



eMedLab

Chairs: David Lomas and Jim Smith

Mike Barnes, Ines Barroso, Ewan Birney, Phil Beales, Mark Caulfield, Dave Hawkes,
Harry Hemingway, Martin Hibberd, Tim Hubbard, Nick Luscombe, Seb Ourselin,
Charlie Swanton, Richard Trembath

UCL Partners, Crick, Sanger, EBI, QMUL, LSHTM, KCL

Medical Bioinformatics: Data-Driven Discovery for Personalised Medicine

P.L. Beales (UCLP), M. Caulfield (UCLP), P.V. Coveney (UCLP), D. Hawkes (UCLP),
H. Hemingway (UCLP), T.J. Hubbard (Sanger), D.A. Lomas (UCLP), N.M. Luscombe (UCLP, Crick),
J.P. Overington (EBI), L. Smeeth (UCLP), J.C. Smith (Crick), C. Swanton (UCLP, Crick)

1. Objectives	6. Coordinating Analytics Research: Academy Labs
2. The Partnership	7. Strategic Issues
3. Disease Types	8. Costs
4. eMedLab Hardware Infrastructure	9. Metrics for Success
5. Research and Training Academy	

1. Objectives

Our vision is to maximise the gains for patients and for medical research that will come from the explosion in human health data. To realise this potential we need to accumulate medical and biological data on an unprecedented scale and complexity, to coordinate it, to store it safely and securely, and to make it readily available to interested researchers. It is vital to develop people with the skills and expertise to exploit these data for the benefit of patients. Together, *UCL Partners*, the *Francis Crick Institute*, *Sanger Institute* and the *European Bioinformatics Institute* shall deliver the following:

1.1 Create a powerful eMedLab E-infrastructure (lead: Smith)

We are hampered in our work to generate new medical insights because of the fragmented accessibility of fundamental clinical and research data, and the lack of a high-performance computing (HPC) facility in which to analyse them. We shall build eMedLab, a shared computer cluster to integrate and share heterogeneous data from personal healthcare records, imaging, pharmacoinformatics and genomics. Through co-location, we will eliminate the delays and security risks that occur when data are moved. It also provides a platform to develop analytical tools that allow biomedical researchers to transform raw data into scientific insights and clinical outcomes. eMedLab will store data securely and its modular

“We will maximise the gains for patients and for medical research that will come from the explosion in human health data”

How do we achieve this?



eMedLab infrastructure

- Shared computer cluster
- Integrate exchange heterogeneous data
- Methods and insights across diseases



eMedLab academy

- Recruit outstanding Career Dev Fellows
- Training workshops and courses
- Promote collaborations via “Labs”

Research operations

Medical Bioinformatics: Data-Driven Discovery for Personalised Medicine

P.L. Beales (UCLP), M. Caulfield (UCLP), P.V. Coveney (UCLP), D. Hawkes (UCLP),
H. Hemingway (UCLP), T.J. Hubbard (Sanger), D.A. Lomas (UCLP), N.M. Luscombe (UCLP, Crick),
J.P. Overington (EBI), L. Smeeth (UCLP), J.C. Smith (Crick), C. Swanton (UCLP, Crick)

1. Objectives	6. Coordinating Analytics Research: Academy Labs
2. The Partnership	7. Strategic Issues
3. Disease Types	8. Costs
4. eMedLab Hardware Infrastructure	9. Metrics for Success
5. Research and Training Academy	

1. Objectives
Our vision is to maximise the gains for patients and for medical research that will come from the explosion in human health data. To realise this potential we need to accumulate medical and biological data on an unprecedented scale and complexity, to coordinate it, to store it safely and securely, and to make it readily available to interested researchers. It is vital to develop people with the skills and expertise to exploit these data for the benefit of patients. Together, UCL, Partners, the Francis Crick Institute, Sanger Institute and the European Bioinformatics Institute shall deliver the following:

1.1 Create a powerful eMedLab E-infrastructure (lead: Smith)
We are hampered in our work to generate new medical insights because of the fragmented accessibility of fundamental clinical and research data, and the lack of a high performance computing (HPC) facility in which to analyse them. We shall build eMedLab, a shared computer cluster to integrate and share heterogeneous data from personal healthcare records, imaging, pharmacogenomics and genomics. Through co-location, we will eliminate the delays and security risks that occur when data are moved. It also provides a platform to develop analytical tools that allow biomedical researchers to transform raw data into scientific insights and clinical outcomes. eMedLab will store data securely, and its modular

“We will maximise the gains for patients and for medical research that will come from the explosion in human health data”

How do we achieve this?

eMedLab enables projects through infrastructure and **people**

- Integrate exchange heterogeneous data
- Methods and insights across diseases

eMedLab academy

- Recruit outstanding Career Dev Fellows
- Training workshops and courses
- Promote collaborations via “Labs”

Support team

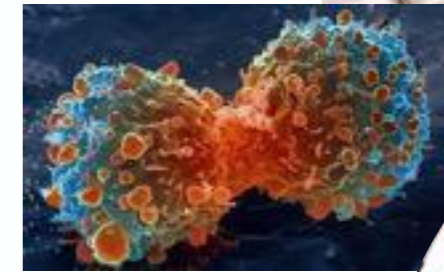
eMedLab

is a hub



3 disease areas

>6M patients
rare



cancer



cardio

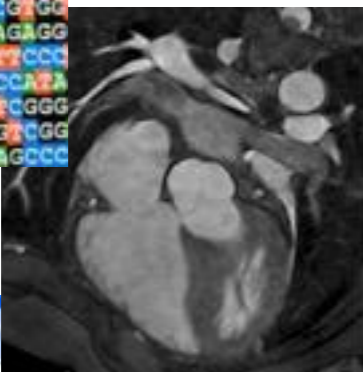
7 partners



3 data types



genomic



images



electronic
health
records

clinician
scientists



analytics

basic
science



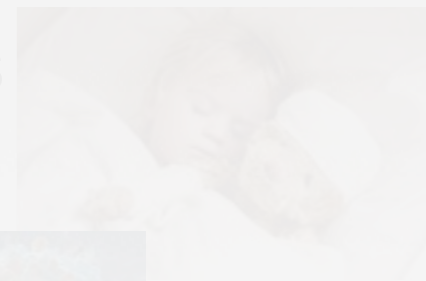
3 expertises

eMedLab

is a hub

3 disease areas

>6M patients
rare



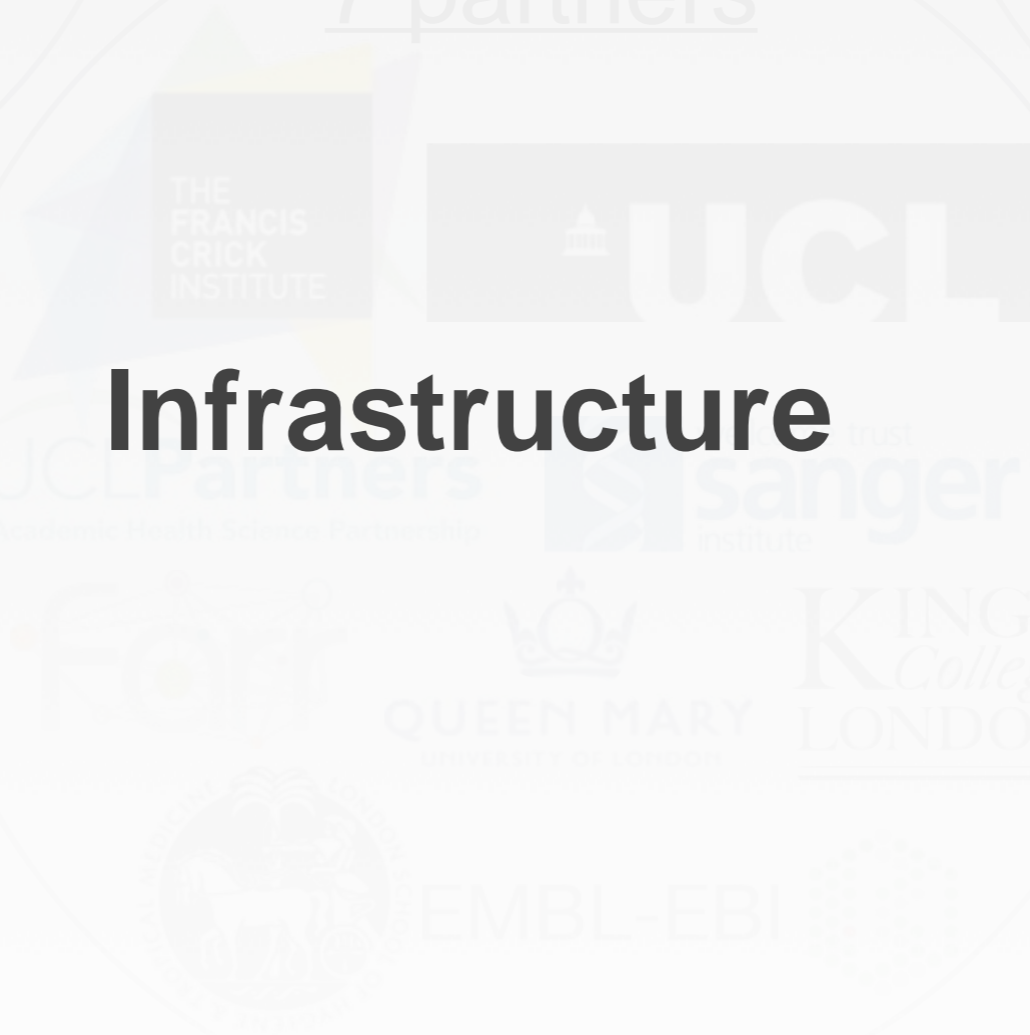
cancer



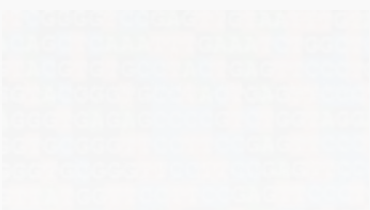
cardio

7 partners

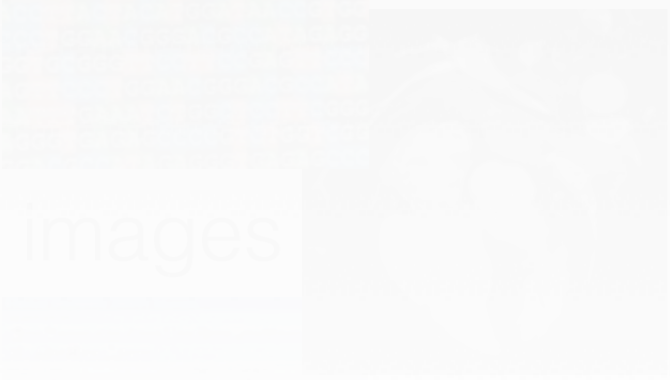
Infrastructure



3 data types



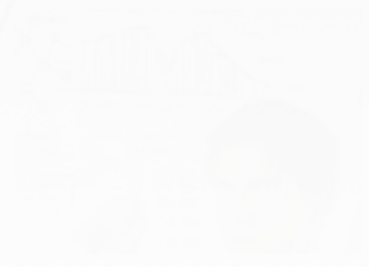
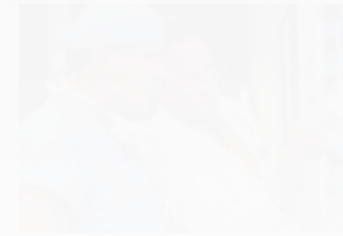
genomic



images

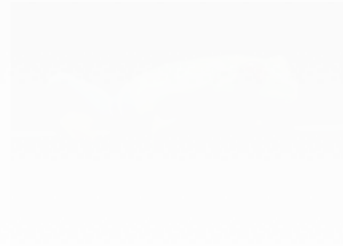
electronic
health
records

clinician
scientists



analytics

basic
science



3 approaches

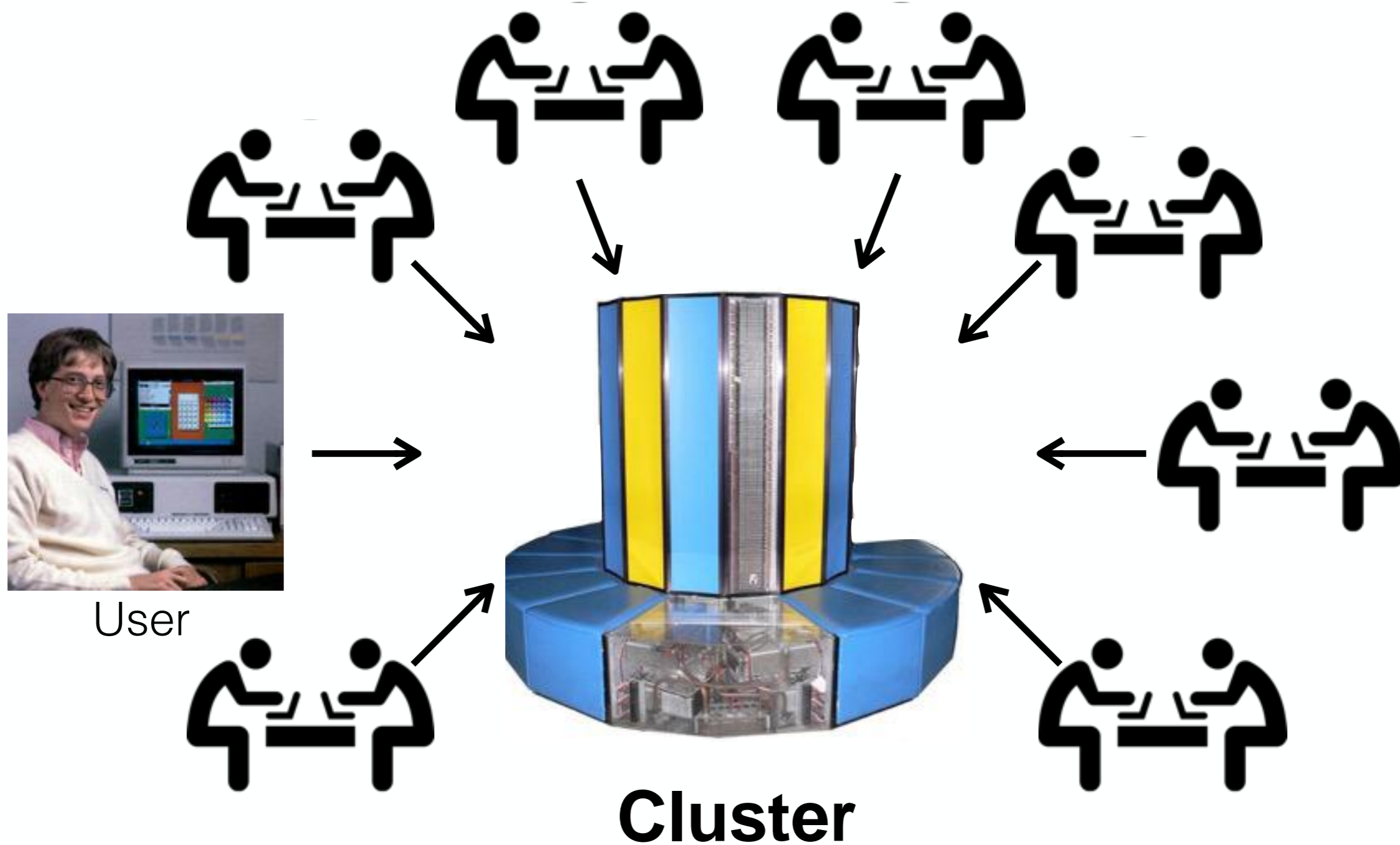
eMedLab infrastructure

Private cloud computing
for medical bioinformatics

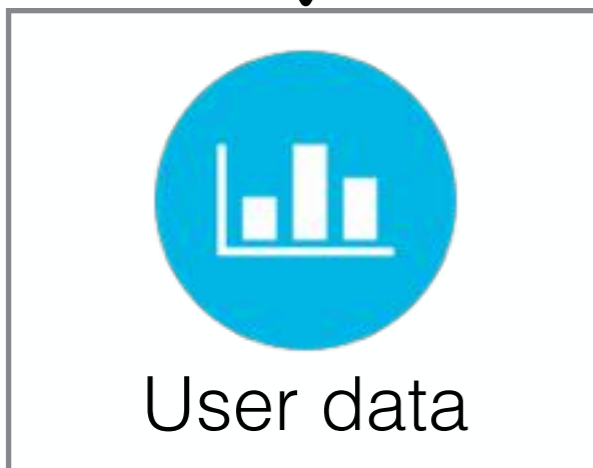
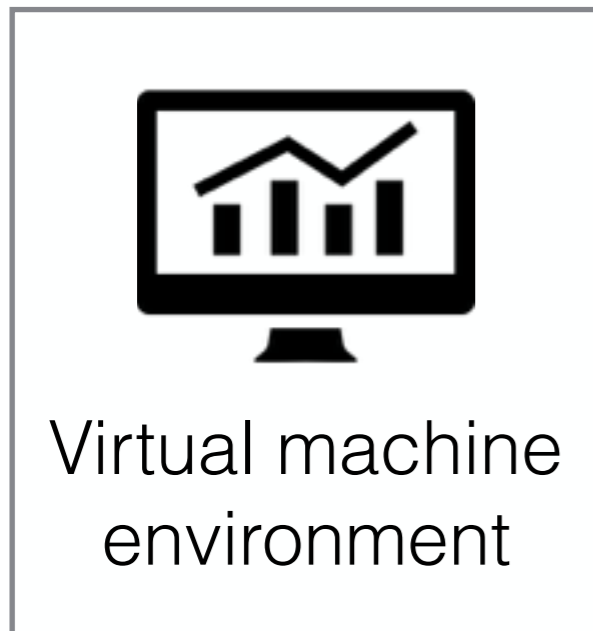
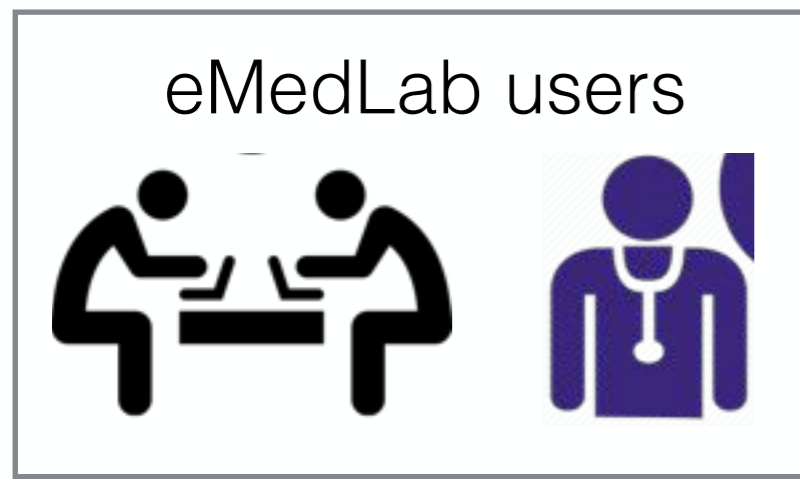
- **One of the largest high-spec private cloud computing systems**
- Very high specifications
 - 6048 cores
 - 5.5PB data storage
 - Fast interconnects
- Flexible, transferrable, accessible



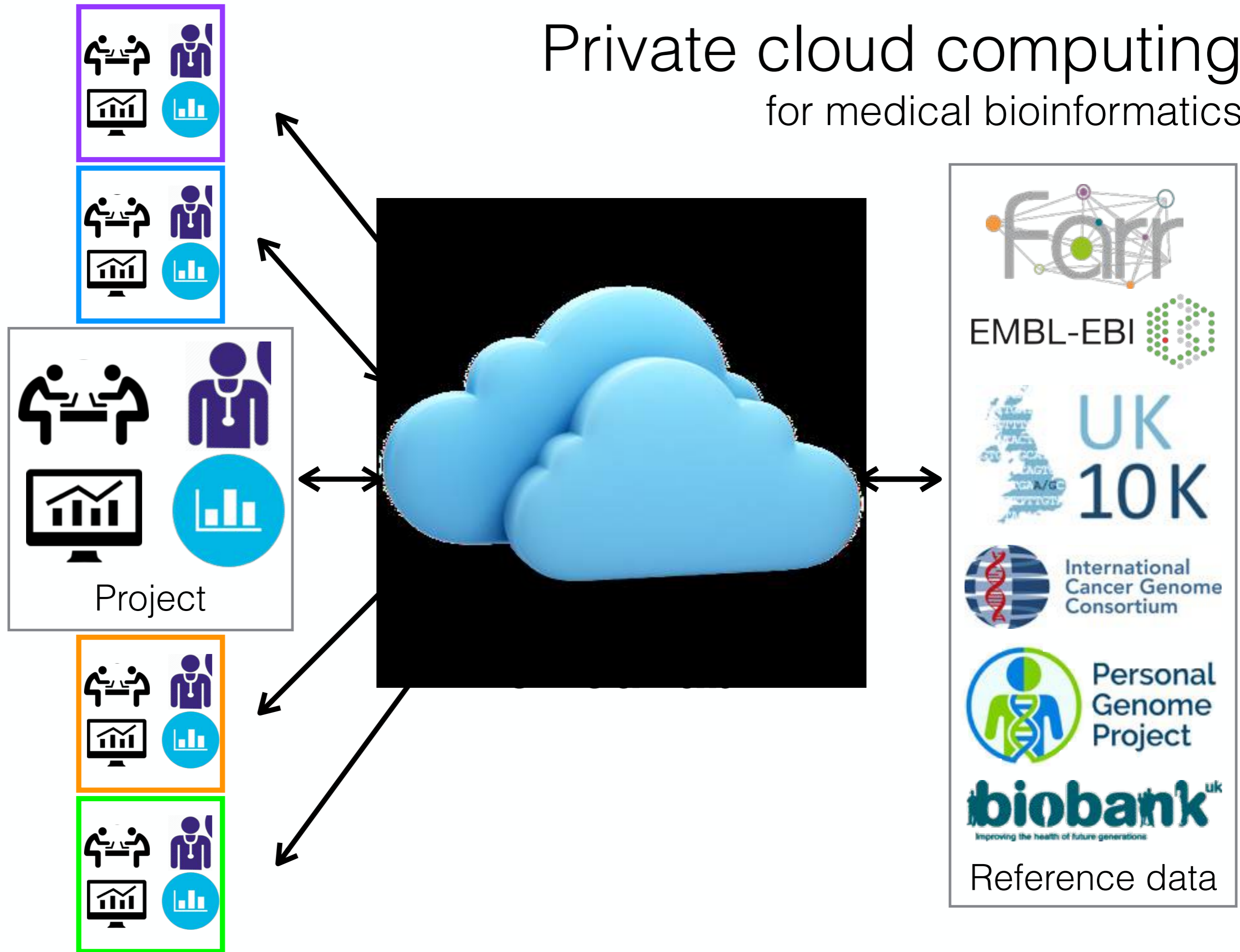
Traditional HPC cluster

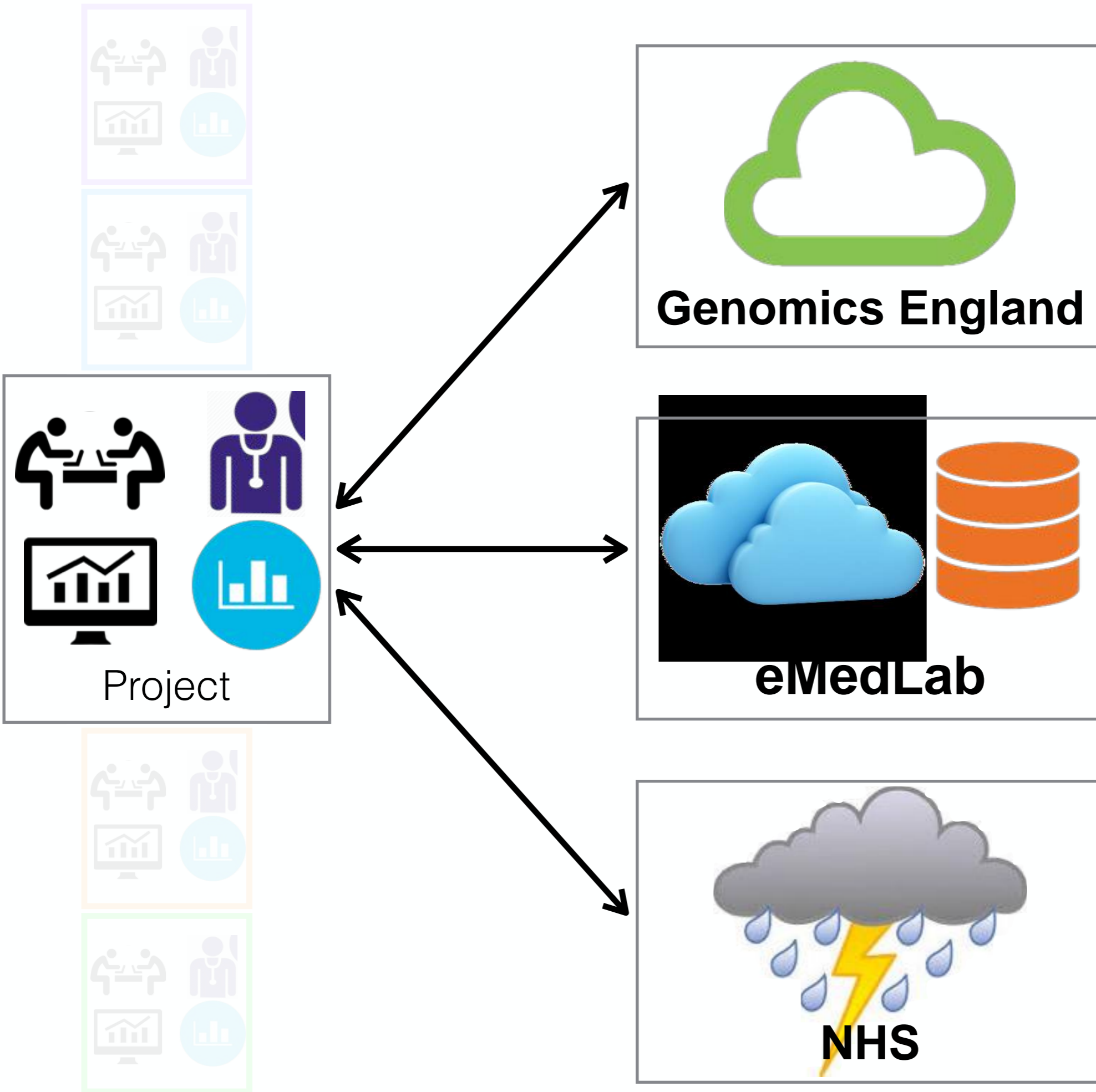


Private cloud computing for medical bioinformatics

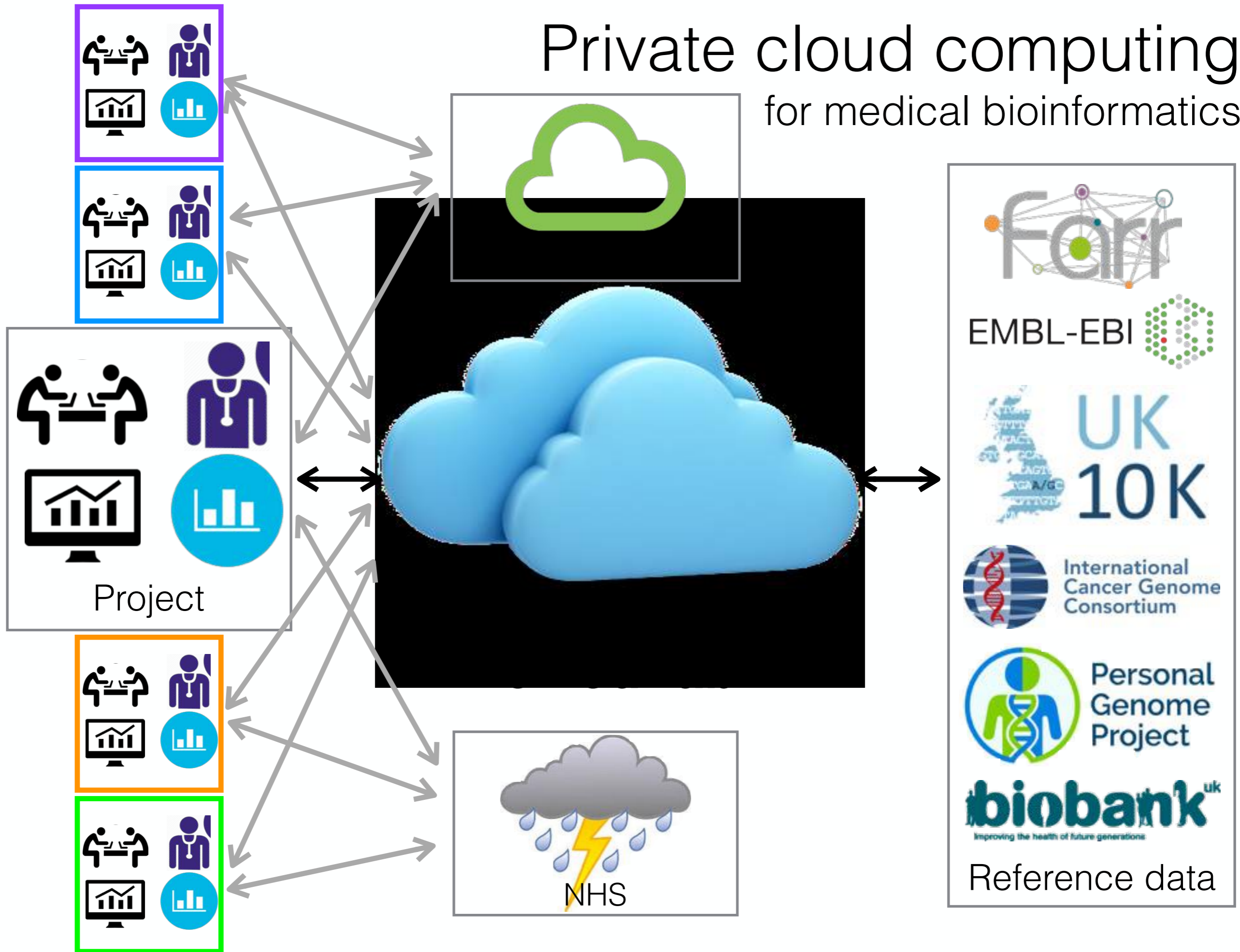


Private cloud computing for medical bioinformatics





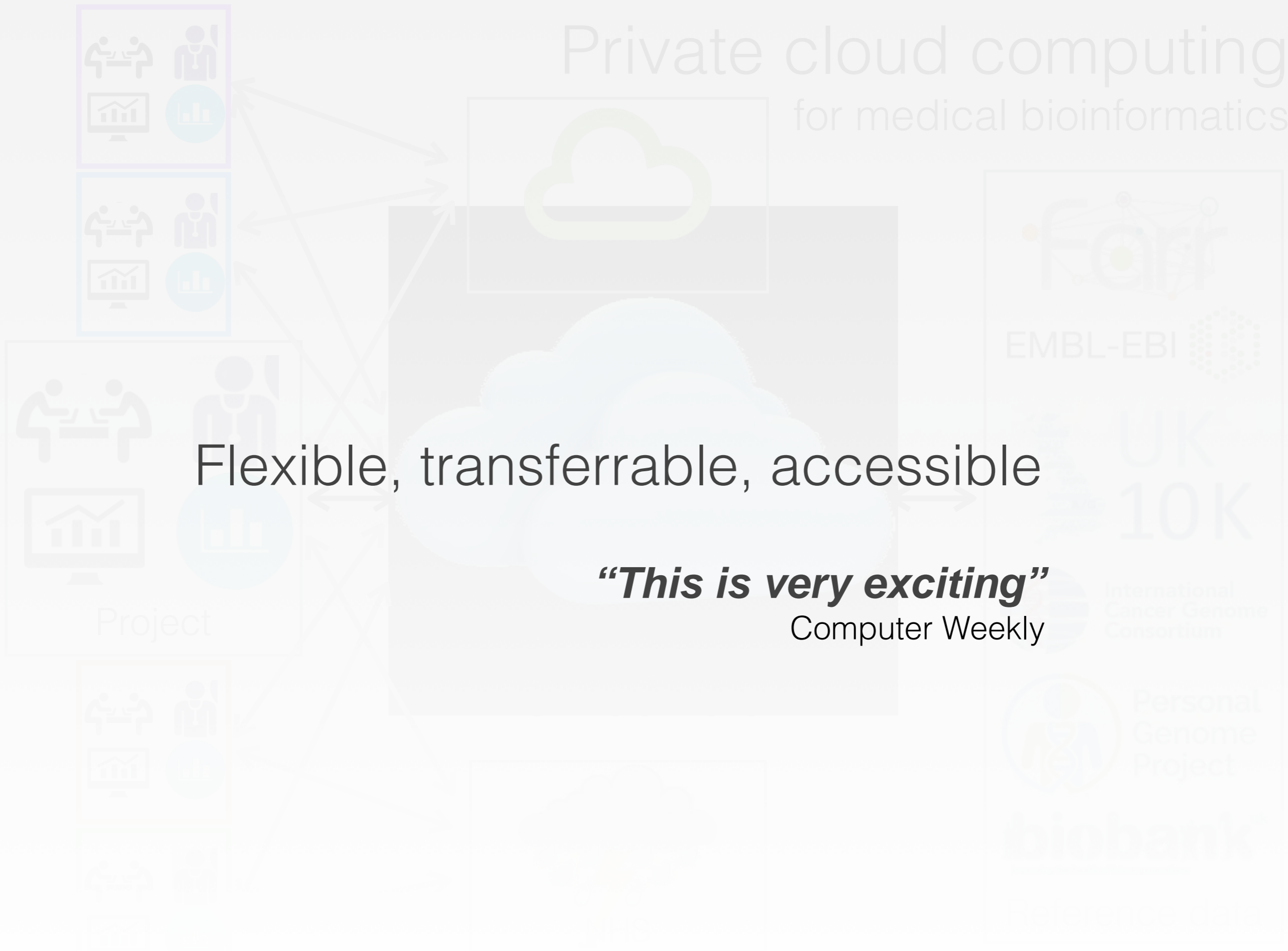
Private cloud computing for medical bioinformatics



Private cloud computing for medical bioinformatics

Flexible, transferrable, accessible

“This is very exciting”
Computer Weekly

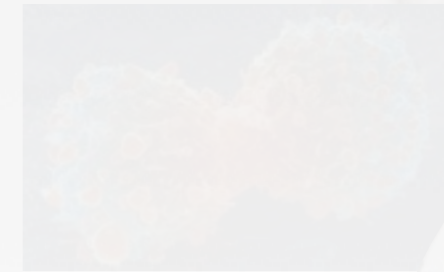
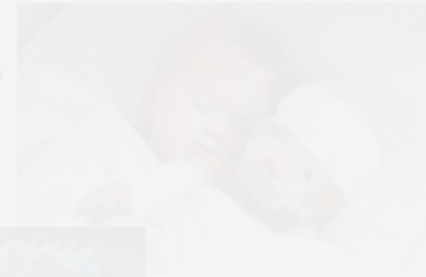


eMedLab

is a hub

3 disease areas

>6M patients
rare



cancer



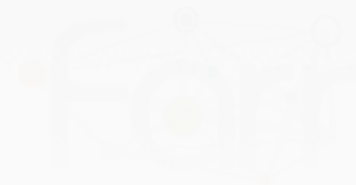
cardio

7 partners

THE FRANCIS CRICK INSTITUTE



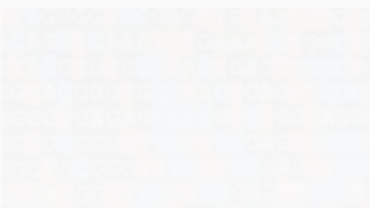
UCL Partners
Academic Health Science Partnership



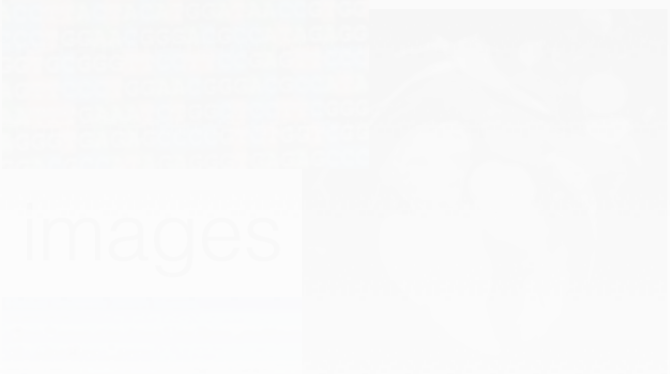
EMBL-EBI



3 data types



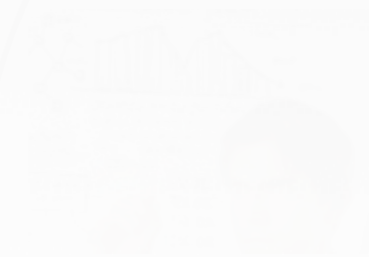
genomic



images

electronic
health records

clinician
scientists



analytics

basic
science



3 approaches



eMedLab data

Principles

- **Anonymised data:**
 - eMedLab can host anonymised clinical data subject to local ethics approval
- **Pseudoanonymised data:**
 - eMedLab can host pseudoanonymised data from patients who have given consent for research

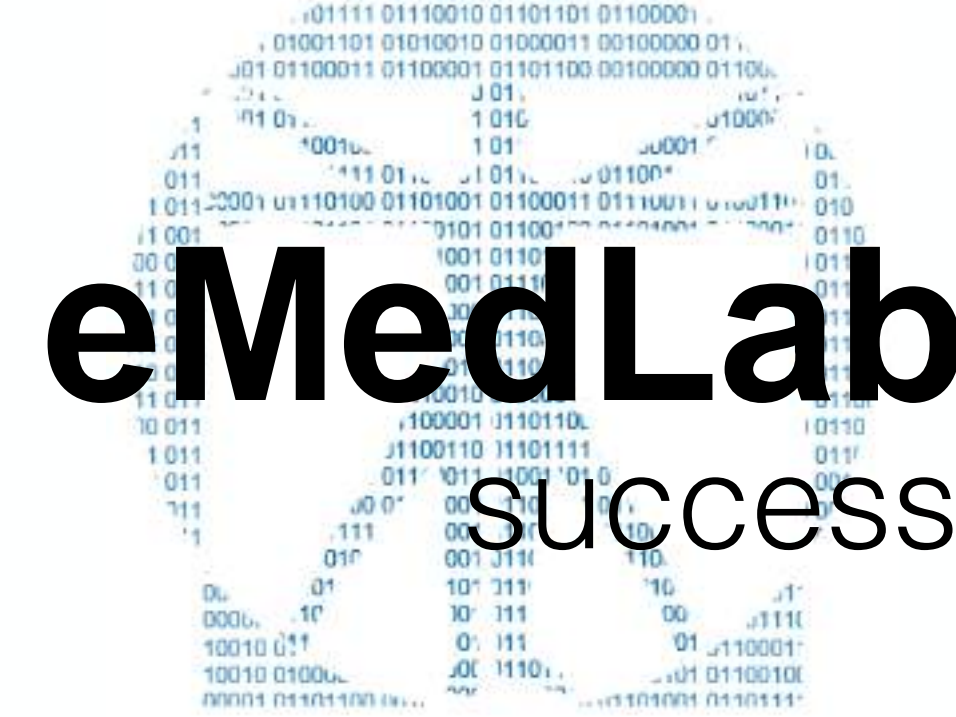


eMedLab

data

Principles

- **Reference data:**
 - Suitable for data with high levels of commonality across projects
 - Users are responsible for approvals to access
 - Access managed through VM configurations
- **Specific data:**
 - Uploaded to eMedLab via project-specific VMs
 - Extend access to other projects through VM cross-talk
- **Data manager:**
 - Full-time manager for curation and access of reference data



eMedLab

Success

Challenges

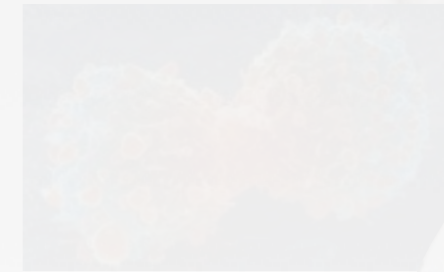
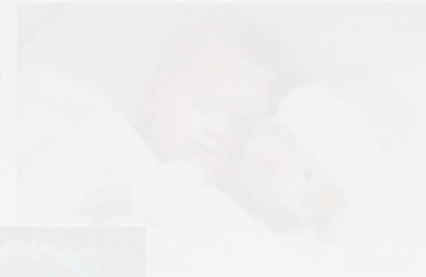
- **Linking with identifiable data:**
 - eMedLab will not host identifiable data
 - Technical solutions:
 - Virtual Machines go to the data
 - Link remotely with identifiable data in Farr Safe Haven and others
 - A Safe Haven Virtual Machine that hard-codes ISO 27001 requirements?
- **Disparate data models:**
 - Harmonise data models across centres
 - Global Alliance for Genomics & Health (<https://genomicsandhealth.org/>)
 - Observational Medical Outcomes Partnership (<http://omop.org/>)

eMedLab

is a hub

3 disease areas

>6M patients
rare

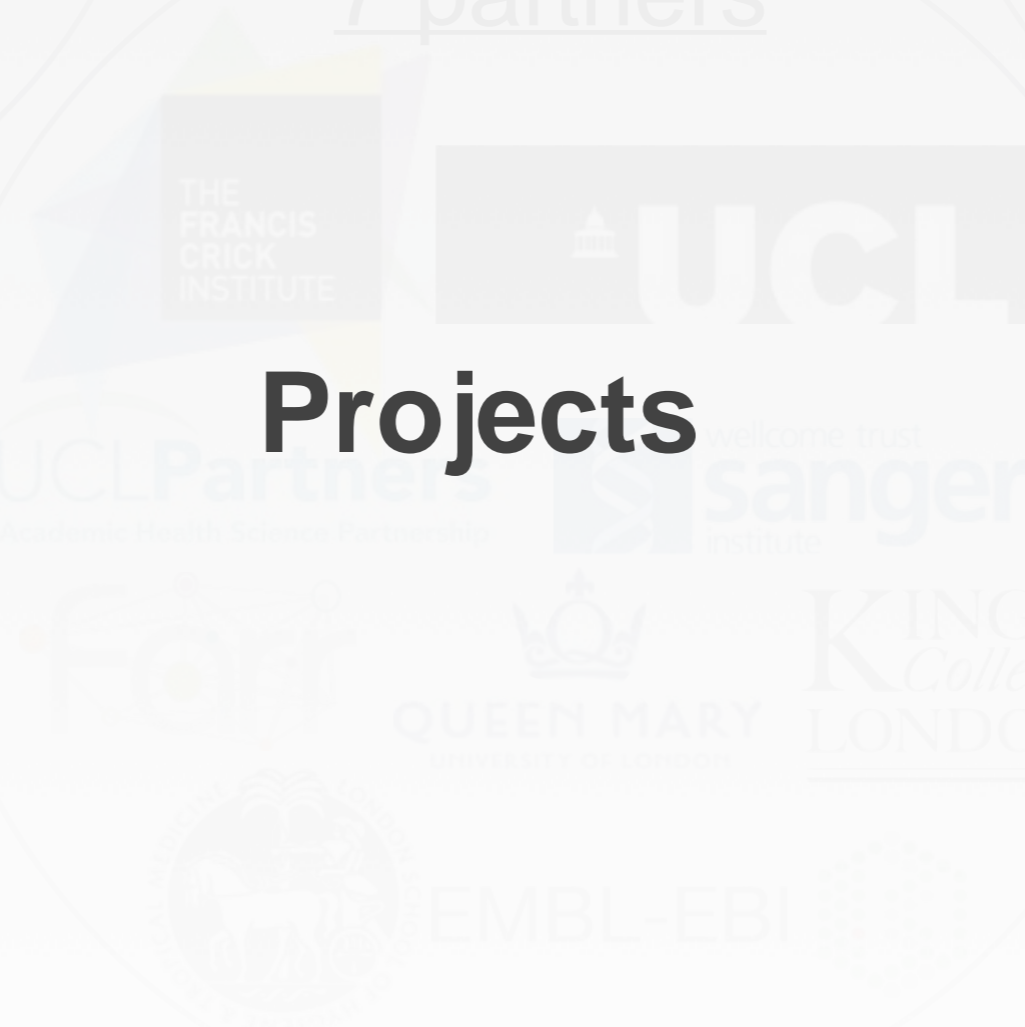


cancer



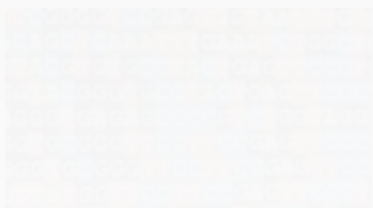
cardio

7 partners

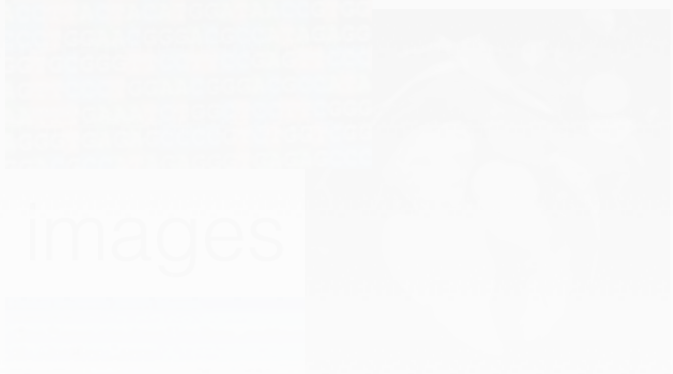


Projects

3 data types



genomic



images

electronic
health
records

clinician
scientists



analytics

basic
science



3 approaches

eMedLab

enables
projects

StratMed platform
MRC Matura Consortium

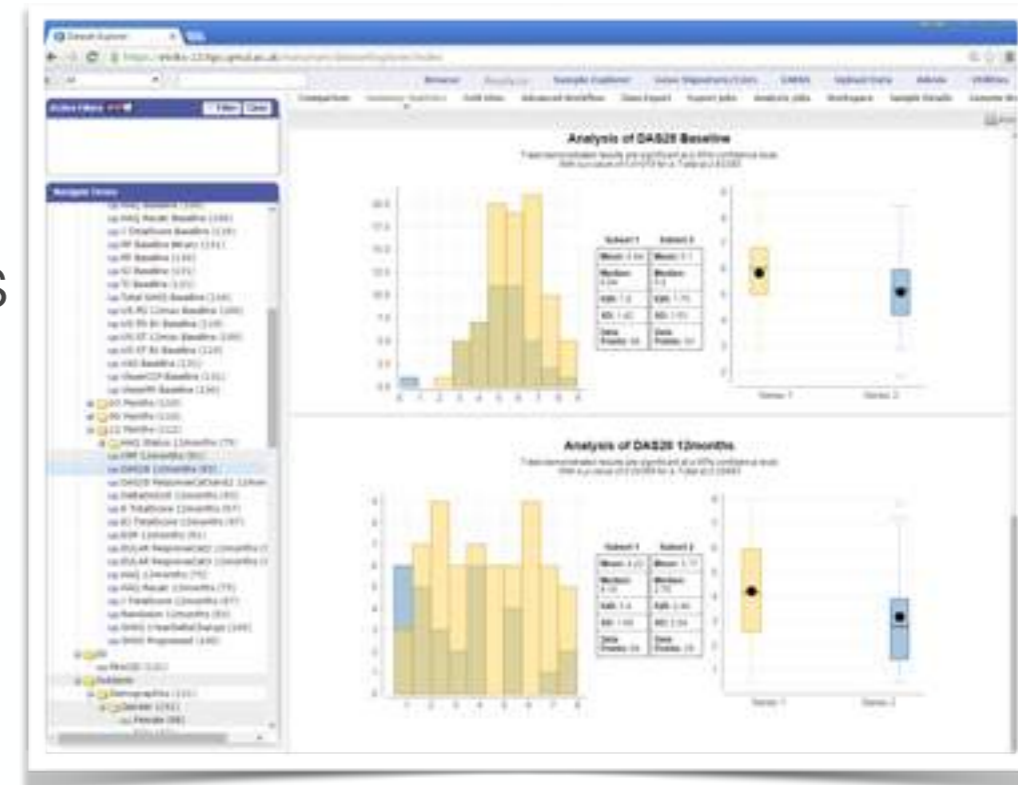


- Rheumatoid arthritis doesn't always respond to biologics
- **Data:** -omic and clinical trial data for 10,000 patients across 21 partners
- **How:** Integrate and present information to clinician researchers
- **Who:** clinicians, bioinformaticians, pharma

- **Outcomes**

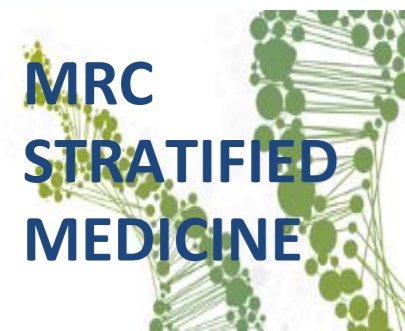
Biomarkers to predict responses to biologics

Effective treatment programmes for patients



StratMed platform

Joined up MRC Infrastructure



RA-Map, MATURA & PSORT adopt a federated TranSMART/i2b2 infrastructure

1

Strat Med

2

Farr London

Hosting and prototyping enabled by Farr-funded Virtual Machine (VM) cluster at QMUL



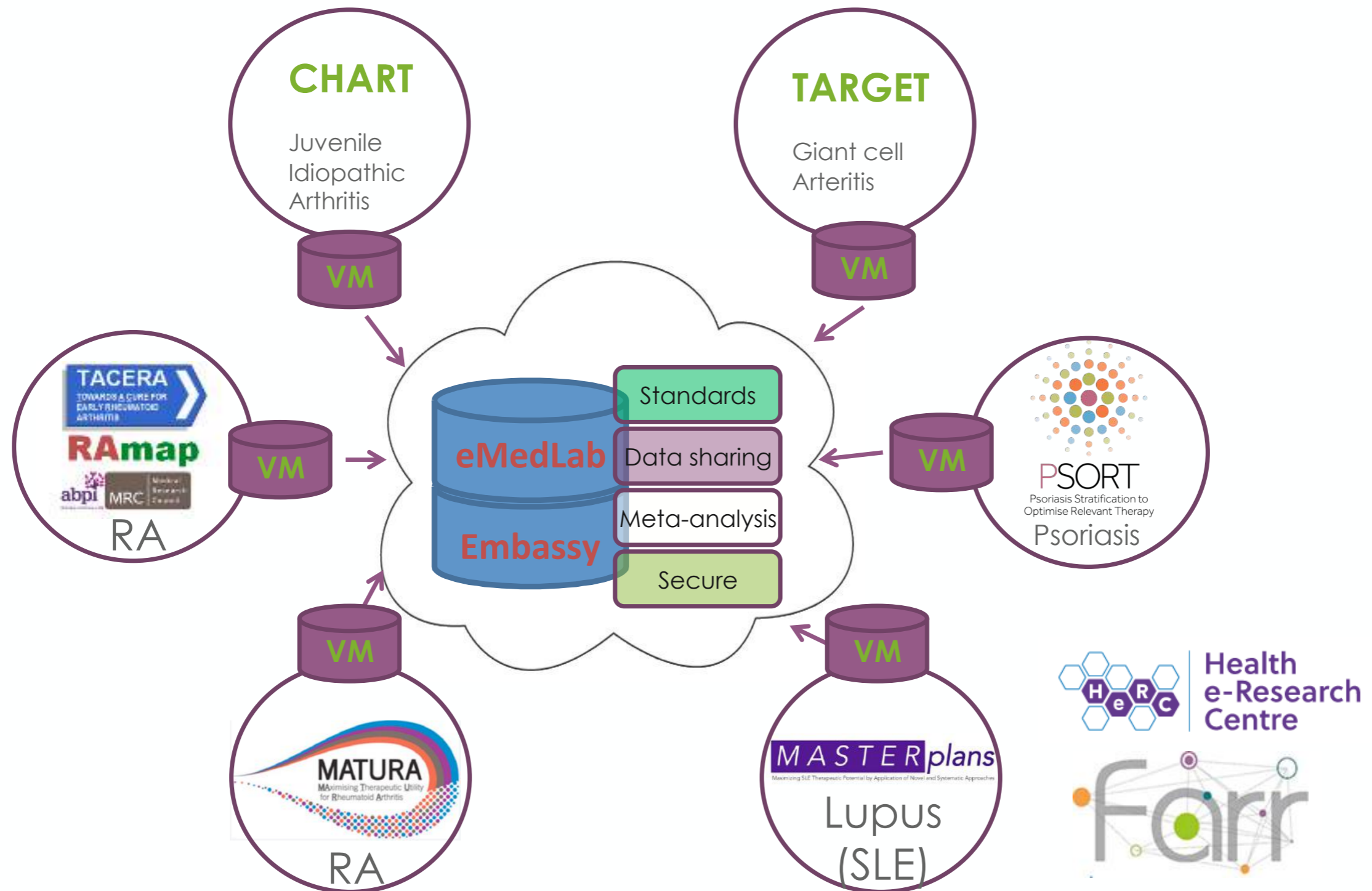
Medical Bioinformatics Awards
MRC eMEDLAB

Stratified medicine VMs Transferred to eMedLab to enable HPC and Machine Learning applications

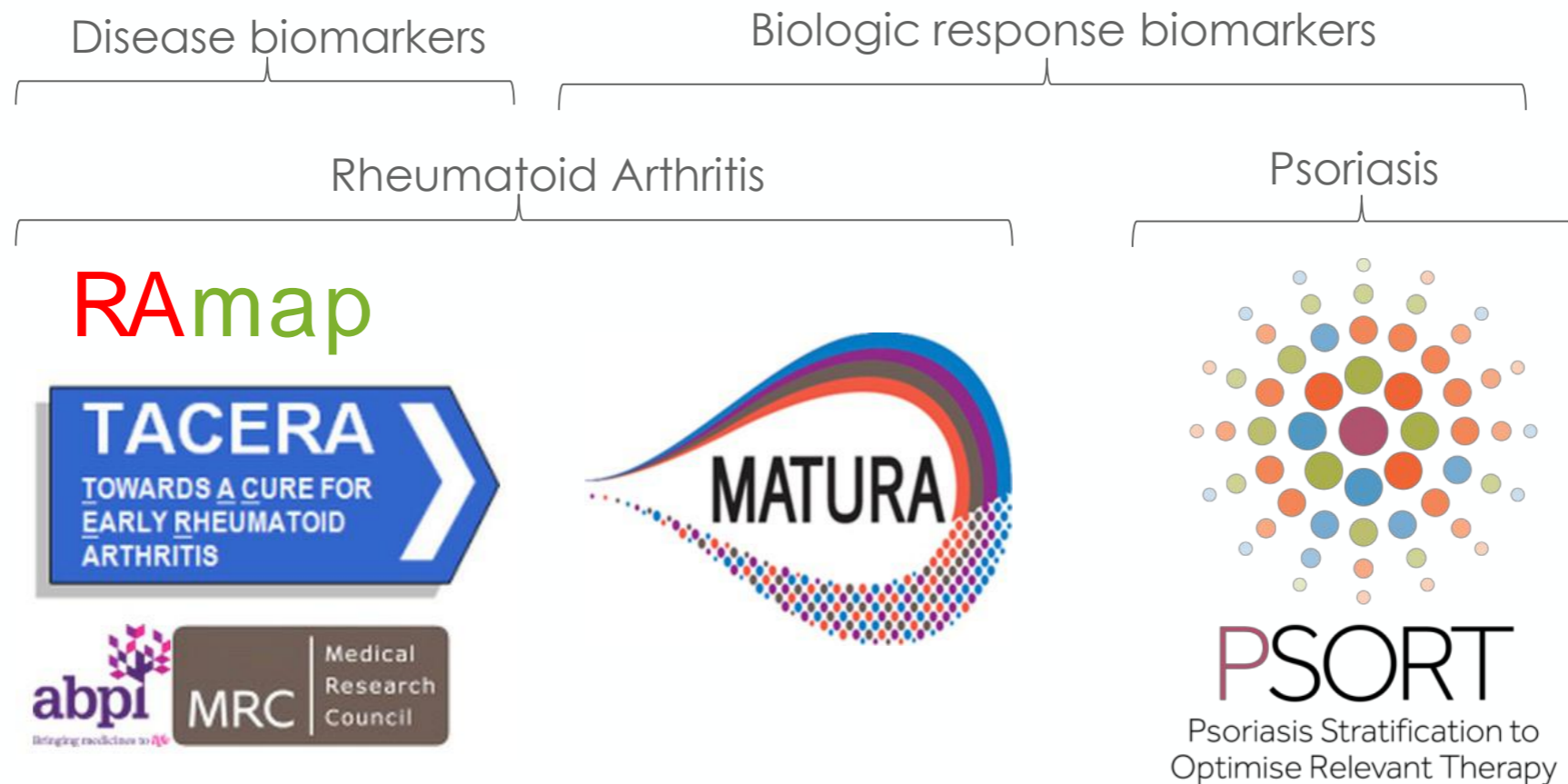
3

eMedLab

eMedLab is building synergy and facilitating data sharing



eMedLab is building synergy and facilitating data sharing



3
Projects

MANY PARTNERS

Academic



Informatics



Pharma



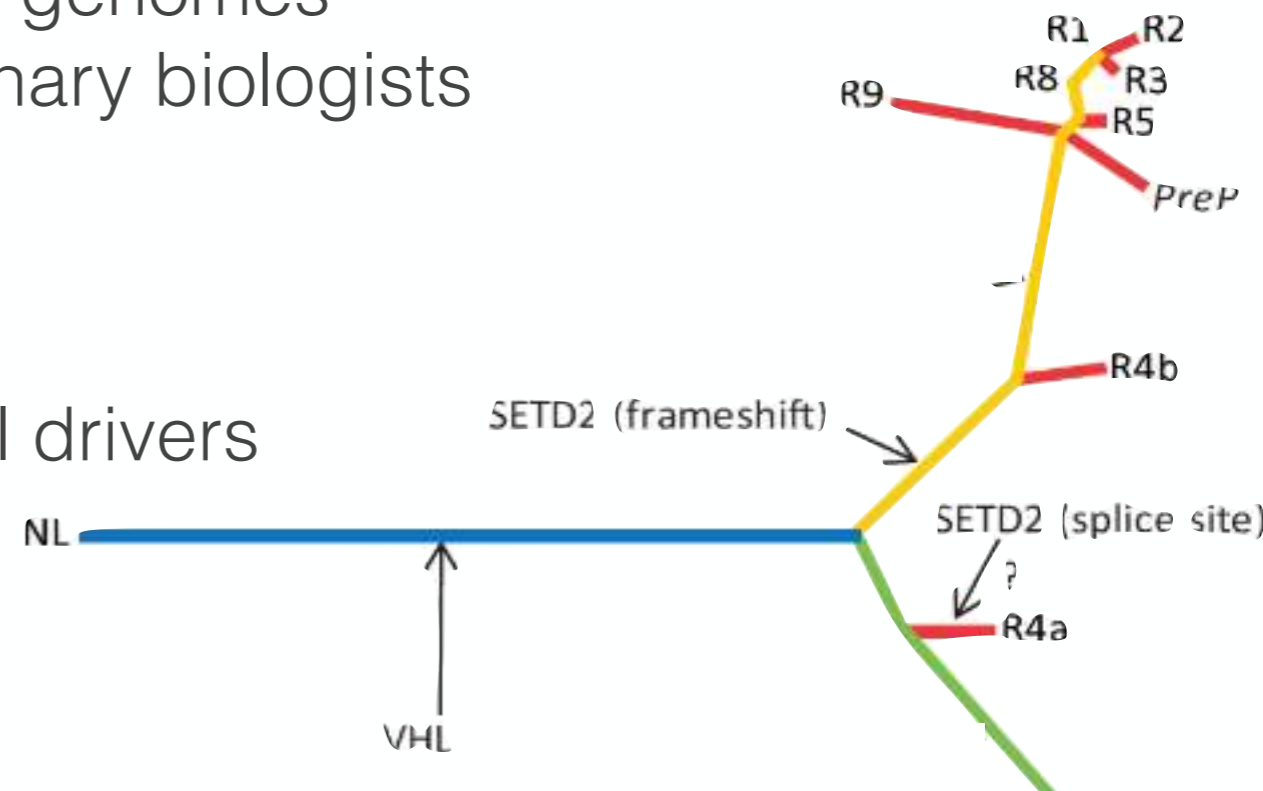
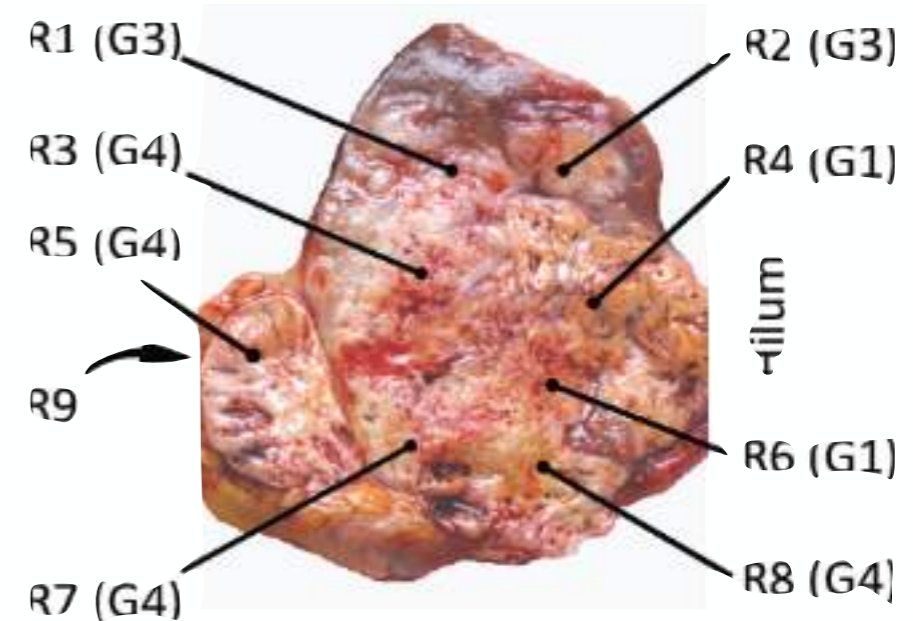
eMedLab

enables projects

- Cancers evolve heterogeneously
- **Data:** genomes, MRI, histopathology
- **How:** Phylogenetic analysis of cancer genomes
- **Who:** clinicians, statisticians, evolutionary biologists
- **Outcomes**
Mechanisms of cancer progression
DARWIN II clinical trial to target clonal drivers

Cancer genomics

Swanton & Van Loo

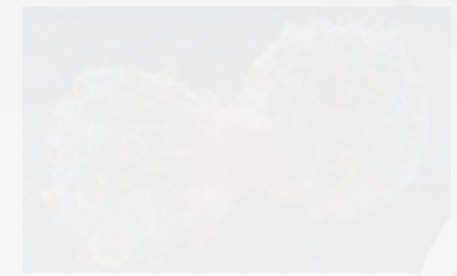
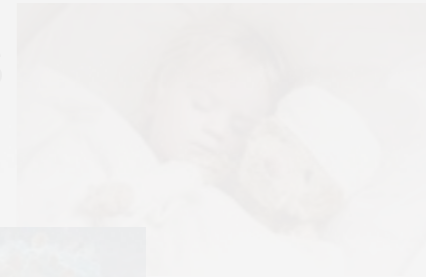


eMedLab

is a hub

3 disease areas

>6M patients
rare

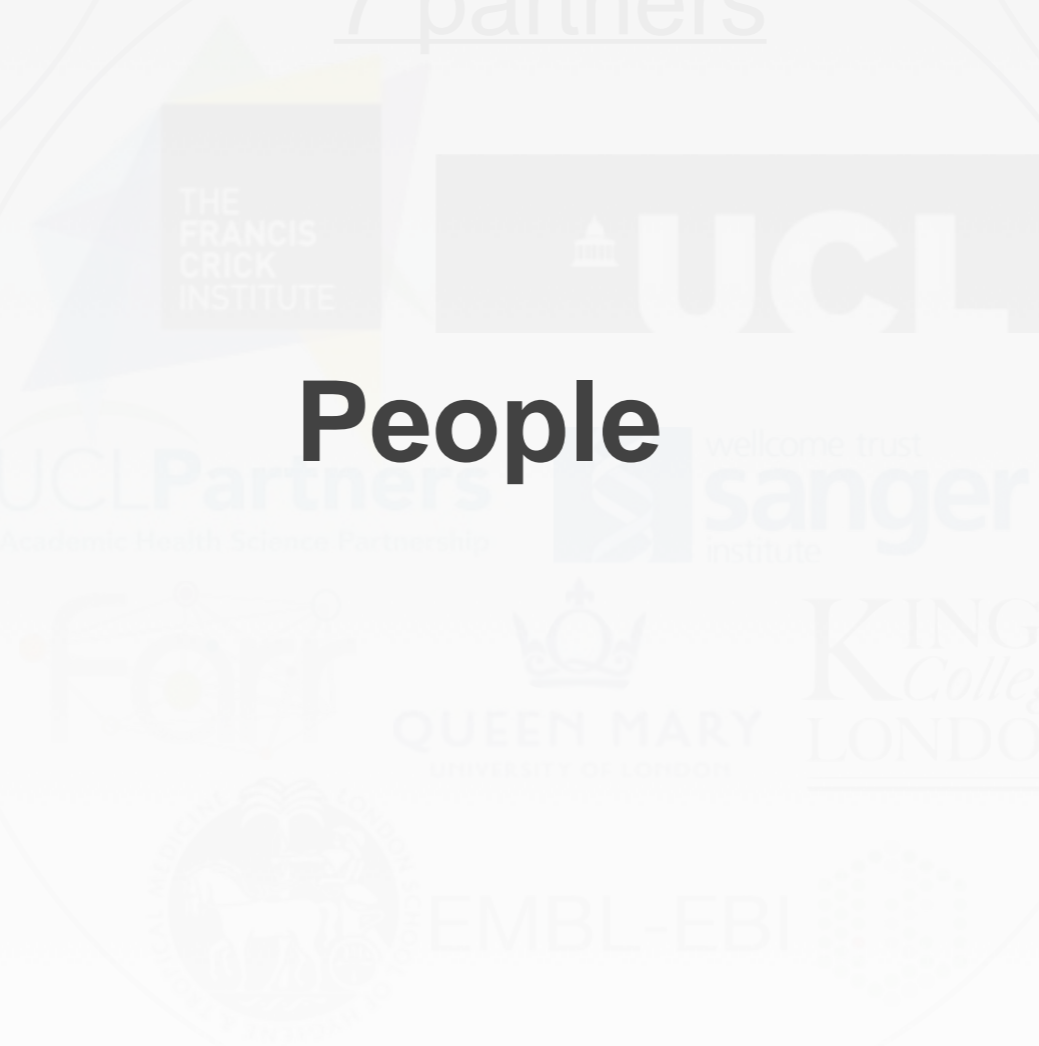


cancer



cardio

7 partners

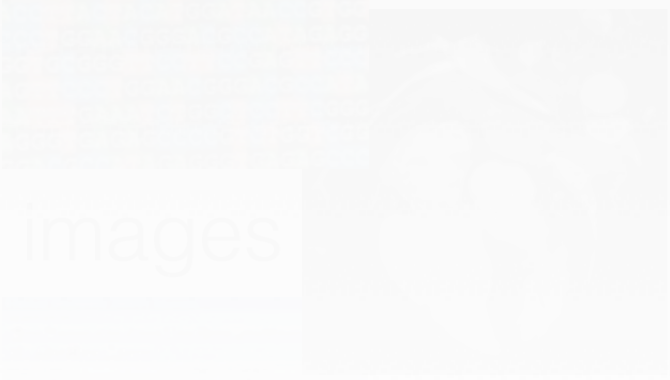


People

3 data types



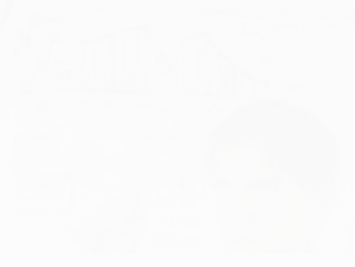
genomic



images

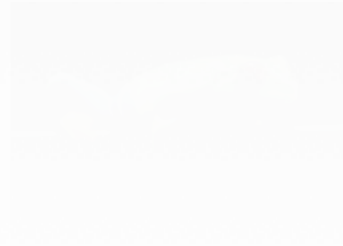
electronic
health
records

clinician
scientists



analytics

basic
science



3 approaches

eMedLab people

career development fellows have
been recruited internationally



Andre Altman

- Neurodegenerative diseases
- Stanford to UCL
- Altmann, Science 2015



Helena Kilpinen

- Rare diseases and HiPSCI
- EBI to GOSH/Sanger
- Kilpinen, Science 2013



Alan Hodgkinson

- Mitochondrial diseases
- McGill to KCL
- Hodgkinson, Science 2014



Borbala Mifsud

- Non-coding mutations
- Crick to QMUL
- Mifsud, Nat Genet 2015

eMedLab people

Research operations underpinning eMedLab



Bruno Silva

Kuba Purebski	Richard Christie	
Tom King	David Fergusson	
David Ocana	John Bouquiere	
David Wong	Mike Atkins	
Rich Boyce	Andy Cafferkey	
Ric Passey	Faruque Sarker	Jackie Stewart
Luke Sudbury	Tom Jones	Gianni Dalla Torre
Pete Clapham	James Beale	Steve Whitbread
Adam Huffman	Tim Cutts	Luke Raimback
Stefan Boeing	Josh Randall	William Hay

- Federated support across partners
 - procurement
 - core infrastructure
 - data management
 - virtual machines...

eMedLab

people

UK technical network
for national infrastructure

- Future proofing technologies
 - team coordination
 - compatibility
 - secure data transfer
 - regulated data access...

Crick, EMBL-EBI, Sanger,
Farr, Medical
Bioinformatics, Genomics
England, ADRC, JISC



Jacky Pallas

Simon Thompson
Tim Hubbard
Jeremy Sharp
Alan Real
John Ainsworth
Mark Parsons

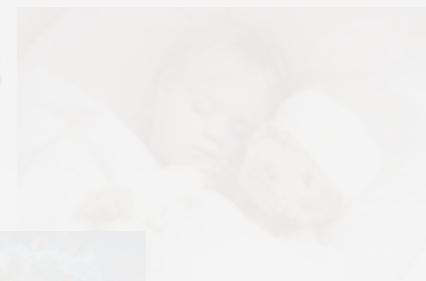
Malcolm Teague
Steve Pavis
Steve Newhouse
Tim Cutts
David Fergusson

eMedLab

is a hub

3 disease areas

>6M patients
rare

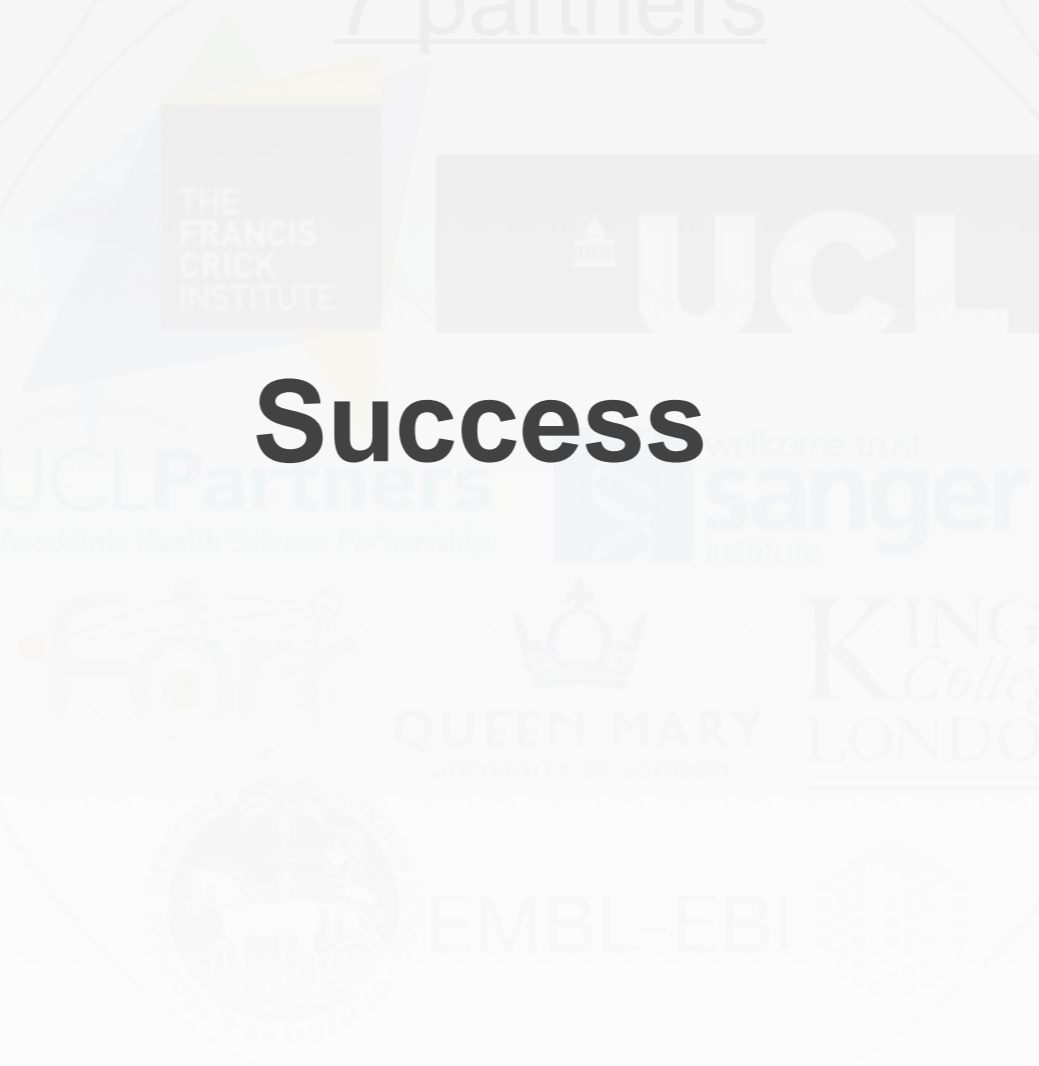


cancer



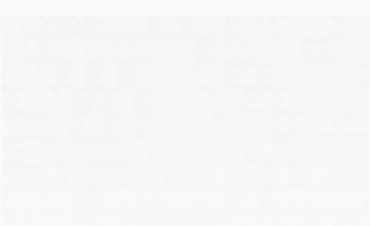
cardio

7 partners

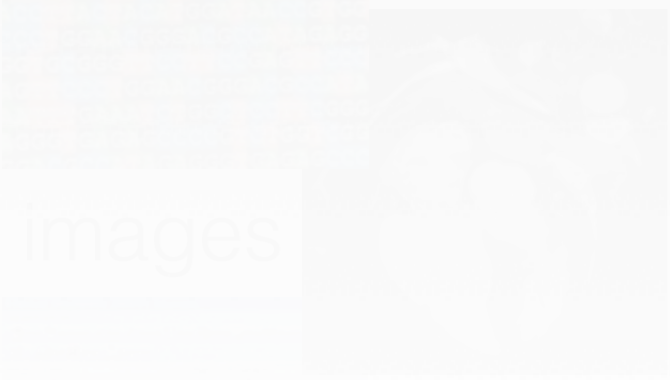


Success

3 data types



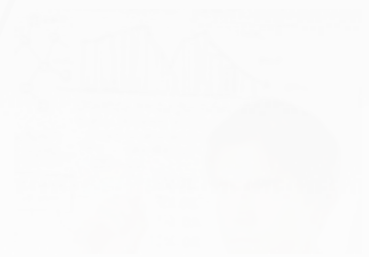
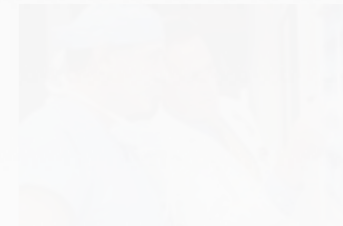
genomic



images

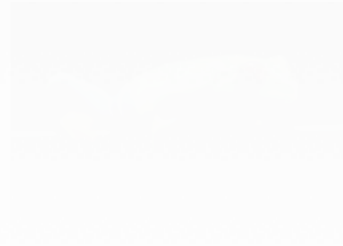
electronic
health
records

clinician
scientists

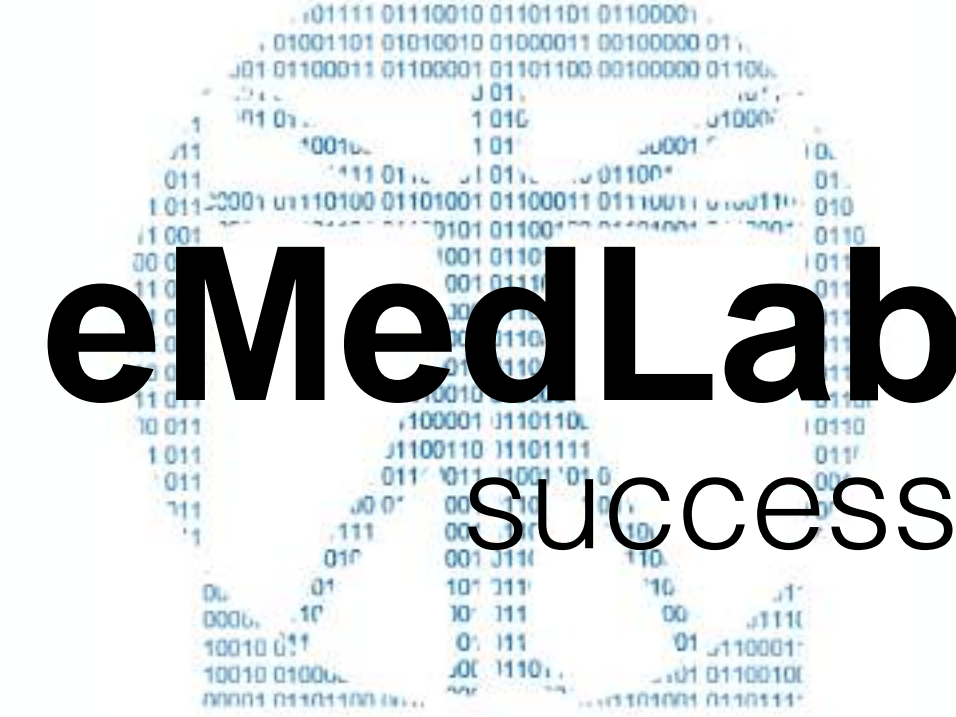


analytics

basic
science



3 approaches



eMedLab

success

we've made progress
but it's not enough!

- **Sustainable infrastructure**: enable data intensive clinical research
- **Capacity**: network of scientists, clinicians and research operatives
- **Cross disease boundaries**: exploit commonalities in data analysis
- **Strategic alliances**: **Farr UK**, Genomics England, Turing etc...

Clinical outcomes: scientific & clinical success
through genomic, imaging and EHR data



eMedLab

Chairs: David Lomas and Jim Smith

Mike Barnes, Ines Barroso, Ewan Birney, Phil Beales, Mark Caulfield, Dave Hawkes, Harry Hemingway, Martin Hibberd, Tim Hubbard, Nick Luscombe, Seb Ourselin, Charlie Swanton, Richard Trembath

UCL Partners, Crick, Sanger, EBI, QMUL, LSHTM, KCL

- **Contact:** exec@emedlab.ac.uk