

H3Africa Data sharing, access and release policy

DSAR Working group

H3Africa Funded projects

1 Bioinformatics Network

4 Pilot Biorepositories

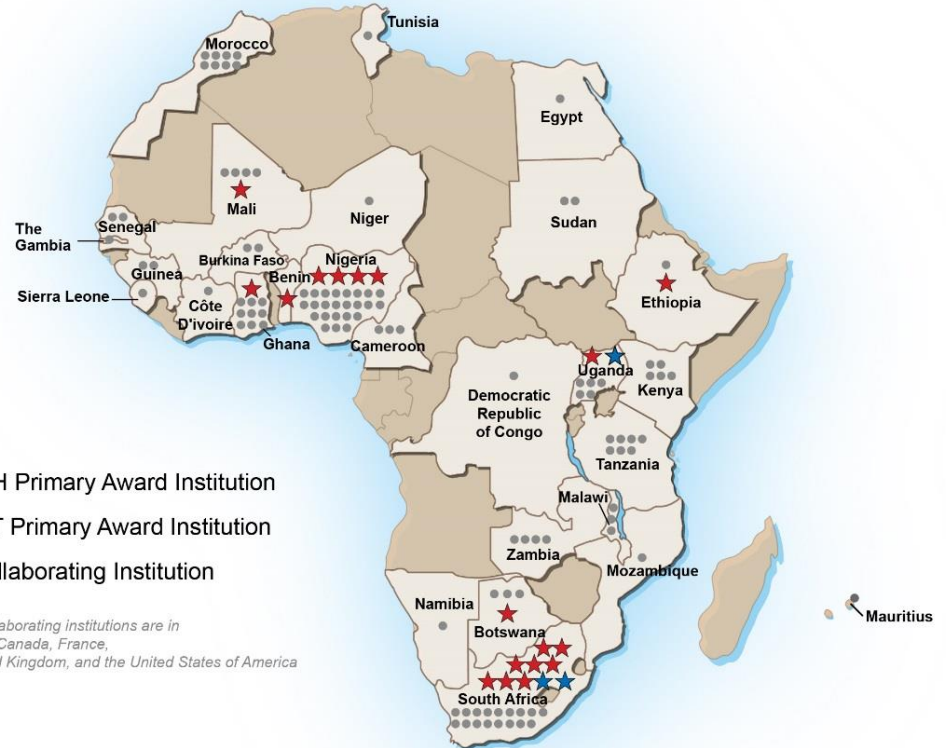
16 Research Projects on:

- **Obesity**
- **Diabetes**
- **Heart Disease**
- **Stroke**
- **Kidney Disease**
- **Sickle Cell (Ethics)**
- **Microbiome (2)**
- **Trypanosomes (2)**
- **TB (2)**
- **Schizophrenia**
- **Rare neurological disorders**
- **Fevers of unknown origin**

Partner institutions & project sites



National Institutes of Health - Wellcome Trust H3Africa Research Network



H3Africa data expected

- GWAS arrays:
 - 30Mb/cell file (per sample), 150Gb per 96 well plate processed.
 - Samples planned >50,000
 - Storage required for processed data: >50TB
- Sequencing NGS:
 - Exome = Illumina, ~45 depth, 90bp paired end, 17GB raw fastq, Full = Illumina, ~33 depth, 101 bp paired end, 250GB raw fastq (per sample)
 - Samples planned in total: exome ~1,800, full genome ~3,000
 - >1000 microbiome samples
 - Total storage for raw reads: >100TB

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

Policy documents

- **H3Africa Consortium Data Sharing, Access and Release (DSAR) Policy**
- Related documents (appendices):
 - A. Data Quality Control and Validation Guidelines
 - B. H3ABioNet Data Submission Guidelines
 - C. EGA Submission Guidelines
 - D. Data and Biospecimen Access Committee Guidelines
 - E. Data Access Request Form (Form for researchers to request access to data)
 - F. Data Access Agreement (Form for researchers who have access to data)
 - G. 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”

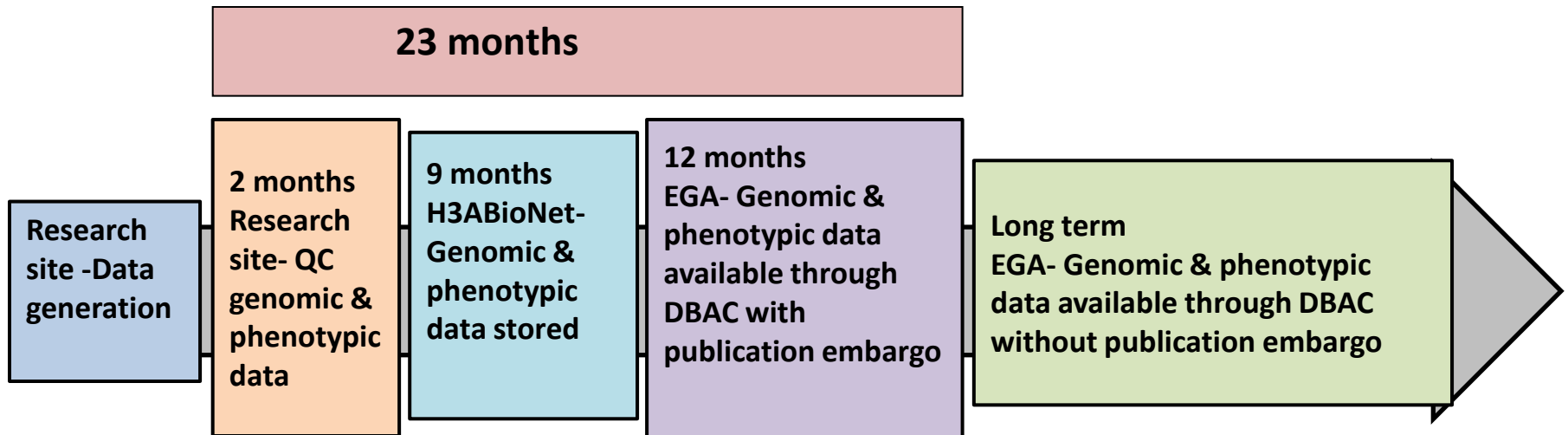
Principles of DSAR

- 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 microbiome
 - Possible human sequence contamination

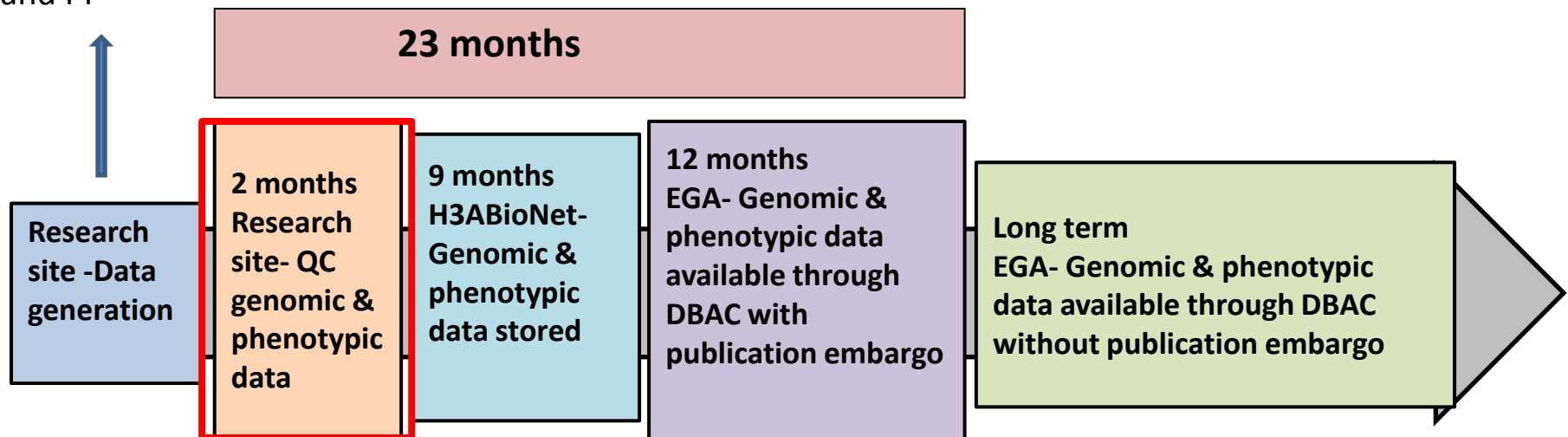
Process for data submission



Process for data submission

Data Quality Control

Publishable dataset –can be subset of complete project, but standalone.
To be agreed between grant officer and PI



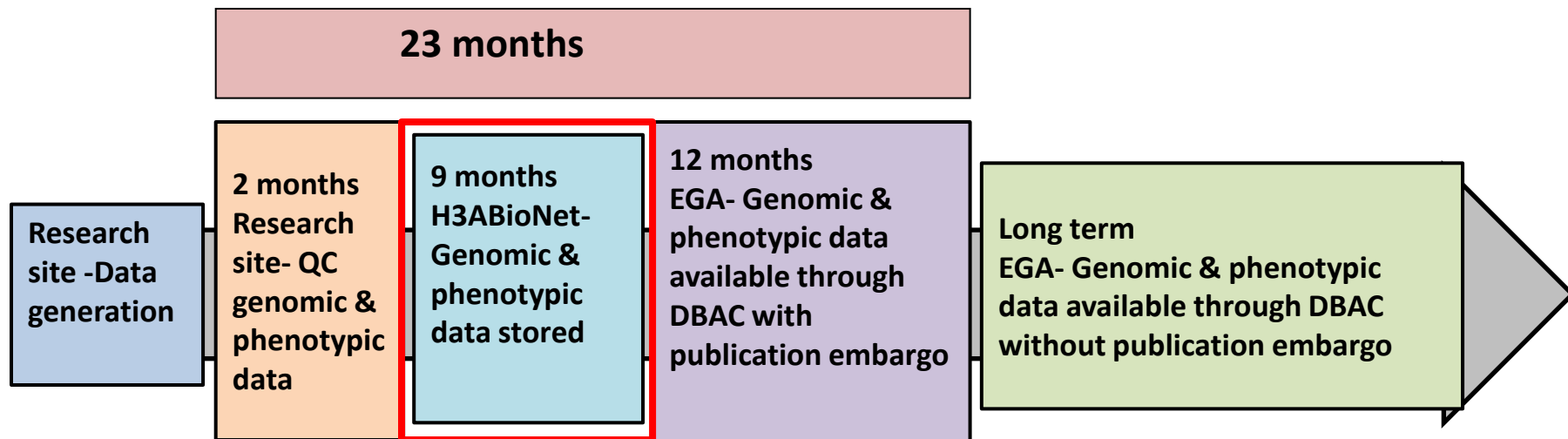
QC for arrays = data ready for genotype calling
QC for sequence = data ready for alignment to reference

QC for very large datasets may take longer –to be discussed with grant officer

Appendix A

Process for data submission

Data submission to H3ABioNet



Appendix B

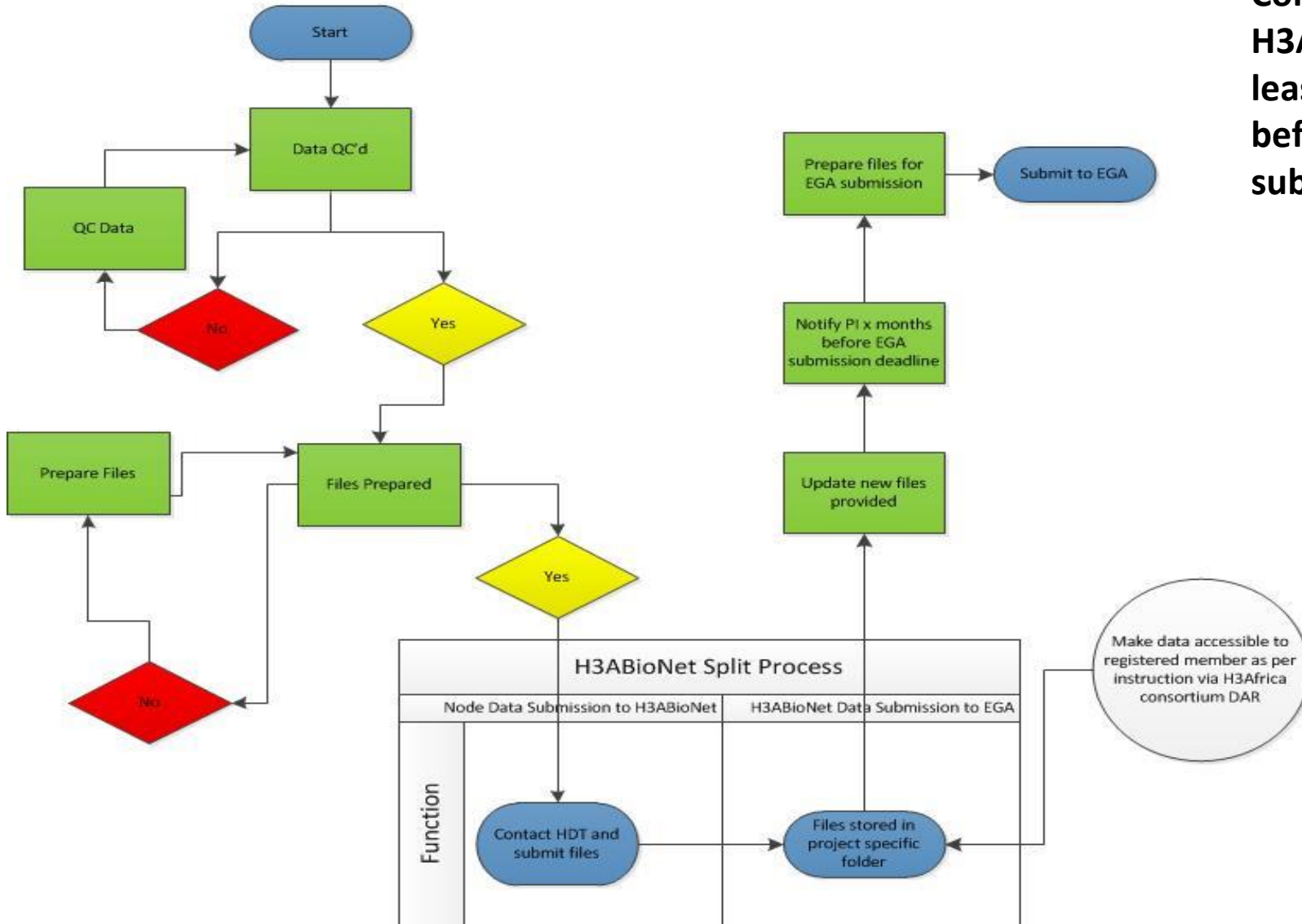
Files to be submitted determined by EGA requirements

Only PI and those who PI grants access to from within H3Africa consortium can access data

PI to ensure relevant ethics consent in place

H3ABioNet data submission guidelines

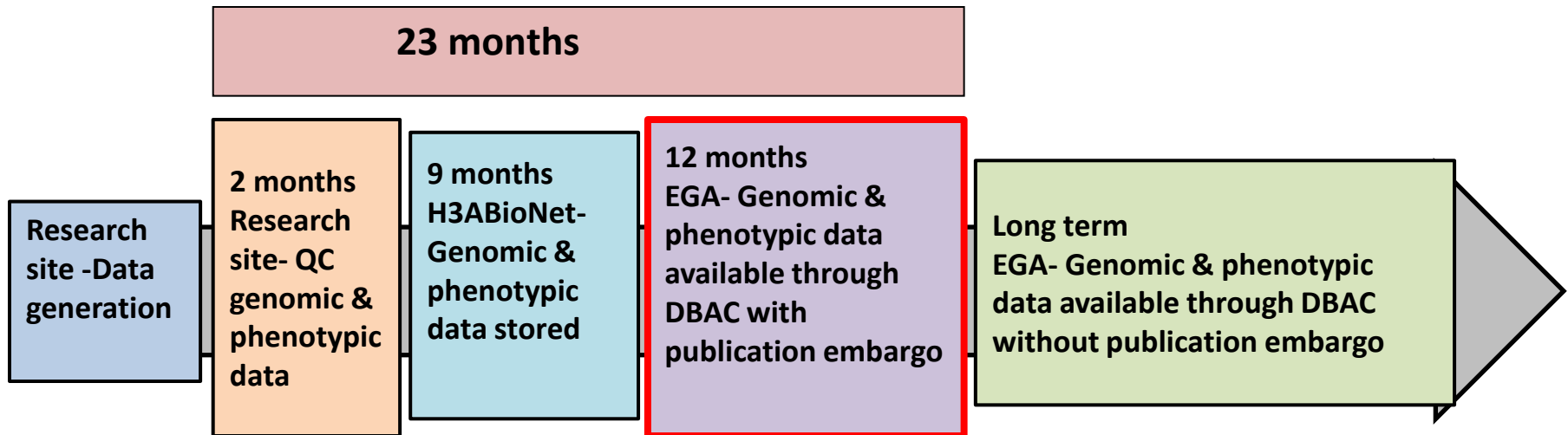
H3ABioNet Data Submission Workflow



Contact
H3ABioNet at
least 1 month
before
submission

Process for data submission

Submission to EGA –publication embargo



H3Africa projects to put their aims on the H3Africa website. Where projects may overlap suggest possibility of collaboration

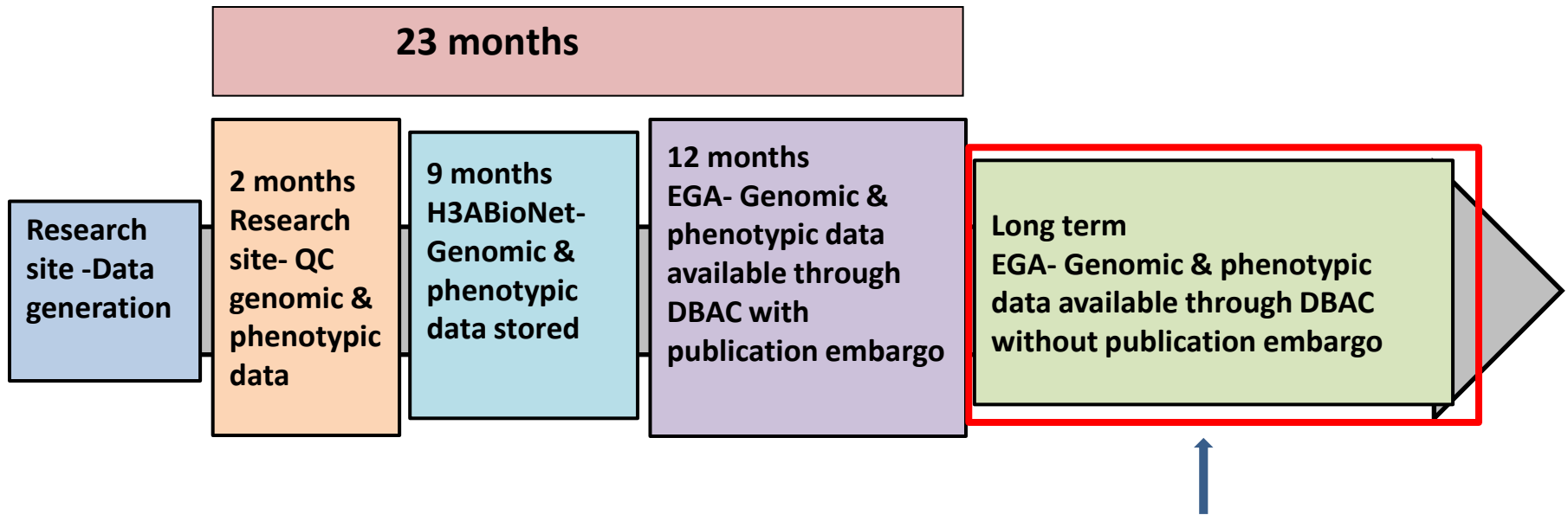
Embargo falls away after first publication on data

External access to data is only via DBAC

Appendix C

Process for data submission

Long term storage in EGA



H3Africa projects to put their aims on the H3Africa website. Where projects may overlap suggest possibility of collaboration

Appendix D, E, F

External access to data is only via DBAC

CC monitors access requests and progress reports

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

Policy documents

- H3Africa Consortium Data Sharing, Access and Release (DSAR) Policy
- Related documents (appendices):
 - A. Data Quality Control and Validation Guidelines
 - B. H3ABioNet Data Submission Guidelines
 - C. EGA Submission Guidelines
 - **D. Data and Biospecimen Access Committee Guidelines**
 - E. Data Access Request Form (Form for researchers to request access to data)
 - F. Data Access Agreement (Form for researchers who have access to data)
 - G. Policy Statement & Institutional Certification (Form for PIs to submit to EGA)

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

Policy documents

- H3Africa Consortium Data Sharing, Access and Release (DSAR) Policy
- Related documents (appendices):
 - A. Data Quality Control and Validation Guidelines
 - B. H3ABioNet Data Submission Guidelines
 - C. EGA Submission Guidelines
 - D. Data and Biospecimen Access Committee Guidelines
 - **E. Data Access Request Form (Form for researchers to request access to data)**
 - F. Data Access Agreement (Form for researchers who have access to data)
 - G. Policy Statement & Institutional Certification (Form for PIs to submit to EGA)

Data access request form

- Name and details
- Project title
- Dataset required
- Research question
- Feasibility
- Research ethics
- Benefits to Africa

Will be merged with sample request form

Policy documents

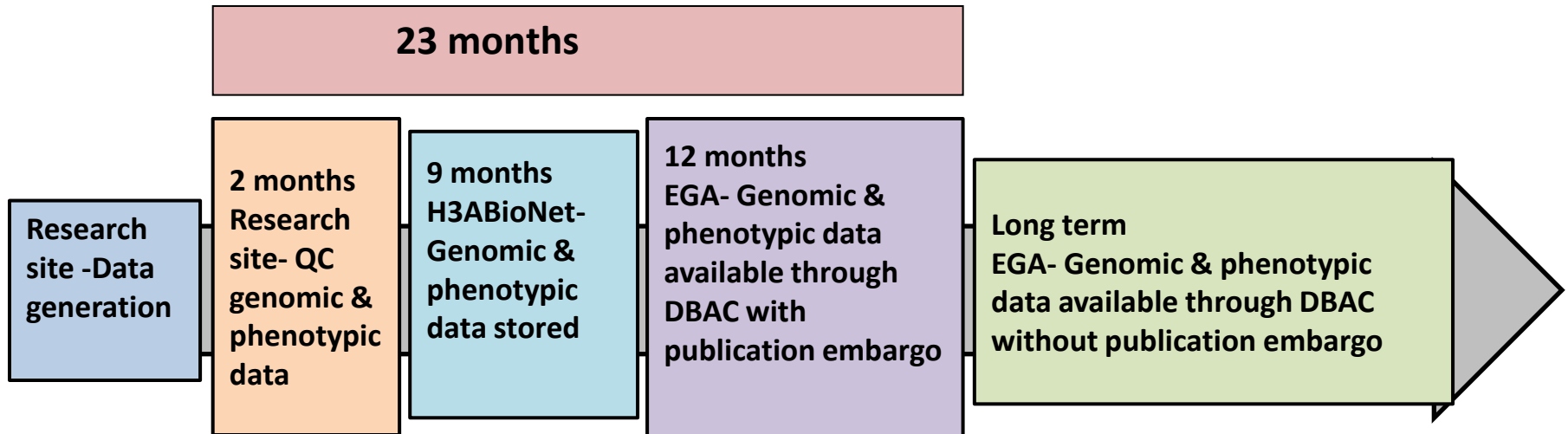
- H3Africa Consortium Data Sharing, Access and Release (DSAR) Policy
- Related documents (appendices):
 - A. Data Quality Control and Validation Guidelines
 - B. H3ABioNet Data Submission Guidelines
 - C. EGA Submission Guidelines
 - D. Data and Biospecimen Access Committee Guidelines
 - E. Data Access Request Form (Form for researchers to request access to data)
 - **F. Data Access Agreement (Form for researchers who have access to data)**
 - G. Policy Statement & Institutional Certification (Form for PIs to submit to EGA)

Data access agreement

- Agreement with institution, PI and researchers named
- Agreeing to:
 - H3Africa not liable for use of data
 - Only use data for agreed purpose
 - Maintain data confidentiality
 - Make sure data is secure
 - Acknowledge source of data
 - Submit annual reports
 - Project put onto website
 - Access is granted for 1 year

Summary

DSAR Policy, which applies to all projects



Data Sharing, Access and Release Working Group Members

Maria Giovanni	Chester Wayne Brown	Annette Macleod
Audrey Duncanson	Marape Marape	Maia Lesosky
Julia Puzak	Gaolathe Thobokwe	Jantina de Vries
Alan Christoffels	Christian Happi	Nicky Mulder
Catherine Rossouw	Onikipe Folarin	Dean Everett
Akin Abayomi	Pardis Sabeti	Kais Ghedira
Beverley van Rooyen	Lwanga Newton	Yasmina Jaufeerally Fakim
Tim Reid	Musinguzi Henry	Winston Hide
Musa Gabadi	Guida Landoure	Judit Kumuthini
Sunji Nadoma	Mathurin koffi	Patrick Musicha
Emmanuel Oga	Gustave Simo	Benjamin Kumwenda
Patrick Musonda	Godfrey Tangwa	Silindile (Slee) Mbhele
James Okuma	Ebony Madden	Mayowa Owolabi
Sally Anthony	Mark Guyer	Kwamena Sagoe
Nicki Tiffin	Robin Mason	Rufus Akinyemi
Bamidele Tayo	Quynh Nguyen	Mulugeta Gebregziabher
Aurel C Allabi	Jennifer Troyer	Albert van Eck
Thuli Mthyane	Geetha Senthil	Catherine Kyobutungi
Gezahegne Mamo	Jeff Struewing	Shane Norris
Marcus Jones	Sudha Srinivasan	Scott Hazelhurst
Kenneth Ekoru	Ken Wiley	Adam Baldinger
Segun Fatumo	Abigail Abrahams	Jantina de Vries
Michelle Skelton		

H3ABioNet data submission doc

- Guidelines for file naming
PI_ShortDate_SubmissionNumber
- Submit files + Readme (list of files required for different data types)
- Include phenotype data
- Submit list of people who should have access
- Two transfer options –online or courier
- Info on submissions will be on internal website
 - Notify funders of date of submission
 - Notify PIs 2 months before EGA submission is due

Responsibilities

- H3ABioNet responsible for:
 - Keeping data secure
 - Will have copy + replicated sites
 - Helping with transfer of data
- H3ABioNet not responsible for:
 - Deciding on who gets access
 - Ensuring ethics is in place
 - Ensuring validity of data
- H3ABioNet can help with:
 - Accessibility for data analysis/mining
 - Preparation for submission to EGA