
The Cambridge Bio-Medical-Cloud

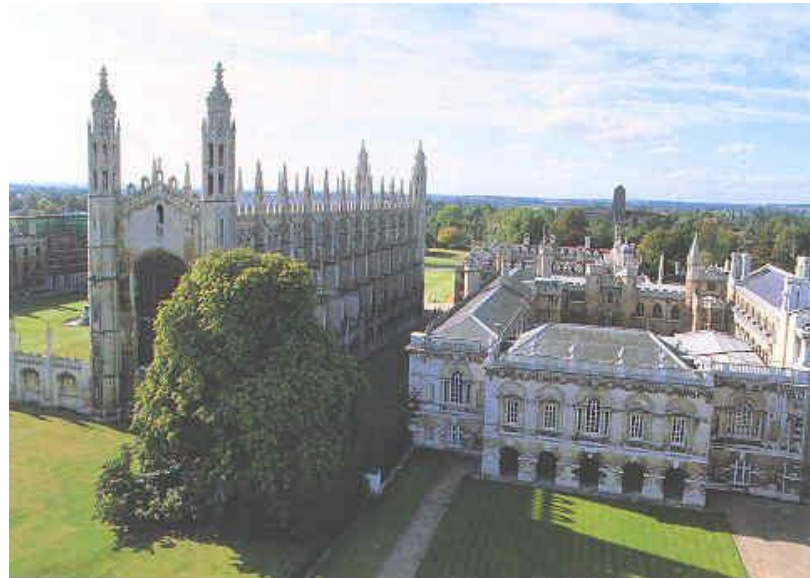
An OpenStack platform for medical analytics
and biomedical research



Dr Paul Calleja
Director of Research Computing
University of Cambridge

Global leader in science & technology innovation

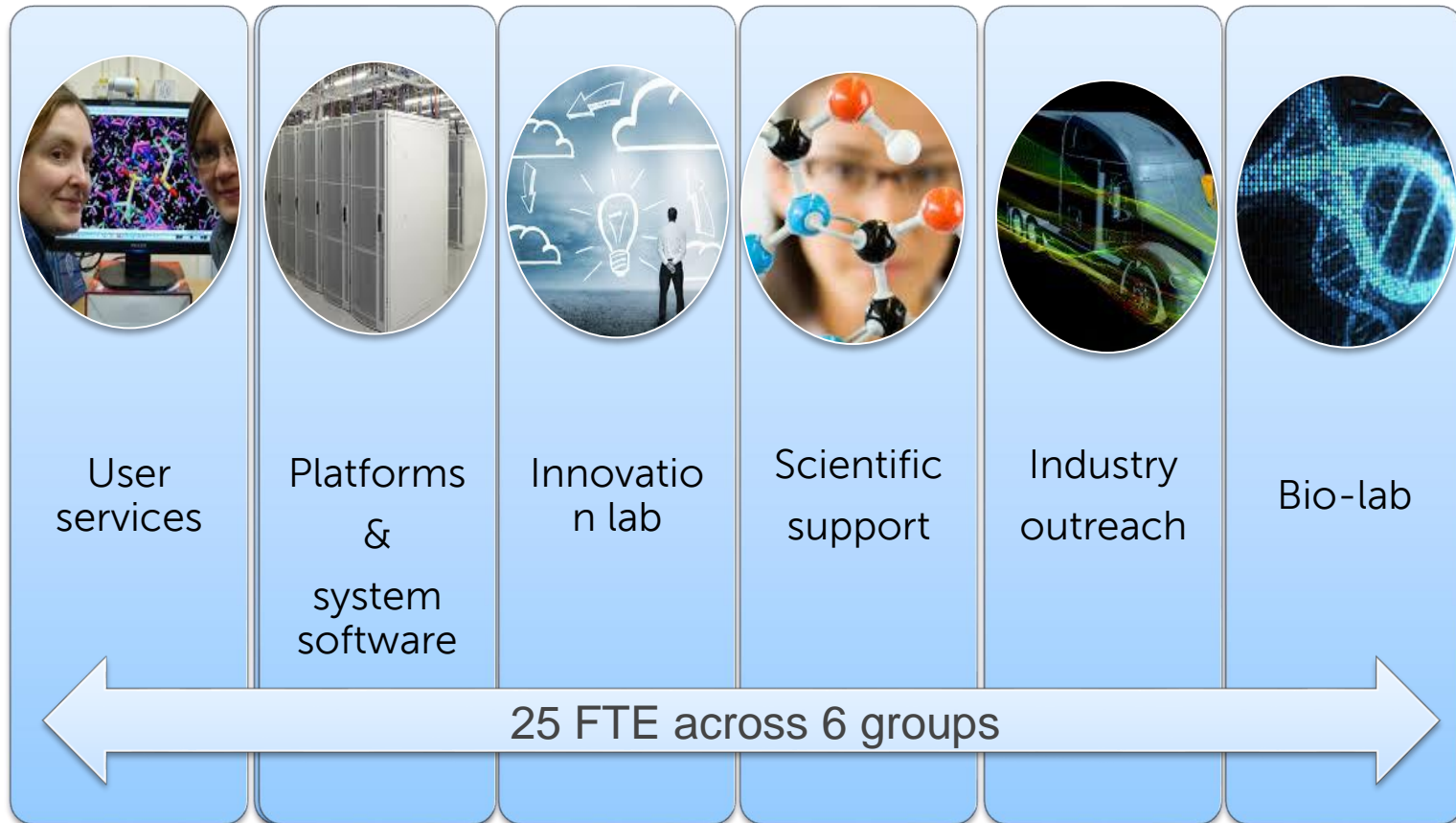
- One of the worlds leading research intensive Universities in terms of research outputs and impact, 10,000 staff £1.2B turn over
- Over 800 years old with 92 Nobel Laureates
- The Cambridge Cluster
 - 1535 technology companies in surrounding science parks
 - 27,000 staff, £13B turn over



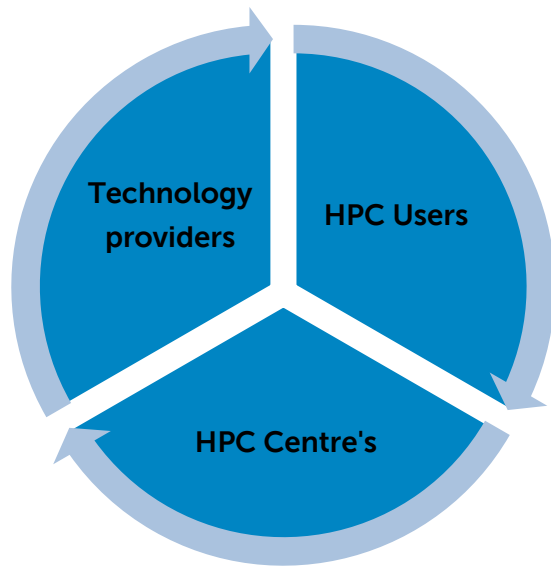
Research computing focus



Service structure



Dell-Intel-Cambridge HPC Solution Centre



POC's



Customer engagement



White papers



Cambridge HPC investment

Highly resilient DC

- 100 Cabinets, 2000Kw IT Load
- 30 Kw water cooled racks
- Second site multi 10Gb connected



People

- 25 FTE technical team
- Strong skills in HPC system integration
- Large scale storage
- Openstack development & deployment
- Scientific support



Systems

- 500 TF compute performance (1000 servers X86 +GPU),
- 200 node Hadoop system
- 15 PB storage + Intel Lustre & tape
- Q1 2017 upgrade will add an additional 3PF performance and 30 PB storage
- **Largest academic supercomputing facility in UK**
- **Run rate cost £4M per year**



Research computing usage and outputs

- 1016 active from 272 research groups from 42 University departments
- 80% system utilisation
- Long tail appearing - significant usage by over 300 users who consumed 200 workstation days of usage in last 12 months
- New user growth rate is 28% CAGR year on year for last 9 years, growth rate is expected to grow with Openstack usage models
- Research computing services support a current active grant portfolio of £120 – which represents 8% of the Universities annual grant income
- Underpinning 1400 publications over the last 9 years, current output ~300 per year



Why OpenStack in research computing



Makes computing, data and applications more accessible, flexible and secure.

Makes research computing & data easier to use and easier to share.

Decreasing the time to science and increasing innovation



OpenStack activities @ Cambridge

- Development and deployment of Openstack in the broad research computing environment with a particular focus on bio-medical computing
- Proof of concept work for OpenStack for use within the SKA, contracted by Astrophysics under the direction of Prof Paul Alexander
- The work being undertaken by the Research Computing Service is an active collaboration with the following partners:-
 - Dell
 - Intel
 - Redhat
 - Mellanox
 - Nexenta
 - StackHPC
 - TACC, Monash University, CHPC
 - Emerging scientific OpenStack community



OpenStack use cases within research computing

1) Research computing & storage infrastructure as a service (IaaS)

- *Provides central IT infrastructure, for departmental local IT staff to construct persistent IT services. Local IT staff build and support platforms*
- *Enfranchises local IT staff, drives efficiencies, increases agility*

2) Research computing & storage platform as a service (PaaS)

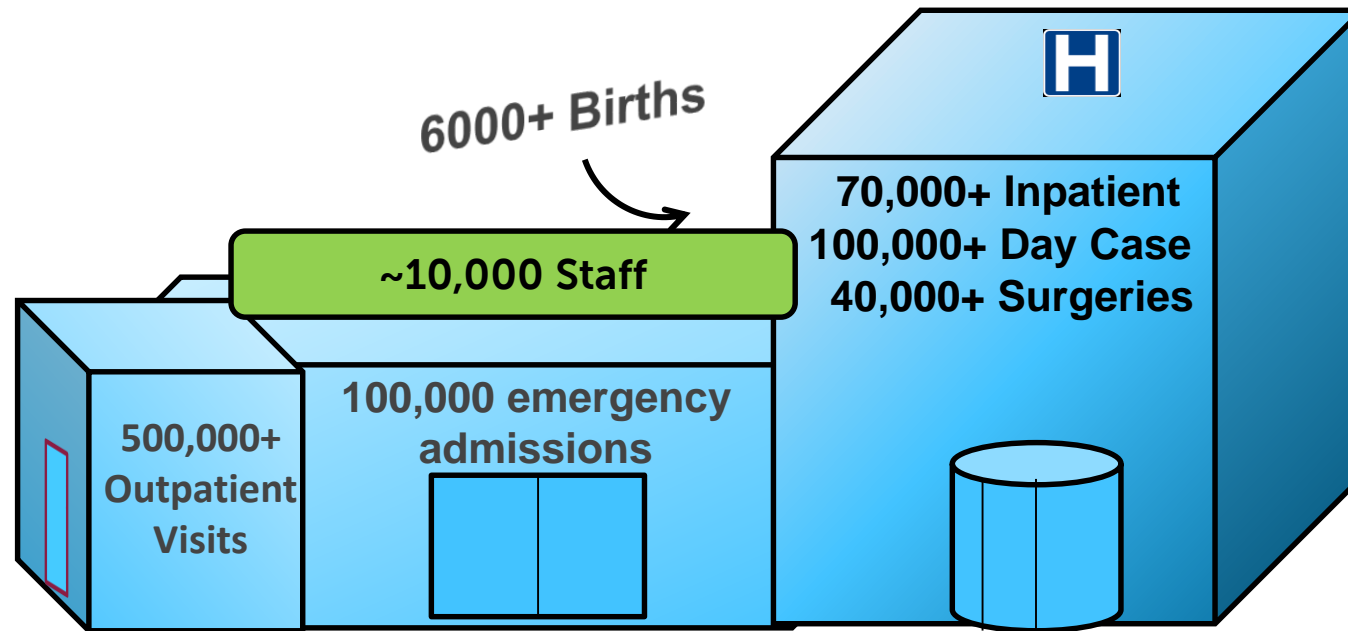
- *Provides central IT platforms, for research group or departmental local IT staff to build persistent IT services. Local IT staff build and support platforms*
- *Enfranchises research staff and local IT staff drives efficiencies increases agility*

3) Science as a service (SciaaS)

- *Allowing easy access to virtualised or containerised compute infrastructure, long tail science, new user communities, science portals, applications and workflows*
- *Can also include traditional HPC architectures, Data analytics, Hadoop, machine learning*
- *Easy access to compute, analytics, data, applications & workflows greatly democratising research computing capability driving discovery*



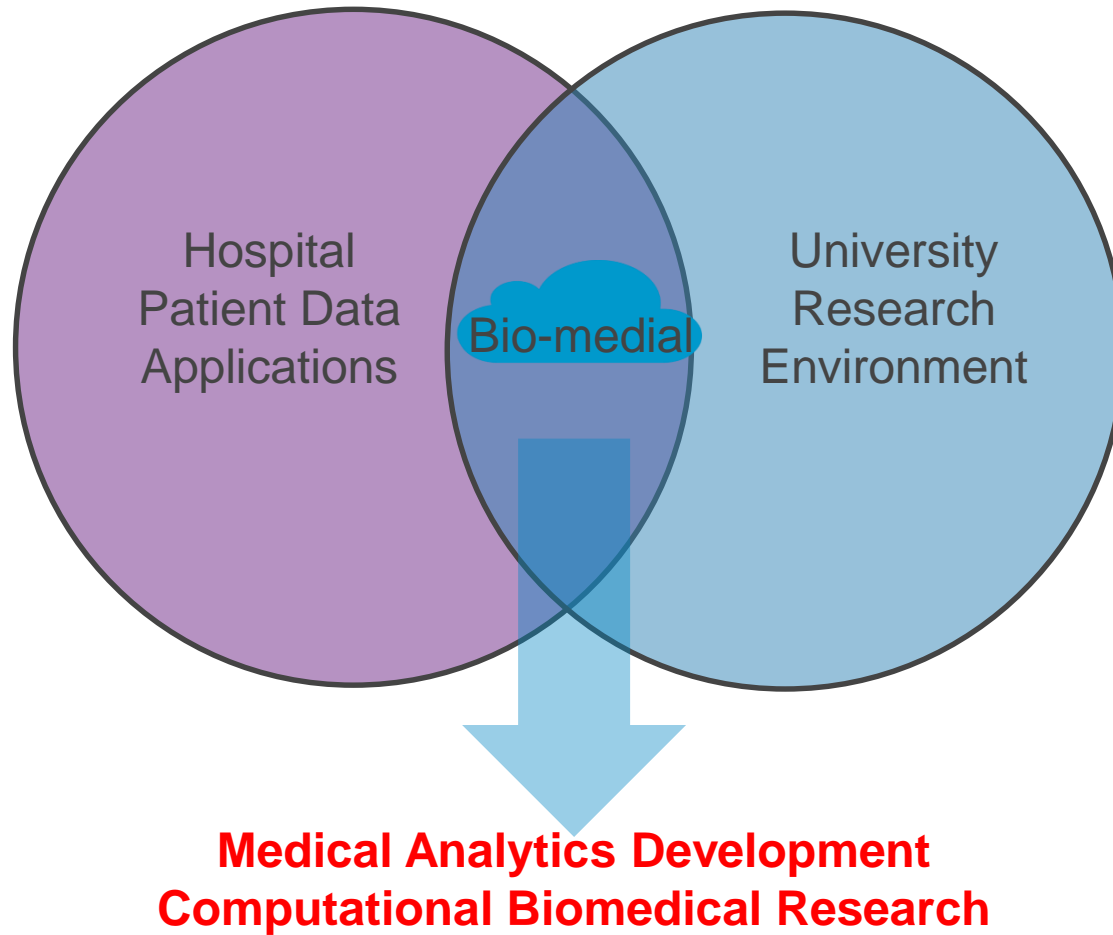
Why OpenStack in health care



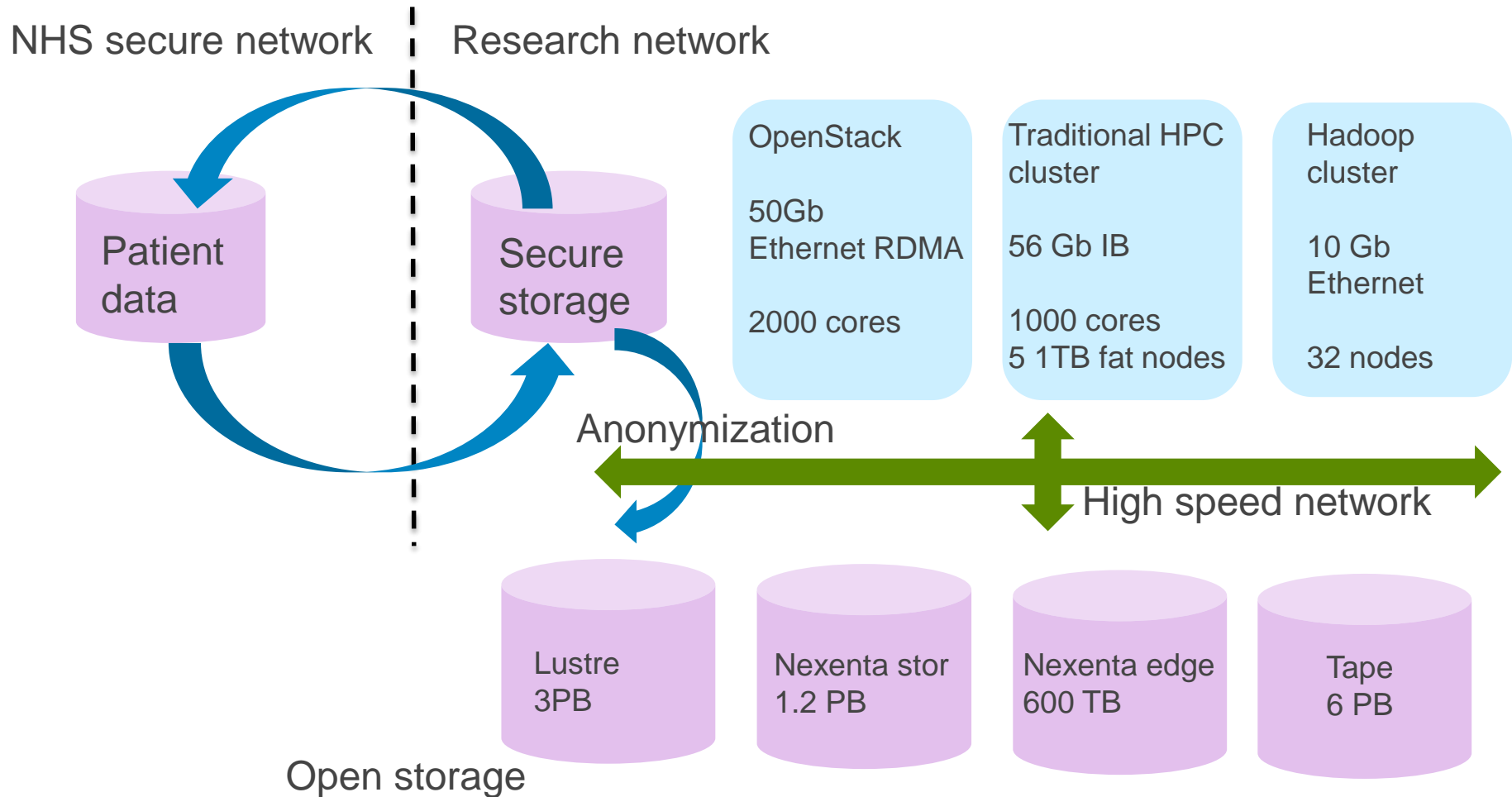
- **Huge amount of data** all held in a coordinated **electronic records system**
- Ideal opportunity to apply **big data techniques for improved health** outcomes
- Need **secure, flexible, elastic** IT platform, that suits it self to **sand-boxed research computing**



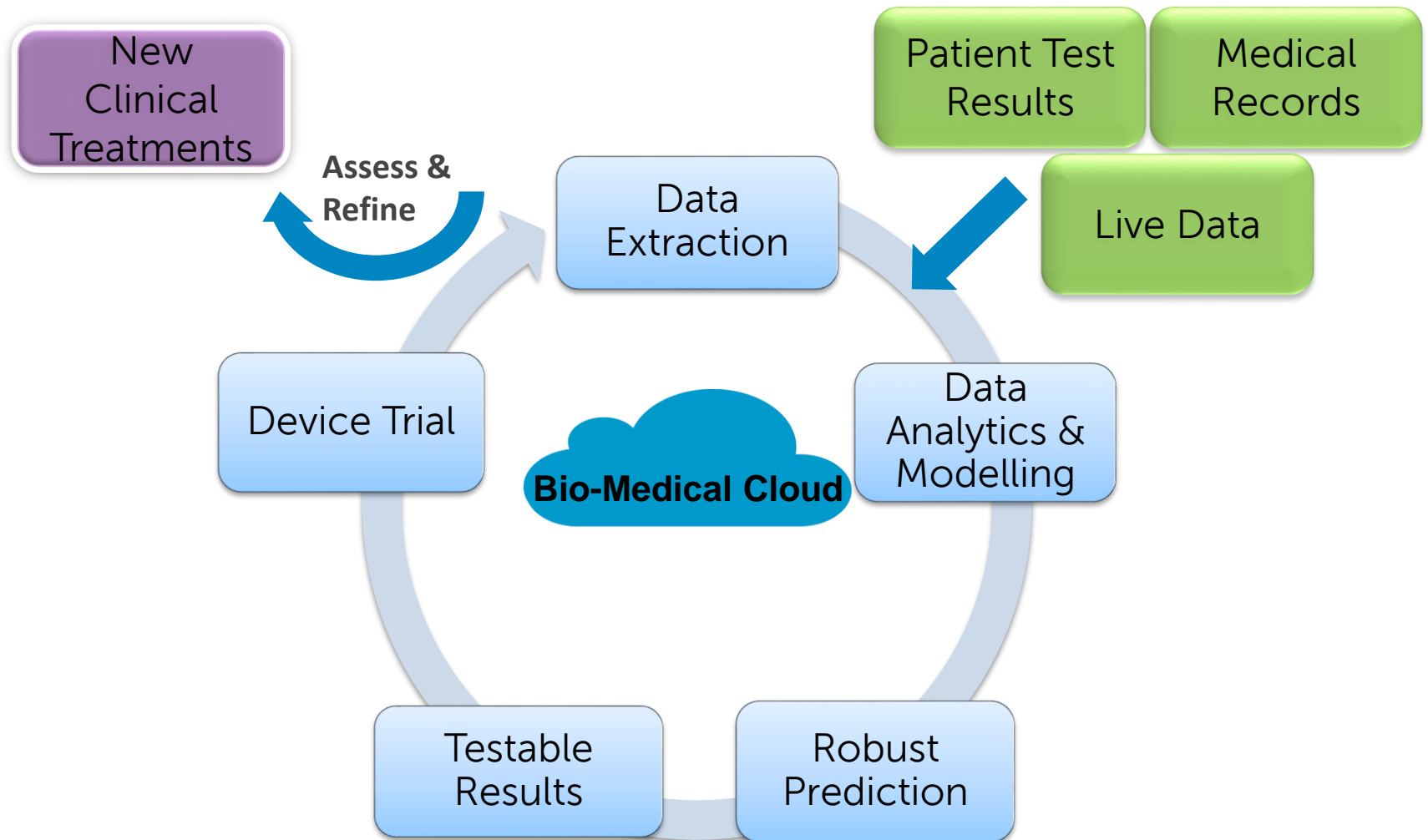
Bio-Medical-Cloud



Bio-Medical-Cloud



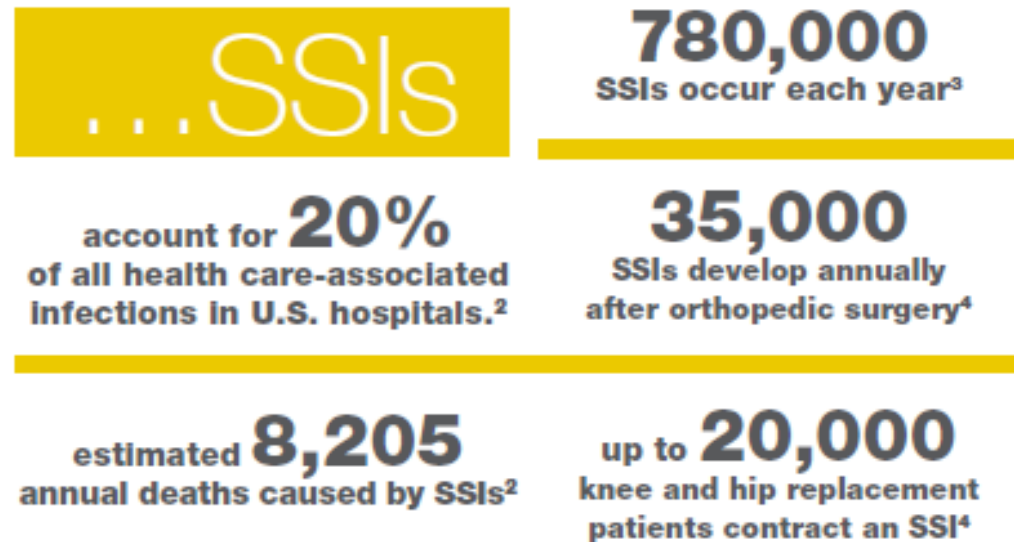
Bio-Medical-Cloud & predictive medical analytics



Surgical site infection reduction

Dr John Cromwell from Iowa University Hospital developed a new statistical model that takes patient medical records, live feeds from operating room runs a real time statistical model

Cuts surgical site infection rates by 58%



Large scale NGS sequencing & analytics

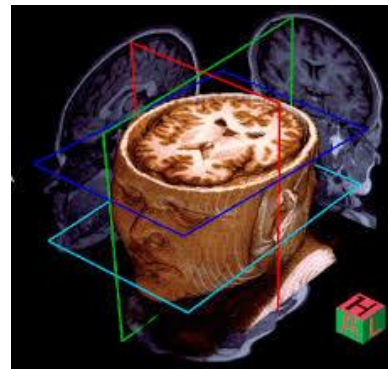
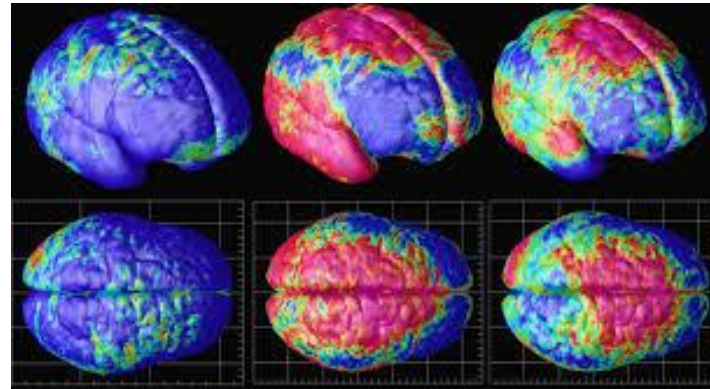
- “OpenCB” a next generation big data analytics platform for population scale genomics analysis.
- Will be used by Genomics England for the UK 100K genome study the largest study of its kind anywhere in the world
- OpenCB is already deployed on the Bio-Medical-Cloud driving the “Bridge study” to analyse the genomes of 10,000 rare disease patients **(in production)**



Medical imaging @ Wolfson Brain Imaging Centre

New state of the art brain scanning facility, needed step change in computational and data storage capability.

OpenStack image analysis VM's provide that step change



In beta test

