



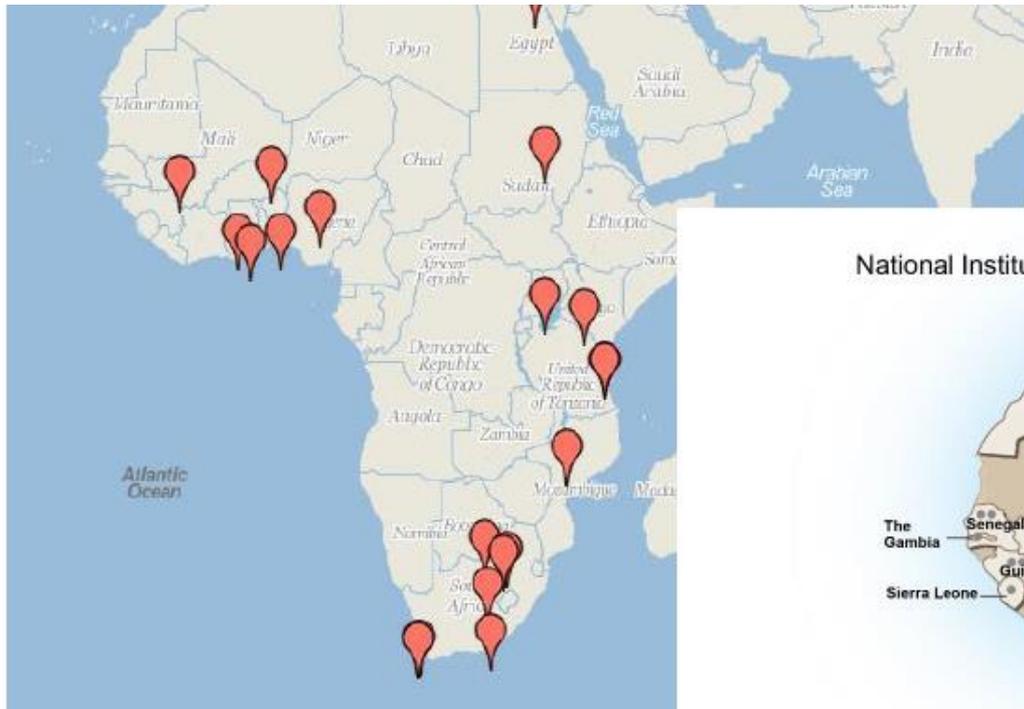
H3ABioNet

Pan African Bioinformatics Network for H3Africa

H3Africa Archive and EGA submission



Background - Project sites



National Institutes of Health - Wellcome Trust H3Africa Research Network





Background

- The NIH and Wellcome Trust have funded multiple projects on genomic research
- Data is required to be deposited in public domain – selected EGA
- European Genome-phenome Archive: <https://www.ebi.ac.uk/ega/> is repository for sequence and genotype data
- Data access is controlled by access committees



EGA -



International
Cancer Genome
Consortium



CANCER
RESEARCH
UK

ICR The Institute of
Cancer Research

Wellcome Trust Case Control Consortium





What is EGA controlled access data anyway?

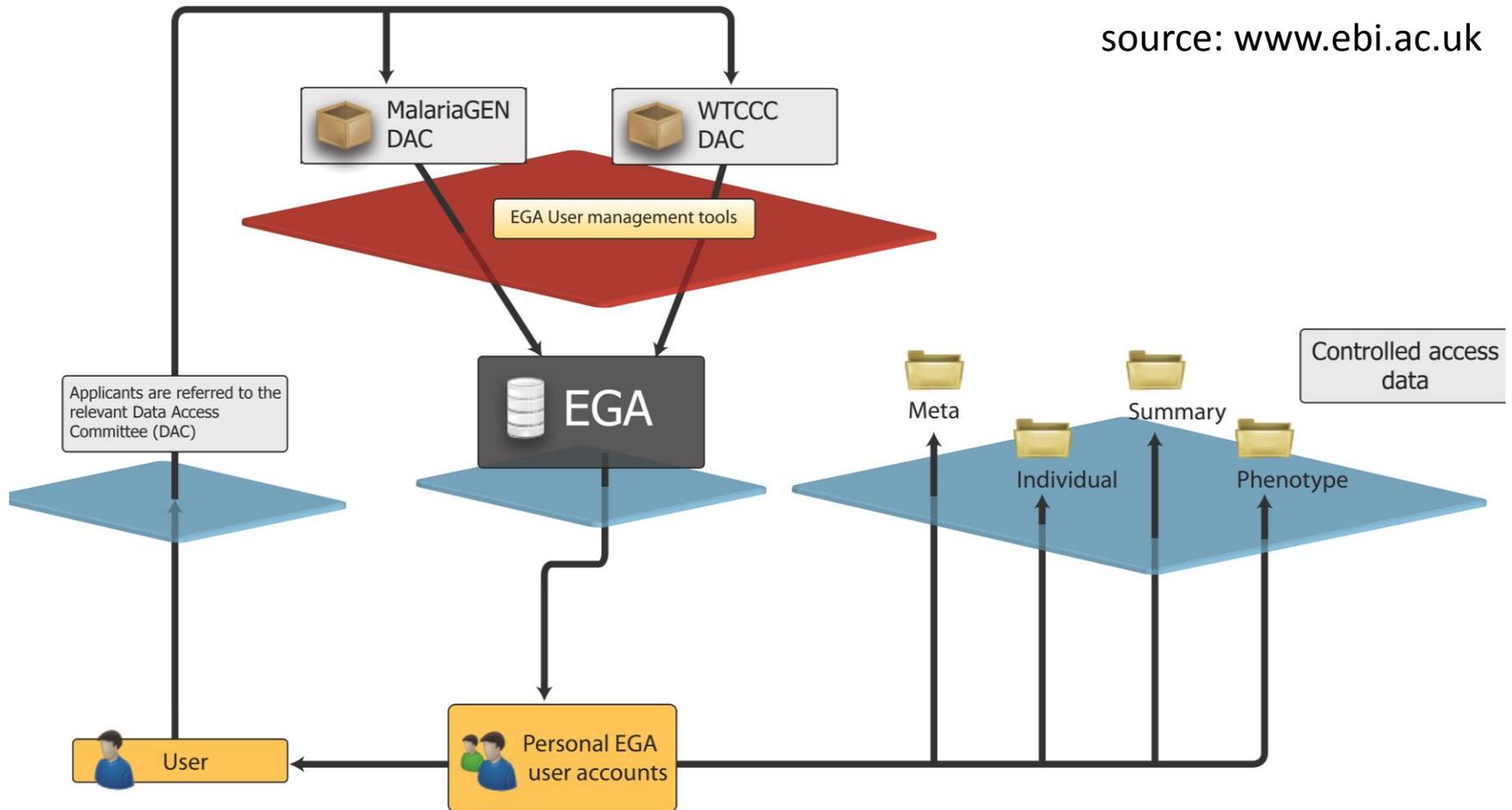
- Usually human, personally identifiable (Genetic, phenotypic)
- Affiliated to medical research or consortium projects
- Access determined by formal application procedure
- Requires secure storage and distribution
- **Informed consents specifying controlled release requirements**

source: www.ebi.ac.uk



EGA Distribution model

source: www.ebi.ac.uk



Using Your EGA Account



The European Genome-phenome Archive



- EGA Home
- Submit to EGA
- Information
- Help
- Contact Us

EXPLORE THE EGA

- Browse studies
- Browse datasets
- Browse data access committees
- Browse data providers

jeffalmeidaking@googlemail.com

EGA

EGA Status

User: jeffalmeidaking@googlemail.com

Date	Action	Dataset	Packet	Status	Status History
2013-03-15 10:18:21	download	EGAD00010000280	Raw data File (e.g. CEL/IDAT)	7-10 days until data access	View status history
2013-03-15 10:18:21	download	EGAD00010000280	Signal intensity file	Data access provided - check your email	View status history
2013-03-15 10:25:29	download	EGAD00010000238	Raw data File (e.g. CEL/IDAT)	3-5 days until data access	View status history

- Download accounts require manual set-ups
- Distribution requires data preparation
- Download accounts expire in 14 days
- Data must be decrypted using keys supplied offline

Secure EGA Downloader (Pilot stage)

EMBL-EBI European genome-phenome archive

Version: 0.1.37 [Contact the EGA Helpdesk](#)

Download / Slice Download File Slice Decrypt / Index

EGA User Name: jeff@ebi.ac.uk EGA User Password: ***** Logged in. Logout Download Protocol: Standard FTP

Local Encryption Key: ***** Verify Key: ***** Slice to Download: Invalid Region.

Local Download Directory: /Users/jeff/Desktop

Remote Files

Download	Name	Size	Date	Progress
<input checked="" type="checkbox"/>	172TR.CEL.gpg	4.31 MB	08-Oct-2013	Download Complete
<input checked="" type="checkbox"/>	110TR.CEL.gpg	4.41 MB	08-Oct-2013	Download Complete
<input checked="" type="checkbox"/>	761TR.CEL.gpg	4.40 MB	08-Oct-2013	Download Complete
<input checked="" type="checkbox"/>	680TR.CEL.gpg	4.45 MB	08-Oct-2013	Download Complete
<input checked="" type="checkbox"/>	785TR.CEL.gpg	4.28 MB	08-Oct-2013	87%
<input type="checkbox"/>	876TR.CEL.gpg	4.32 MB	08-Oct-2013	0%
<input type="checkbox"/>	029TR.CEL.gpg	4.43 MB	08-Oct-2013	0%
<input type="checkbox"/>	175TR.CEL.gpg	4.50 MB	08-Oct-2013	0%
<input type="checkbox"/>	016TR.CEL.gpg	4.52 MB	08-Oct-2013	0%
<input type="checkbox"/>	166TR.CEL.gpg	4.41 MB	08-Oct-2013	0%
<input type="checkbox"/>	186-01-8TR.CEL.gpg	4.32 MB	08-Oct-2013	0%
<input type="checkbox"/>	100TR.CEL.gpg	4.43 MB	08-Oct-2013	0%
<input type="checkbox"/>	032TR.CEL.gpg	4.36 MB	08-Oct-2013	0%
<input type="checkbox"/>	019-01-1TR.CEL.gpg	4.66 MB	08-Oct-2013	0%
<input type="checkbox"/>	009TR.CEL.gpg	4.14 MB	08-Oct-2013	0%
<input type="checkbox"/>	040TR.CEL.gpg	4.42 MB	08-Oct-2013	0%
<input type="checkbox"/>	184-01-5TR.CEL.gpg	4.32 MB	08-Oct-2013	0%
<input type="checkbox"/>	039TR.CEL.gpg	4.31 MB	08-Oct-2013	0%
<input type="checkbox"/>	022TR.CEL.gpg	4.39 MB	08-Oct-2013	0%
<input type="checkbox"/>	012TR.CEL.gpg	4.37 MB	08-Oct-2013	0%
<input type="checkbox"/>	290TR.CEL.gpg	4.38 MB	08-Oct-2013	0%
<input type="checkbox"/>	157TR.CEL.gpg	4.44 MB	08-Oct-2013	0%
<input type="checkbox"/>	030TR.CEL.gpg	4.35 MB	08-Oct-2013	0%

Select All 4 files selected. Total size: 17.40 MB Display Files: Request Dataset Filter

Cancel Download

Status: Downloading file 5 of 5 File: 1837 KB/s

Log in/select transfer protocol



Create key



Select destination directory



Filter and select files

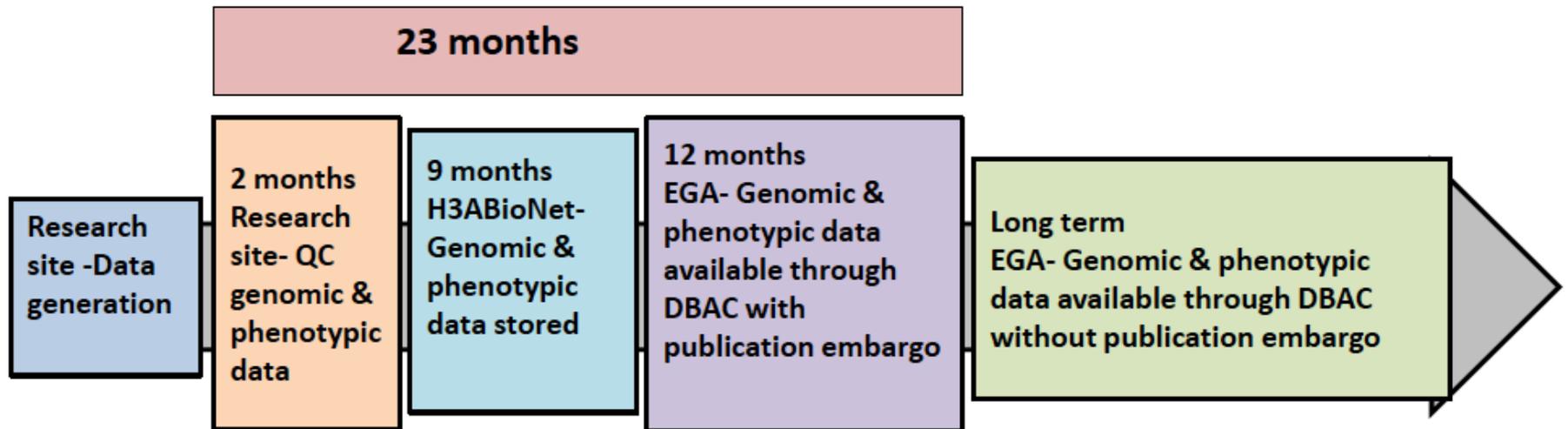


Download!





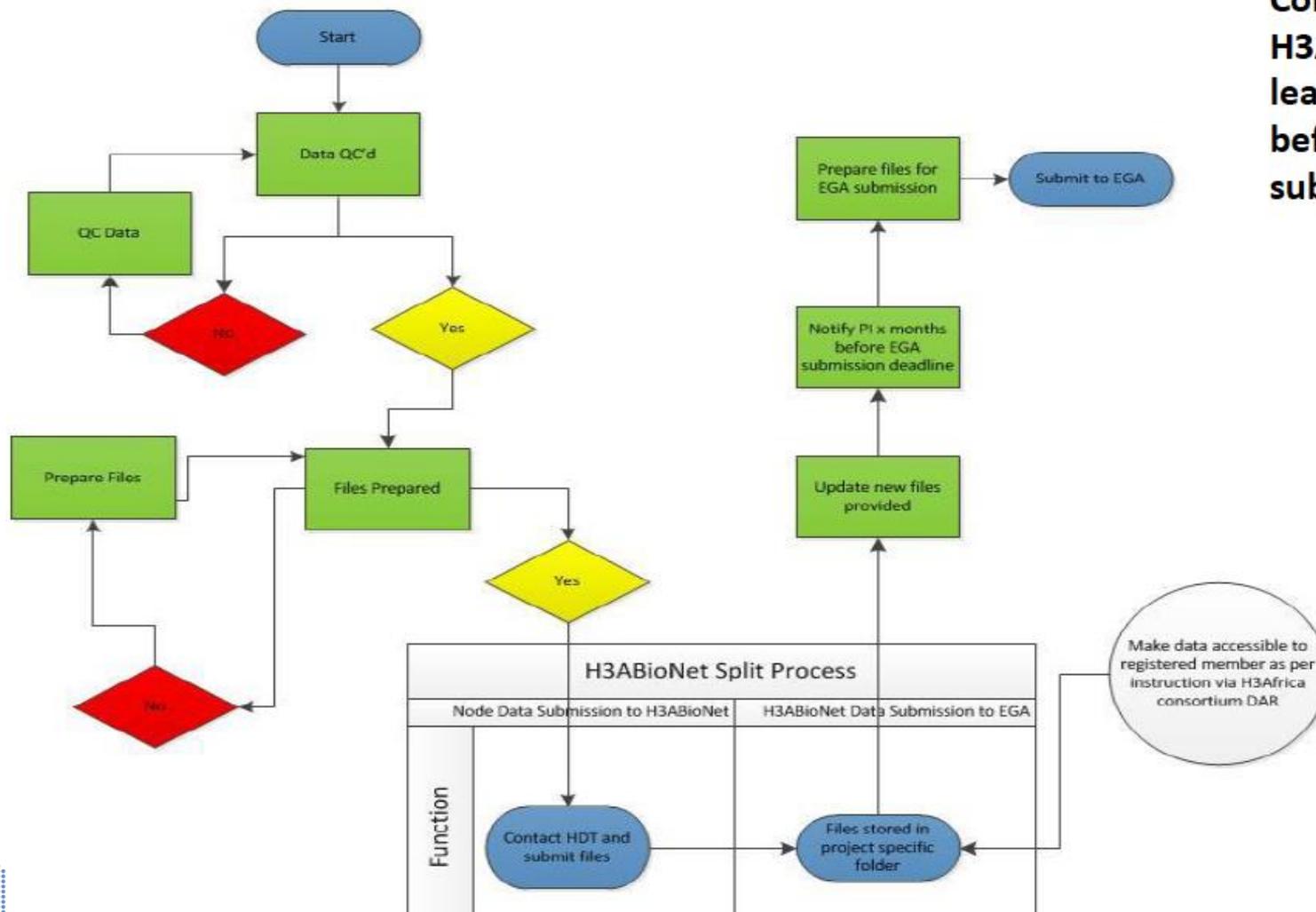
H3A Archive submission timeline





H3A Archive- workflow

H3ABioNet Data Submission Workflow



Contact
H3ABioNet at
least 1 month
before
submission



H3A Archive



H3Africa Data Archive

[Projects](#)

[Submissions](#)

[Funders](#)

[Help](#)

ayton ▾

1. Pre-register

2. Receive Submission pack

3. Upload

4. Confirm validation

Pre-register a submission

Your name:

Ayton Meintjes

Project:

AWIGEN

Data types:

- Human whole genome sequence data
- Human exome sequence data
- Genotyping array data
- Metagenomics sequence data
- Pathogen sequence data

Estimated date of submission:

2016/06/01

test submission

Save



H3A Archive



[H3Africa Data Archive](#)

[Projects](#)

[Funders](#)

[Help](#)

[Register](#)

[Log in](#)

Projects

TGEN - TrypanoGEN: An integrated approach to the identification of genetic determinants of susceptibility to trypanosomiasis

KDRN - H3Africa Kidney Disease Research Network

AWIGEN - Genomic and Environmental Risk Factors for Cardiometabolic Disease in Africans

RHDGen - The RHDGen Network: Genetics of rheumatic heart disease and molecular epidemiology of *Streptococcus pyogenes* pharyngitis

T2DSSA - Burden, spectrum and etiology of type 2 diabetes in sub-Saharan Africa

RAFAGENE - RAFAGene: Contribution of genetic variation to pharmacokinetic variability and toxicity in patients undergoing multi-drug tuberculosis treatment in Sub-Saharan Africa

GSSAX - Identification of genes important for schizophrenia based on genomic analysis of the South African Xhosa population

ACCME - African Collaborative Center for Microbiome and Genomics Research

CAfGEN - Collaborative African Genomics Network

NMRDAC - The Nasopharyngeal Microbiome and Respiratory Disease in African Children

HMGWA - Host and Microbial Genetic Determinants of febrile illness in West Africa

SIREN - Stroke Investigative Research & Educational Network

SBMTb - Systems Biology for Molecular Analysis of Tuberculosis in Ethiopia

CGSHND - Clinical and Genetic Studies of Hereditary Neurological Disorders in Mali



H3A Archive



H3Africa Data Archive

[Projects](#)

[Submissions](#)

[Funders](#)

[Help](#)

ayton -

AWI-GEN

Genomic and Environmental Risk Factors for Cardiometabolic Disease in Africans

Owner

Michele Ramsay

Submitters

Ayton Meintjes

Submissions

Date started	Submitter	Status
May 11, 2016, 7:10 p.m.	ayton.meintjes	New



H3A Archive

- Don't register directly at EGA
- Pre-register your submissions at least 6 weeks in advance



Policy documents

H3Africa Consortium Data Sharing, Access and Release (DSAR) Policy

- Related documents (appendices):
 - Data Quality Control and Validation Guidelines
 - H3ABioNet Data Submission Guidelines
 - EGA Submission Guidelines
 - Data and Biospecimen Access Committee Guidelines
 - Data Access Request Form (Form for researchers to request access to data)
 - Data Access Agreement (Form for researchers who have access to data)
 - Policy Statement & Institutional Certification (Form for PIs to submit to EGA)



DSAR Policy

- Outlines the data sharing and access policy of the consortium
- The “principles aim to strike an appropriate balance in ensuring that adequate safeguards are in place to protect participants, while maximizing the ability of investigators to advance research”



Principles of DSAR

- Maximizing the availability of research data, in a timely and responsible manner.
- Protecting the rights and privacy of human subjects who participated in research studies.
- Recognizing the scientific contribution of researchers who generated the data.
- Considering the nature and ethics of the research proposed in establishing the timely release of data, and mechanisms of data sharing.
- Promoting deposition of genomic data in existing community data repositories whenever possible.



DSAR Policy –data types

- **Phenotype data (associated with genotype data)**
 - Demographic information
 - Anthropometric data
 - Disease and health related phenotype data
- **Genetic Variation data human & pathogen**
 - Sequencing data (whole genome, exome, targeted)
 - Primary data following QC
 - Variant calls following QC
 - Genotype data from arrays
 - Primary array data following QC
 - Variant calls following QC
- **Microbiome sequence data**
 - Patient/sample phenotypes
 - Non-human 16S rRNA sequence data for microbiome
 - Non-human full genome sequence data for



DBAC Guidelines

- 8 members unrelated to H3Africa consortium
 - Scientists with relevant research experience
 - Biobanking expertise
 - Data expert
 - Ethics expert
 - Legal representative
 - community representative, advocacy group or nurse
 - Ex officio H3Africa members
- Quorum is 5/8 members
- Appointed for 3 years + 3 extra possible
- 3 should stay on with any change in committee



Microbiome/non-human data

- EGA takes human genotypes and phenotypes
- For non-human data sequences should be submitted to European nucleotide archive (ENA) –open access
- EGA can still store phenotypes (controlled access) and link these to sequences in ENA
- Timelines and process is same as for other data



Questions