Helix Nebula - The Science Cloud
The EMBL Experience and other bits

Dario Vianello (@vianello_d)
Cloud Bioinformatics Application Architect
Technology and Science integration team
EMBL-EBI
What is EMBL-EBI?

- Europe’s home for biological data services, research and training
- A trusted data provider for the life sciences
- Part of the European Molecular Biology Laboratory, an intergovernmental research organisation
- International: 600 members of staff from 57 nations
- Home of the ELIXIR Technical hub.
HNSciCloud Joint Pre-Commercial Procurement

**Procurers:** CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, STFC, SURFSara  
**Experts:** Trust-IT & EGI.eu

The group of procurers have committed:

- Procurement **funds**
- **Manpower** for testing/evaluation
- **Use-cases** with applications & data
- **In-house IT resources**

Resulting services will be made available to end-users from many research communities

Co-funded via H2020 Grant Agreement 687614

Total procurement budget >5M€

e-IRG Workshop  
8-9th June 2017 - Malta
The Hybrid Cloud model

Brings together:

- Research Organisations
- Data providers
- Publicly funded e-infrastructures
- Commercial cloud service providers

In a hybrid cloud with procurement and governance approaches suitable for the dynamic cloud market
The Pre-Commercial Procurement process
3 EMBL use-cases:

- **PanCancer**
  *Builds a dataset to enable researchers to compare data across cancer types, ~ 2800 samples*

- **EuroBioImaging**
  *Image data repositories and analysis tools*

- **ELIXIR - Long Tail of Science**
  *To simplify access to quality-controlled data, services and tools for researchers in all life-science disciplines.*
Strategically: why clouds?

- As data increases, have the community bring their compute to data
  • *But not all their compute to our data centre!*

- Need to push relevant data sets and services out to cloud providers
  • *EMBL-EBI Embassy Cloud → ELIXIR → EOSC → AWS/GCP/MSA → ?*

- Hybrid cloud an approach to optimise EMBL-EBI CapEx vs. OpEx
  • *Allow CapEx to lag demand & use OpEx to manage peaks*

How to make data and workloads fly to the clouds?
EMBL-EBI & Public clouds

In the last 2 years:

- Established an EMBL-EBI Benchmarking Suite
- Basic and now *advanced* PoCs with several Cloud Providers

**Summer 2016**, first EMBL-EBI “Cloud Tender”: GCP, MSA, UKCloud

**September 2016**, Hybrid Cloud Working group is established

**May 2017**, first results coming out of the *(many)* pilots
EMBL-EBI & The Clouds

Our “Science” take home messages

➢ The Cloud is a cool place: virtually unlimited resources & unlimited scalability

➢ True, but easy to achieve only for cloud-native workloads

➢ Porting “ground-flying” pipelines isn’t that easy

➢ Need DevOps / ResOps approach

   Pipelines deployment must be on demand, as compute

➢ Build a playground, monitor everything, and then iterate
HNSciCloud - what next?

- **Now in Prototype** phase (end December 2017)

- Requires an *interesting* amount of involvement from all the parties

- Once completed, will hopefully lower the entry barriers to the Cloud:
  - *Data transparency layer*
  - *Federated AAI*
  - *Procurement frameworks*
EMBL-EBI - what next?

From an EMBL-EBI perspective

- **Expand** the breadth of our pilots

- Training, training, and **some more training**

- **Fully** port applications to the cloud: *investment* needed!

- Contribute & pull in any outstanding results out of:
  - *Helix Nebula - The Science Cloud*
  - EOSCpilot
Thank you!
The Embassy Cloud

- Built & operated by EMBL-EBI
- 92 compute nodes
- Which provide a total of 6,000 vCPU
- 4GB RAM per vCPU
- All hosts have 2x10Gb network in chassis
- 40Gb network from chassis to storage networks
The European Molecular Biology Laboratory

80+ nationalities

Heidelberg, Germany
Main Laboratory
Tissue Biology, Disease Modeling

Barcelona, Spain

Hinxton, Cambridge, UK
Bioinformatics
Mouse Biology

Monterotondo, Rome, Italy

Grenoble, France
Structural Biology

Hamburg, Germany

>1600 personnel

8-9th June 2017 - Malta
EMBL-EBI & The Clouds
Our “Science” take home messages

EMBL-EBI Grid (1 sources) (tree view)

- CPUs Total: 646
- Hosts up: 82
- Hosts down: 0

Current Load Avg (15, 5, 1m):
- 1%, 0%, 0%

Avg Utilization (last custom):
- 20%

Localtime:
- 2017-04-27 14:57

EMBL-EBI Grid Load last custom

EMBL-EBI Grid Memory last custom

EMBL-EBI Grid CPU last custom

EMBL-EBI Grid Network last custom

---
e-IRG Workshop
8-9th June 2017 - Malta
EMBL-EBI & The Clouds

Our “Science” take home messages

SRP005784

e-IRG Workshop
8-9th June 2017 - Malta