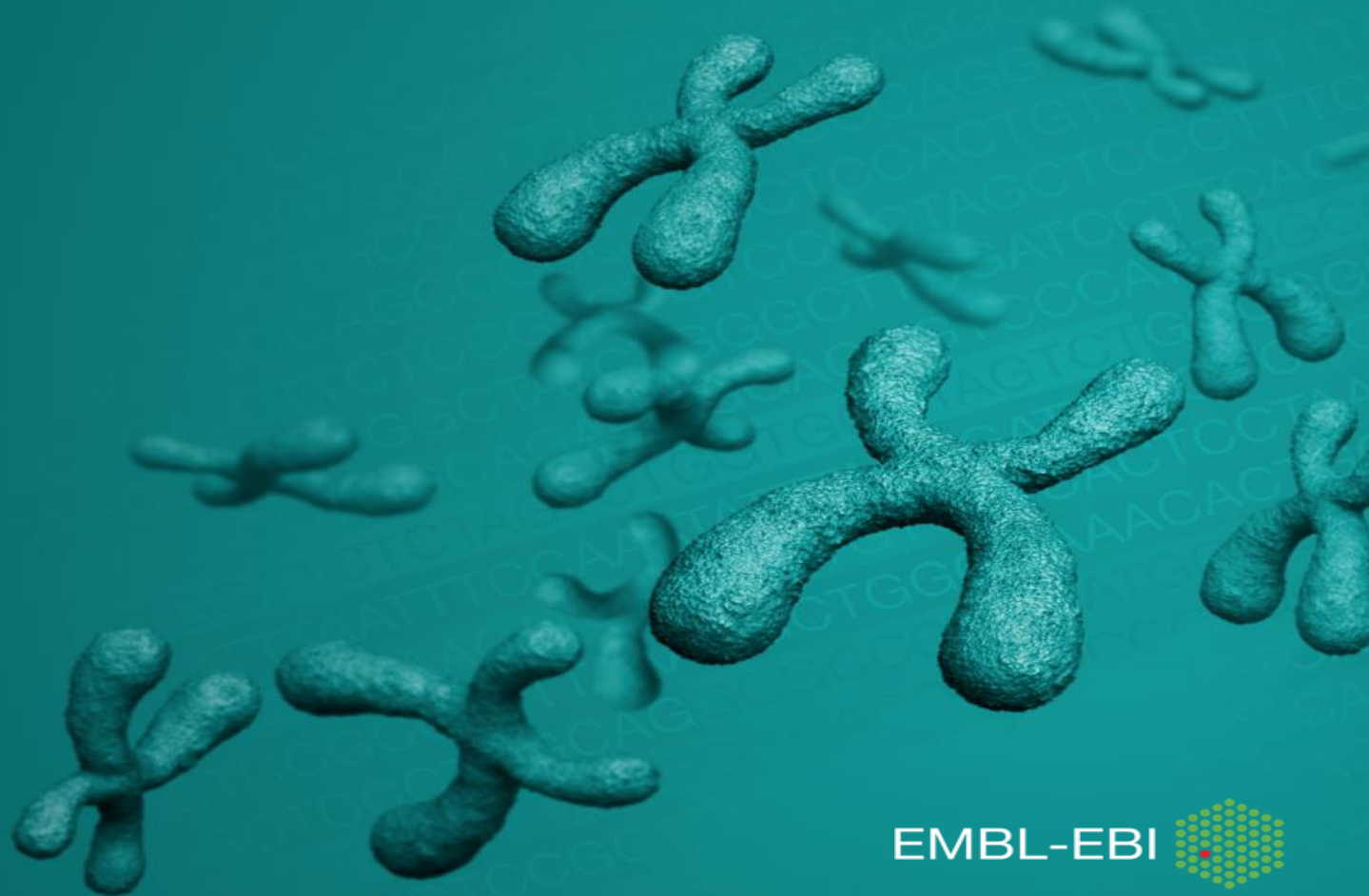


# Embassy Cloud

Andy Cafferkey



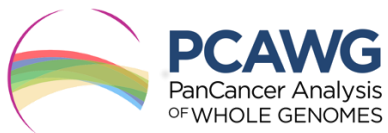
# Introduction to the Embassy Cloud

- Embassy Cloud is the cloud infrastructure owned & operated by EMBL-EBI
- It provides a secure collaborative workspace with devolved administration
- Embassy provides access to EMBL-EBI services & data
  - ...and tenants create their own datasets that they share with each other
- Embassy is a cloud designed for bioinformatics workloads
- Developed with EMBL Research & Service teams

EMBASSY  cloud

# What drove the development of the Embassy Cloud?

- Driven by our users:
  - Initially hosting Europe PMC on a VMware platform
  - Working with GSK and establishing the Embassy model
  - Supporting Pan-Cancer in VMware vCloud
  - OpenStack supporting Pan-Cancer expansion
  - Now 60+ tenants
- Making workloads portable between infrastructures
- Sharing data in distributed collaborations (Image Data Resource, Pan-Cancer, COMPARE)



# Embassy Cloud Service team use case examples

- Image Data Resource
  - European project to build image data repositories for life science
  - Accessing imaging data that has been processed at EBI
  - source data is thousands of small (0.5-1MB) .tif image files
  - Size of the current dataset is 10TB, stored at EMBL-EBI
- COMPARE (a collaboration of 29 partners)
  - Disease outbreak and response tool
  - Accesses 100TB of the ENA directly
  - Deploys resources elastically to deal with bursty workload



# How research tenants have driven Embassy development

- Pan-Cancer analysis required 1PB reference data, 200TB output, 1000 cores
- Data was simultaneously used by a second tenant to serve as the European hub of the project
  - All this using an application stack that the consortium was developing whilst being deployed
  - Constant interaction between us and the research team dev/ops
  - And from us into the other technical TSC teams
  - The limits of the software, hardware & networking were tested
- Through this Embassy Cloud has been tuned to support bioinformatics workloads

# Helping users adapt to cloud

- Users have a steep learning curve with IaaS
  - Requires sysadmin skills to manage VMs & automation
  - Skills not always readily available in research groups or service teams
  - Successful cloud use is most likely when the skills above are present
- TSC provides a cloud application consultancy service
- We ensure every tenant has access to a central support hub and the option of ongoing meetings with the team to provide input to their tech stacks, etc

# Future research interactions

- Research teams are already working on the next generation of large scale workloads including Pan-Paediatic Cancer and Marine MetaGenomics
- We are working much more closely with these tenant admins to help bridge the skills gap
- For Pan-Paediatic the Gerstung/Zerbino/Systems Applications teams have collaborated to make a joint appointment for the tenant admin role
  - The tenant admin spent the first 3 months working in the Systems Applications team
  - The teams continue to work closely together to develop Embassy Cloud

# Future service interactions

- Service team collaborations like IDR & COMPARE suit Embassy well, requiring delegated administrative responsibility that can be easily adapted to match the collaboration organization, no need for institutional user accounts.
- These collaborations scale when they need to using the API driven software defined infrastructure.
- These are the bulk of our 60+ tenants, requiring smaller resources per tenant compared to research.



# Future of Embassy Cloud

- Adapting to Cloud is challenging
  - New workloads can be developed straight into a cloud environment
  - Moving existing workloads to cloud is a lot of work
- Success requires a DevOps skillset
  - Infrastructure provider and tenant need to interact closely
- Establishing the cloud is only part of the job
  - Initially we focused on building & establishing the infrastructure, but...
  - Adapting workloads to the Cloud environment is a complex and lengthy task
  - Tenants needed support in developing the software layers between their workload and IaaS. - Containers as a Service offers a potential solution

# EMBL-EBI Embassy Cloud



## Co-located with EBI services

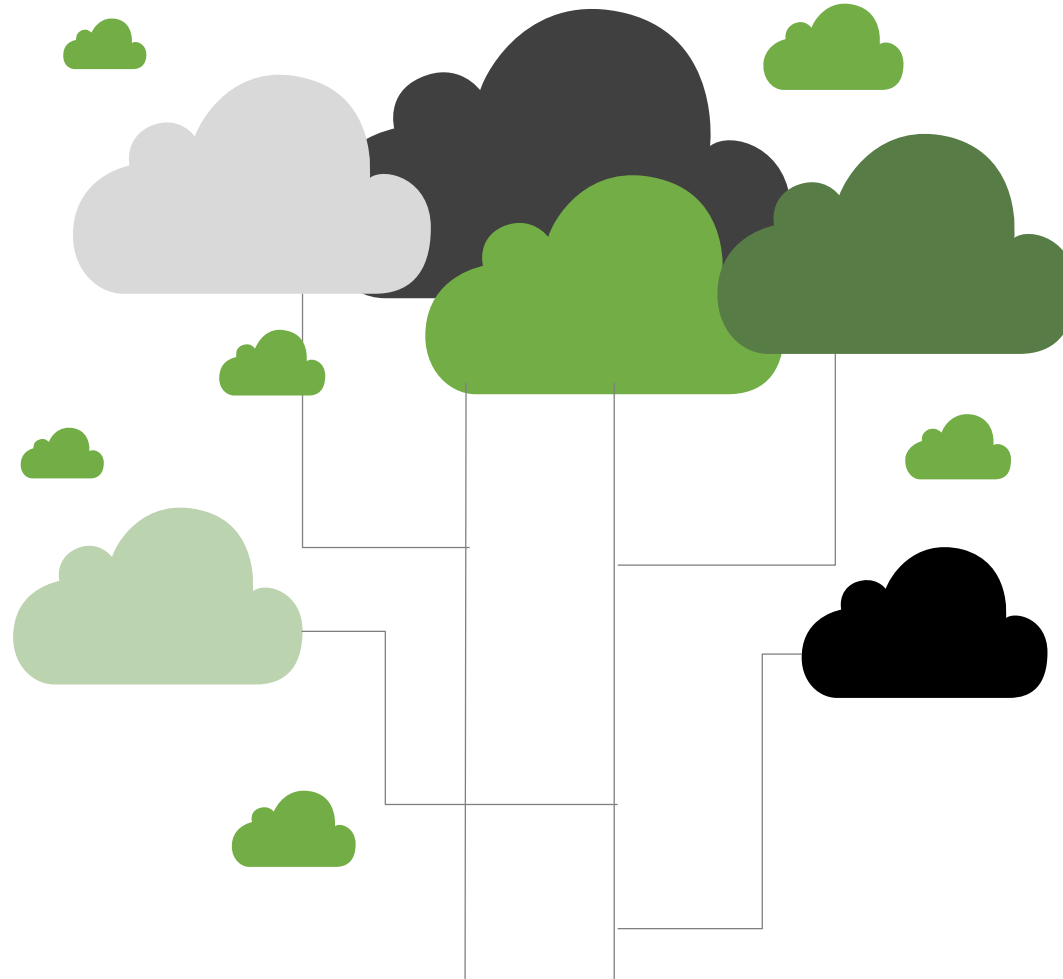
Embassy provides access to EMBL-EBI services & data.

Tenants can also create their own datasets that they share with each other



## User Support

We ensure every tenant has access to a central support hub and the option of ongoing meetings with the team to provide input to their technology stacks



## Designed for Bioinformatics

Embassy has been developed with EMBL-EBI's research and service teams to provide an infrastructure tuned for bioinformatics workloads.



## Secure

Embassy provides a secure collaborative workspace with devolved administration.

