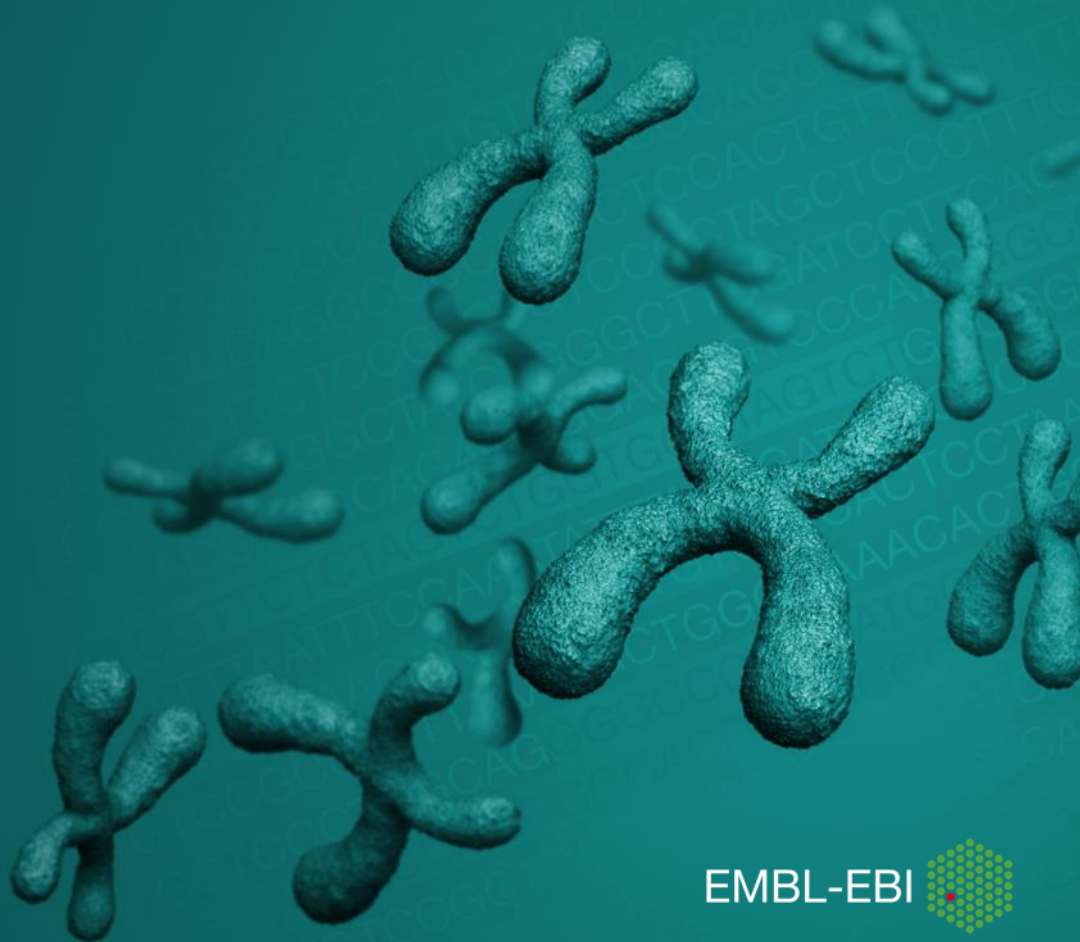


Virtualisation and Cloud Computing at

EBI

Steven Newhouse, Head of Technical Services



European Bioinformatics Institute

- Outstation of the European Molecular Biology Laboratory
- International organisation created by treaty (cf CERN, ESA)
- 20 year history of service provision and scientific excellence
- EMBL-EBI has 500+ Staff & €50 Million Budget
- Provide services to a wide range of users using an “as-possible” usage model
- Thin-client model
 - Web browser & web services
 - Equivalent to SaaS



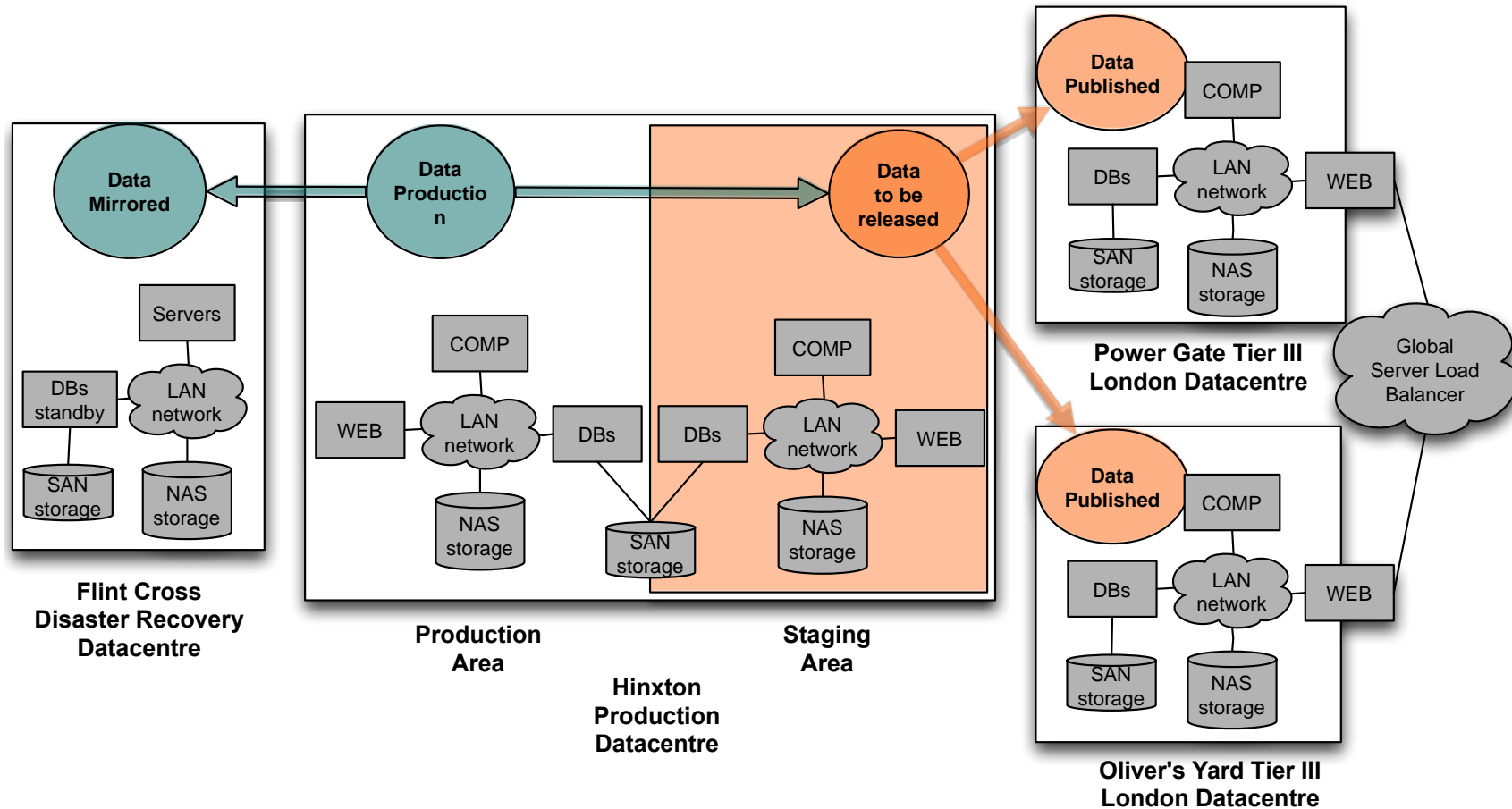
The Challenge Facing Bioinformatics

- Volume and variety of genomic data expanding
 - Data at EBI doubling every year - replication is challenging
 - >10,000 CPUs & 30PB (but need more!)
- Complex analysis
 - Access to both public and managed access data sets
 - Bespoke workflows and tools across a variety of domains
 - Issues with disk to memory bandwidth
- EMBL-EBI Provides
 - Public & restricted data sets
 - Web and programmatic access to services (3M unique users)

Impact on EMBL-EBI's Infrastructure

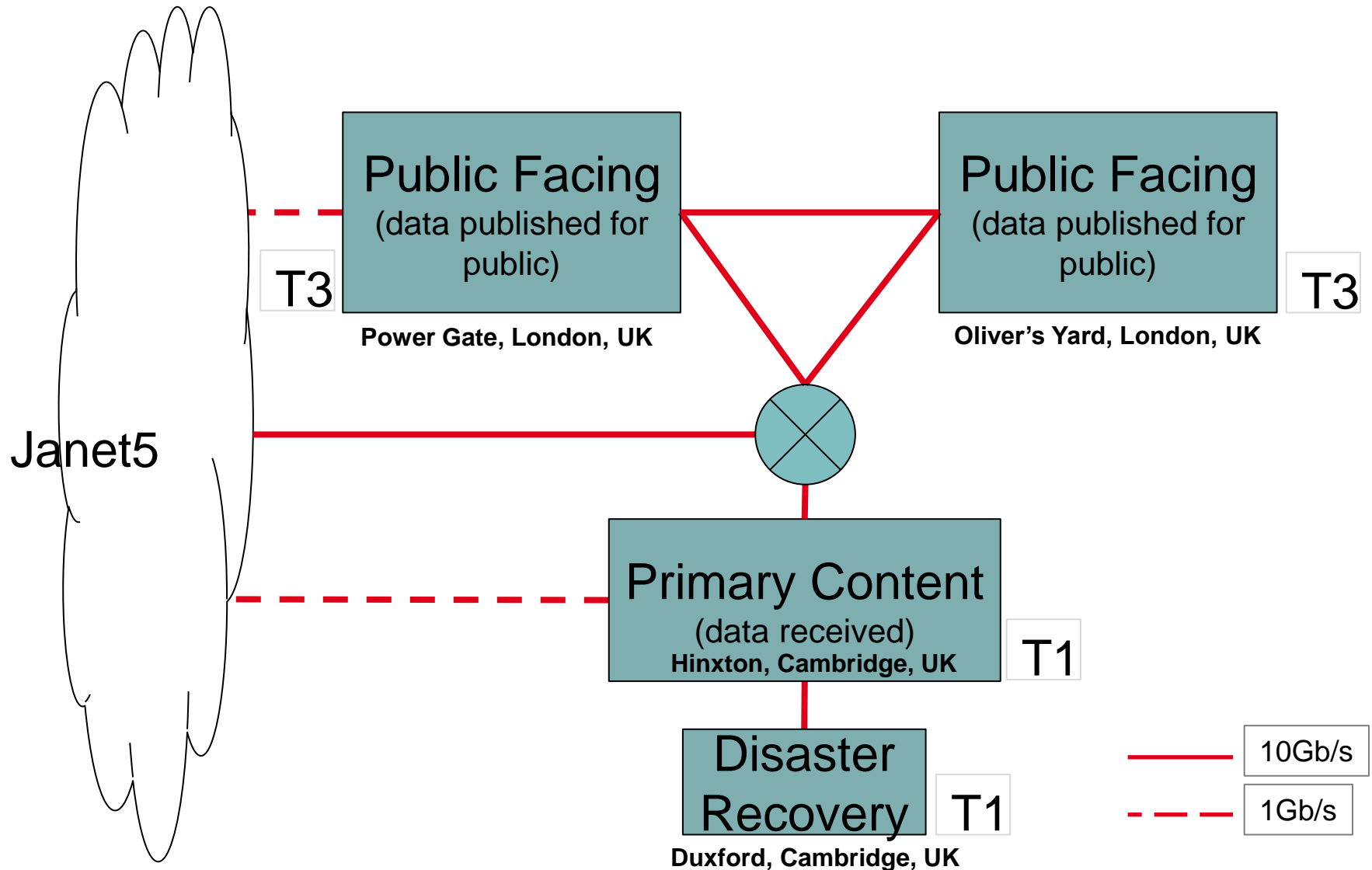
- Grow the capacity of the current data centres
 - Commodity infrastructure – blades and NAS (50 racks)
 - RDBMS and SAN for high throughput transaction processing
 - Tape backup is no longer feasible
- Provide a resilient topology by geographical separation
 - Against local & regional disaster in the UK
 - Against national disaster through international collaboration
- Enable new easier science through the cloud
 - Provide access to the increasingly hard to replicate data sets
 - Embassy Cloud: IaaS service coupled to public data sets

Overview EMBL-EBI IT infrastructure

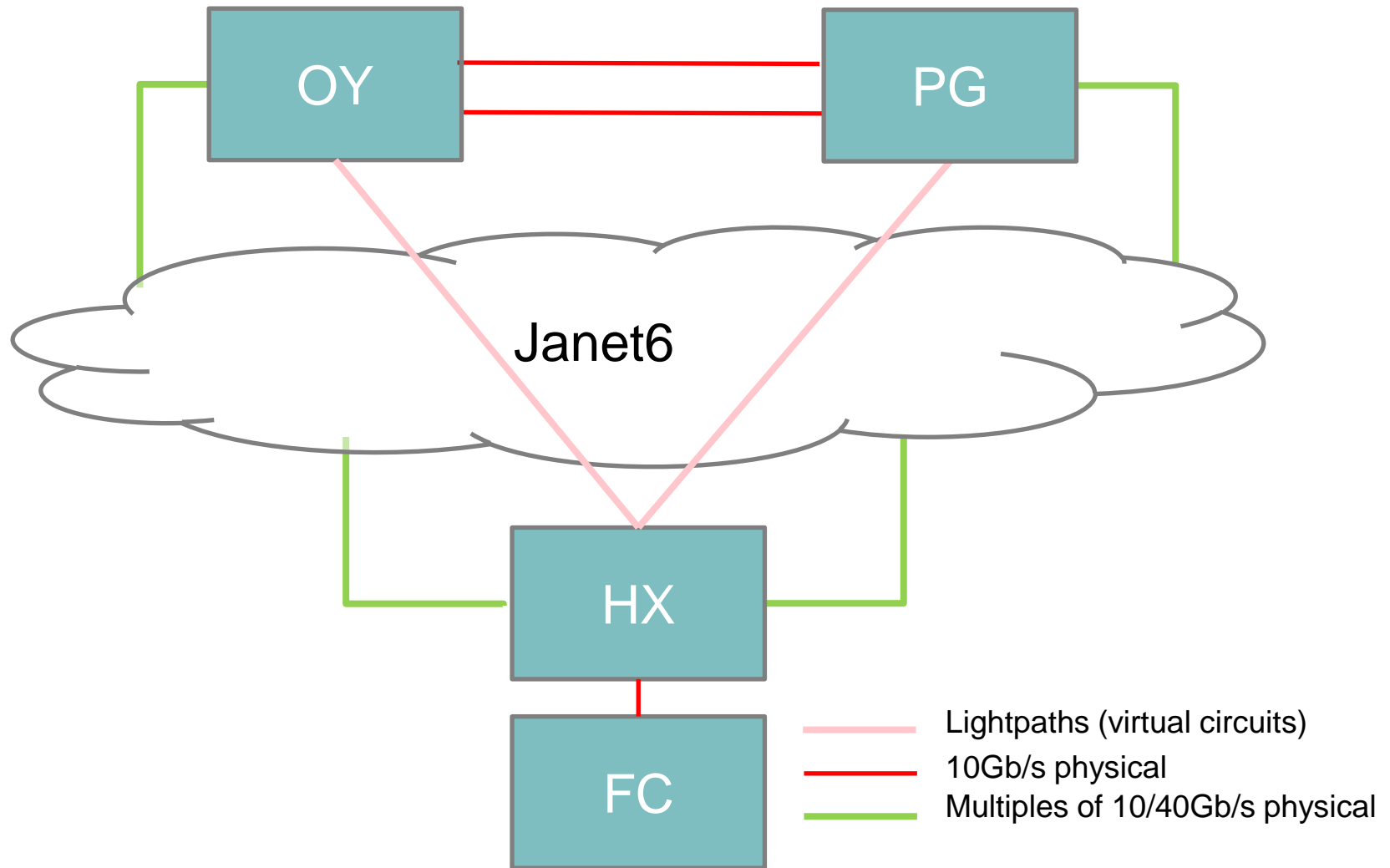


Data centre virtualised throughout with VMWare

Overview Datacentre facilities and function



Upgraded WAN topology from Jan 2014



EMBL-EBI Embassy Cloud

- Pilot service hosted at EMBL-EBI data centres
 - Logically isolated outside EBI's LANs
 - Secure flexible infrastructure for both tenant and host
 - File based access to EBIs' data sets
 - Currently, only the 1000 Genomes dataset exposed
- Expect both academic and commercial users
 - Wishing to move their compute and data to EBI's 'big-data'
- Resources exposed using VMware's vCloud Director
 - SSL Connections to the web management interface
 - Provide isolated IaaS clouds to multiple tenant organisations

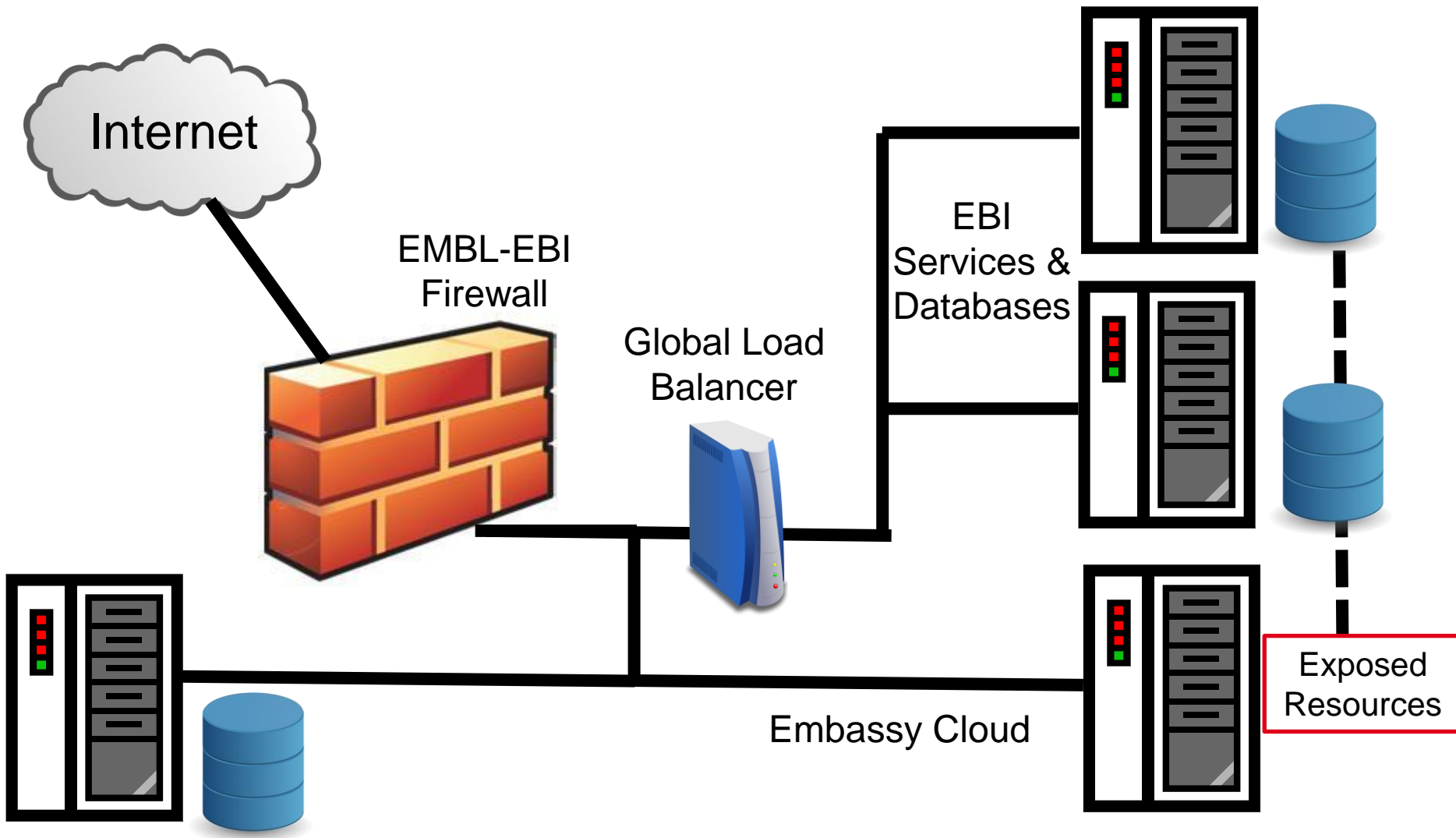
Why 'Embassy' Cloud?

- An embassy is sovereign territory in a host country
 - Host Country: EMBL-EBI Data Centre
 - Sovereign Territory: Host Country not allowed to enter
- Virtualisation provides the protection for 'tenant' and 'host'
 - Host puts boundaries in place to protect it from the tenant
 - Tenant has freedom and control within those boundaries
- Added value from EMBL-EBI over other clouds:
 - Machines and data hosted in known jurisdiction
 - File access to hosted data sets (public & managed access)
 - Direct network access to public EMBL-EBI services

Adopting an IaaS Model

- Tenant organisations get an empty virtual infrastructure
 - They establish their own VMs and networks
- Tenant organisation establishes their own access rules
 - Firewall to control access and site to site VPN tunnelling
 - Can use LDAP or manually create users
 - Users can be assigned access to specific vApps (VM groups)
- Tenant organisations to the work
 - Run their own services but with fast access to EBI's datasets
 - System administration performed by the tenant
 - EMBL-EBI staff have no access to the VMs

Embassy Cloud



Embassy Cloud – User (Operator) Experience

The screenshot displays the Elixir user interface for an organization administrator. The top navigation bar includes the Elixir logo, the user name 'andy (Organization Administrator)', and links for 'Preferences', 'Help', and 'Logout'. Below the navigation bar are tabs for 'Home', 'My Cloud', 'Catalogs', and 'Administration'. The main content area is divided into a 'Quick Access' section and a vApp management area. The 'Quick Access' section provides instructions on how to start or use a vApp. The vApp management area shows a single vApp named 'vApp_root_13' in a 'Running' state, with a thumbnail image and control buttons (play, pause, stop). A sidebar on the right contains a list of management options under three categories: 'Organizations' (Org Settings, Manage vDCs, Manage Networks), 'Content' (Manage vApps, Add vApp, Build new vApp, Manage Catalogs, New Catalog), and 'Users & Groups' (Administer Users, Notify Users). At the bottom, a status bar shows '0 Running' and '0 Failed' vApps, and the interface is powered by VMware vCloud Director.

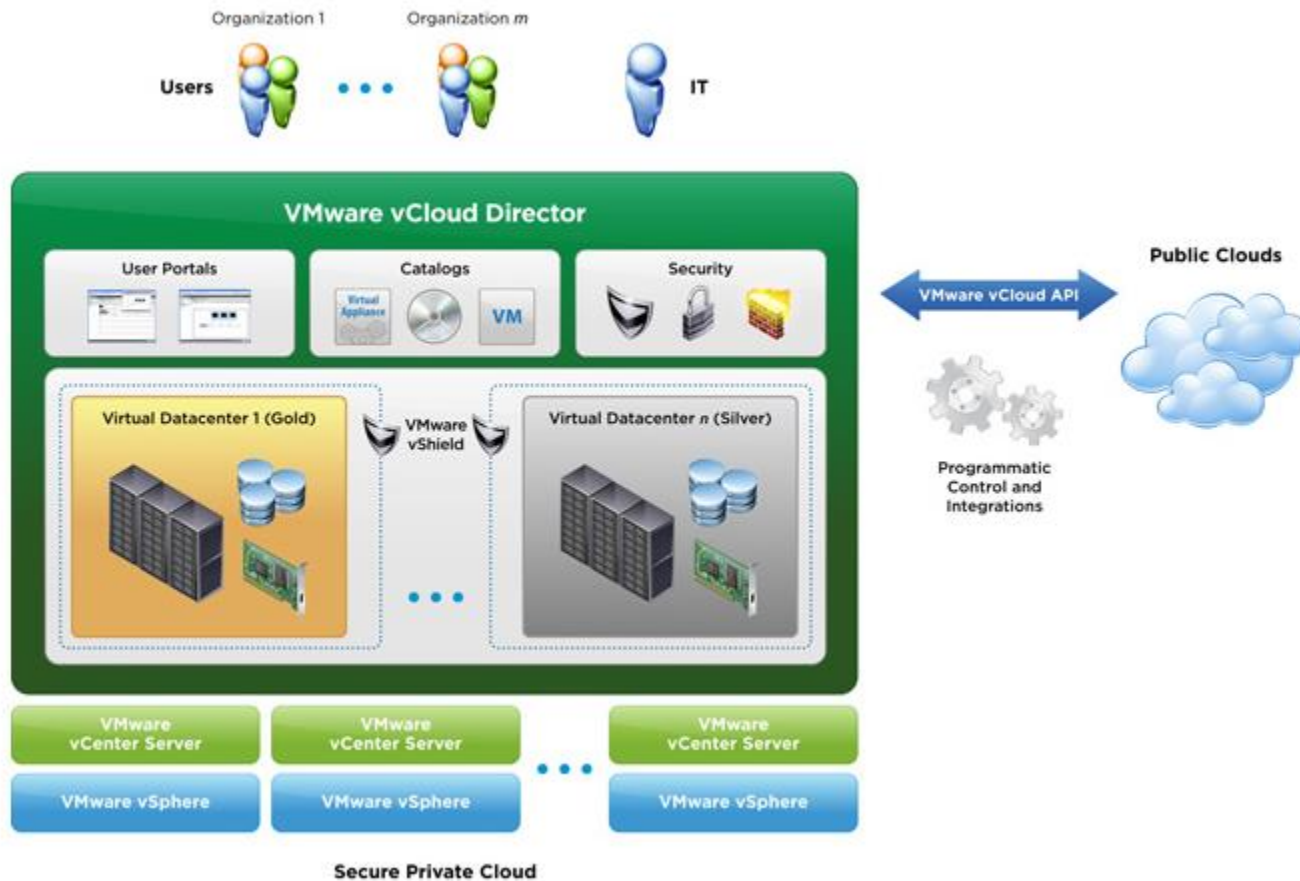
Adding a preconfigured Application

The screenshot displays a cloud management interface with a navigation bar at the top containing 'Home', 'My Cloud', 'Catalogs', and 'Administration'. Below the navigation bar, there is a 'Quick' section with a 'To start' list. The 'To start' list includes 'Add vApp from Catalog', 'Select vApp Template', and 'Name this vApp'. A 'Select vApp Template' dialog box is open, showing 'Add vApp from Catalog' as the selected option. In the background, the 'Add vApp from Catalog' wizard is visible, with the 'Configure Virtual Machines' step selected. The wizard shows a 'my_vApp' instance in a 'Pending' state, with a 'Creating...' progress indicator. The main window displays a virtual desktop environment with a top bar containing 'Applications', 'Places', and 'System'. The desktop background features a blue cloud and a green leaf. A window titled '1000-Genomes-Data' is open, showing a red and white striped icon and a spring icon.

Networking

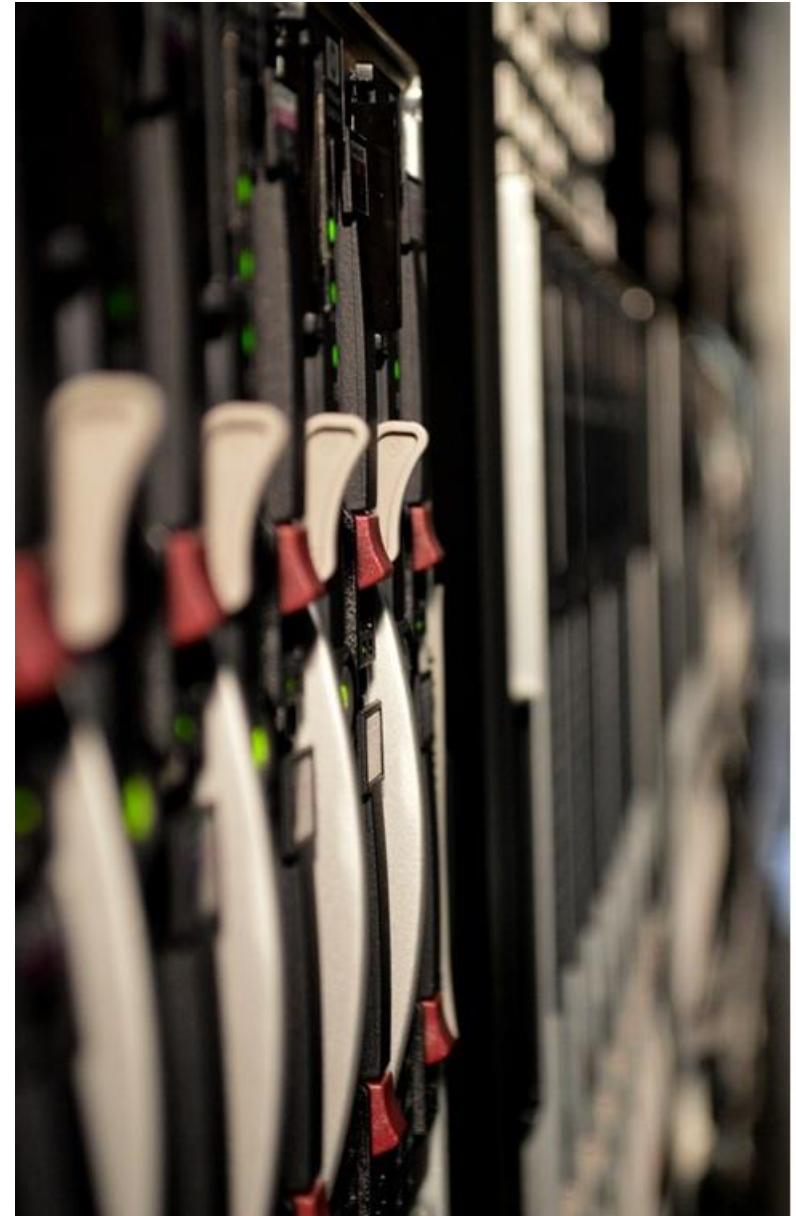
The screenshot displays a multi-layered configuration window for 'Configure Services: Elixir routed'. The top layer shows tabs for DHCP, Firewall, NAT - External IPs, NAT Mapping, Site-to-Site VPN, and Static Routing. A second layer is partially visible behind it, and a third, more detailed layer is in the foreground. This foreground layer has the 'Site-to-Site VPN' tab selected. It contains a checkbox for 'Enable site-to-site VPN' which is currently unchecked. Below this, there is a field for 'External IP address' with the value '10.3.3.103' and a descriptive note: 'This is the IP address of the external interface of the vShield edge router deployed for this network.' There is also a 'Public IP address' field which is empty, with a note: 'This field is optional. It is required so that entities from different external networks (across WAN/Internet) can reach the external interface of the vShield edge router.' At the bottom of this layer, the text 'Tunnels to other networks' is visible.

Technical Solution



Hardware and Software

- Hardware
 - 349 GHz CPU
 - 2.26 TB RAM
 - 33TB HDD
- Software
 - VMware ESXi 5 installed
 - Managed by vCloud Director
 - Provides the cloud layer & automates provision of the physical resources to tenants

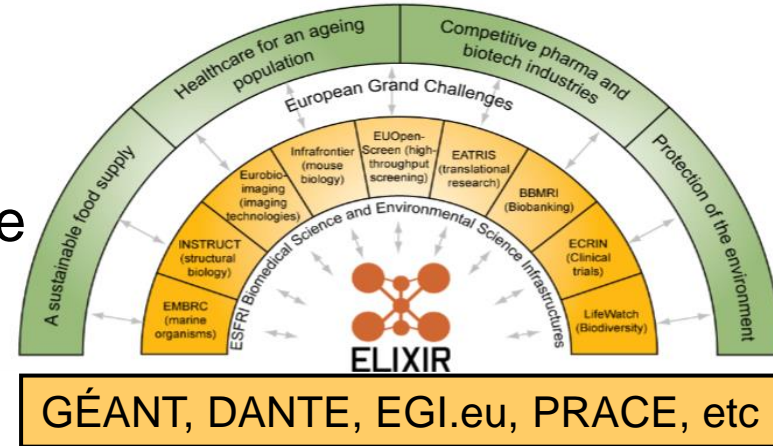


Other Cloud Activity at EMBL-EBI

- Use Amazon to provide geographical distribution
 - Direct link to globally replicate databases
- HelixNebula
 - Integration of commercial cloud providers with big research
- Benefit of additional security assurances
 - For use by pharmaceutical companies
 - For on-demand personalised medicine
- Explore using IaaS to supplement/replace data centres
 - Put DC on cloud, scale out services (service + database), etc.

The Future

- Exploitation by ELIXR
 - An e-Infrastructure for Life Science
 - Develop the Embassy Cloud
 - Commercial Use
 - Secure access to restricted datasets
 - Open up access to more external users
 - Explore use for internal service delivery teams
 - Assess mixed model
 - Use of commercial IaaS and public sector resources
 - Use of OpenStack



Any questions?

- Contact Points
 - steven.newhouse@ebi.ac.uk
 - embassycloud@ebi.ac.uk
- Acknowledgements
 - Andy Cafferkey
 - Pete Jokinen
 - EMBL-EBI Systems Team