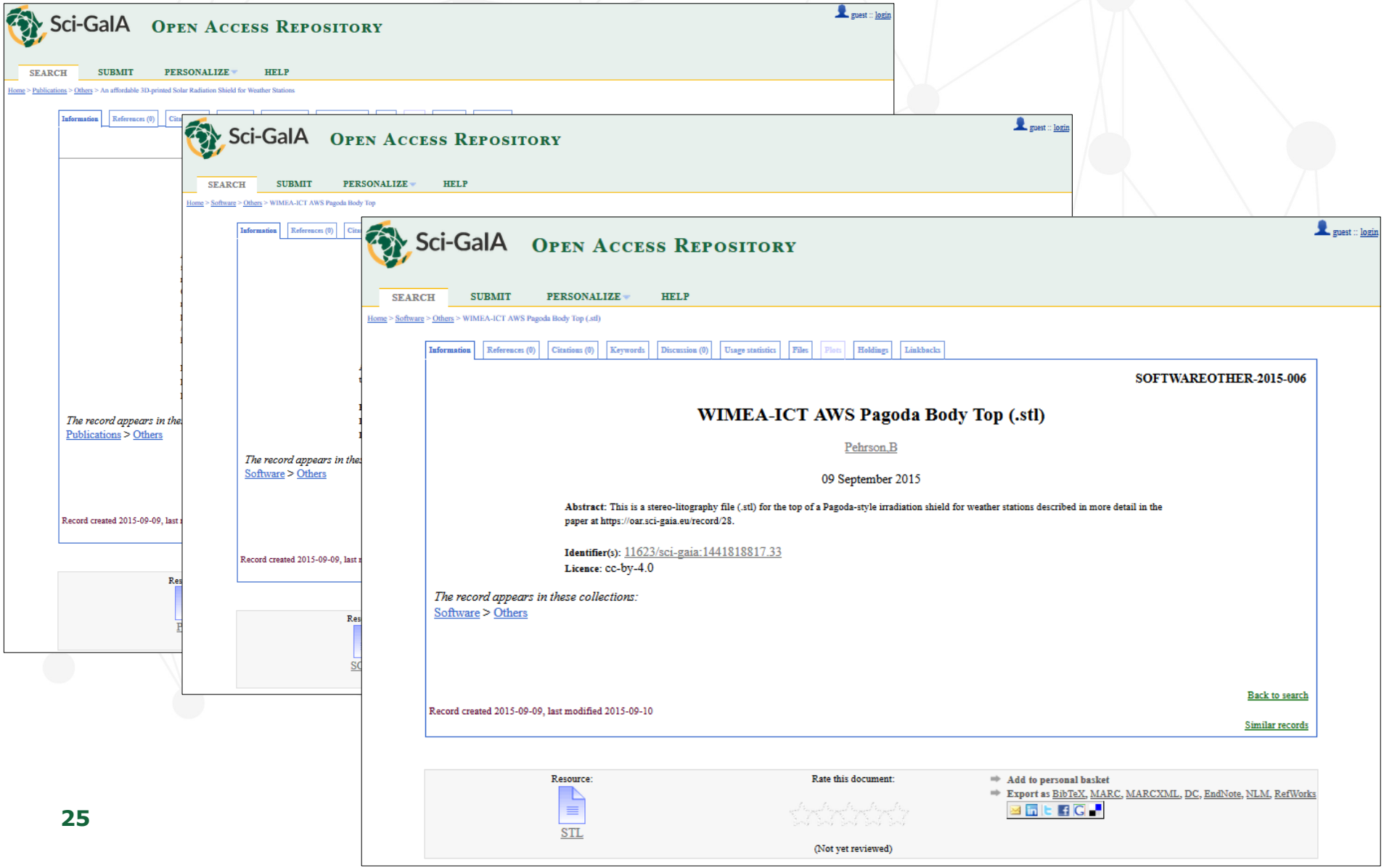
(http://oar.sci-gaia.eu)

Alternative reputation systems: possibility to add researcher ID's



(Single) resource upload and DOI registration

Examples of (cross-referenced) resources



Visibility and compliance

(Full conforming with Open Archive Initiative's standards)

Registration Record

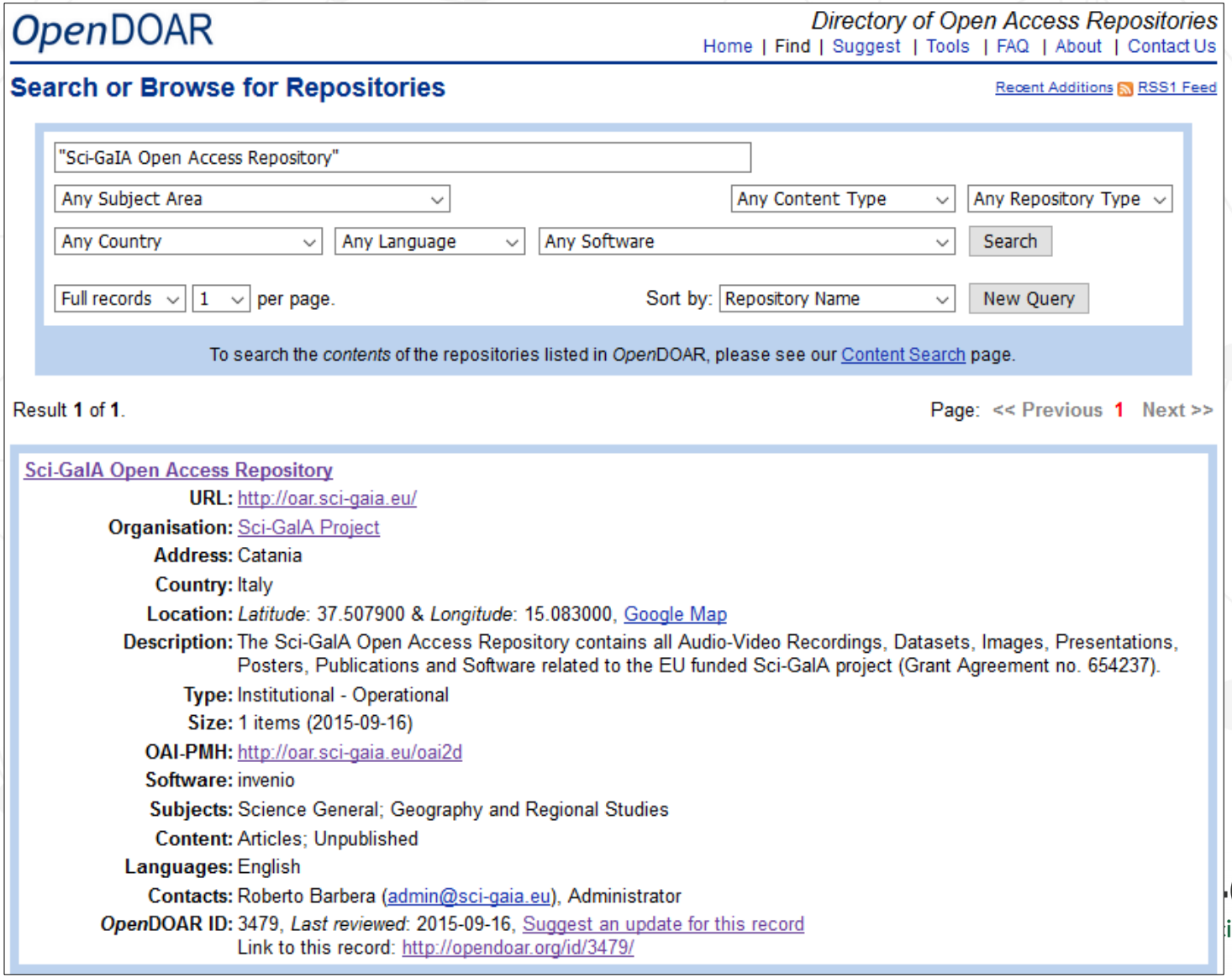
element name	element value
Base URL	http://oar.sci-gaia.eu/oai2d
Repository Name	Open Access Repository
Protocol Version	2.0
Email	admin@sci-gaia.eu
Registration Date	2015-08-12T17:48:44Z
Date Last Validated	Wed Aug 12 17:48:44 2015
OAI Repository ID	oar.sci-gaia.eu

If you are the maintainer of this repository, you may to update the information recorded to match new information exposed via the Identify response by running the validation/registration process again. Go to the validation page and select "Register this site".

Sat Aug 15 09:34:50 2015

Visibility and compliance

(Registered as an OpenDOAR data provider)



OpenDOAR

Directory of Open Access Repositories

Home | Find | Suggest | Tools | FAQ | About | Contact Us

Search or Browse for Repositories

Recent Additions | RSS1 Feed

"Sci-GaIA Open Access Repository"

Any Subject Area | Any Content Type | Any Repository Type

Any Country | Any Language | Any Software | Search

Full records | 1 | per page. | Sort by: Repository Name | New Query

To search the *contents* of the repositories listed in *OpenDOAR*, please see our Content Search page.

Result **1** of **1**.

Page: << Previous 1 Next >>

Sci-GaIA Open Access Repository

URL: http://oar.sci-gaia.eu/

Organisation: Sci-GaIA Project

Address: Catania

Country: Italy

Location: *Latitude*: 37.507900 & *Longitude*: 15.083000, Google Map

Description: The Sci-GaIA Open Access Repository contains all Audio-Video Recordings, Datasets, Images, Presentations, Posters, Publications and Software related to the EU funded Sci-GaIA project (Grant Agreement no. 654237).

Type: Institutional - Operational

Size: 1 items (2015-09-16)

OAI-PMH: http://oar.sci-gaia.eu/oai2d

Software: invenio

Subjects: Science General; Geography and Regional Studies

Content: Articles; Unpublished

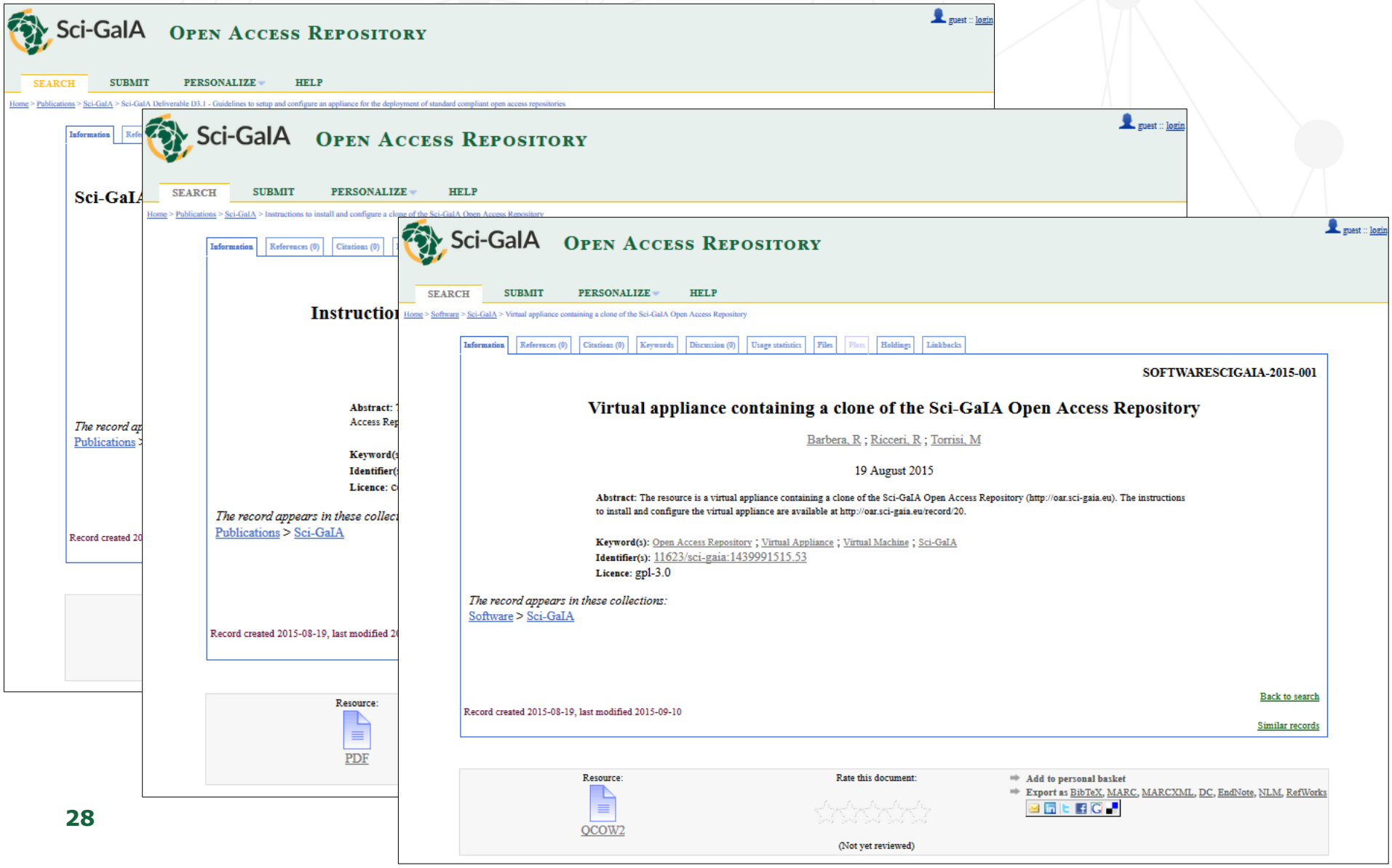
Languages: English

Contacts: Roberto Barbera (admin@sci-gaia.eu), Administrator

***OpenDOAR* ID:** 3479, *Last reviewed*: 2015-09-16, Suggest an update for this record

Link to this record: http://opendoar.org/id/3479/

Sci-GaIA OAR can be “cloned” wherever and whenever it is needed



The Africa Grid Science Gateway
(http://sgw.africa-grid.org)

CSGF Components

Users having different roles and privileges

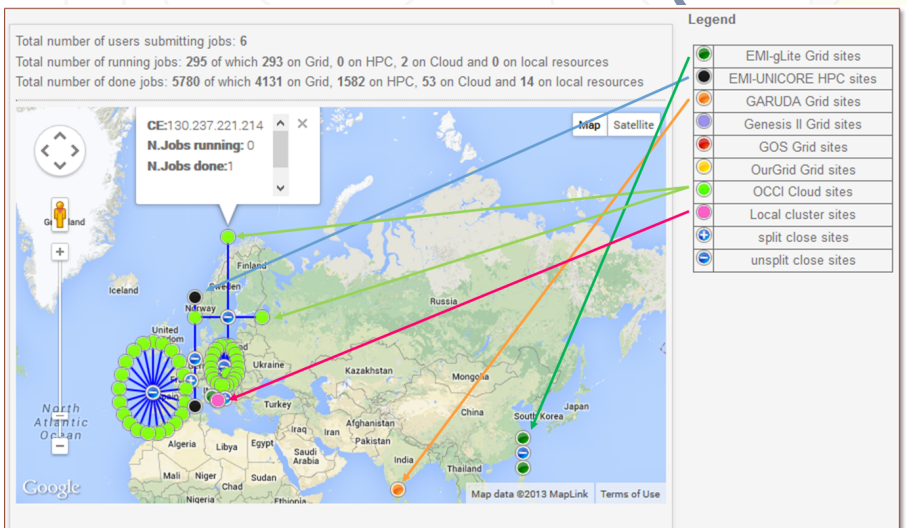
- Administrators
- Power users
- Basic users
- VRC members
- etc,

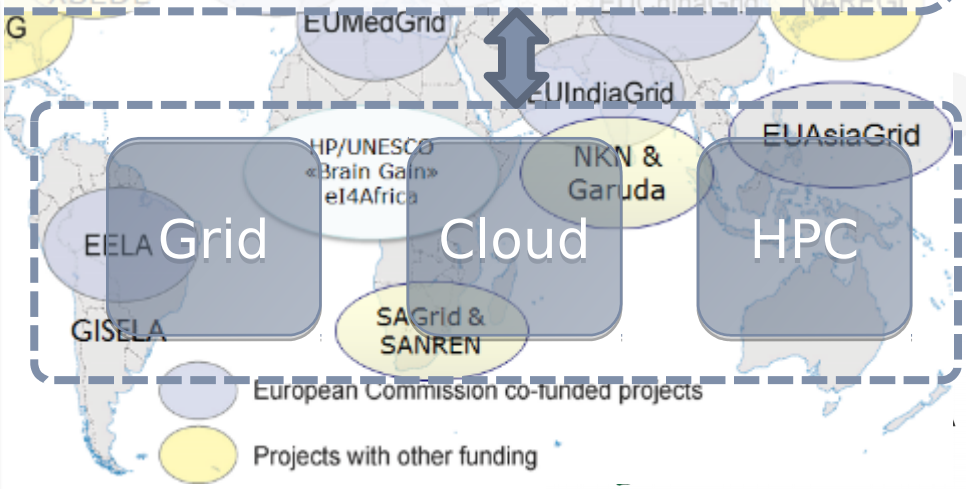
LIFERAY
Enterprise. Open Source. For Life.

AAI

Portlets

Grid&Cloud Engine





The Africa & Arabia Regional Operation Centre (ROC)

(www.africa-grid.org)

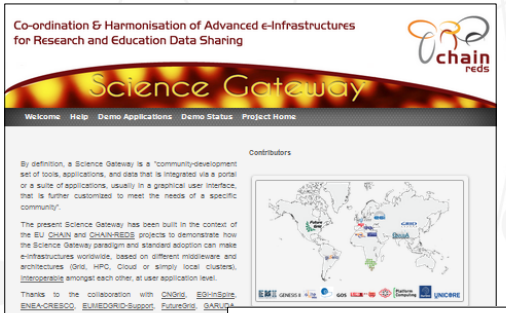
Services

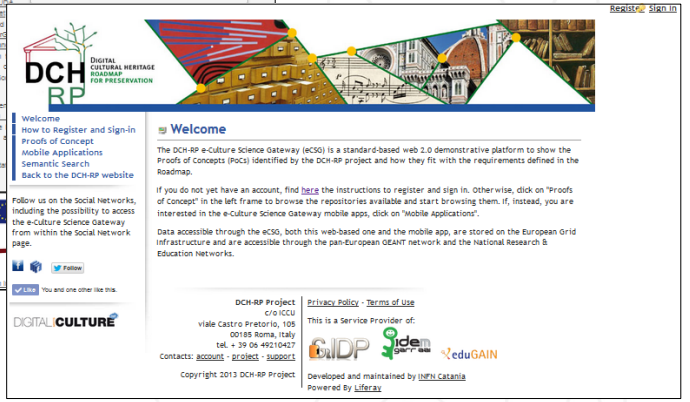
- Applications
- Infrastructure
- Security
- Compute & Data
- Science
- Identity
- Support
- Dev & CI

- Unique entry point to explore/monitor the African e-Infrastructure and/or to join it with HPC, Grid and Cloud sites
- Fully compliant with DevOps paradigm:
  - e-Infrastructure services (IdP, SG, OAR, etc.) and applications can be

Sci-GaIA
www.sci-gaia.eu

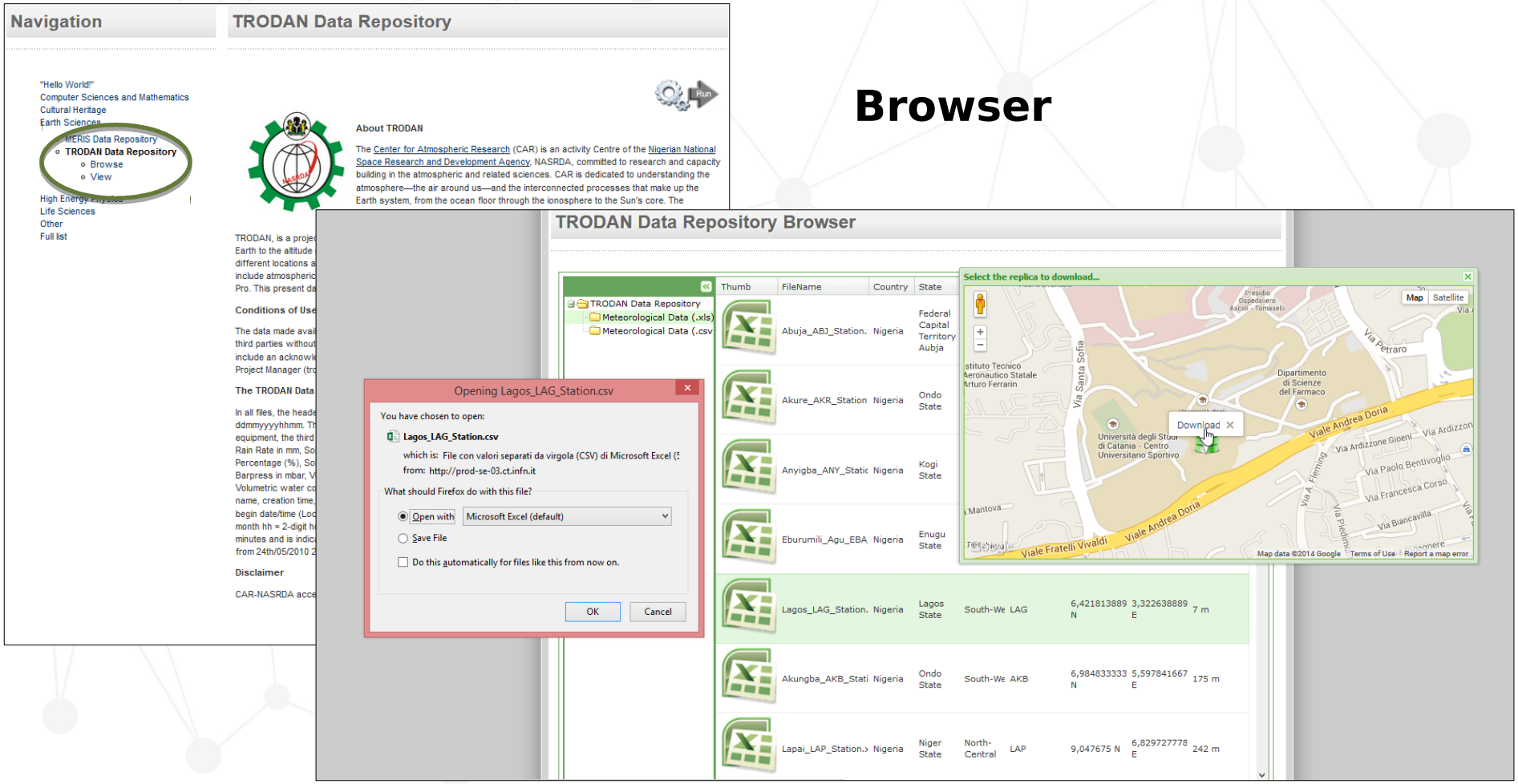
Some of the SGs implemented with the CSGF





www.sci-gaia.eu

Examples of applications available in the Africa Grid SG: The TRODAN data repository

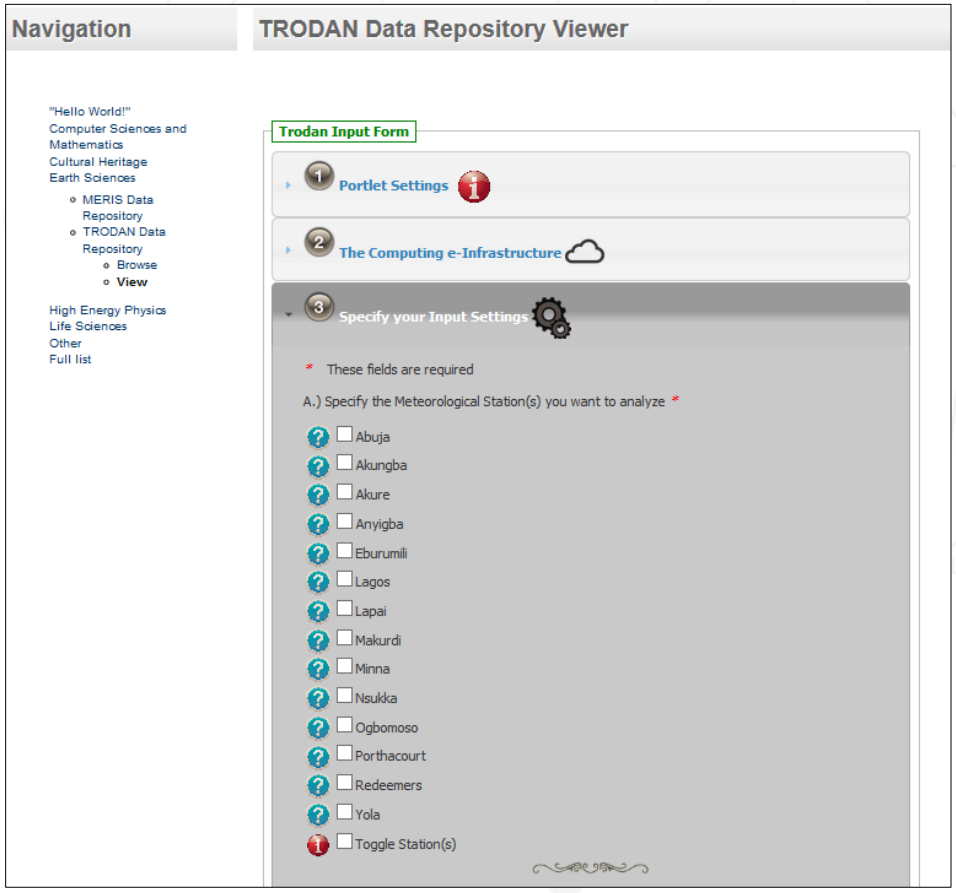


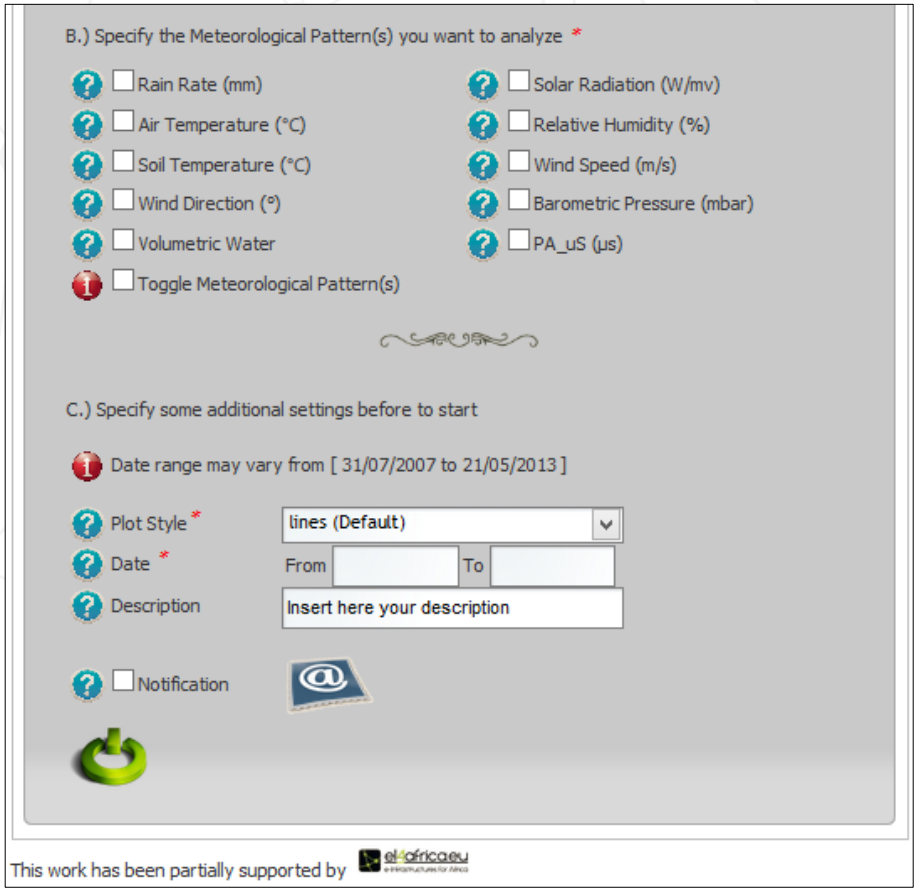
Browser

Sci-GaIA
www.sci-gaia.eu

Examples of applications available in the Africa Grid SG:
The TRODAN data repository

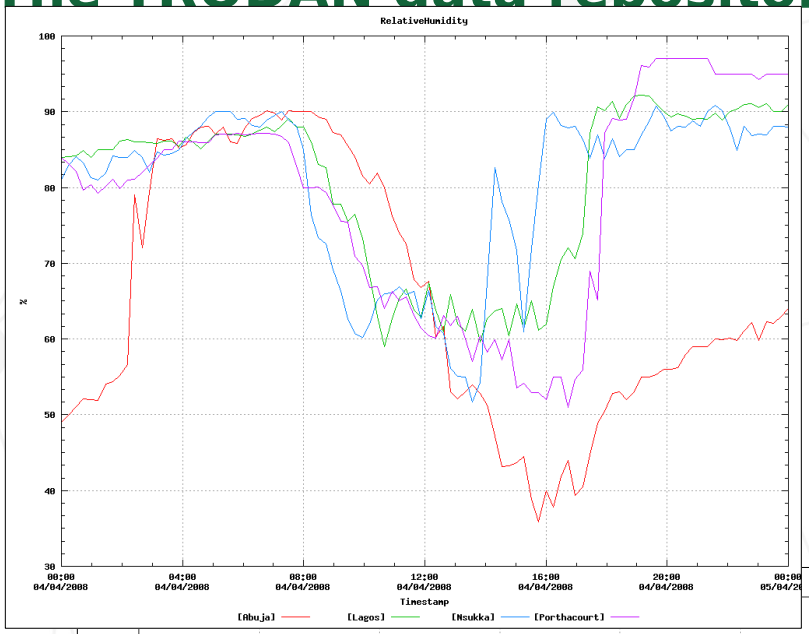
Viewer

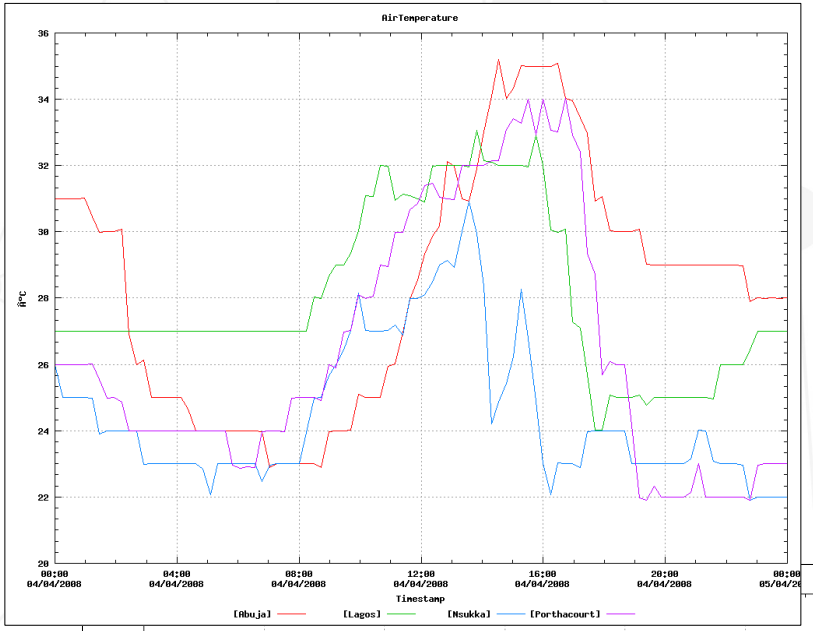


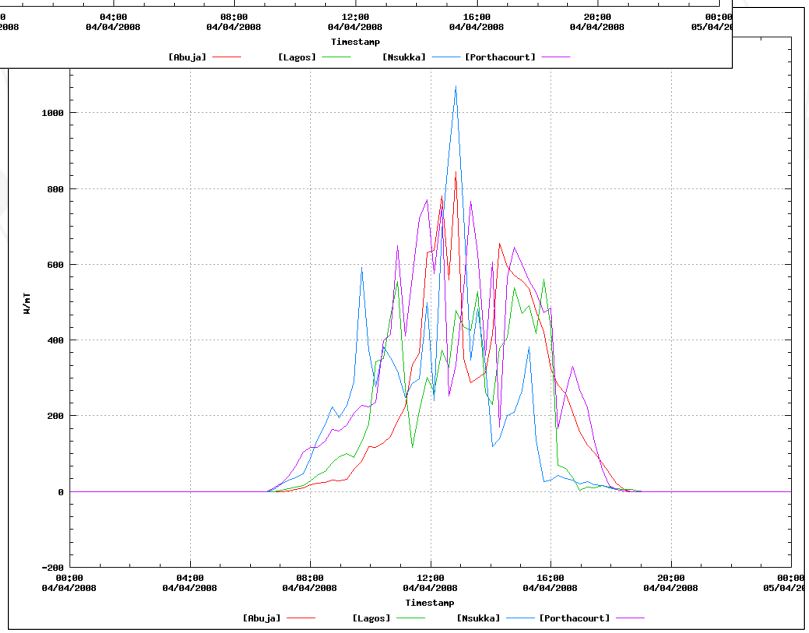


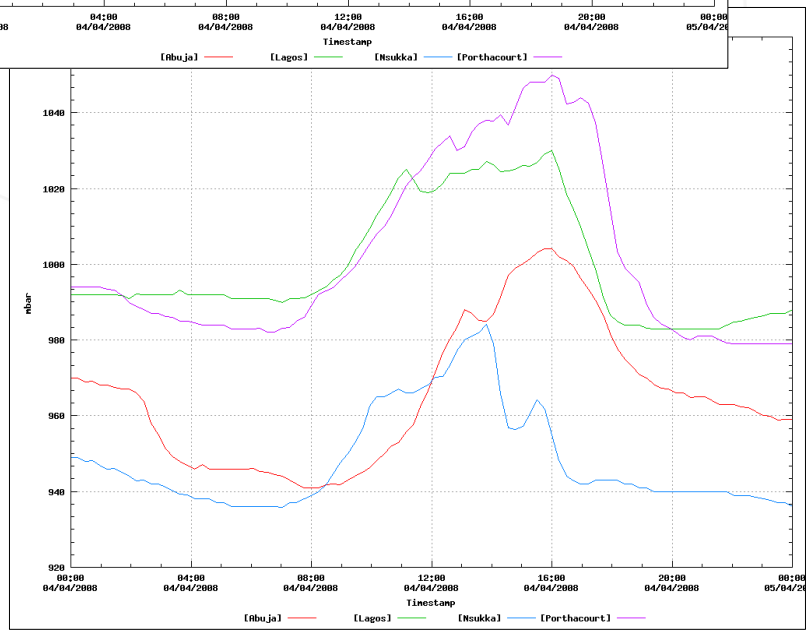
Sci-GaIA
www.sci-gaia.eu

Examples of applications available in the Africa Grid SG:
The TRODAN data repository

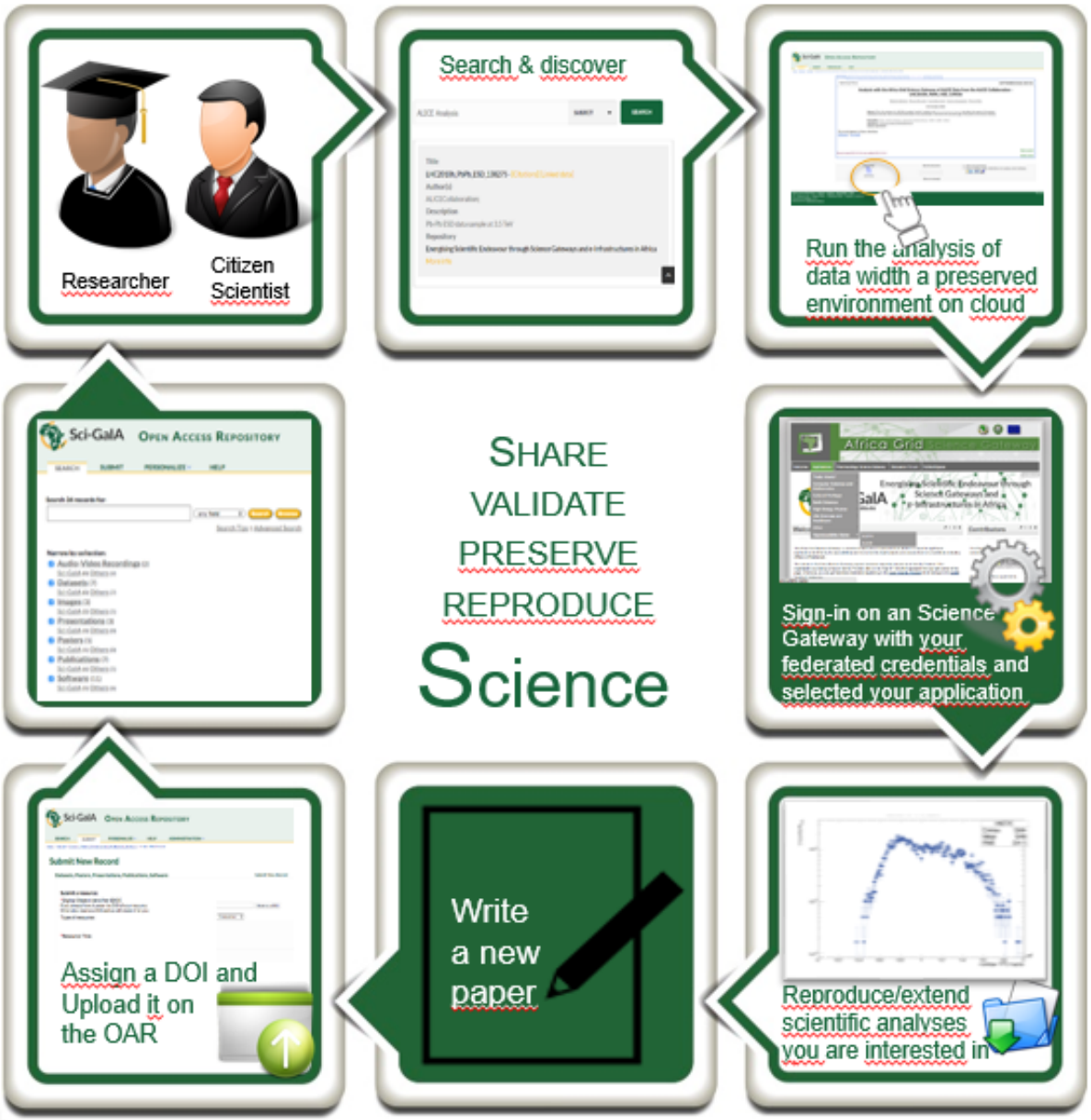






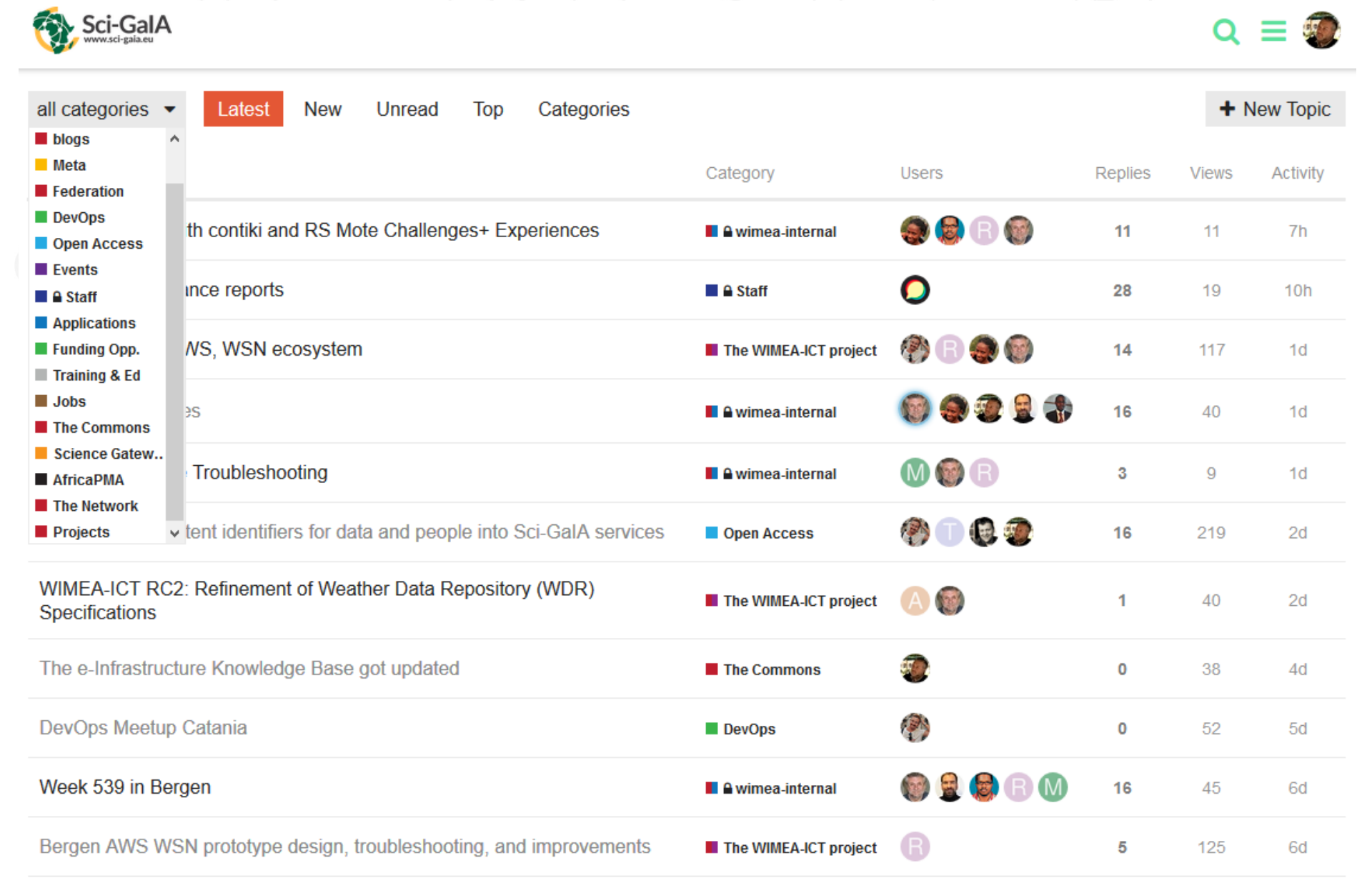


The Open Knowledge workflow: we do know how to implement it … and we are actually doing it!

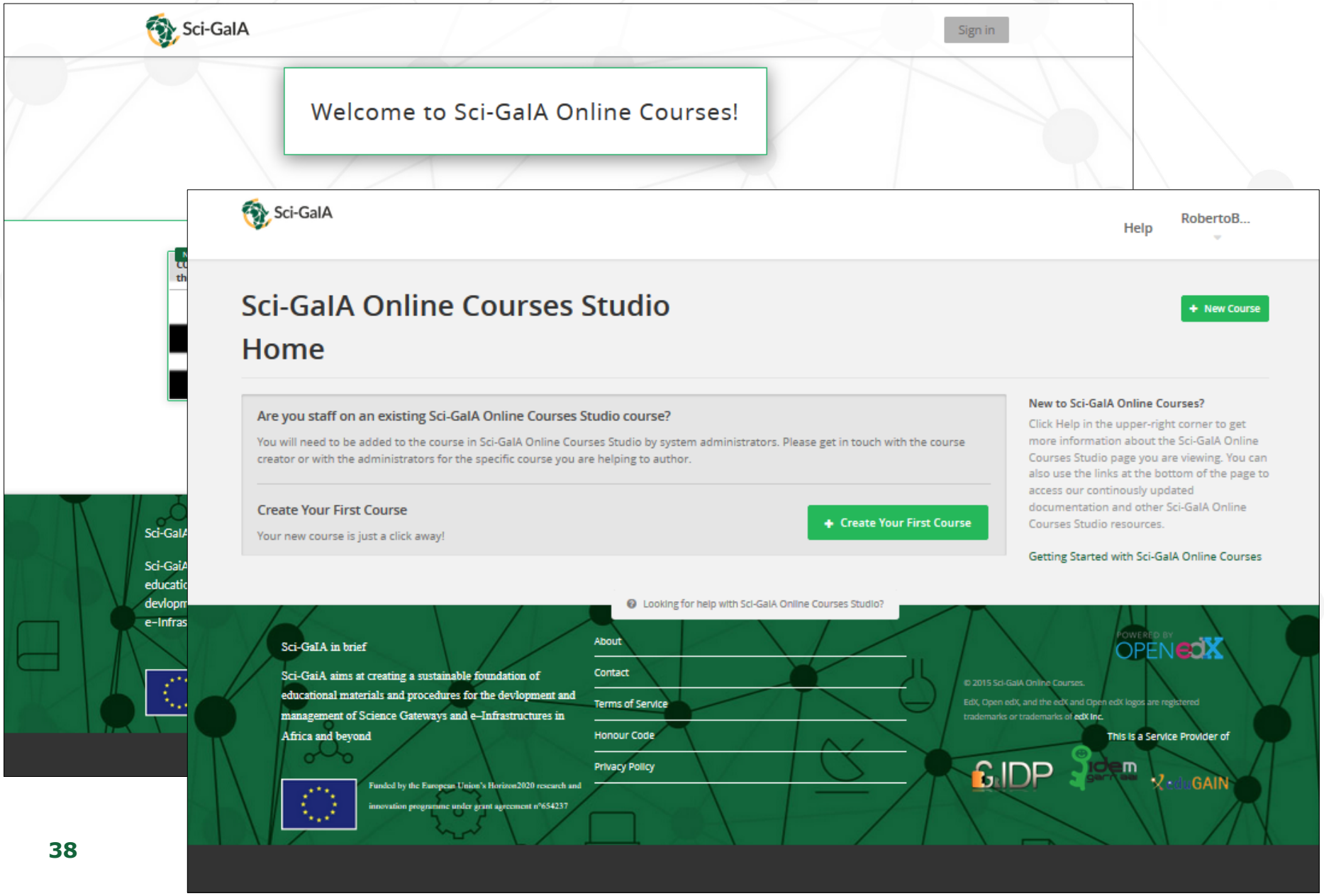


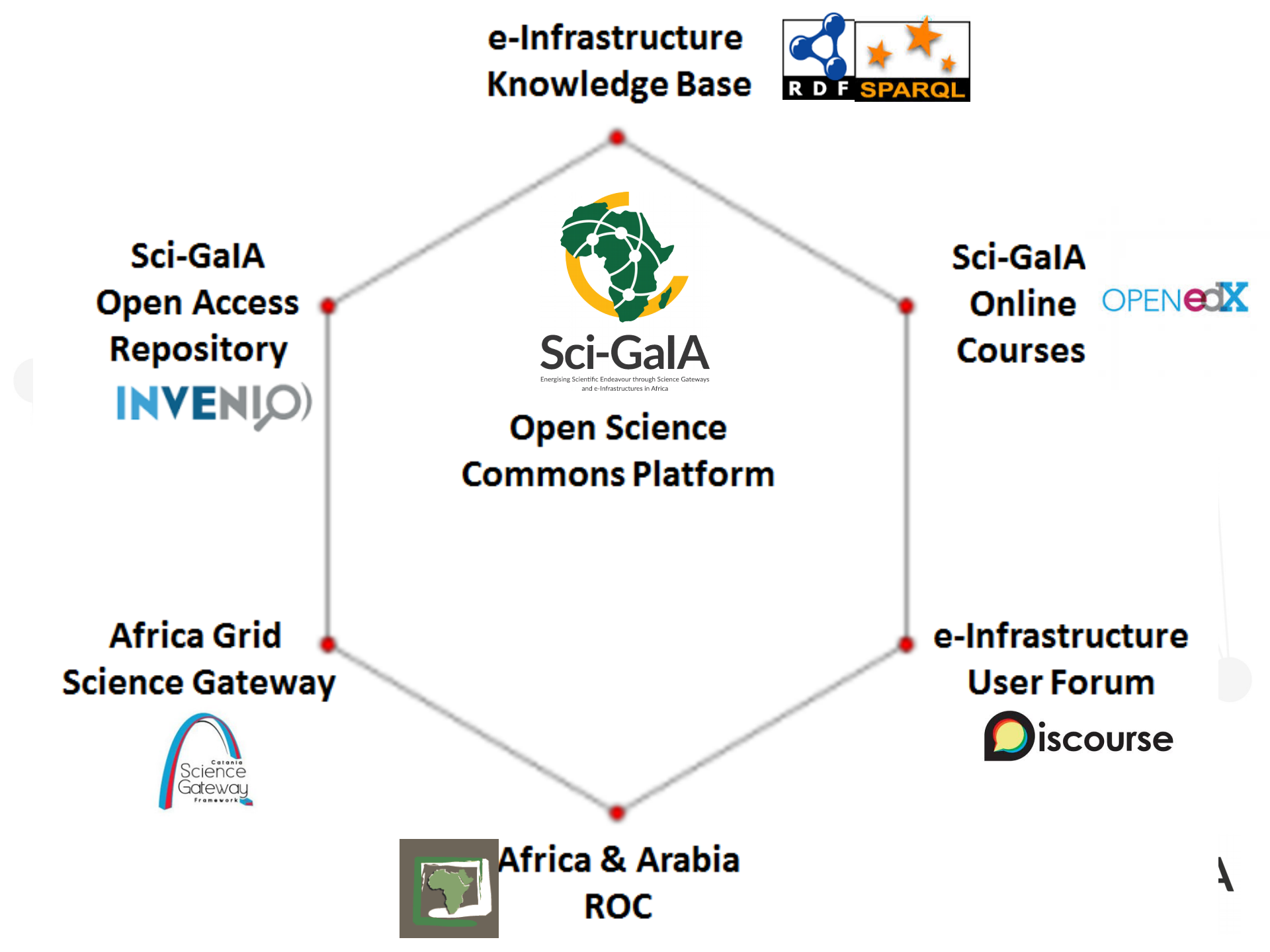
Sci-GaIA
www.sci-gaia.eu

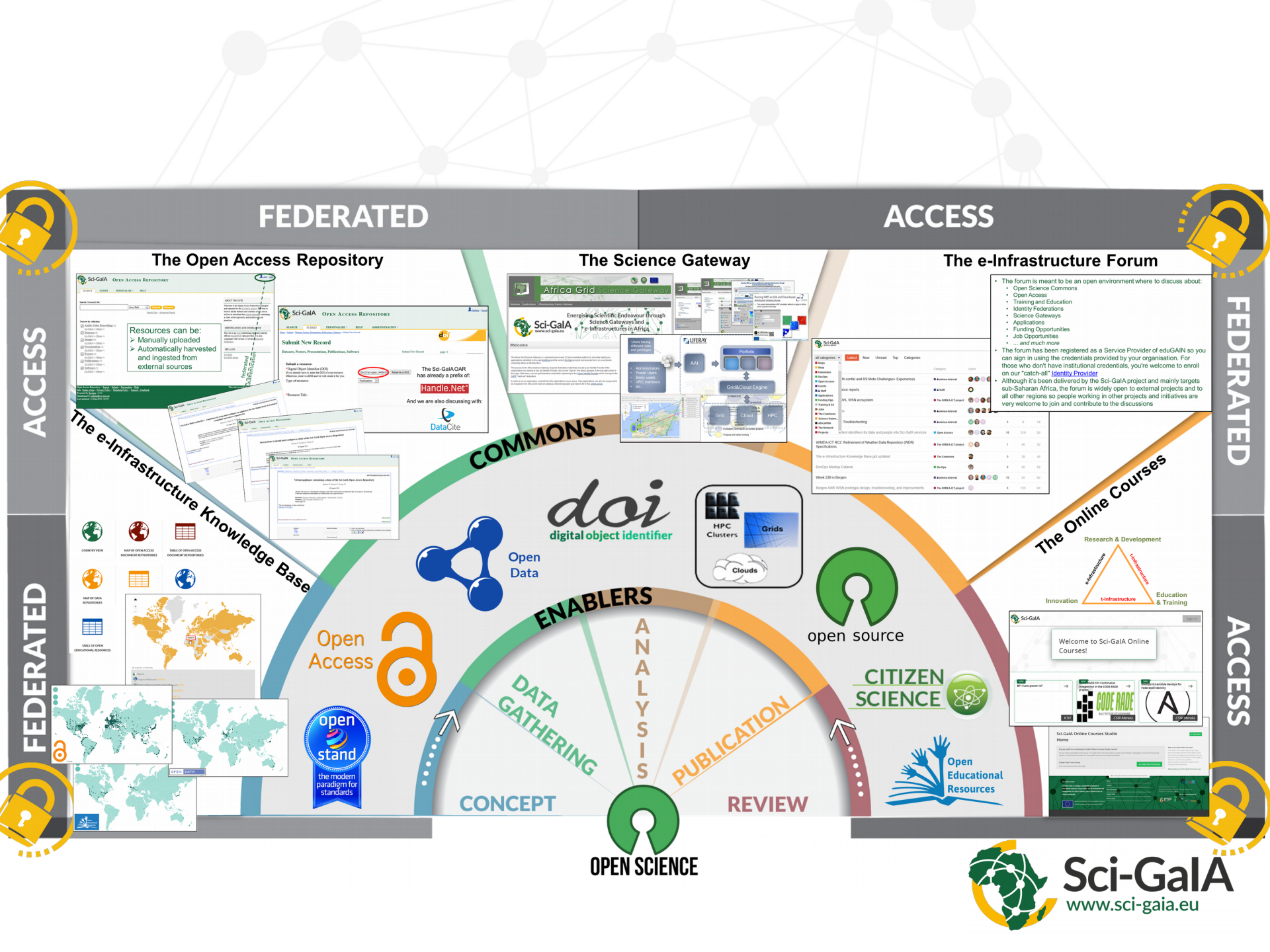
The e-Infrastructures discussion forum
(http://discourse.sci-gaia.eu)



Sci-GaIA Online Courses
(http://courses.sci-gaia.eu)




e-Infrastructure
Knowledge Base
RDF SPARQL
Sci-GaIA
Open Access
Repository
INVENIO
Sci-GaIA
Energising Scientific Endeavour through Science Gateways
and e-Infrastructures in Africa
Open Science
Commons Platform
Sci-GaIA
Online
Courses
OPENedX
Africa Grid
Science Gateway
Catania
Science
Gateway
Framework
e-Infrastructure
User Forum
Discourse
Africa & Arabia
ROC


FEDERATED
ACCESS
The Open Access Repository
The Science Gateway
The e-Infrastructure Forum
Resources can be:
Manually uploaded
Automatically harvested and ingested from external sources
Federated Authentication
The Sci-GaIA OAR has already a prefix of:
Handle.Net
And we are also discussing with:
DataCite
Africa Grid Science Gateway
Sci-GaIA
www.sci-gaia.eu
Energising Scientific Endeavour through Science Gateways and e-Infrastructures in Africa
Portals
AAI
Grid&Cloud Engine
Grid
Cloud
HPC
The forum is meant to be an open environment where to discuss about:
Open Science Commons
Open Access
Training and Education
Identity Federations
Science Gateways
Applications
Funding Opportunities
Job Opportunities
... and much more
The forum has been registered as a Service Provider of eduGAIN so you can sign in using the credentials provided by your organisation. For those who don't have institutional credentials, you're welcome to enroll on our "catch-all" Identity Provider
Although it's been delivered by the Sci-GaIA project and mainly targets sub-Saharan Africa, the forum is widely open to external projects and to all other regions so people working in other projects and initiatives are very welcome to join and contribute to the discussions
The e-Infrastructure Knowledge Base
COMMONS
doi
digital object identifier
Open Data
HPC Clusters
Grids
Clouds
The Online Courses
Research & Development
Innovation
Education & Training
e-Infrastructure
t-Infrastructure
Open Access
ENABLERS
DATA GATHERING
ANALYSIS
PUBLICATION
REVIEW
CONCEPT
open source
CITIZEN SCIENCE
Open Educational Resources
Welcome to Sci-GaIA Online Courses!
open stand
the modern paradigm for standards
OPEN SCIENCE
Sci-GaIA
www.sci-gaia.eu

Summary and conclusions

- Reproducibility, re-usability and extensibility, are key to walk through the “knowledge path” in both directions
- The Sci-GaIA project is committed to the uptake of the Open Science paradigm and is building a federated platform for an Open Science Commons
- Research is made more robust, cost-effective and impactful when done in an Open Science paradigm.
- Researchers are better evaluated for **all** of their contributions
- Access to e-Infrastructures and the adoption of services within the e-Science Commons are crucial to improving the **visibility and recognition of African Research**


What are you waiting for ?
www.africa-grid.org
www.sci-gaia.eu
discourse.sci-gaia.eu
github.com/AAROC
@ei4africa
bbecker@csir.co.za