
# Phylogenetics and Networks for Generalised HIV Epidemics in Africa

# PANGEA 2

# Principles of Research Conduct, Data Sharing, Accreditation, and Publication

## Version 1.0
## Date: 15th November 2018

**Effective Date**: 15th November 2018

Revision History

| Version | Suggested by | Reason for Revision | Date of approval | Approved by |
|---|---|---|---|---|
| 1.0 | Christophe Fraser / Deenan Pillay | Draft document | 15/11/2018 | PANGEA 2 Steering Committee |
| 1.0 | | No changes made | 17/09/2019 | Text confirmed by the PANGEA 2 Steering Committee |

## Table of contents

## 1. Purpose

The aim of this document is to provide the bases for making optimal use of the data generated within the PANGEA 2 Consortium (PANGEA), in terms of public health benefits and scientific discovery, to give credit to the people who were involved in the work, and to encourage collaboration between the different groups in PANGEA. This SOP therefore describes the principles and procedures for sharing of data from the PANGEA 2 Consortium. This document will be reviewed annually at the consortium meeting. Changes require a two-third majority vote.

## 2. Definitions

| Term | Definition |
|---|---|
| Consortium | Members of the PANGEA 2 Steering Committee. |
| Steering Committee | Members as listed in Appendix 1 |
| Executive committee | Members as listed in Appendix 1 |
| Research Institute | Member institution, this can be a Field Research Institute where a study is performed and data are analysed or a Research Institute where data are analysed. |
| Sequencing data | Unprocessed sequences (unassembled raw reads), contigs (short fragments of the genome obtained by combining overlapping reads) and assembled genomes |
| Consortium data | The sequencing and linked clinical/epidemiological dataset which is provided through the Consortium |
| PANGEA core variables | Metadata variables that all Field Research Institutes agree to contribute to the PANGEA database if available. PANGEA core variables are specified in the document “PANGEA core variables”. |
| Data Sharing | Data release and data transfer to external or internal applicants |
| Internal researcher | Member of the PANGEA 2 Steering Committee or researcher directly line-managed by a member of the PANGEA 2 Steering Committee |
| External researcher | All researchers that are not internal researchers |
| Working group | Group of researchers working on one or more analysis that is expected to become a manuscript. |

| | |
|---|---|
| PANGEA Forum | Internet forum of the Consortium where abstracts, presentations, manuscripts can be posted and discussed; url: http://pangea.community |

## 3. Responsibility and roles

The process of receiving and processing applications, developing agreements, producing and transferring data, and responding to subsequent requests for clarification of data involves a broad range of functions across the Consortium.

Principal investigators of PANGEA 1 and PANGEA 2, together with the representatives from the field research institutes, agree that from the start of PANGEA 2, new applications will be reviewed by the new PANGEA 2 Steering Committee; details are provided in Appendix 2.

| Role | Responsibility |
|---|---|
| Database Manager | • Secure storage of Consortium data<br>• Release of Consortium data as approved by the Steering Committee |
| Project Manager | • Receive requests and ensure requests are provided to the Executive Committee and the Steering Committee for review in a timely manner |
| Executive Committee | • Review and discuss requests, provide background information to the Steering Committee where appropriate |
| Steering Committee | • Review and approve requests to release Consortium data.<br>• Review and approve applications for accreditation<br>• Authorise Consortium Database Manager to release Consortium data, contingent on study-specific approvals where required |
| Field research institute representatives | • Review and, where applicable, approval of requests concerning additional metadata. |

## 4. Consortium data storage and classification

Consortium data is sequencing data in conjunction with linked clinical/epidemiological data. All Consortium data will be stored in a secure database at the Big Data Institute in Oxford. Variables are classified into three categories.

The following data will not be accepted into the PANGEA database: Any unique patient identifier that is linked to external systems (health system ID, social security number, identifier used for other studies). These fields should never be submitted to the common database. They should stay at the field research institutes and be replaced with a study-specific identifier (PANGEA ID) before submission.

Security level 1 (sensitive):
**Fields or combination of fields that would potentially enable somebody to identify individuals with some effort (e.g. fine geographical location)**
These fields will be stored in a separate table on a different server and will only be shared for specific analyses that require this information and only with analysis-specific permission from the field research institutes (and can only be accessed on the BDI server). In the main database, these will be replaced with proxies that are sufficient for most analyses (newly generated categorical code instead of fine geographical location, e.g. geo-1 to geo-9 for nine geographical locations in a dataset, to be precise enough for analyses but reduce identifiability).

**Fields relating to sexual contact (sexual partners, phylogenetic linkage)**
These fields will be stored in a separate table and will however only be shared for specific analyses where these fields are needed.

Security level 2 (less sensitive)
**All other fields**
All PANGEA core variables and most of the additional data that field research institutes are happy to share, unless they fall under the fields specified under level 1. PANGEA core variables are that all Field Research Institutes agree to contribute to the PANGEA database if available. PANGEA core variables are specified in the document “PANGEA core variables”. These fields are visible to all Consortium members and will be shared upon request with accredited external researchers and external researchers whose concept sheets have been approved for access to these fields.

Security level 3 (public)
**Consensus sequences, country, month and year of sampling**
These fields will be submitted to public databases along with the consensus sequences.

Field research centres can decide into which category their data should fall. Submitted data will be classed as level 2 unless level 1 is requested by the research centre. Please add one line of explanation to each field which is classed as level 1.

## 5. Working with Consortium data

The Consortium aims to make data access as straightforward as possible to further scientific discovery and provide up-to-date data for public health decision-making, while at the same time protecting the privacy of the study participants and acknowledging the contribution of everybody involved in data generation and processing.

Field research centres can request 12 months exclusive access to data sets that are generated during PANGEA 2. During this time period which starts when sequence files are shared with the field research institutes, data can be shared with PANGEA members for joint analysis, but there is no obligation to share with all PANGEA members or with external accredited researchers. Sequences will only be uploaded to public databases after this 12 months period.

The Consortium encourages collaboration within the Consortium and with accredited collaborators. Groups of researchers working on similar projects within the Consortium will be encouraged to join forces but will not be obliged to do so. However, subgroups must regularly update the Steering Committee on how their research is progressing.

If a group is planning to embark on an analysis that seems identical to an analysis that has been presented before by another group and is actively being worked on, the Steering Committee may ask the second group to join forces with the first group or explain in which way their analysis differs from or adds to the first analysis and how the two projects can lead to separate publications in their own right.

## 6. Sharing of consortium data

The aim of the PANGEA data sharing policy is to allow external researchers to make use of the PANGEA data and draw them in to become part of PANGEA rather than to just hand over data. It is anticipated that this will lead to more collaboration between external researchers and Consortium members and ensure that those who generate and process the data will be acknowledged in the publications that build on their work.

Studies that use clinical/epidemiological data without any associated sequence data are outside the remit of PANGEA and researchers need to go through cohort-specific approval procedures to obtain the data.

Consortium data can be accessed either by submitting concepts sheets to the Steering Committee or by becoming an accredited researcher. Internal researchers i.e. researchers who are members of the Steering Committee or are directly line-managed by members of the Steering Committee can apply for accreditation straight away. External researchers are required to apply via the concept sheet route first and can apply for accreditation after six months if all requirements are met.

Consortium data received must be held securely on password protected computer, or on approved encrypted cloud services, and access limited to internal members of the analysis group. Consortium data received for specific analyses should be destroyed within two years after completion of the project (publication of manuscript) unless an extension has been agreed by the Steering Committee. Of note, all Consortium data will be centrally archived and can be re-requested for any re-analysis if required.

Consortium data must not be shared with non-accredited individuals or organisations without the permission of the Steering Committee. This includes non-accredited members of the same group. Security level 1 data obtained upon request must not be shared with researchers not named on the request without permission of the Research Institute which granted the request. This includes all other accredited PANGEA researchers.

After the end of PANGEA, either at the end of this funding period in 2021 or at the end of the last funding period, all sequence data and all level 2 and 3 data will be handed over to an external data broker agreed on by the Steering Committee. The broker is expected to honour the most current version of this document. All level 1 data will be destroyed.

### 6.1 Data access via a concept sheet

External researchers can obtain access after submitting a proposal via a concept sheet (details in Appendix 3). They will be bound by all clauses of the most current version of this document.

Before obtaining access to the data, external researchers are required to outline the scientific proposal, submit a risk-benefit mitigation table (example in appendix 5), provide a CV and a certificate for standard course on human subject research (e.g. CITI Biomedical Basics) which is less than three years old, and agree to familiarize themselves with the studies by reading any study-specific information supplied by PANGEA for this purpose.

The concept sheet is assessed by the Steering Committee and if complete should be approved unless the applicants lack academic training in a relevant discipline, the proposal is scientifically not sound or the proposed research puts participant privacy at risk.

After obtaining the data, external researchers are encouraged to collaborate where possible with members of the consortium where aspects of the proposal overlap with already existing strands of research.

External researchers are required to communicate progress every three months to the Steering Committee, through short written reports, shared slides and/or teleconferences as requested. The first presentation should include a brief introduction about themselves and their scientific interests.

If external researchers miss four quarterly updates or should the Steering Committee deem that no progress has been made on the project for 12 months, the Steering Committee can withdraw access to the data.

Six months after their proposal has been accepted, and after two quarterly updates, external researchers can apply for accreditation which will give them the same access rights as internal researchers. They will also be bound by the same obligations.

### 6.2 Data access via accreditation

Internal researchers or external researchers that fulfil the requirements outlined under section 6.1 can apply to become accredited PANGEA researchers (details in appendix 4). A decision will be taken by the PANGEA Steering Committee.

In order to get accredited, researchers are required to sign a code of conduct, provide a CV and a certificate for standard course on human subject research (e.g. CITI Biomedical Basics) which is less than three years old, and agree to familiarize themselves with the studies by reading any study-specific information supplied by PANGEA for this purpose.

Application forms should be submitted to the project manager who will pass them on to the Steering Committee. If no objections are raised within two weeks, the project manager will inform the researcher that the application has been successful. Objections should only be raised if not all parts of the application have been completed, the applicant has no academic training in a relevant discipline or has been previously found guilty of scientific misconduct.

Accredited researchers are bound by all clauses of the most current version of this document. Within this framework, they are granted access to level 2 and 3 data. Level 1 data will be accessible upon justified request and subject to the agreement of the Field Research Institutes.

Before embarking on a new analysis, accredited researchers are required to send a short abstract (200-400 words) outlining the proposed work to the Steering

Committee and post it on the forum before embarking on an analysis. The abstract should contain the research question, why it is of interest, the methods and if the expected results will fit into the remit of PANGEA.

Accredited researchers are required to communicate progress every three months to the Steering Committee, through short written reports, shared slides and/or teleconferences as requested. The first presentation should include a risk-benefit mitigation table (example in appendix 5).

If accredited researchers miss four quarterly updates or should the Steering Committee deem that no progress has been made on the project for 12 months, their access is withdrawn and they need to reapply for accreditation.

The Steering Committee reserves the right to withdraw accreditation from individual researchers that do not comply with the code of conduct. Cases of gross misconduct will be brought to the attention of the researcher's host institution. Researchers who have lost their accreditation are obliged to delete all Consortium data immediately.

## 7. Authorship

The main analysis activity is expected to take part in the working groups and the PANGEA Forum. Working groups are group of researchers working on one or more analysis that is expected to become a manuscript. Members of the working groups are expected to collaborate on questions proposed in the working group and conduct analyses that meet the milestones as set out in the grant proposal. Members of the working groups are expected to contribute and critically review analyses conducted in the group and satisfy themselves that the results and interpretation are sound. They are expected to contribute to drafting and reviewing manuscripts in a timely fashion, and to be prepared to sign statements requested by journals to vouch for the data and their contribution to the study.

Through the process of actively participating in a working group, the members of the group are likely to fulfil requirements for authorship on manuscripts drafted by the group. Authorship should also be granted to individuals who advance the analysis significantly by suggestions on the PANGEA Forum.

There are three tiers for authorship that relate to different types of publication:

1) Descriptive analyses of the Consortium data should be published as:

The PANGEA Consortium*

2) Specific analyses of PANGEA data carried out by a subgroup of PANGEA researchers with or without external collaborators should be published as:

Names of the people who contributed measurably to the study or the manuscript, on behalf of the PANGEA Consortium*

3) Studies that use PANGEA data as part of a larger data set to study questions outside the remit of the PANGEA analysis goals should be published as:

Names of the people contributed measurably to the study or the manuscript, with the PANGEA Consortium*

*All Steering Committee members, other members of the working groups and/or field research institute teams as appropriate

Members of the Consortium who generated (sequenced), processed and managed the Consortium data should be recognised by co-authorship or acknowledgement, as appropriate and as discussed with the Steering Committee. Details with regards to the Project Management Team are provided in Appendix 1.

Furthermore, whilst the Consortium is managed by the PANGEA 2 teams, the members of the PANGEA 1 Steering Committee and the PANGEA 1 Management team should be acknowledged for use of any data that they participated in generating, processing, or managing. This should take the form or co-authorship or named acknowledgement as appropriate, and as discussed with the Steering Committee. Members of these groups are listed in Appendix 2.

## 8. Publication

Though individual study teams may present and publish analyses and results based on the phylogenetic data from their own study/studies without obtaining prior approval from the Steering Committee, this is discouraged. The Consortium expects recognition for having generated the data and encourages collaboration. Any activity which results from bringing together analysts and Field Research Centres through the Consortium should be regarded as a Consortium and not a Field Research Centre-led activity, unless agreed by the Consortium Steering Committee. The Consortium encourages the Field Research Institutes to lead analysis groups within the Consortium.

All publications should contain a risk-benefit mitigation table in the supplement / appendix.

All publications should be submitted via the Chronos, the publication management system of the Bill & Melinda Gates Foundation. Any papers derived from PANGEA should be published in open access journals. If open access cannot be negotiated, the Gates Foundation will pay the associated fees for publications submitted via Chronos. All publications related to PANGEA are required to list the two PANGEA

grants including grant numbers (PANGEA 1: OPP1084362, PANGEA 2: OPP1175094)

### 8.1 Approval of conference abstracts

The first author or analysis lead of a conference abstract should submit the abstract to the Steering Committee and the PANGEA Forum no less than two weeks before the abstract is submitted at the conference. If no concerns or questions are raised during this two-week period, the authors are free to submit the abstract to the conference.

If concerns or questions are raised, the authors are responsible for discussing the matter with the concerned member of the Steering Committee and are encouraged to modify the abstract accordingly. If the matter extends beyond scientific disagreement (e.g. when patient privacy is at stake) and agreement cannot be reached, as a last resort, the Steering Committee member concerned can ask the Steering Committee to block the abstract from submission. After a discussion, the matter is settled by the Steering Committee by majority vote if required.

### 8.2 Approval of manuscripts

The first author or analysis lead of a manuscript should submit the manuscript to the Steering Committee no less than three weeks before the manuscript is submitted to a journal. If no concerns or questions are raised during this three-week period, the authors are free to submit the manuscript to the journal.

If concerns or questions are raised, the authors are required to compile all comments into a single document and provide responses to the Steering Committee. It is expected that the authors are responsive to feedback. Where there are differences with regards to the interpretation of the data a suitable compromise should be found, e.g. by stating different interpretations in the discussion/conclusion of the manuscript.

The Steering Committee will discuss the responses and either approve submission, approve submission with changes, or, ask to review the manuscript again after changes have been made by the authors. If the matter extends beyond scientific disagreement (e.g. when patient privacy is at stake) and agreement cannot be reached, as a last resort, the field research institutes have the right to withdraw their data from a manuscript before publication.

### 8.3 After submission

Once an abstract or manuscript has been approved for submission, the analysis lead is responsible for submission for peer review to the appropriate meeting / journal. The analysis lead should notify the Steering Committee once the paper has been submitted and when reviews have been received.

When a manuscript has been accepted for publication, the analysis lead forward an electronic copy to the Steering Committee.

# 9. Appendices

## 9.1 Appendix 1: PANGEA 2 Steering Committee

## 9.2 Appendix 2: PANGEA 1 Steering Committee
(to be acknowledged for use of data generated by PANGEA 1)

## 9.3 Appendix 3: Application for data access via a concept sheet proposal

## 9.4 Appendix 4: Application to become an accredited PANGEA researcher

## 9.5 Appendix 5: Example of a risk mitigation table

## 9.1 Appendix 1: PANGEA 2 Steering Committee

| Name | Institution | Representatives & roles | Email address[1] |
|---|---|---|---|
| Lucie Abeler-Dörner | University of Oxford | Project manager | lucie.abeler-dorner@bdi.ox.ac.uk |
| Helen Ayles | PopART/ Zambart | PopART | helen@zambart.org.zm |
| David Bonsall | University of Oxford | Sequencing lab | david.bonsall@bdi.ox.ac.uk |
| Rory Bowden | University of Oxford | Sequencing lab | rbowden@well.ox.ac.uk |
| Vincent Calvez | Institut Pasteur | TasP trial | vincent.calvez@me.com |
| Max Essex | Harvard Botswana | Botswana studies | messex@hsph.harvard.edu |
| Sarah Fidler | PopART/Imperial College London | PopART | s.fidler@imperial.ac.uk |
| Christophe Fraser | University of Oxford | Principal Investigator PANGEA 2, Executive Committee and PopART Phylogenetics | christophe.fraser@bdi.ox.ac.uk |
| Kate Grabowski | Johns Hopkins University | Executive Committee and Rakai | mgrabows@jhu.edu |
| Tanya Golubchik | University of Oxford | Data manager | tanya.golubchik@bdi.ox.ac.uk |
| Richard Hayes | PopART/LSHTM | PopART | Richard.Hayes@lshtm.ac.uk |
| Josh Herbeck | University of Washington | Partners PrEP; Partners in Prevention | herbeck@uw.edu |
| Joseph Kagaayi | Rakai Health Sciences Program | Rakai Health Sciences Program | jkagayi@rhsp.org |
| Pontiano Kaleebu[1] | MRC/UVRI Uganda | MRC studies | pontiano.kaleebu@mrcuganda.org |
| Jairam Lingappa[1] | University of Washington | Partners PrEP; Partners in Prevention | lingappa@uw.edu |
| Vladimir Novitsky[1] | Harvard University | Botswana studies | vnovi@hsph.harvard.edu |
| Deenan Pillay[1] | Africa Health Research Institute / University College London | Principal Investigator PANGEA-1 / Executive Committee, Africa Health Research Institute studies | dpillay@ahri.org |
| Thomas Quinn | Johns Hopkins University | Rakai Health Sciences Program | tquinn2@jhmi.edu |
| Andrew Rambaut | University of Edinburgh | Executive Committee | a.rambaut@ed.ac.uk |
| Oliver Ratmann | Imperial College London | Analysis | oliver.ratmann@imperial.ac.uk |
| Janet Seeley | MRC/UVRI Uganda / LSHTM | MRC Uganda | Janet.Seeley@LSHTM.ac.uk |
| Deogratius Ssemwanga | MRC/UVRI Uganda | MRC studies | Deogratius.Ssemwanga@mrcuganda.org |
| Maria Wawer[1] | Johns Hopkins University | Rakai Health Sciences Program | mwawer1@jhu.edu |

[1] It is the responsibility of each Steering Committee member to inform the PANGEA 2 Project Manager of any changes to the contact details.

## 9.2 Appendix 2: PANGEA 1 Steering Committee

| Name | Institution | Study team representatives | Email address[3] |
|---|---|---|---|
| Myron Cohen | University of North Carolina | -------- | myron_cohen@med.unc.edu |
| Tulio D'Oliveira | University of KwaZulu-Natal | -------- | tuliodna@gmail.com |
| Ann Dennis | University of North Carolina | -------- | ann_dennis@med.unc.edu |
| Max Essex | Harvard Botswana | Botswana studies | messex@hsph.harvard.edu |
| Sarah Fidler | PopART/Imperial College London | PopART Phylogenetics | s.fidler@imperial.ac.uk |
| Dan Frampton[2] | University College London | -------- | d.frampton@ucl.ac.uk |
| Christophe Fraser | University of Oxford | PopART Phylogenetics | christophe.fraser@bdi.ox.ac.uk |
| Tanya Golubchik | University of Oxford | | tanya.golubchik@bdi.ox.ac.uk |
| Richard Hayes[1] | PopART/LSHTM | PopART Phylogenetics | Richard.Hayes@lshtm.ac.uk |
| Josh Herbeck | University of Washington | Partners PrEP; Partners in Prevention | herbeck@uw.edu |
| Anne Hoppe[2] | University College London | Project Manager PANGEA 1 / EARNEST | a.hoppe@ucl.ac.uk; hoppe.anne@gmail.com |
| Pontiano Kaleebu[1] | MRC/UVRI Uganda | MRC studies | pontiano.kaleebu@mrcuganda.org |
| Paul Kellam | Cambridge University | -------- | paul.kellam@kymab.com |
| Cissy Kityo | EARNEST/JCRC Uganda | EARNEST | ckityo@jcrc.org.ug |
| Andrew Leigh-Brown | University of Edinburgh | -------- | A.Leigh-Brown@ed.ac.uk |
| Jairam Lingappa[1] | University of Washington | Partners PrEP; Partners in Prevention | lingappa@uw.edu |
| Vladimir Novitsky[1] | Harvard University | Botswana studies | vnovi@hsph.harvard.edu |
| Nick Paton[1] | EARNEST / University of Singapore | EARNEST | nick.paton@ucl.ac.uk |
| Deenan Pillay[1] | Africa Health Research Institute / University College London | Principal Investigator PANGEA 1 / Africa Health Research Institute studies | dpillay@ahri.org |
| Tom Quinn | Johns Hopkins University | Rakai Health Sciences Program | tquinn2@jhmi.edu |
| Oliver Ratmann | Imperial College London | -------- | oliver.ratmann@imperial.ac.uk |
| Deogratius Ssemwanga | MRC/UVRI Uganda | MRC studies | Deogratius.Ssemwanga@mrcuganda.org |
| Frank Tanser | Africa Health Research Institute | -------- | ftanser@gmail.com |
| Maria Wawer[1] | Johns Hopkins University | Rakai Health Sciences Program | mwawer1@jhu.edu |

[1] To be contacted for authorship recommendations with regards to nominating one or more study group representative for the studies highlighted in column 3.
[2] Member of the PANGEA 1 Project Management Team.
[3] It is the responsibility of each Steering Committee member to inform the PANGEA 2 Project Manager of any changes to the contact details.

## 9.3 Appendix 3: Application for data access via a concept sheet proposal

# Application for PANGEA data access via a concept sheet proposal

PANGEA currently holds over 16,000 next generation sequences of HIV genomes samples between 2005 and 2016 in Eastern and Southern Africa, and associated metadata. The samples have been generated by the African Health Research Institute (South Africa), the Botswana-Harvard AIDS Institute Partnership (Botswana), the HPTN071 / PopART Phylogenetics study (Zambia), the Rakai Health Sciences Programme (Uganda), the MRC/UVRI & LSHTM Uganda Research Unit (Uganda) and University of Washington ICRC / Partners in Prevention (around Lake Victoria and South Africa).

Our aim is to establish an inclusive data sharing policy that at the same time respects the work that has gone into generating these samples. External researchers are welcome to apply for access via this proposal form. Six months after the proposal has been accepted and all documents have been provided, external researchers can apply to become accredited PANGEA researchers. A decision will be taken by the PANGEA Steering Committee.

| Name and affiliation of applicants<br>Please list all researchers who will have access to the data. |
|---|
| |

| Title of proposal |
|---|
| |

| Which data are you requesting access to? Please include cohort and fields of database. |
|---|
| |

| Please briefly outline the work you are planning to undertake, what the milestones are and what you envisage the time frame to be (one page, separately if you prefer). |
|---|
| |

| Please provide a risk-benefit mitigation table (separately if you prefer). |
|---|
| |

| Are you planning to collaborate with any PANGEA members on this proposal? |
|---|
| |

Please include a dated signature of each researcher who will have access to the data.

I agree to be bound by all clauses of the most current version of the PANGEA document “Principles of Research Conduct, Data Sharing, Accreditation, and Publication”. This includes me agreeing to

- collaborate where possible with members of the consortium where aspects of the proposal overlap with already existing strands of research.
- communicate progress every three months to the Steering Committee, through short written reports, shared slides and/or teleconferences as requested.
- follow PANGEA authorship and publication guidelines
- submit conference abstracts two weeks and paper manuscripts three weeks before submission.
- not share PANGEA data with non-accredited individuals or organisations without permission of the Steering Committee.
- not share PANGEA level 1 data obtained upon request with accredited researchers without permission of the institute that granted the request.
- provide a CV.
- provide a certificate for standard course on human subject research (e.g. CITI Biomedical Basics) which is less than three years old.
- familiarize myself with the studies by reading any study-specific information supplied by PANGEA.

Date and signature:

Please attach a CV and a certificate for a standard course on human subject research (e.g. the CITI Biomedical Basic course (https://about.citiprogram.org/en/series/human-subjects-research-hsr/)) less than three years old for each of the researchers that will have access to the data.

Please return the completed form to the PANGEA 2 project manager, Lucie Abeler-Dörner, lucie.abeler-dorner@bdi.ox.ac.uk

## 9.4 Appendix 4: Application to become an accredited PANGEA researcher

# Application to become an accredited PANGEA researcher

Our aim is to establish an inclusive data sharing policy that at the same time respects the work that has gone into generating these samples. External researchers are welcome to apply for access via this proposal form six months after their concept sheet proposal has been approved and they have successful updated the Steering Committee in two quarterly calls. A decision will be taken by the PANGEA Steering Committee.

Accredited researchers actively conduct research in a way that is respectful of Consortium aims and of the ethical requirements for research on HIV infected participants.

| Name and job title of applicant, head of research group if applicable |
|---|
| |

| Address of host institution, email of applicant, webpage of applicant if applicable |
|---|
| |

| Scientific or public health are of interest (e.g. Phylodynamics; Molecular Epidemiology and Modelling; Ethics, Drug Resistance and Clinical Science, Migration, Key Populations, Vaccine Design, …) |
|---|
| |

| **Example projects:** List up to four potential project titles (e.g. Using phylodynamics to estimate the rate of HIV superinfection and recombination; Molecular epidemiology and the role of acute and early HIV in transmission; The role of commercial sex work in sustaining HIV transmission; The fitness costs of efavirenz resistance mutations in different genetic backgrounds; …). Please note that you will need to submit separate abstracts for each project when work commences. |
|---|
| |

Code of conduct

I agree to be bound by all clauses of the most current version of the PANGEA document “Principles of Research Conduct, Data Sharing, Accreditation, and Publication”. This includes me agreeing to

- collaborate where possible with members of the consortium where aspects of the proposal overlap with already existing strands of research.
- specify my research area and example projects of interest.

- send a short abstract outlining the proposed work to the Steering Committee and post it on the forum before embarking on an analysis.
- communicate progress every three months to the Steering Committee, through short written reports, shared slides and/or teleconferences as requested.
- follow PANGEA authorship and publication guidelines
- submit conference abstracts two weeks and paper manuscripts three weeks before submission.
- not share PANGEA data with non-accredited individuals or organisations without permission of the Steering Committee.
- not share PANGEA level 1 data obtained upon request with accredited researchers without permission of the institute that granted the request.
- provide a CV.
- provide a certificate for standard course on human subject research (e.g. CITI Biomedical Basics) which is less than three years old.
- familiarize myself with the studies by reading any study-specific information supplied by PANGEA.

Date and signature:

Please attach a CV and a certificate for a standard course on human subject research (e.g. the CITI Biomedical Basic course (https://about.citiprogram.org/en/series/human-subjects-research-hsr/)) less than three years old.

Please return the completed form to the PANGEA 2 project manager, Lucie Abeler-Dörner, lucie.abeler-dorner@bdi.ox.ac.uk

## 9.5 Appendix 5: Creating a risk-benefit-mitigation table

Structure of a risk-benefit-mitigation table:

| Risk | Benefit | Mitigation strategy |
|---|---|---|
| | | |

If you are using fully anonymized data, there might be no risk and therefore no need for a mitigation strategy. However, most projects will use at least some epidemiological data. We would like to invite you to start by considering the risks listed below. What benefits do you expect from the analysis? Do the benefits justify the risks? How can you minimise the risks without compromising the benefits? It is very likely that your specific analysis will require the consideration of additional risks.

Potential risks:

- Loss of laptop with data
- Misinterpretation of epidemiological context
- Findings that result in potential harm to individuals or populations, e.g. a transmission analysis that finds a large cluster of men in a country in which homosexuality is illegal
- Data are request by local police
- Misinterpretation of results in a newspaper article, e.g. on migration, reinforcing stigmatisation