

Ethical and Equitable Data Sharing: Navigating the Benefits and Challenges

Nicki Tiffin
SANBI, UWC
ntiffin@sanbi.ac.za



Overview:

- Why should we do data sharing?
- When can we do data sharing?
- How can we do data sharing?
- How can we do benefit sharing?
- Introduction to the African Data and Biospecimen Exchange

Why should we do data sharing?

Participants make health research possible

- We ask them to invest time and energy:

Clinical appointments, filling in surveys, answering questions, follow up visits

- We ask them to take risks:

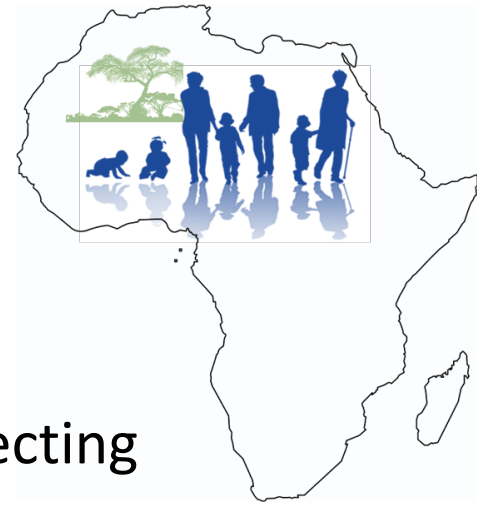
Data privacy, new treatments, health interventions, donate samples



Why should we do data sharing?

It's an ethical requirement to ensure benefits

- Use the data and specimens well
- Do as much research as possible, whilst respecting consents and ethical approvals
- Ethically share data and biospecimens in order to further research



Why should we do data sharing?

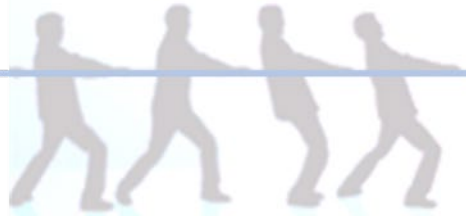
It's better for our research

- Larger sample size, better significance
- Validation datasets
- Generalisable findings



Finding a balance

Benefits



Public Health
Common Good
Scientific Progress

Risks



Autonomy
Privacy

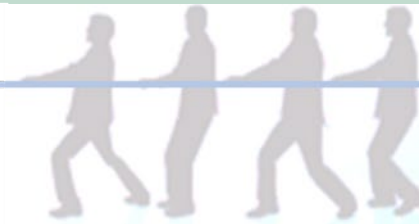
Finding a balance

Benefits



Public Health
Common Good
Scientific Progress

Risks



Autonomy
Privacy

SDGs

3 GOOD HEALTH
AND WELL-BEING



UNHR

Protection from discrimination
Protection of privacy, reputation



When can we do data sharing?

When the law allows

- Health Act, patient confidentiality
- Protection of Personal Information Act
- In addition, various other Acts e.g. protection of minors, protection of people living with disability

When can we do data sharing?

In clinical research:

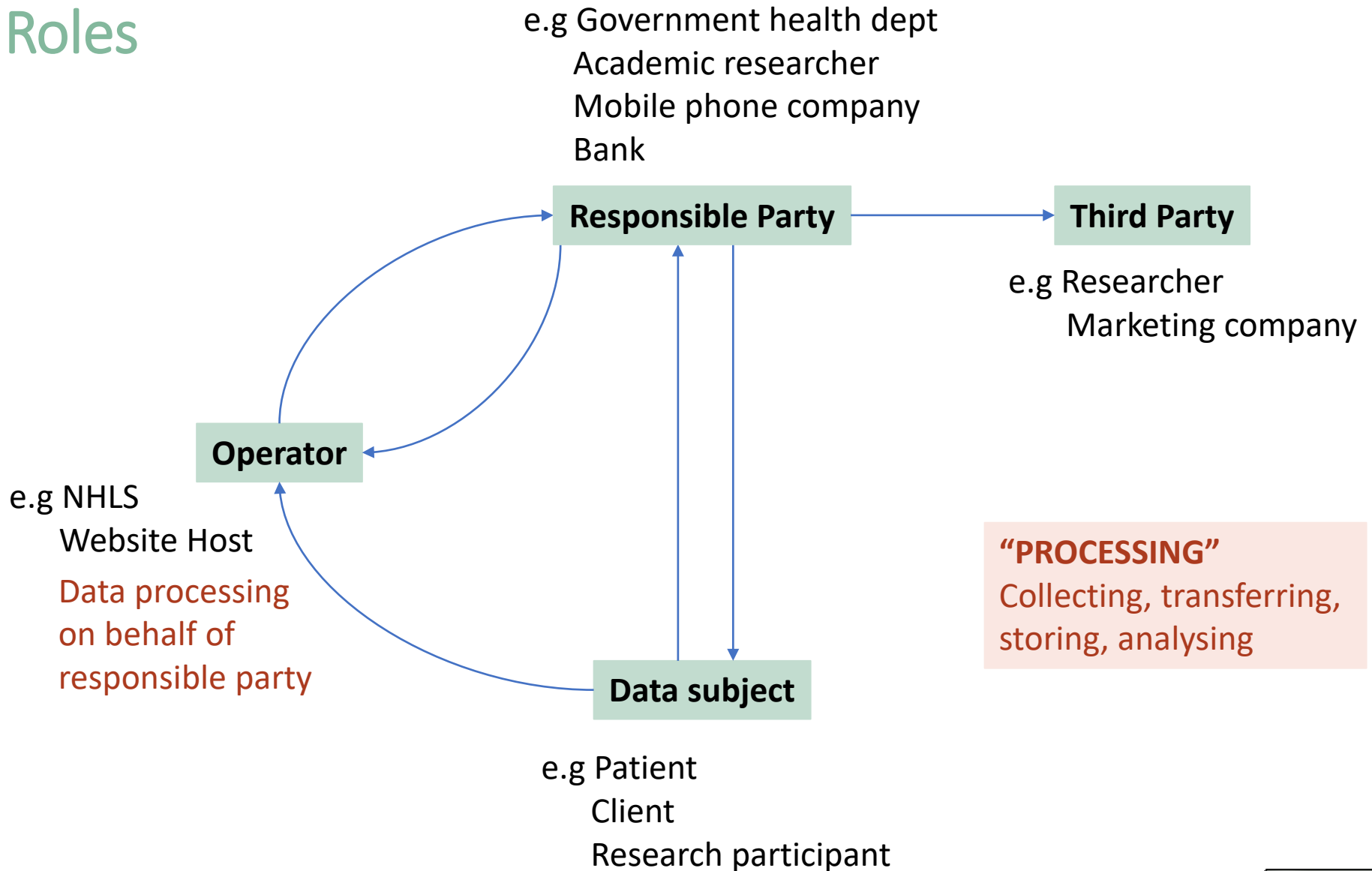
- Health data are collected at a health facility where a client has a consultation with a health provider.
- The primary reason that the data are collected is to ensure that the best possible care is provided to the client.
- The health care provider is bound by legislation about patient confidentiality, e.g. Health Act

The POPI Act in South Africa

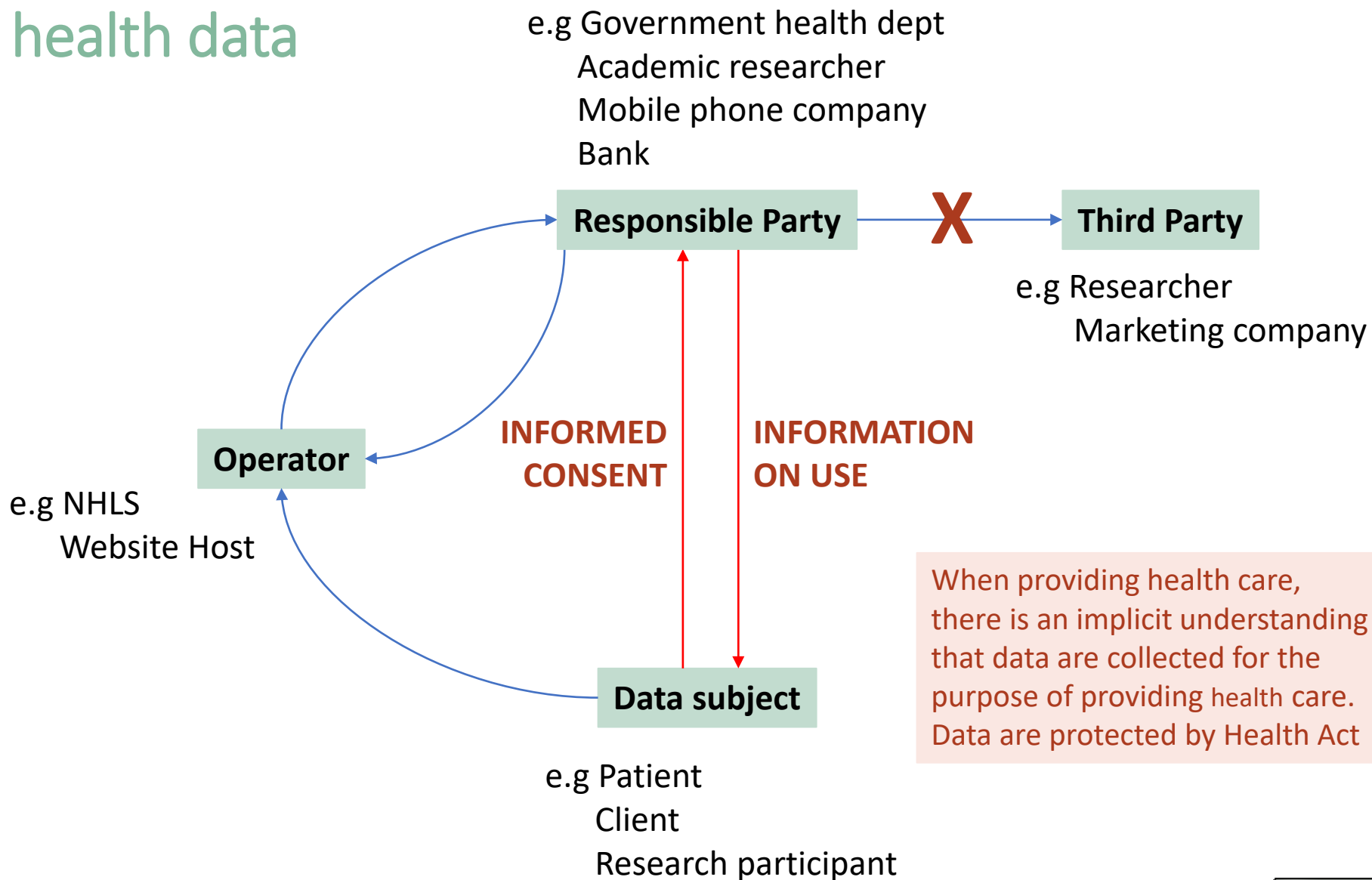
The **Protection of Personal Information Act (POPI)**, South Africa (equivalent to GDPR in EU)

- Governs the use of personal information
- Synergises with the Health Act, upholds patient confidentiality
- Identifies “special” data, which include health data
- Governs international transfer of data
- Distinguishes identified data vs anonymised data
- Recognises informed consent

The POPI Act Roles



Identified health data



When providing health care, there is an implicit understanding that data are collected for the purpose of providing health care. Data are protected by Health Act

When can we do data sharing?

When the participant agrees

Informed consent processes

- Many debates about broad, tiered, dynamic consent
- Generally accepted broad consent is no longer sufficient
- Dynamic consent hard to do in low-resourced populations

Tiered consent increasingly being used

Especially for health data: must adhere to POPI Act



Tamuhla, T. **An e-consent framework for tiered informed consent for human genomic research in the global south, implemented as a REDCap template.** *BMC Med Ethics* **23**, 119 (2022). <https://doi.org/10.1186/s12910-022-00860-2>

DATABASE

Open Access



An e-consent framework for tiered informed consent for human genomic research in the global south, implemented as a REDCap template

Tsaone Tamuhla¹, Nicki Tiffin^{1,2,3*}  and Taryn Allie¹

Abstract

Research involving human participants requires their consent, and it is common practice to capture consent information on paper and store those hard copies, presenting issues such as long-term storage requirements, inefficient retrieval of consent forms for reference, and the risk of lost or damaged consent forms. This study describes an e-consent framework for capturing informed consent. The

Primary consent for collecting biospecimens and health data for specific disease in current study
Consent for access to medical records
Consent for return of individual results
Consent for return of individual results that are actionable and/or treatable
Consent for return of individual results that are NOT actionable and/or treatable
Consent for inclusion of individual data in genetic summary data
Consent for use of genetic and health data for future studies on specific disease
Consent for use of genetic and health data for future studies on other health conditions or related health processes
Consent to re-contact for future studies
Consent for use of genetic and health data in international studies
Consent for use of genetic data in population origins and ancestry studies



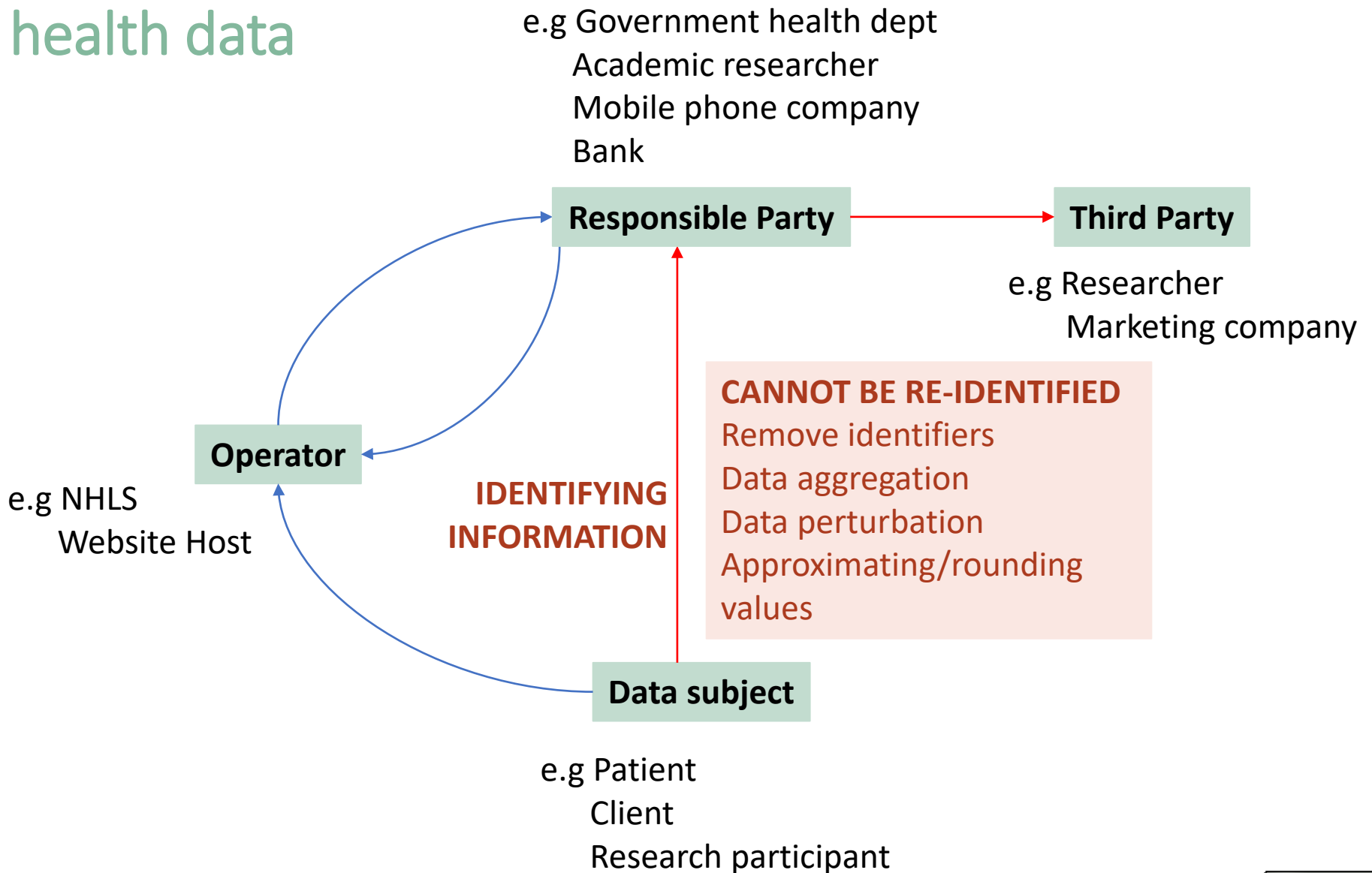
Nicki Tiffin

When can we do data sharing?

When the participant can't be identified/re-identified

- Data are **de-identified**: names, ID numbers, folder numbers are removed
- Data are **anonymised**: de-identified and can't ever be re-linked to individuals
- Re-identification is a real risk for granular data (and genomic data):
- Use of data perturbation

Anonymised health data



Can data be re-identified?

A 37-year old lady with epilepsy attending a particular clinic on given dates.

A 12 year old, female, grade 6 learner who lives at 1 Green Street, Townsville.

Female baby born on 7 December 2021 at Mowbray Maternity clinic, recorded birthweight 3.4921 kg

- Do not manage de-identified data in the same way as truly anonymised data, because they can be re-identified.

Can data be re-identified?

Data perturbation

- Adding/deleting integers from dates
- Hide, round off or alter dates e.g. Year of birth
- Age scale e.g. days for neonates, weeks for newborns, months up to 2yrs, years thereafter
- 'time to' events in days from index event. e.g. Time in days to death after admission
- Round off numbers e.g. birthweights, VL or CD4 counts
- **K ANONYMITY:** For every identifying attribute a person has, there are at least $k-1$ others with the same value e.g. If at least ten people have a value of $X_1=0.15$, then k is 10



Can data be re-identified?

Trusted third party stewardship

- Independent, trusted third party joins datasets, then anonymises and perturbs before returning combined dataset
- Binding MoU/agreement for third party to delete all data

Data aggregation

- Rule of thumb: minimum 15 – 20 counts per aggregated data

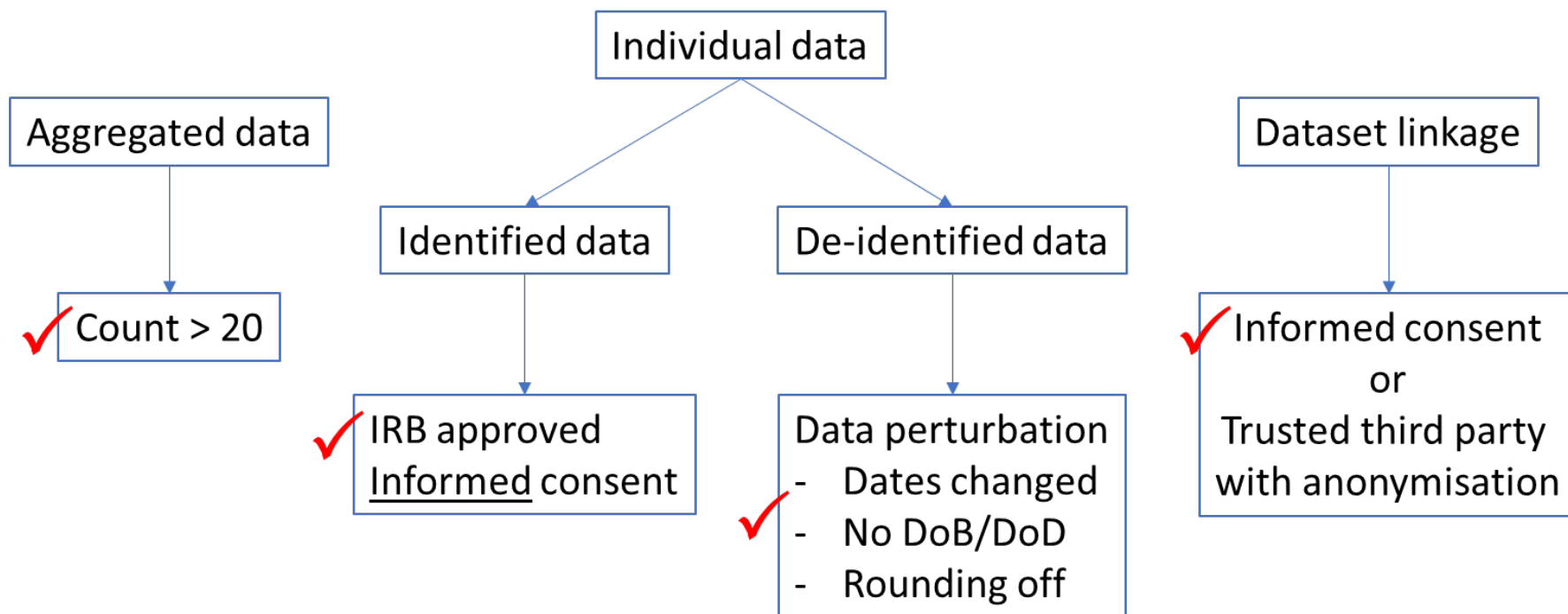
Geographical data

- Geocoding can be identifying, use geographic regions, shapes, suburbs, and show aggregated data

NO DOTS ON MAPS



When can we do data sharing?

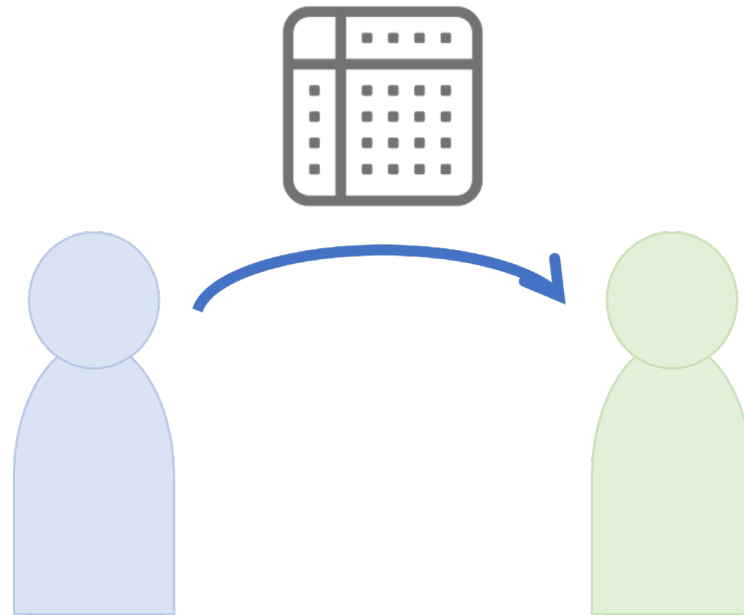


How can we do data sharing?

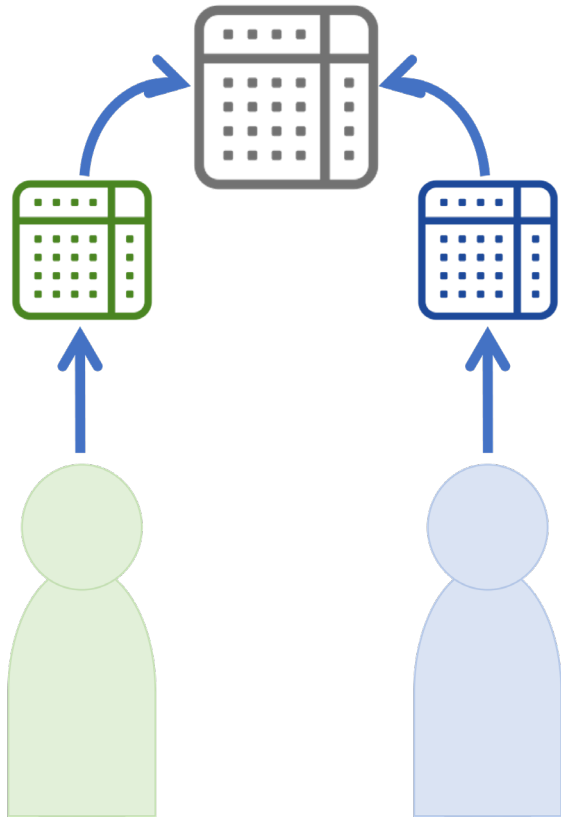
- Direct sharing
- Collaborative analysis
- Federated analysis
- Trusted Research Environments

Direct data sharing

- One party provides data to another party
- Unidirectional



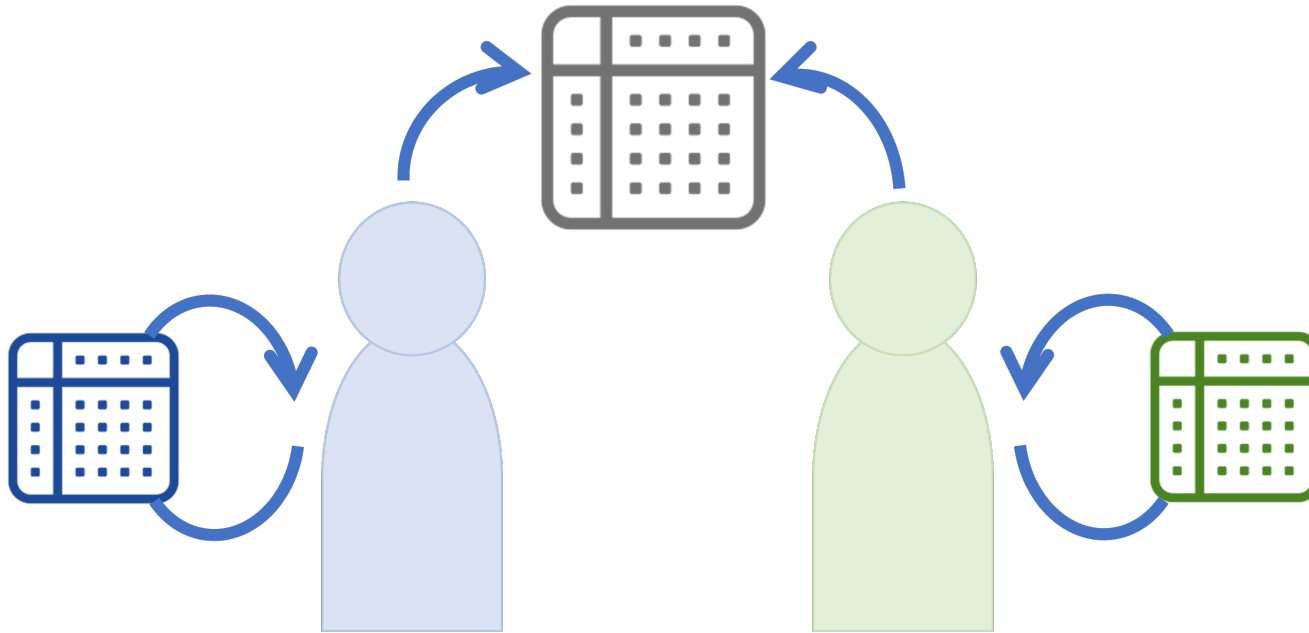
Collaborative analysis, meta-analysis



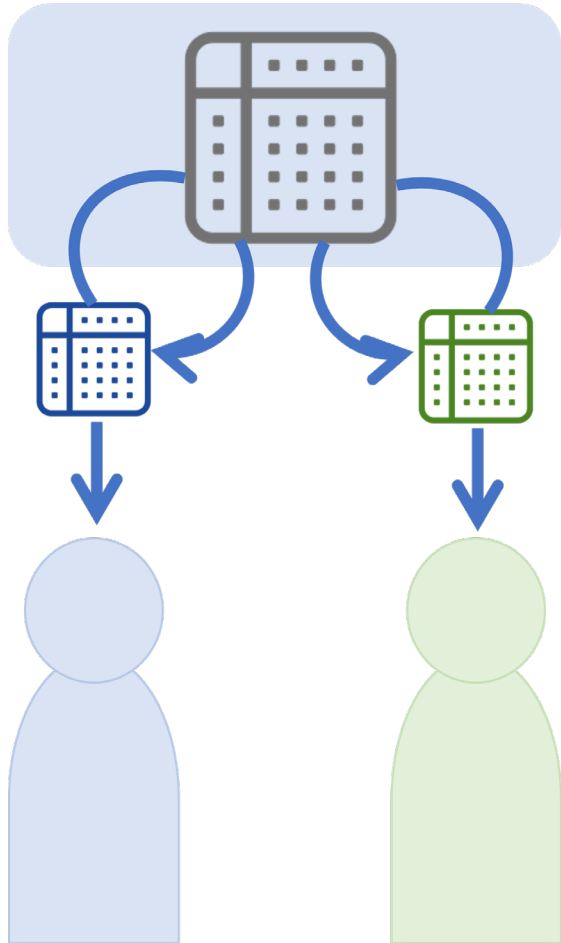
- Data from two sources are combined
- Analysed as a single dataset: meta-analysis

Federated analysis

- Datasets are held separately by collaborating parties
- Data are independently analysed in the same way
- Findings are combined and reported jointly

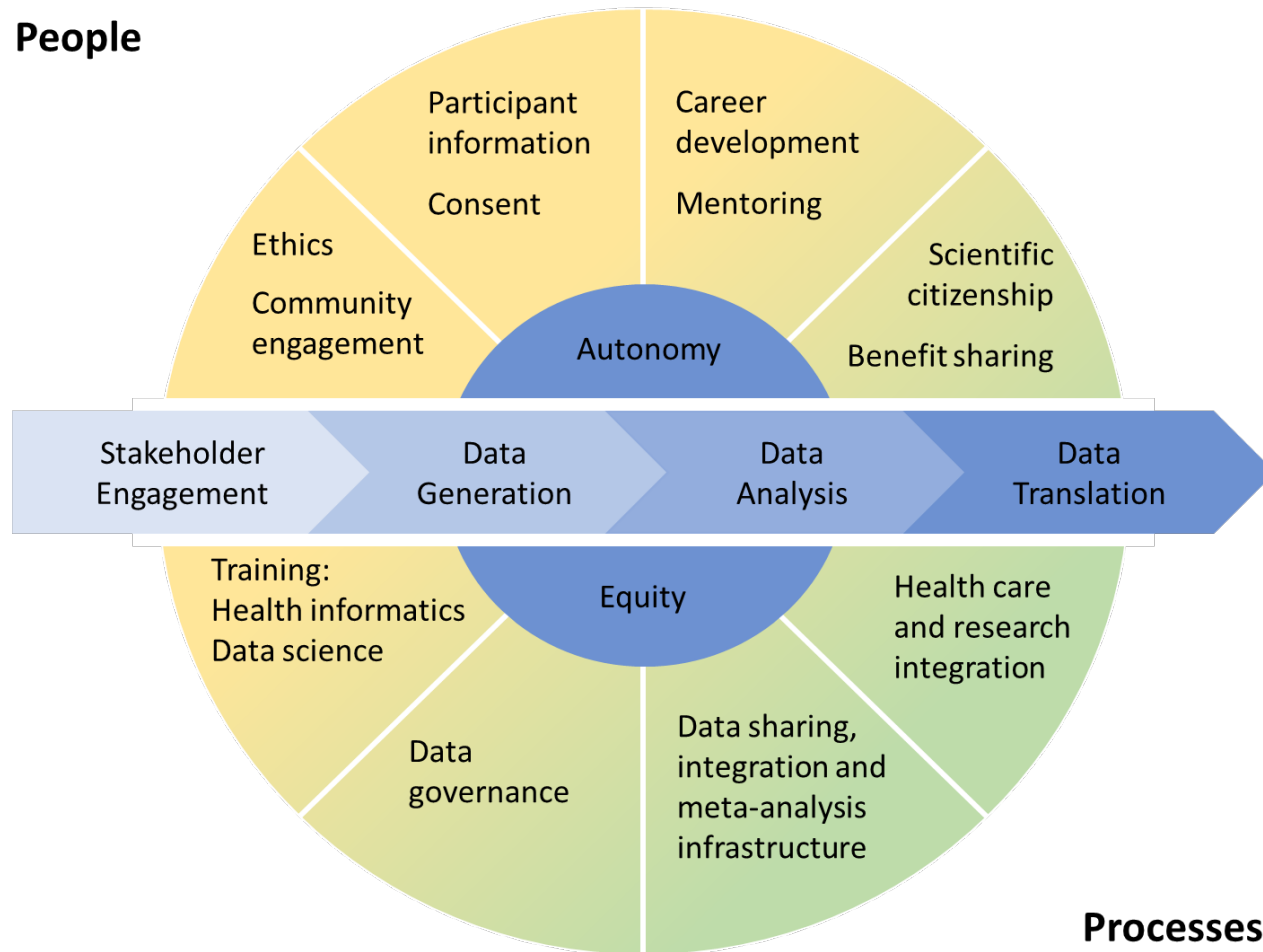


Trusted research environment



- Analysis is run on a secure platform
- Actual data may not be visible
- Data may not be downloaded
- Only results of analysis may be downloaded

Data and biospecimen governance occurs across the data ecosystem



Data/biospecimen sharing in the global South



Inequitable power relationships

- Funders and collaborators from the global North
- Inequitable access to resources and infrastructure
- Helicopter science and “global health” agenda
- Unidirectional flow of samples and data to the global North
- Funding requirements mandate centralising samples and data in global North
- No benefit-sharing from secondary use
- Researchers can't access data or samples they collected

What about data and biospecimens from Africa?

Historical inequities, current barriers

- Reluctance to send data and biospecimens off-Continent

- Legal limitations

- Risk of being scooped by better resourced labs

- No oversight of onward data use

- Lose access to own data/biospecimens

- Informed consent may not be sufficient

- Legacy data without explicit consent

- Anonymised routine health data without informed consent

What about data and biospecimens from Africa?

Historical inequities, current barriers

- Data are too sensitive for open sharing

Genomic data

Clinical trial data

Personal health data

- Research sustainability

No mechanisms for cost recovery

Benefit sharing is not inculcated in resource sharing



The African Data and Biospecimen Exchange



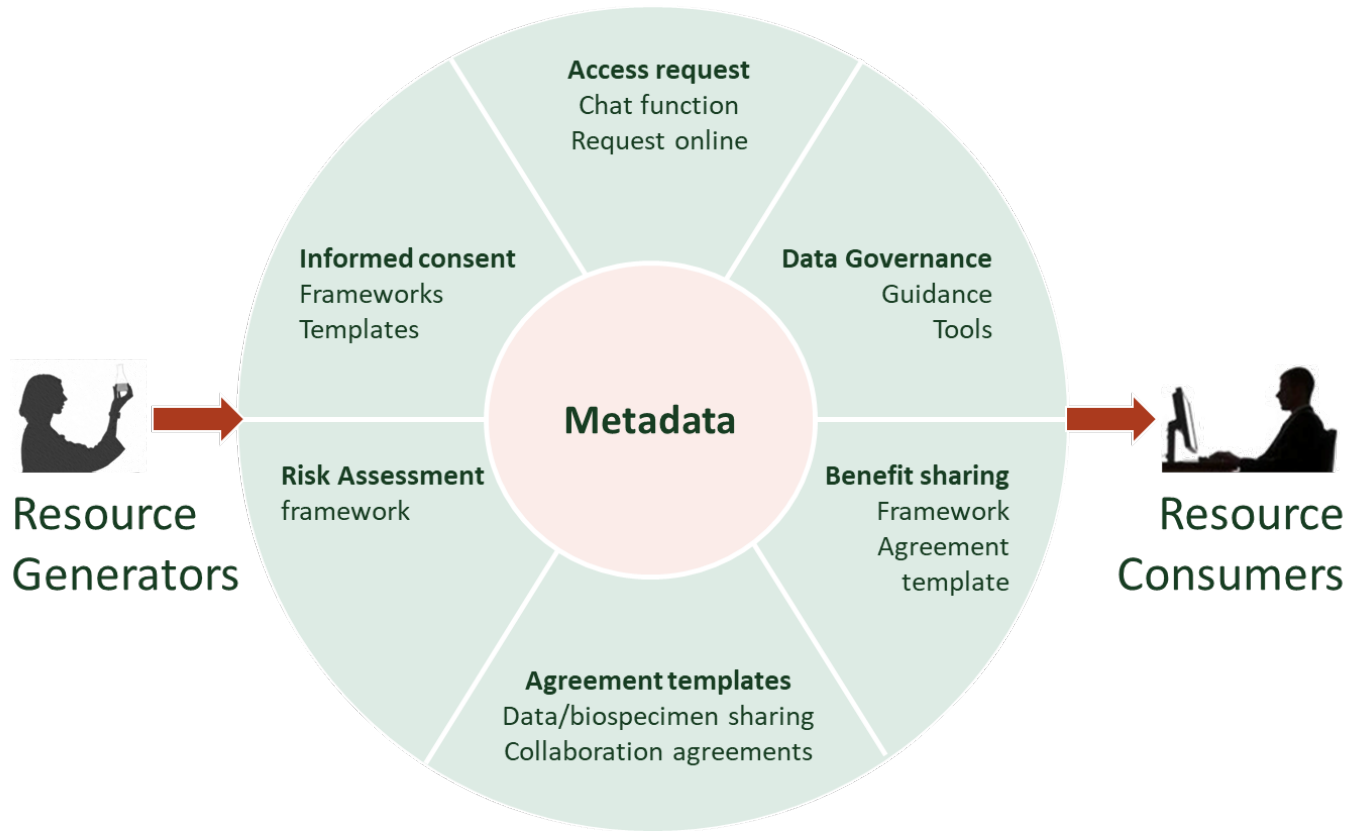
ADBEEx

- Online platform to connect resource generators and consumers
- Address barriers to sharing data and biospecimens from Africa
- Ensure equitable sharing agreements and benefit-sharing



The African Data and Biospecimen Exchange

- Catalogue datasets and biospecimens (resources) by metadata only
- Connect resource generators directly to consumers via online platform
- Provide resources to assist users with governance, sharing agreements



No need to send data and biospecimens away

- No centralisation
- **Only metadata uploaded**
- Resources remain with researcher/ institution
- Direct transfer, only after agreement reached
- Create listings of data or biospecimens you want to share

The screenshot shows a web interface for creating a new listing. The main form is titled "Create new listing" and includes a sub-header: "Add a new post and we will alert those who showed interest in your shared tags." The form is divided into several sections:

- Title:** A text input field containing "Human DNA".
- Status:** A dropdown menu set to "Draft - Ready to publish", with "Preview" and "Publish" buttons.
- Progress Indicators:** A vertical list of steps with checkboxes: "Essential details", "Sample and provider details", "Participant(s) data details", "Biospecimen and data set details" (selected), and "Governance checklist".
- Biospecimen Section:**
 - Consent type:** Radio buttons for "Specific consent" (checked), "Unknown", "General consent", and "N/A".
 - Sample size:** A text input field containing "Aliquot no" and a dropdown menu set to "cm" with the label "Size, volume or mass".

A detailed sidebar on the right, titled "Add to your listing", provides further options for each section:

- Biospecimen:** Checkboxes for "Consent type" and "Sample size", with "Read more" and "Remove from listing" buttons.
- Specimen collection data:** Checkboxes for "Sample count", "Collection institution", "Collection dates", "Sample type", "Morphological code", and "Topographical code", with "Read more" and "Add to listing" buttons.
- Sample processing data:** Checkboxes for "Processing specimen category", "Processing temperature conditions", "Sample type", "Sample container", "Fixation/ Sstabilization type/ Freezing mode", "Sample quality methods", and "Sample QC", with "Read more" and "Add to post" buttons.
- Storage data:** Checkboxes for "Storage date & time" and "Storage temperature", with "Read more" and "Add to listing" buttons.
- Distribution data:** Checkboxes for "Retrieval date", "Retrieval reason(s)", "Project destination", and "Delivery temperature", with "Read more" and "Add to listing" buttons.

A "Next" button is located at the bottom right of the sidebar.

No need to send data and biospecimens away

- No centralisation
- **Only metadata uploaded**
- **Resources remain with researcher/ institution**
- **Direct transfer, only after agreement reached**
- Create listings of data or biospecimens you want to share

Create new listing
Add a new post and we will alert those who showed interest in your shared tags.

Title
Human DNA

Status
Draft - Ready to publish
[Preview](#) [Publish](#)

Essential details
Please provide essential details for your post.

Sample and provider details
Details on the sample and provider

Participant(s) data details
Start collaborating with your team

Biospecimen and data set details
Start collaborating with your team

Governance checklist
Share posts to your social accounts

Biospecimen

Consent type
 Specific consent Unknown
 General consent N/A

Sample size
Aliquot no
cm Size, volume or mass

Add to your listing
Lorem ipsum dolor sit amet.

Biospecimen

Consent type
 Sample size
[Read more](#) [Remove from listing](#)

Specimen collection data

Sample count
 Collection institution
 Collection dates
 Sample type
 Morphological code
 Topographical code
[Read more](#) [Add to listing](#)

Sample processing data

Processing specimen category
 Processing temperature conditions
 Sample type
 Sample container
 Fixation/ Sstabilization type/ Freezing mode
 Sample quality methods
 Sample QC
[Read more](#) [Add to post](#)

Storage data

Storage date & time
 Storage temperature
[Read more](#) [Add to listing](#)

Distribution data

Retrieval date
 Retrieval reason(s)
 Project destination
 Delivery temperature
[Read more](#) [Add to listing](#)

[Next](#)

Search metadata to find data and biospecimens

Search filters
Keyword searches
Review returned listings

Start chat

The screenshot shows a search results page with a search bar at the top containing the text "Ebola". Below the search bar are navigation tabs for "Popular", "Human", "Microbial", "Animal", and "Other". A list of search results is displayed, each with a profile picture, author name, date, title, and a brief description. The first result is "SLE, whole genome dataset for 80 South Africans" by Olivia Rhye, dated 20 Jan 2022. Other results include "M.tb cultures, Ghana" by Kagee Jain, "WGS data, chronic kidney disease" by Lame Moses, and "Maternal outcomes, Tanzania" by Theo Phapani. Each result has tags for "SARS 19", "South Africa", "Ghana", "Burundi", "Chronic Kidney Disease", "Tanzania", "DNA", and "Dataset". A "Filters" panel on the right allows filtering by "Type" (Dataset, Biospecimen) and "Category" (Human). A pagination bar at the bottom shows "Previous" and page numbers 1 through 10.

The screenshot shows a detailed view of the "SLE, whole genome dataset for 80 South Africans" dataset. The page header includes the dataset title, author "@Olivia Rhye, 12 January 2022", and action buttons for "Agreements" and "Message". The main content is divided into sections: "Overview" (Specifications, Governance), "Research interests" (Human, South Africa, DNA), "Data set overview" (This dataset contains Whole Genome Sequence for 80 participants from South Africa who have symptomatic systemic lupus erythematosus.), "About the data set" (DNA samples were prepared from whole bloods donated by 80 study participants over the age of 18 years who have symptomatic SLE. Participants were recruited at rheumatology and nephrology clinics at tertiary hospitals in the Western Cape, South Africa. Some clinical data on disease severity is available for the participants. 59 of the dataset are women DNA was sequenced on the Illumina platform to an average depth of 15x. Informed consent has been given by participants for use of these samples in secondary studies of SLE.), and "Further studies" (An analysis of the genetic variants found in autoimmunity-associated genes has been published in Brown et al. 2022, Journal name, PMID: XXXX. An ongoing collaboration is exploring novel aetiological variants in these patients in a case-control study). A "Popular" sidebar on the right lists other datasets like "Gorilla beringei, DRC, DNA" and "Genetic modification study". A "Share this listing" section at the bottom provides a share link: "abdex.co.za/listing/sle_whole...".

Search metadata to find data and biospecimens

Search filters
Keyword searches
Review returned listings
Start chat

The screenshot shows a search interface with a search bar containing the text "Ebola". Below the search bar are tabs for "Popular", "Human", "Microbial", "Animal", and "Other". A "Filters" panel is open, showing options for "Type" (Dataset, Biospecimen) and "Category" (Human). Search results are listed below, including "SLE, whole genome dataset for 80 South Africans" by Olivia Rhye and "M.tb cultures, Ghana" by Kagee Jain.

The screenshot shows a detailed listing for the "SLE, whole genome dataset for 80 South Africans" by Olivia Rhye. The page includes a "Data set overview" section with a description: "This dataset contains Whole Genome Sequence for 80 participants from South Africa who have symptomatic systemic lupus erythematosus." It also features an "About the data set" section, "Further studies" section, and a "Popular" sidebar with related datasets like "Gorilla beringei, DRC, DNA" and "Genetic modification study". A "Share this listing" section is also present.

Dr Eddie Lulamba



Chat function: discuss and agree sharing mode

Templates

You choose and customise a template that fits your agreement, following prompts and using agreement elements provided through the platform.

As you compile your agreement you can:

- Import information from profiles and listings on the platform
- Access resources and information to assist writing the agreement
- Invite collaborators and institutional representatives to view, edit and/or sign the agreement

Direct transfer

The data or biospecimens will be transferred to the requestor and the data provider has no further involvement in the onward analysis.

Get started

Preview

- Unidirectional share
- Review ethics approvals
- Agree permitted onward use
- Negotiate benefit sharing
- Agree acknowledgements

Collaborative

Agree to combine data or samples and undertake a joint analysis of the combined resource.

Get started

Preview

- Bi-directional sharing
- Review ethics approvals
- List participating researchers
- Identify researcher roles
- Agree analysis methodology
- Negotiate benefit sharing
- Define outputs

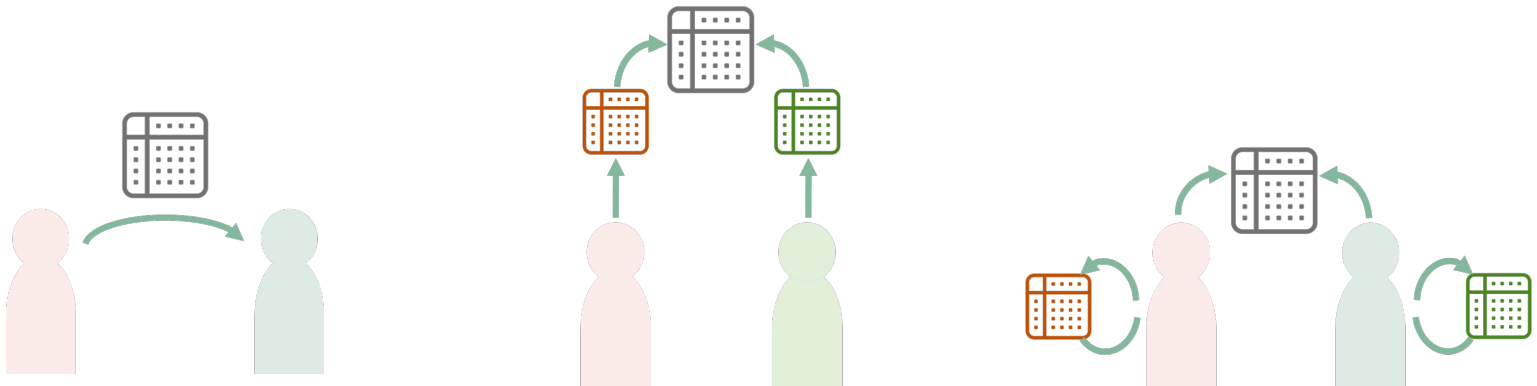
Federated

Agree to combine standardised analytical output from analysis of separate datasets or biospecimens.

Get started

Preview

- No sharing of data or biospecimens
- Agree analytical approach
- List participating researchers
- Identify researcher roles
- Define results to be combined
- Define IP and outputs
- Negotiate benefit sharing



Build a data- or biospecimen-sharing agreement

- Elements to create a customised data- or biospecimen-sharing agreement.
- Invite institutional representatives
- Negotiate benefit sharing
- Build a benefit sharing agreement

Data Sharing Agreement
You are able to comment, edit and sign. [Request edit access](#)

0/5 Signature

Project Title
Diabete type II in Human DNA from ...
8 comments

Prepared by
Margaret Leblanc John Koffi
Who is compiling this agreement
8 comments

Prepared by
Margaret Leblanc John Koffi
Who is compiling this agreement

Proposed Research Output
Enter a description...
Max 500 characters

Intellectual Property Owner
UCT UWC
Who the intellectual property for this project resides with

Ethics Considerations
Click to upload or drag and drop
SVG, PNG, JPG or GIF (max. 800x400px)

Send

Customise Sign Agreement

Customise your Agreement
Lorem ipsum dolor sit amet.

Project specific aim
Read more Add to MOU

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?
Odo dictumst tempus magna elit cras posuere cursus pulvinar id. Facilisis at eu amet ornare enim arcu malesuada rutrum a.

Proposed research output
Read more Remove

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?
Odo dictumst tempus magna elit cras posuere cursus pulvinar id. Facilisis at eu amet ornare enim arcu malesuada rutrum a.

Intellectual property owner
Read more Remove

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?
Odo dictumst tempus magna elit cras posuere cursus pulvinar id. Facilisis at eu amet ornare enim arcu malesuada rutrum a.

Ethics considerations
Read more Remove

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?

Build a data- or biospecimen-sharing agreement

- Elements to create a customised data- or biospecimen-sharing agreement.
- Invite institutional representatives
- Negotiate benefit sharing
- Build a benefit sharing agreement

(Dr) Tsaone Tamuhla



Data Sharing Agreement
You are able to comment, edit and sign. [Request edit access](#)

0/5 Signature

Project Title
Diabete type II in Human DNA from ...
8 comments

Prepared by
Margaret Leblanc, John Koffi
Who is compiling this agreement
8 comments

Prepared by
Margaret Leblanc, John Koffi
Who is compiling this agreement

Proposed Research Output
Enter a description...
Max 500 characters

Intellectual Property Owner
UCT, UWC
Who the intellectual property for this project resides with

Ethics Considerations
Click to upload or drag and drop
SVG, PNG, JPG or GIF (max. 800x400px)

Send

Customise Sign Agreement

Customise your Agreement
Lorem ipsum dolor sit amet.

Project specific aim
Read more Add to MOU

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?
Odo dictumst tempus magna elit cras posuere cursus pulvinar id. Facilisis at eu amet ornare enim arcu malesuada rutrum a.

Proposed research output
Read more Remove

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?
Odo dictumst tempus magna elit cras posuere cursus pulvinar id. Facilisis at eu amet ornare enim arcu malesuada rutrum a.

Intellectual property owner
Read more Remove

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?
Odo dictumst tempus magna elit cras posuere cursus pulvinar id. Facilisis at eu amet ornare enim arcu malesuada rutrum a.

Ethics considerations
Read more Remove

What is it for?
Ipsum dolor sit amet, consectetur adipiscing elit.

When to use it?

Progress to date

End of 2022 Completed detailed specification:

Front end design – Hominum Global team

- User interface design and functionality, wireframing
- Metadata structure
- Detailed metadata structure with ontologies for human and microbial data
- Search term functionality
- User testing

Back end specification – MethodLab team

- Containerised (docker), can be redeployed and shared when complete
- SQL Relational database, ease of admin and maintenance
- Optimised for restricted internet bandwidth (loaded to user end for search)
- Ilifu local cloud server infrastructure: www.ilifu.ac.za

Next steps

Platform development – MethodLab team

- Started 1 June 2023
- 18 month timeline

First stakeholder workshop

- Entebbe, Uganda 10 – 12 July 2023
- Focus group – genomics researchers
- Open day at MRC Uganda/LSHTM co-hosted with Dr Segun Fatumo

Acknowledgements



ADBEEx

African Data and Biospecimen Exchange

Made possible by a Calestous Juma Fellowship (N Tiffin)
awarded by the **Bill & Melinda Gates Foundation**

Hominum.

www.hominum.global

Hominum Global

User interface and platform development

Nicol Ronga and Maria Cavanna

<https://hominum.global/>

Cape Town

Methodlab.

MethodLab

Platform Software Development

Tim Smith and Brendon Joseph

info@methodlab.io

Cape Town

**Also recognising funding from:
UKRI/MRC UK (MC_PC_22007)**

