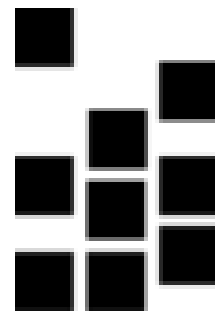


Continuous Integration and Delivery of Research Applications in a Distributed Environment

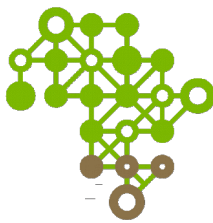


CODE RADE
BUILD TEST EXECUTE EVERYWHERE

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

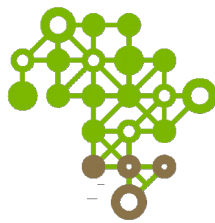
Sean Murray
murrays@cern.ch
Senior Sys Admin/Researcher,
CHPC/ALICE@CERN





Outline

- Background :
 - Physics ♥ computing
 - The reality of computational infrastructure: It's a jungle out there
 - The problem CODE-RADE solves
 - When researchers talk to research software engineers (Note to sensitive viewers: it gets ugly.)
 - 7 Hypotheses of scientific computing
- CODE-RADE design, actors, workflow
- Enough with the talking – take me to the science !

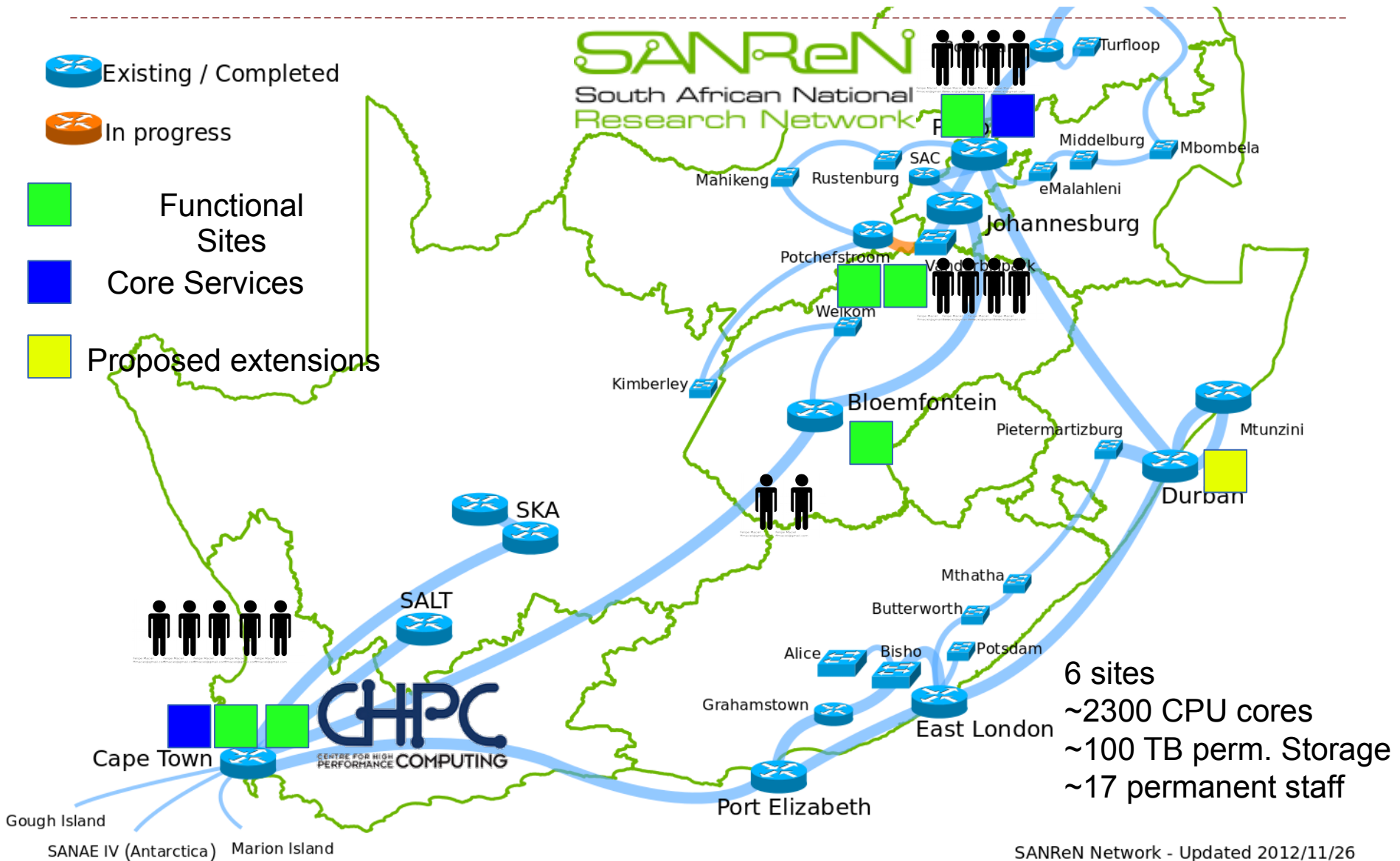


Physics ♥ computing

- The field of physics has had a huge impact on computing and vice versa
- A huge library of physics applications has been developed by many
 - Just HEP : e.g. <https://www.hepforge.org/projects>
- Many different *kinds* of applications :
 - Data analysis, Monte-Carlo
- Physics applications usually make good use of massively-distributed infrastructures.
 - Basically we built grid computing ***because*** of physics !



So, we built this...



SANReN Network - Updated 2012/11/26



ZA-UCT-ICTS



ZA-CHPC



ZA-UFS

www.africa-grid.org/sites



ZA-UJ



ZA-WITS-CORE



ZA-MERAKA



MA-01-CNRST



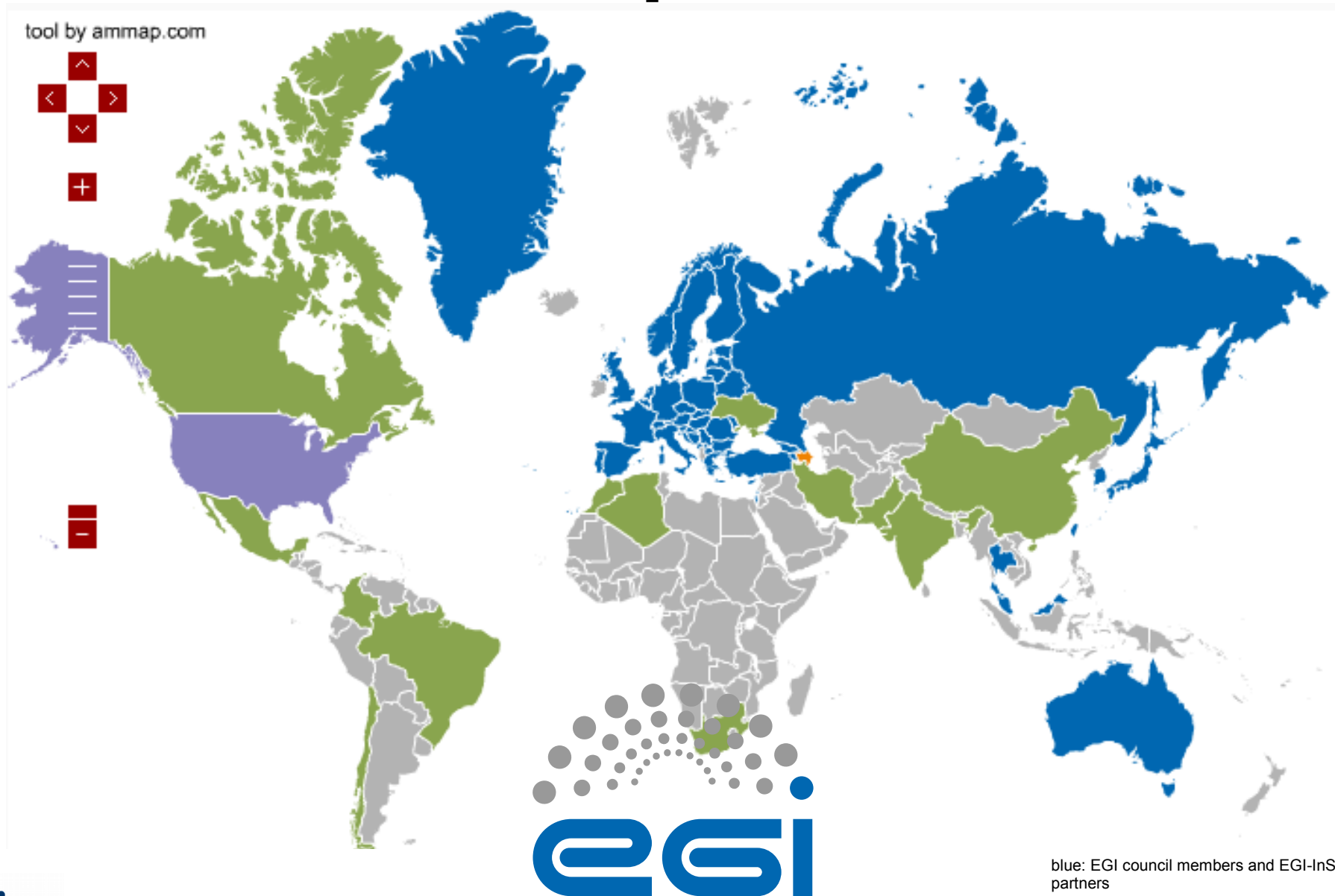
EG-ZC-T3



DZ-ARN-01



And it is part of this





Great, right ?!

**ONE DOES NOT SIMPLY RUN
APPLICATIONS**

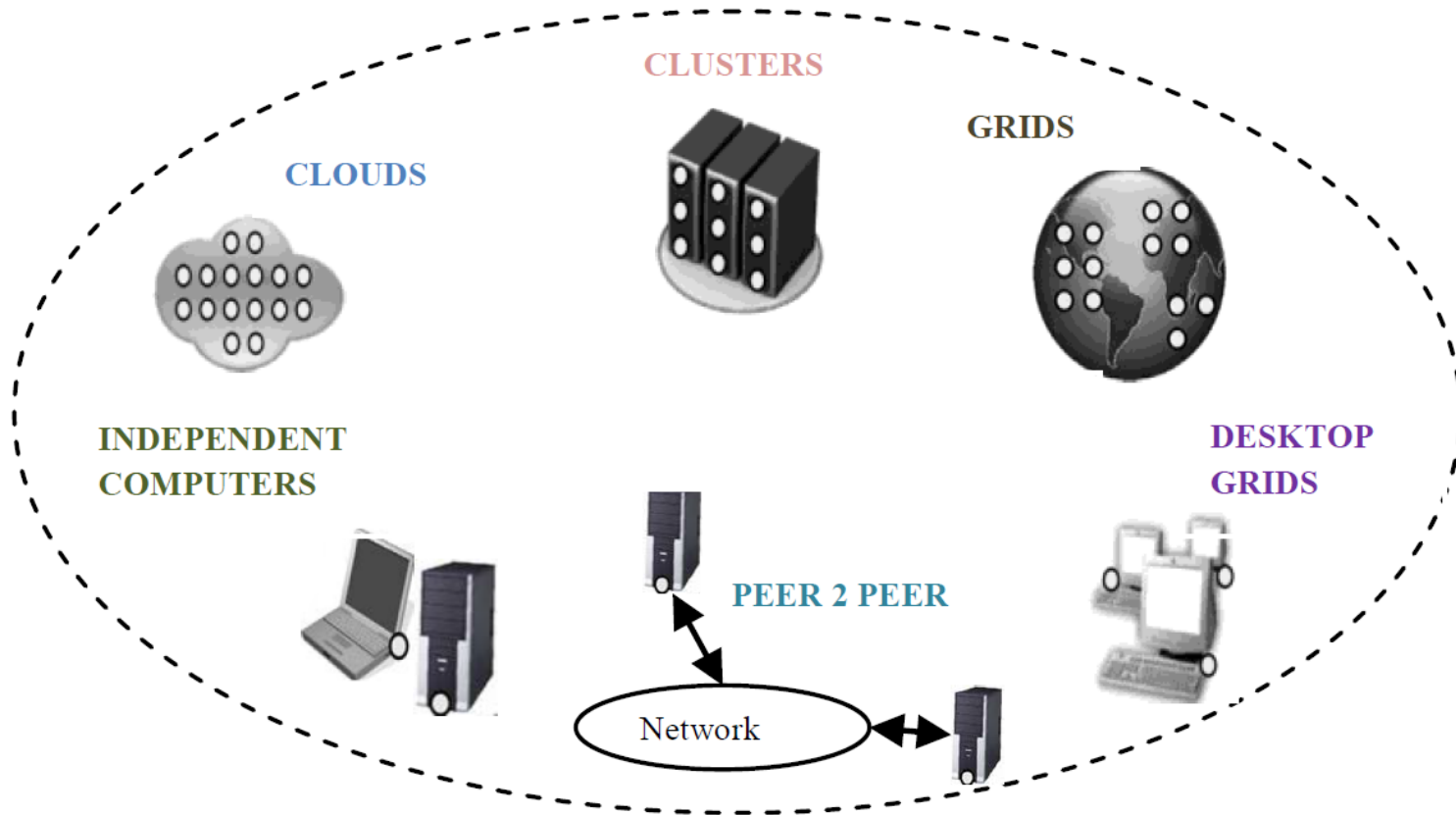


ON THE GRIDCLOUD



It's (apparently) a jungle out there.

http://en.wikipedia.org/wiki/Jungle_computing



Jungle Computing: Distributed Supercomputing beyond Clusters, Grids, and Clouds

Seinstra, F. J., Maassen, J., van Nieuwpoort, R.-V., et al. 2011, Grids, Clouds and Virtualization, 167



Where does your code run ?

- On your laptop ?
 - Congrats ! Uh oh, you just ran out of X
 - You want to do what !?
- Local cluster ?
 - hm. different setup... <hack hack>
hey it's working ! ... kinda.
 - Permissions, environment,
data staging, etc.
- CHPC ?
 - There's more than one
 - ALL the Optimisations !!



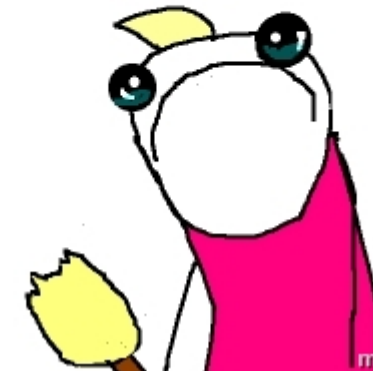


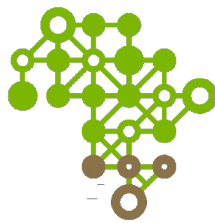
Are you even talking about the same application ?

- Apart from performance, are you sure that you are getting the same results on different architectures and platforms ?
- Is it even worth the effort ?



ONLY RUNS ON QUANTUM
COMPUTERS?





The problem CODE-RADE solves

The researcher is faced with a paradox :



There are a huge amount of computing resources available, and a rich, well-maintained, interoperable infrastructure of services...

But

Getting applications onto it requires too much effort.

When Researchers talk to Research Software Engineers



Hi, I heard about AfricaGrid – sounds grrrrreat ! Can I Use it ?

I've got beakers ...
And glasses !



Are you a legit scientist ?



Ok, sounds good
what applications
are you running ?
We can run
almost anything !



Awesome ! Well, first we do some preprocessing with RaNDo, then we run a massively parallel pipeline with XBOT, then blah and blah then blah blah blah and blah, we also need blah blah and sometimes our colleagues from FarAwayU need to access the data with StrangeProtocolx1, and of course blah blah, not to mention blah



.....

We'll see
what we
can do





Actual workflow

- Find the source code
 - Unpack, ./configure && make ... Error 1 missing x.h
 - ... hack hack hack ... hm. Missing dependencies
- Recursively* determine the dependencies**
 - *Yes, recursively means curse, curse then curse again.
 - **The *actual* dependencies, not the ones that were written on the web page from 5 years ago.
- ... mutter mutter ...
- ????? Why are you compiling **now** !?





Actual "communication"

Hey, how's it going ?
you done with my porting yet ?

You done yet ?

You *done* ?

Have you actually *seen* this code ?

I just...
It looks like Cthulhu's beard....

Ok, I got it to compile,
but only with if you
type with your left toes.

.... bloody IT people !

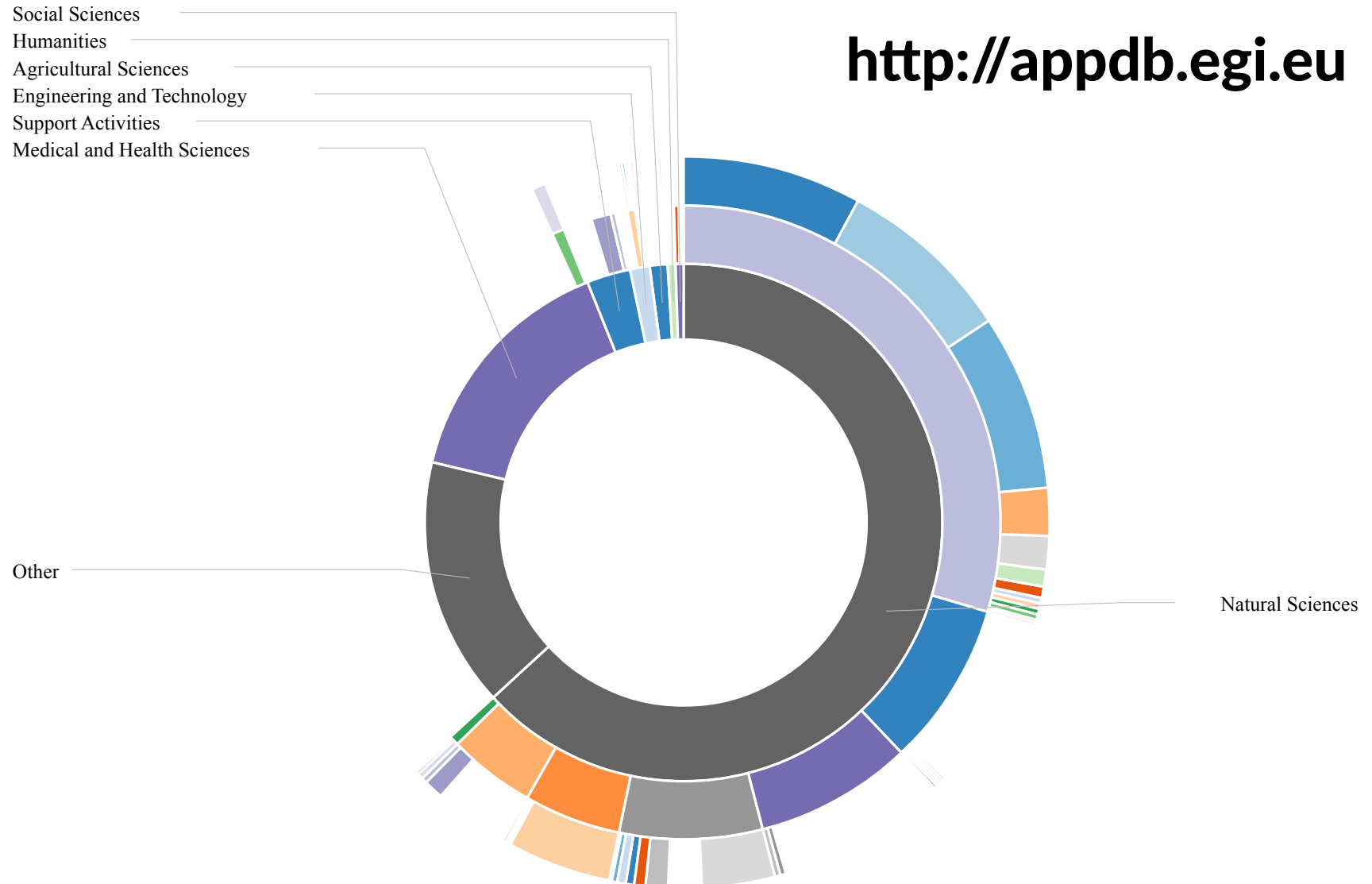
AAArrrgh, I'm a *scientist*,
not an IT people !

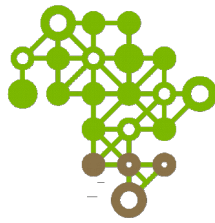




How many applications are there ?

<http://appdb.egi.eu>



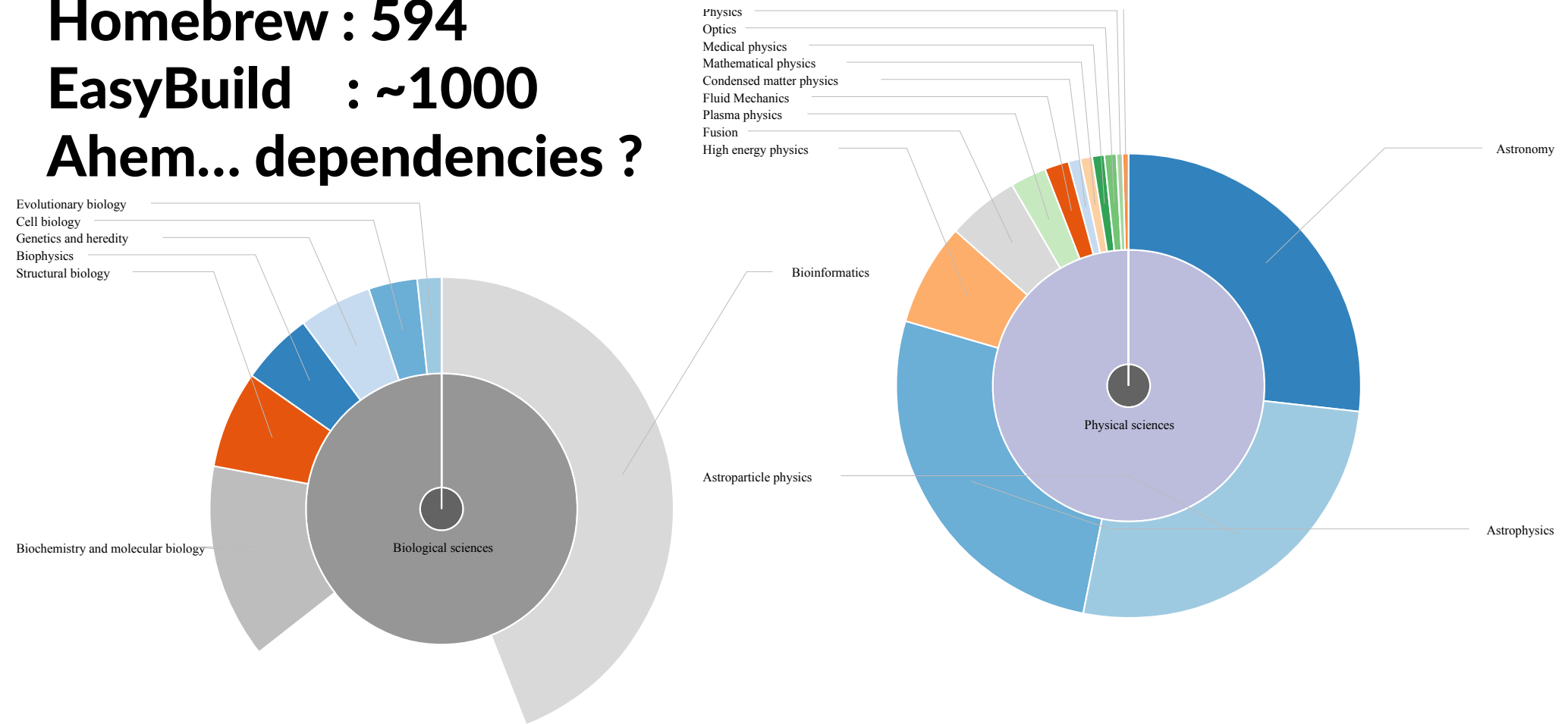


Way too many !

Homebrew : 594

EasyBuild : ~1000

Ahem... dependencies ?

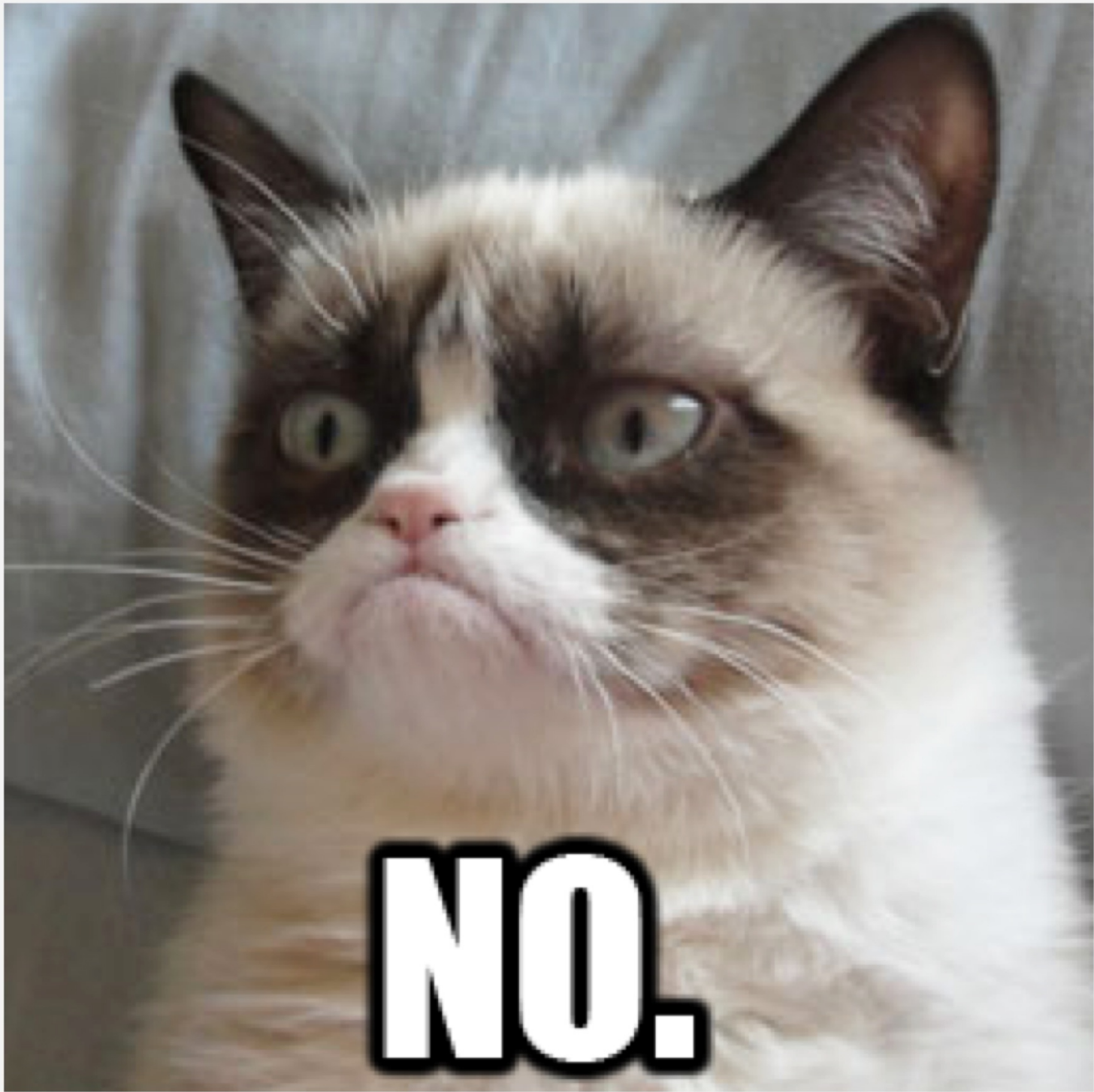


Can we do this for every application ?

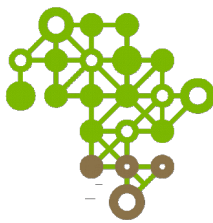


H
you do

e ?

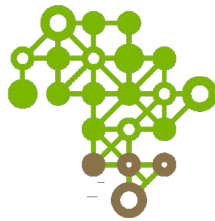


CODE-RADE Design Philosophy

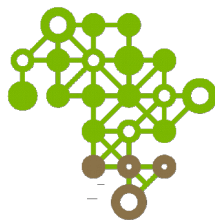


- Hypothesis 1: It always comes down to an application
 - At the end of all first-contact you get a list of applications that users want
 - *Let's **focus** on that, shall we ?*
- Hypothesis 2: No software is an island
 - Everything has dependencies, even if it's just **glibc** or **libgfortran**
 - *Let's make those **explicit**, shall we ?*

CODE-RADE Design Philosophy



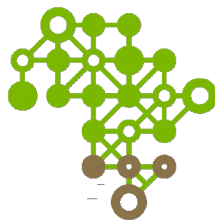
- Hypothesis 3: Applications need an environment
 - Compiler ?
 - Architecture ?
 - Dependencies ?
 - Middleware ?
 - Etc....
 - *Let's make those **explicit** shall we ?*



CODE-RADE Design Philosophy

- Hypothesis 4: there is more than one environment.
 - Nothing is going to ever be really useful if it can only run in one place.
 - Nobody wants to be forced into a single environment (grid/HPC/cloud, etc)
 - *Let's try to **simulate** as many target environments (sites) as we can, shall we?*

CODE-RADE Design Philosophy



- Hypothesis 5: Solutions decay
 - The solution – as well as the dead ends – are part of the process.
 - If it's not in a version-controlled repository, it's no good.
 - If it's not executable (in a wiki/document, etc), then you have to convert it back into something a computer understands – why not just leave it in executable format ?
 - *Let's **recognise** that solutions **decay** and code for the future, shall we ?*



CODE-RADE Design Philosophy

- Hypothesis 6: Humans need not apply
 - Most of the work of porting and supporting applications can be automated
 - Reduce the load on the user and the operator, by automating common tasks
 - **Integrate** the various software, middleware, service and infrastructure components as far as possible,
 - Keep it **Open**



CODE-RADE Design Philosophy

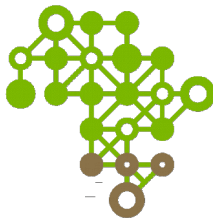
- Hypothesis 7: This stuff's not hard
 - Continuous Integration has been around for a long time
 - Continuous Delivery is widely-used practice (2012)
 - If it can be done by hand, it can be scripted; if it can be scripted, it can be automated properly.
- *adopt widely-used **tools** and **methodologies**, ensure sustainability*



The problem CODE-RADE solves

- *Lowers the barrier to entry* to the grid or cloud infrastructure, or single HPC sites
- The application expert can **prove** to the resource provider that *the application will actually run* on the execution environment of the site
- Easily manages the **lifecycle** of applications across multiple versions, architectures, configurations
- Ensure that once applications are certified, they are **actually available** on as many sites as possible ?
- Allows better collaboration between researcher, research software engineer and infrastructure provider.

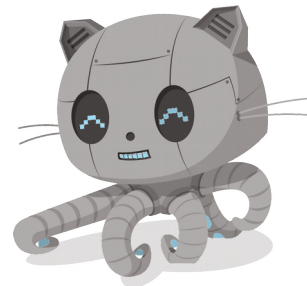
CODE-RADE actors



Researcher



Research
Software
Engineer

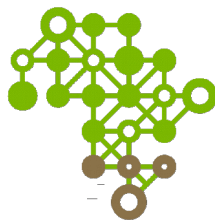


Automated
Agents



Infrastructure
Operations

CODE-RADE is an Open Platform



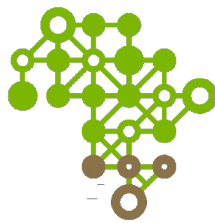
- **CODE-RADE components :**

- Version Control (Github)
- Continuous Integration (Jenkins)
- Automated Delivery (CVMFS)



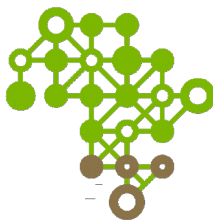
- **CODE-RADE workflow**

- Build : User/expert-defined means to produce executable applications or libraries
- Test : Infrastructure operator-defined targets. means to ensure viability
- Deliver : Ensure that once tests pass, the application is available

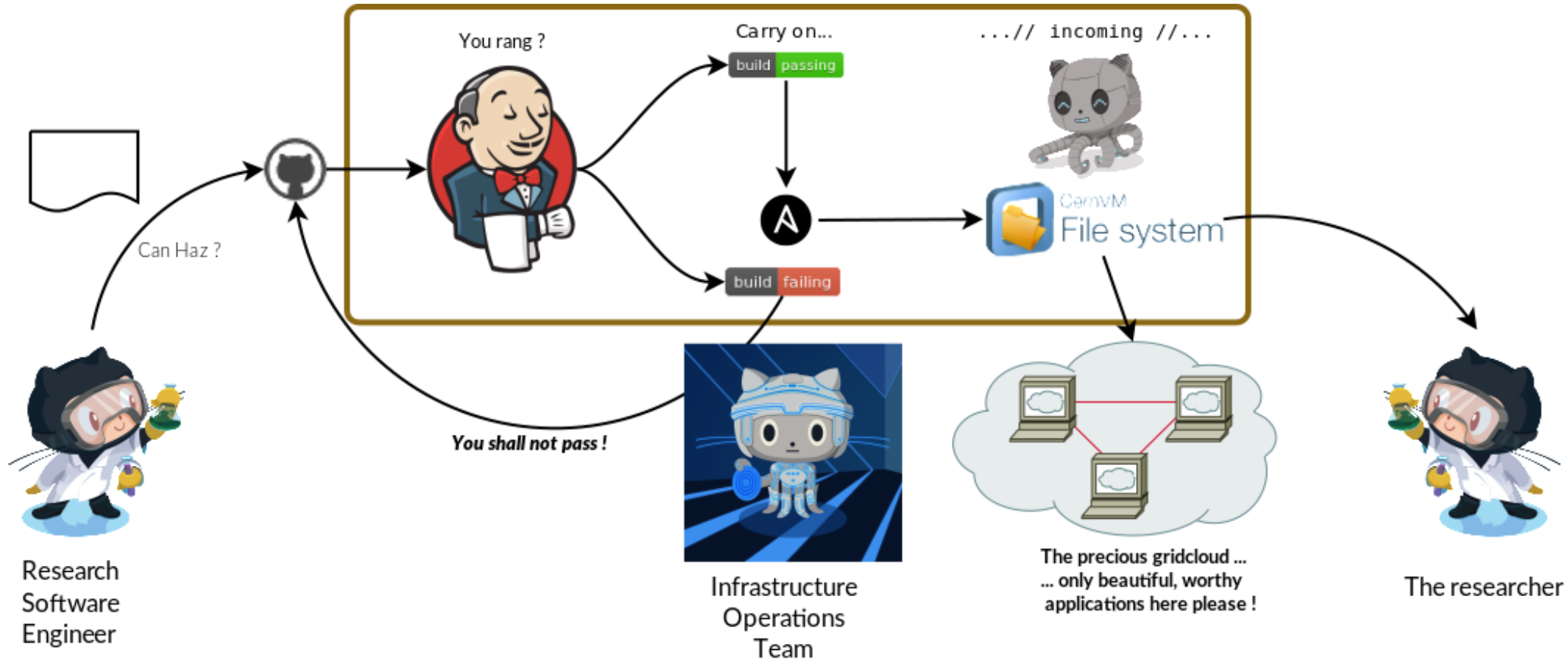


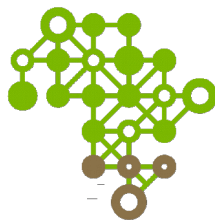
CODE-RADE is :

- **Cross-platform :**
 - Build and test artifacts for an arbitrary set of targets.
 - Promote diversity in computing platforms.
 - Ensure proper optimisations and application portability
- **Atomic :**
 - Fine-grained control over dependencies, versions and targets
 - Relevant action taken on each event
- **Community :**
 - No restriction on the applications that can be integrated.
 - Anyone can contribute applications, resources, code review, etc
- **Automated**
 - Heavy reliance on automated agents to reduce bias, lead time
 - User-driven





CODE-RADE works



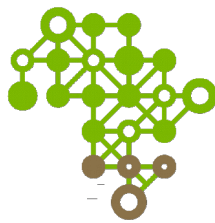


SHOW ME THE SCIENCE

- For the impatient :
 -  github.com/AAROC/CODE-RADE
 -  www.africa-grid.org/CODE-RADE
www.africa-grid.org/cvmfs
www.africa-grid.org/applications
- For the curious :
 - <http://ci.sagrid.ac.za/>
- You can try it on your laptop and run the applications everywhere ...
e.g. on the grid.
<https://github.com/AAROC/CODE-RADE/tree/master/grid>

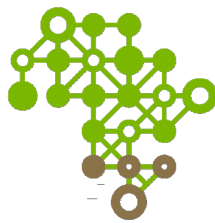
Example : SVM

DOI 10.5281/zenodo.51591



- Problem area :
 - Classification in Human Language, speech recognition
 - Massively-parallel data analysis (National Centre for Human Language Technologies), NWU
- Integration time – 3 hours (no prior experience)
 - github.com/SouthAfricaDigitalScience/libsvm-deploy
 - ci.sagrid.ac.za/job/libsvm-deploy

Build History			trend
✓ #5	Apr 26, 2016 3:56 PM		
✓ #4	Apr 26, 2016 12:18 PM		
! #3	Apr 26, 2016 12:09 PM		
! #2	Apr 26, 2016 12:06 PM		
! #1	Apr 26, 2016 12:05 PM		
			RSS for all RSS for failures



Using CODE-RADE

```
export CVMFS_DIR=/cvmfs/fastrepo.sagrid.ac.za/
```

```
# Add modules
```

```
module use $CVMFS_DIR/modules/compilers
```

```
module use $CVMFS_DIR/modules/libraries
```

```
module use ${CVMFS_DIR}/modules/bioinformatics
```

```
module use ${CVMFS_DIR}/modules/astro
```

```
module use ${CVMFS_DIR}/modules/physical-sciences
```

```
module use ${CVMFS_DIR}/modules/chemistry
```

```
$ module add libsvm
```

```
$ which svm-train
```

```
/cvmfs/fastrepo.sagrid.ac.za//generic/u1404/x86_64/libsvm/3.21/bin/svm-  
train
```


Example : Fermionic Molecular Dynamics

- Problem area : nuclear physics
 - Simulation of nuclear phenomena
 - Ithemba LABS/GSI
- Integration time – 21 hours (no prior experience)
 - github.com/SouthAfricaDigitalScience/FMD-codes-from-T.-Neff-/
 - ci.sagrid.ac.za/view/All/job/Fermionic-Molecular-Dynamics-deploy/

Configurations

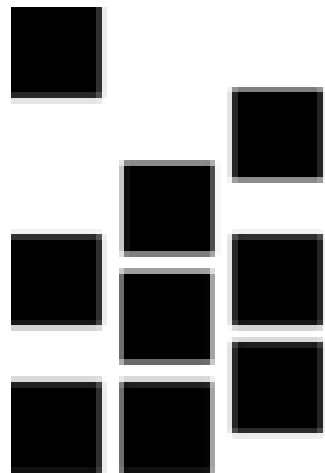
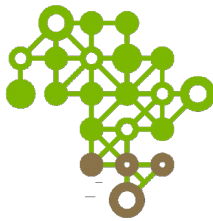
✓ [OS=u1404](#) ✓ [OS=sl6](#)

Upstream Projects

✓ [OpenMPI](#)
✓ [lapack-deploy](#)
✓ [ncurses-deploy](#)

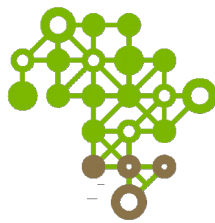
Build History		trend
✓ #13	Jul 4, 2016 9:22 PM	
✓ #12	Jul 4, 2016 8:43 PM	
✓ #11	Jul 4, 2016 8:37 PM	
✓ #10	Jul 4, 2016 7:31 PM	
! #9	Jul 4, 2016 4:58 PM	
! #8	Jul 4, 2016 4:47 PM	
PR #11 : Test case input and LICENSE...		
! #7	Jul 4, 2016 1:44 PM	
✓ #6	Jul 4, 2016 1:34 PM	
! #5	Jul 4, 2016 1:21 PM	
! #4	Jul 4, 2016 1:01 PM	
! #3	Jul 4, 2016 12:49 PM	
! #2	Jul 4, 2016 12:06 PM	
! #1	Jul 4, 2016 12:02 PM	
		RSS for all RSS for failures

Yes, it works



CODE RADE

BUILD TEST EXECUTE EVERYWHERE



Acknowledgements

- Many thanks to contributors for ideas, discussion and code
 - Original idea co-developed with Fanie Riekert (UFS)
 - Input and critique: Dane Kenndy, Sakhile Masoka (CHPC) and Peter van Heusden (SANBI)
 - EGI CVMFS Task Force support: Catalin Condurache (EGI, STFC)
 - Design and extension discussions : Timothy Carr (UCT)
 - CODE-RADE is supported by the [Sci-GalA project](#) under grant 654237 of the European Commission's Horizon 2020 programme



Summary

- Making the best use of computational infrastructure comes down to running applications
- Maintaining and porting them is hard and tricky. Communication is the hardest part.
- We've built an automated porting system, which
 - will deliver functional, tested, relevant software to your site.
 - Can ease the communication blocks in collaborative work
- Oh and it works :)
- Come on in and help us build it.

We want an Open e-Science Commons



Thanks ! Now, let's get to work...

- **Feedback, discussion:** discourse.sci-gaia.eu
- **Join the organisations:**
 -  [http://github.com/AAROC/](https://github.com/AAROC/)
 -  <https://github.com/SouthAfricaDigitalScience>
- **Come hack with us :**
 - Project board : <https://waffle.io/AAROC/CODE-RADE> 
 - Control Room : <https://africa-arabia-roc.slack.com> 
- **Let's talk : www.africa-grid.org**
 -  @TheSAGrid
 -  fb.com/SAGrid