Sci-GaIA contributions to Open Science in Africa

Bruce Becker, C.S.I.R. Meraka Institute, Cyberinfrastructure | bbecker@csir.co.za
orcid.org/0000-0002-6607-7145
Barend Taute, C.S.I.R. Meraka Institute, Manager: ICT Contract R&D | btaute@csir.co.za
DST Open Science Workshop, Pretoria, 10 May 2017

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement nº 654237
Outline

- Energising Scientific Endeavour – the road to Sci-GaIA
- The Sci-GaIA Open Science Platform
- Opening Science – the Sci-GaIA online courses, hackfests and bootcamps
- Opening Scientists – the Sci-GaIA champions
- What now?
The Sci-GaIA project

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

www.sci-gaia.eu
The main aim of the Sci-GaIA Coordination and Support Action is to create a sustainable foundation of educational material and procedures for the development and management of **Science Gateways** and **e-Infrastructures** to Energise Open Science in Africa.

**WP1**
Promote the uptake of Science Gateways and e-Infrastructures in Africa and Beyond

**WP2**
Support new and already emerging user communities

**WP3**
Strengthen and expand Science Gateway and e-Infrastructure related services

**WP4**
Train and disseminate for long sustainability purposes
Sci-GaIA builds on the results of many e-Infrastructure flagship projects

- Exploiting Research Infrastructures potential for Boosting Research and Innovation in Africa (ERINA4Africa) 2009 – 2011
- Teaming-up for exploiting e-Infrastructures’ potential to boost RTDI in Africa (eI4Africa) 2012 – 2014
- Coordination and Harmonisation of Advanced e-Infrastructures (CHAIN) 2010-2012
- Coordination and Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing (CHAIN-REDS) 2012-2015
- Promoting African European Research Infrastructure Partnerships (PAERIP) 2011-2013
- Supporting the development of cooperation on e-Infrastructures with Africa (el-Africa) 2009-2011
- Exchange Programme for e-Infrastructure Know-How (EPIKH) 2009-2013
Sci-GaIA and the reproducibility crisis
Showed how to interpret the performance of part of the ALICE experiment at CERN, for a few important signals. Done before ALICE produced any data (2007). Wouldn’t it be nice to check, now that the data has been taken?

This is still possible now that the ALICE experiment is active. However, the absence of semantic links between the thesis document, the code used, the data generated, the reference datasets and the execution environment make it a very delicate task.

Much better these days: opendata.cern.ch/about/ALICE
Open Science Reproducibility Challenge
“walk” across the knowledge path both ways

HTC/HPC Clusters, Grids, Clouds

Data Infrastructures
Open Access Doc. Repos.
Data Repos.

Evaluate Existing Data
Create New Data: Measure or Simulate.

Exploit Existing Data
Archive Data

Draw Conclusions

Valid Hypothesis
Hypothesis needs work.

Report Methods & Results
Annotate Archived Data

doi: 10.15169/sci-gaia:1473482338.53

www.sci-gaia.eu
No problem... you just need to do everything
Open Science requires Open Infrastructure

The practice of Open Science is done with relevant tools – these need to inter-operate with

- **other tools**
  - scientific applications, web applications, community meeting points, teaching and learning tools, etc

- **components of e-Infrastructure**
  - computational platforms, data platforms, software repositories, application delivery systems

- **components of the commons**
  - Identity Federations, Open Access repositories, persistence and uniqueness services (DataCite, ORCID), Linked Open Data services, Open Educational material

Integration is counter-productive. Inter-operation encourages collaboration
A platform enables services. Sci-GaIA has focused on bringing real services into researchers’ lives, and making them easy to access, use and extend.

- Brought 6 new services into 9 Identity Federations (including edugain)

- E.g.:
  - Take an online course (implemented with OpenEdX) – [courses.sci-gaia.eu](http://courses.sci-gaia.eu)
  - Run data analysis - [sgw.africa-grid.org](http://sgw.africa-grid.org)
  - Publish results – [oar.sci-gaia.eu](http://oar.sci-gaia.eu)

- All with your institutional credentials!
Sci-GaIA outcomes – WP1 support

• Create educational materials
  – Documentation and automation, OERs
    www.sci-gaia.eu/materials

• Survey e-Infrastructure usage
  surveys.sci-gaia.eu

• Maintain compliance with other international efforts through the Africa & Arabia Regional Operations Centre (AAROC) and EGI Cloud
  www.africa-grid.org

Basic infrastructure for e-Science
Sci-GaIA outcomes – WP2: Community

- A community needs a place to live, engage, and discover themselves.
- A community is a place to grow and share ideas, not just announcements.
- Sci-GaIA created the African e-Infrastructure discussion forum discourse.sci-gaia.eu
- Read more in the deliverable under preparation git.io/SciGaIACommunity
Sci-GaIA courses, hackfests and bootcamps

A new scientific paradigm requires a new educational paradigm – rapid development of courses in relevant technologies and methods. Sci-GaIA developed three different kinds of training formats:

- Online courses – courses.sci-gaia.eu
  Online learning platform, populated with Open Educational Resources, self-paced and traditional courses to research infrastructure developers
- Powered by OpenEdX
- Access via Identity Federation
- Reproducible Configuration: github.com/sci-gaia/configuration
Bootcamps for Infrastructure Engineers
doi: 10.5281/zenodo.242394

- Sci-GaIA developed 2-day bootcamps for the people who are helping to build African e-Infrastructures
- Instructor-led, project driven, intensive format. Modelled on Carpentry
- Covers technology needed to develop reproducible expressions of e-Infrastructure
- 4 events run
  - Catania (With GARR)
  - Lagos (With eko-Konnect and WACREN)
  - Entebbe (With UbuntuNet)
  - Pretoria (With NICIS)
Sci-GaIA Hackfests
www.sci-gaia.eu/summer-hackfest

- Intensive, 2-week event aimed at delivering Science Gateways
- Partnered with Indigo DataCloud and EGI - modern full-stack solutions for e-Research platforms
- Ran 3 hackfests:
  - Catania, Lagos, Addis Abeba

Published everything necessary to reproduce the event
Sci-GaIA Champions
www.sci-gaia.eu/champions

Sci-GaIA’s direct impact on scientists

- 35 Researchers from 6 African countries, each built a new science gateway or service in an Open Science platform.

- Each put front-and-centre e.g. www.sci-gaia.eu/champions/dennis-muoki

- From science gateways for citizen-science, crowd-sourced open data, medical imaging, smart-grid technologies, speech recognition, bioinformatics and genomics, open learning and open access data repositories – they built them.

- This community was *not* funded by Sci-GaIA – will live on
Let them speak for themselves
youtu.be/x9_L6Fz02h0

Contact them at discourse.sci-gaia.eu/groups/champions
What’s next

Sci-GaIA held its final event in Pretoria in March
www.sci-gaia.eu/final-event

The project will be reviewed in June – what happens to the products?

- Services → Africa-Arabia Regional Operations Centre
- Community → Self-sustained. Peering with Carpentry movements and Mozilla Science Lab
- Technology → Openly Licensed, maintained by the respective developers (CSIR, University of Catania)
- Infrastructure → ? Complex discussion...
Open Science requires Open Infrastructure

- Sci-GaIA has worked to promote the Commons and re-use of infrastructure services to the research communities and champions.

- We have also worked to promote Open Infrastructure and highlight the pitfalls of working in isolation.
  - Co-operation and co-development with components of NICI, the African Open Science Platform, African Research Cloud, etc. are crucial to ensuring a viable Open Science ecosystem for Africa.
  - Co-ordination with the e.g. European Open Science Cloud and other digital infrastructure providers (DataCite, EGI, EUDAT, OpenAIRE, GEANT) ongoing.
But wait... there’s more!

We’ve only scratched the surface

- discourse.sci-gaia.eu
- github.com/AAROC github.com/sci-gaia
- @ei4africa #scigaia #tandem

Sci-GaIA

oar.sci-gaia.eu