SAMRC/UJ Pan African Centre for Epidemics Research (PACER) Extramural Unit

Refilwe Phaswana-Mafuya, MA(SW), PGD(Epi), MSc(Epi), PhD

Director, SAMRC PACER EMU; Professor, Epidemiology & Public Health

PACER
PAN AFRICAN CENTRE FOR EPIDEMICS RESEARCH

PACER
PAN AFRICAN CENTRE FOR EPIDEMICS RESEARCH

Harmonization of multiple HIV related Data Sources in Sub-Saharan Africa – Lesson Learned from The Boloka Project

RN Phaswana-Mafuya, E Phalane, A Rao , K Willis , K Rucinski, KA Voet, A. Abdulrahman, C. Siyamayambo, B. Sebati, L. Olifant, M. Seloka, M. Jaiteh, R Abdulai, KS Journeay, HI Sisel, F. Wolmarans,L. Motsieloa, M. Rampilo, X Li, B Olatosi, S.D. Baral

Presentation Overview

Context – Data Partners, Data Sources, Data Types

The Boloka Project Data Harmonization Approach

Lessons Learned – Seven lessons so far about harmonization

Closing Remarks

Context – Rationale and Relevance

- This presentation describes the harmonization **process, methods, principles and approaches** used in The Boloka Project to integrate data from different sources as well as **lessons learned**

- The rationale is to enable provision of timely information at **a level and scale** that will improve understanding of HIV heterogeneities for a **targeted HIV response, improved efficiencies, and greater impact** - NSP HIV/TB/STIs 2023-2028

- This includes understanding available data to determine: **HIV size estimates, HIV predictors and risk heterogeneities, as well as models, methods and potential associations of data and precision**

- The need **to prepare for, and respond to, as well as prevent current and future pandemics** and their long-term devastating direct and indirect impacts is apparent.

South African National AIDS Council, 2023; Phaswana-Mafuya et al, 2023

Context – The data for harmonization is being secured from data partners as part of Stage 1 of The Boloka Project
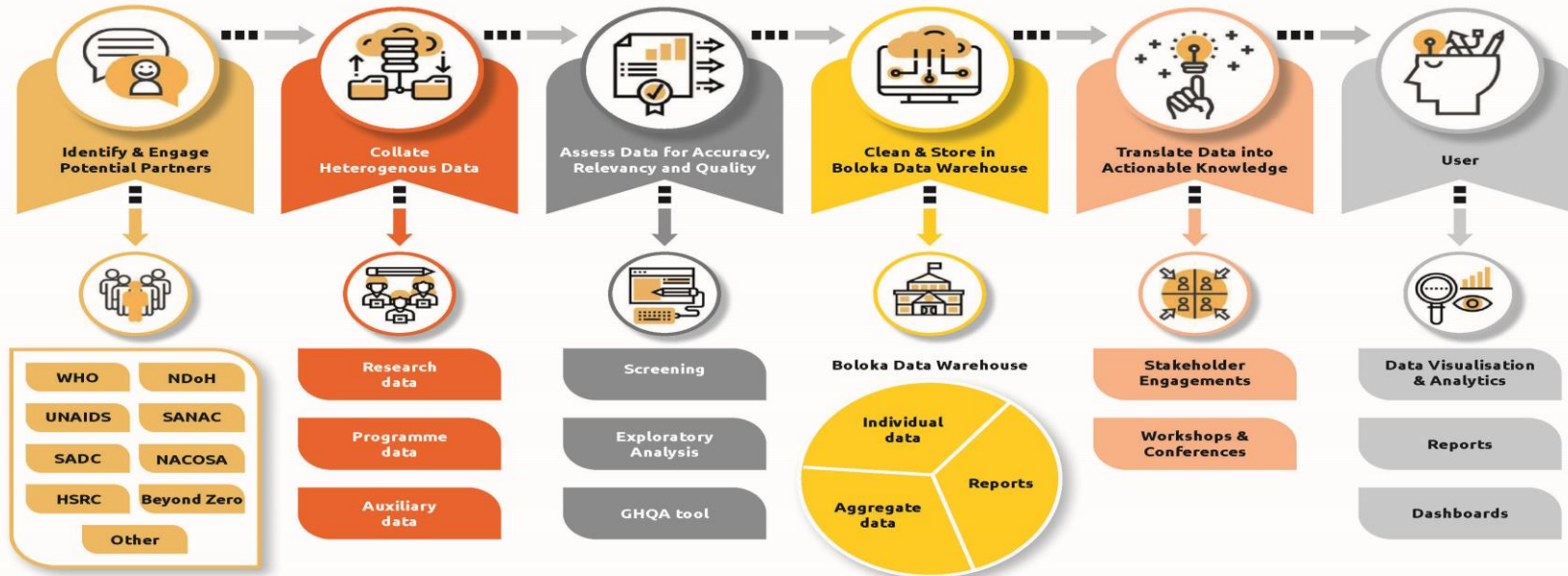

Five stages of the Boloka Data Warehouse

WHO - World Health Organization; NDoH - National Department of Health; UNAIDS - Joint United Nations Programme on HIV/AIDS; SANAC -South African National AIDS Council; NACOSA - Networking HIV and AIDS Community of Southern Africa; HSRC - Human Sciences Research Council; GHQA - Global HIV Quality Assessment Tool

Research data is conducted to answer priority HIV questions for the country, observational, implementation science, and experimental
Programme data -Routinely collected data at the facility, district level and countrywide captured by partners for government reporting, and population size estimate
Auxiliary data-Census data, government department annual reports, national policies, guidelines and strategies

(Wilkinson M, Dumontier M, Aalbersberg I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018)

Phaswana-Mafuya et al, 2023
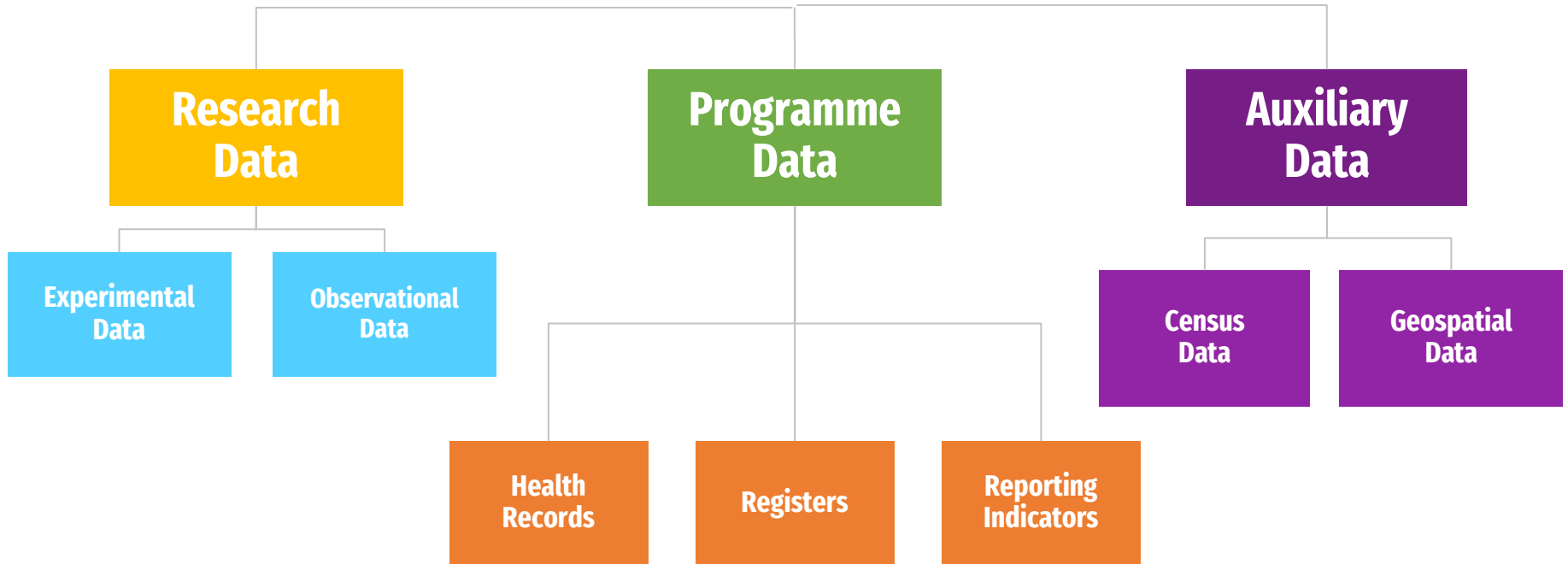
Harmonization Lesson 1:

You have to map your data and clearly understand how your data sources will interact with your data infrastructure.

There should be enough data storage capacity to avoid lagging and crashing of site, overlapping record ID's, etc.,

There should be data security, safety and control mechanisms in place

Considerations regarding affordability and sustainability needs to be made

We developed The Boloka HIV-related Indicator List – UNAIDS Global Monitoring Tool and NDOH Indicator Dataset – Bello et al, 2020

Boloka HIV-related Indicator List

HIV prevalence	Antiretroviral therapy	Sexually transmitted Infections
HIV incidence	Viral load testing	TB Co-infection
HIV testing	Prevention of mother-to-child transmission	Contraception
HIV Prevention	Voluntary male medical circumcision	Mental health
Condom use	Gender-based violence	Substance use
Pre-exposure prophylaxis	Stigma and discrimination	Safe injection

Examples of HIV prevalence, Incidence and Testing Indicators by KPs in The Boloka Indicator list

HIV Prevalence		
HIV Prevalence (total)	Select	Select
HIV Prevalence for MSM	Select	Select
HIV Prevalence for TG	Select	Select
HIV Prevalence for FSW	Select	Select
HIV Prevalence for PWID	Select	Select
HIV Prevalence for Prisoners	Select	Select
HIV Prevalence for AGYW	Select	Select
HIV Prevalence by Age Group	Select	Select
HIV Prevalence by Gender	Select	Select

HIV Incidence		
HIV Incidence (total)	Select	Select
HIV Incidence for MSM	Select	Select
HIV Incidence for TG	Select	Select
HIV Incidence for FSW	Select	Select
HIV Incidence for PWID	Select	Select
HIV Incidence for Prisoners	Select	Select
HIV Incidence for AGYW	Select	Select
HIV Incidence by Age Group	Select	Select
HIV Incidence by Gender	Select	Select

HIV Testing		
HIV Testing Rate (total)	Select	Select
HIV Testing Volume and Positivity	Select	Select
HIV Testing Rate for MSM	Select	Select
HIV Testing Rate for TG	Select	Select
HIV Testing Rate for FSW	Select	Select
HIV Testing Rate for PWID	Select	Select
HIV Testing Rate for Prisoners	Select	Select
HIV Testing Rate by Age Group	Select	Select
HIV Testing Rate by Gender	Select	Select

Perinatal Health Research Unit – Sex Worker Study HIV-related Indicators

Demographics

Age
Gender
Ethnicity
Locality (urban vs. rural, province)
Country of birth
Social grant
Income status
Educational status
Employment status
Dependents
Living conditions

Pregnancy

Age at first sex
Age at first child
Parity
Rape
Contraception and use
PMTCT
Breastfeeding
Childbirth

HIV Knowledge

HIV testing
Knowledge of partner status
HIV disclosure
ART use, initiation
ART type, duration
ART switch, default
PEP use

Gender norm, sex

Sexuality
Marital status
Partner substance use
Condom use
STI
Sex work

Discrimination ,stigma

Intimate partner violence
Emotional abuse
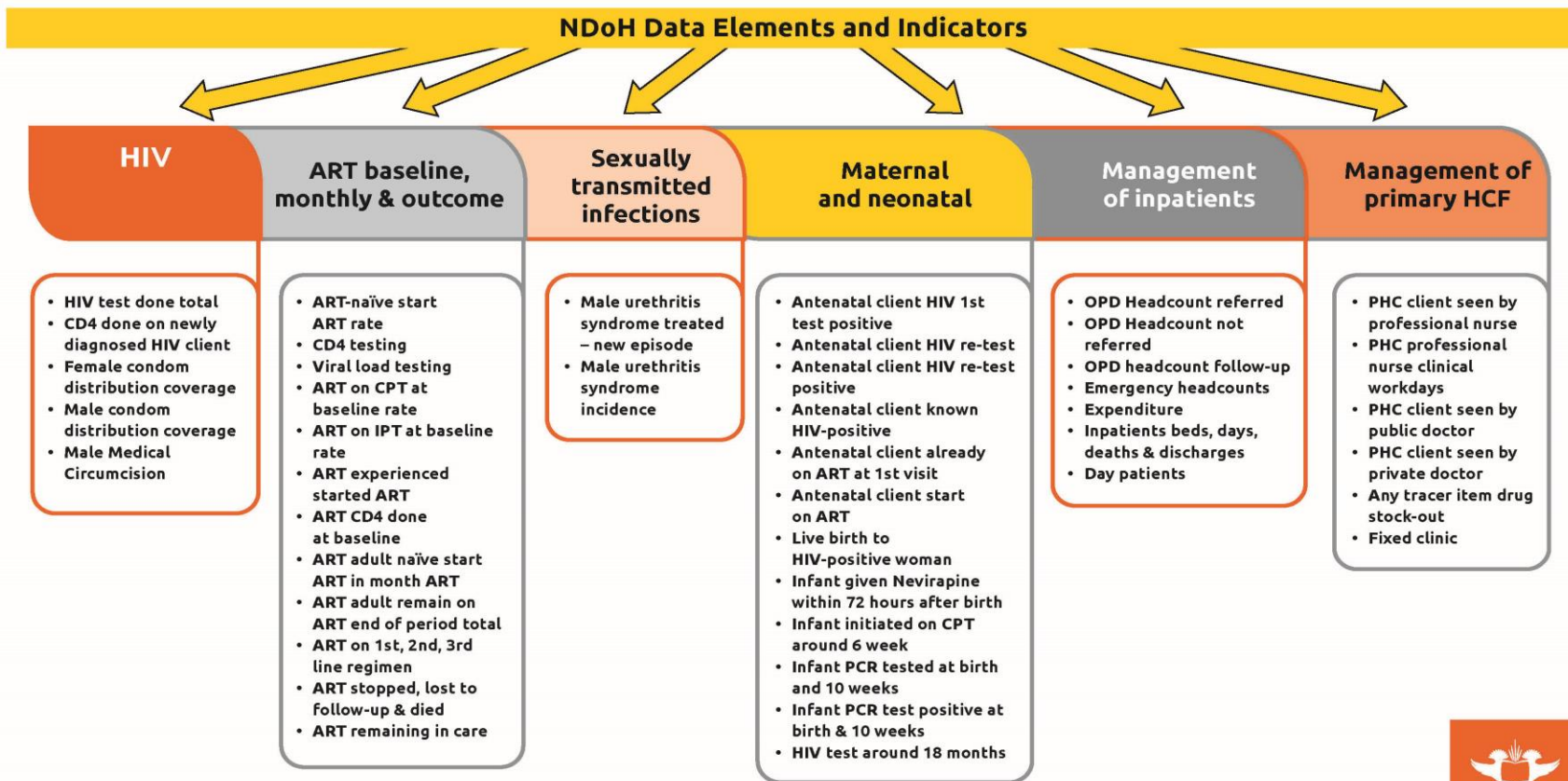Physical abuse
GBV services
Suicide
Self-perception

Other

Smoking

Alcohol & Drugs

Mental health

TB symptoms

NDoH De-identified Routine Programme Data: 2017-2022

NDoH Data Elements and Indicators

HIV
- HIV test done total
- CD4 done on newly diagnosed HIV client
- Female condom distribution coverage
- Male condom distribution coverage
- Male Medical Circumcision

ART baseline, monthly & outcome
- ART-naïve start ART rate
- CD4 testing
- Viral load testing
- ART on CPT at baseline rate
- ART on IPT at baseline rate
- ART experienced started ART
- ART CD4 done at baseline
- ART adult naïve start ART in month ART
- ART adult remain on ART end of period total
- ART on 1st, 2nd, 3rd line regimen
- ART stopped, lost to follow-up & died
- ART remaining in care

Sexually transmitted infections
- Male urethritis syndrome treated – new episode
- Male urethritis syndrome incidence

Maternal and neonatal
- Antenatal client HIV 1st test positive
- Antenatal client HIV re-test
- Antenatal client HIV re-test positive
- Antenatal client known HIV-positive
- Antenatal client already on ART at 1st visit
- Antenatal client start on ART
- Live birth to HIV-positive woman
- Infant given Nevirapine within 72 hours after birth
- Infant initiated on CPT around 6 week
- Infant PCR tested at birth and 10 weeks
- Infant PCR test positive at birth & 10 weeks
- HIV test around 18 months

Management of inpatients
- OPD Headcount referred
- OPD Headcount not referred
- OPD headcount follow-up
- Emergency headcounts
- Expenditure
- Inpatients beds, days, deaths & discharges
- Day patients

Management of primary HCF
- PHC client seen by professional nurse
- PHC professional nurse clinical workdays
- PHC client seen by public doctor
- PHC client seen by private doctor
- Any tracer item drug stock-out
- Fixed clinic

NDoH, National Department of Health; HIV, human immunodeficiency virus; ART, antiretroviral therapy; HCF, Health Care Facility; OPD, outpatients department, CPT, Co-trimoxazole prevention therapy ; PHC, public health care

UNIVERSITY OF JOHANNESBURG

Beyond Zero Indicators

Demographic Information of Clients

Information on HIV Status of Clients

SRHR Education

Information on Gender and Sexual Identity of Clients
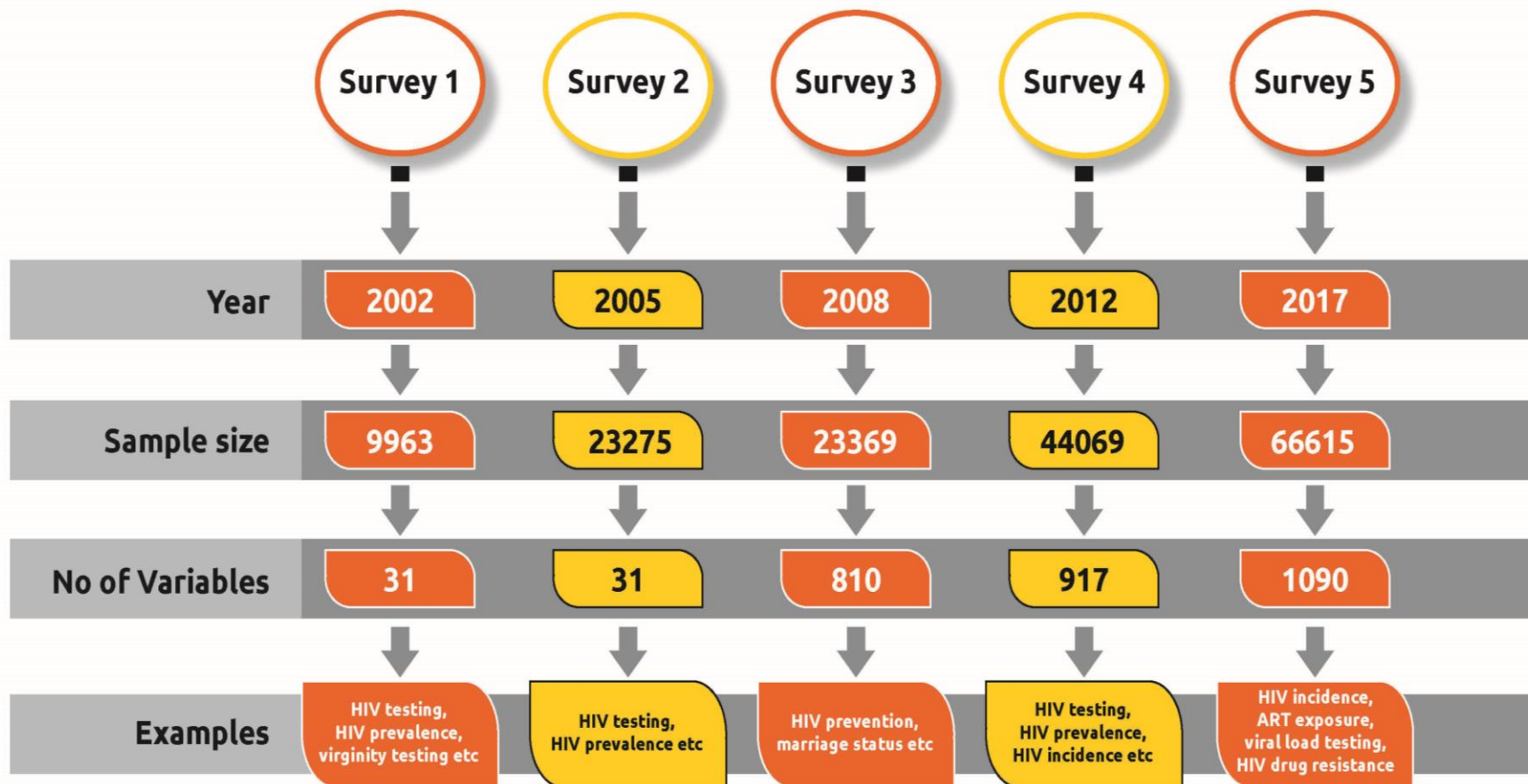
HTS services and clients

STI Screening and Referral

Risk Assessment Psychosocial Assessment

Assessment for PrEP

TB Screening and Referral

SABSSM survey data – 2002, 2005, 2008, 2012, 2017

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5
Year	2002	2005	2008	2012	2017
Sample size	9963	23275	23369	44069	66615
No of Variables	31	31	810	917	1090
Examples	HIV testing, HIV prevalence, virginity testing etc	HIV testing, HIV prevalence etc	HIV prevention, marriage status etc	HIV testing, HIV prevalence, HIV incidence etc	HIV incidence, ART exposure, viral load testing, HIV drug resistance

Harmonization Lesson 2:

A Data Indicator List needs to be developed to guide data partners in providing data on primary and secondary indicators

When developing an indicator list, it is important to have a clear plan and objective about the research question you wish to answer . This will help to break the harmonization process by themes

We identified relevant data sources, selected indicators (prioritization) and variables within indicators for harmonization guided by the master indicator list according to the research question we wanted to explore

The details of each data sources were captured in the Boloka Data Catalogue (adapted from JHU) – Beyond Zero

Study Info	Participation consent form (A	Outreach form (Annexure 4.1	Indicator collection (Annexure 1.4_MSM_TG	Programme Ind
Year				
Eligibility				
Sample Size				
Data Collection Method	interviewer-administered survey	interviewer-administered survey	interviewer-administered survey	
Months of Data Collection				
Study Design	cross-sectional	cross-sectional	cross-sectional	
Sampling Method				
Type of data	individual	individual	percentage (aggregate)	
Region				
Country(ies)				
Domestic				
County(ies)	SA	SA	SA	
(Key) Populations	MSM and transgender people	MSM and transgender people	MSM and transgender people	
Grant Acknowledgement				
Owner	Beyond Zero	Beyond Zero	Beyond Zero	
	Mchenge Ngwane	Mchenge Ngwane	Mchenge Ngwane	

The details of each data sources were captured in the Boloka Data Catalogue (adapted from JHU) – HSRC SABSSM Data

Study Info	HSRC SABSSM 2017	HSRC SABSSM 2012	HSRC SABSSM 2008	HSRC SABSSM 2005	HSRC SABSSM 2002
Year	2017	2012	2008	2005	2002
Eligibility	households, people of all ages living in South Africa				
Sample Size	66,615				
Data Collection Method	interviewer-administered survey using digitally using electronic tablets, dried blood spot (DBS) samples collected for biomarker testing	interviewer-administered survey using digitally using electronic tablets, dried blood spot (DBS) samples collected for biomarker testing	interviewer-administered survey using digitally using electronic tablets, dried blood spot (DBS) samples collected for biomarker testing	interviewer-administered survey using digitally using electronic tablets, dried blood spot (DBS) samples collected for biomarker testing	interviewer-administered survey using digitally using electronic tablets, dried blood spot (DBS) samples collected for biomarker testing

The details of each data sources were captured in the Boloka Data Catalogue (adapted from JHU) – PHRU Studies

Study Info	Klerksdrop Men's Health	Klerksdrop Sex Workers	Soweto Men's Health	Soweto Men's Health - Nurse
Year	2019	2019	2019	2019
Eligibility		biologically females 18 years or older and to have voluntarily sold or transacted during sex for financial gain (not necessarily paid in cash) in the past 6 months. They also needed to work within the site's district.		
Sample Size				
Violence	vio_insult, vio_insult,	included	included	included
Smoking	smoke_ever, sm_type, bum_hh,	included	included	included
Alcohol Use	drunk, alcohol_often,	included	included	included
Drug Use	drugs, audit_painkiller,	included	included	included
Health Screening	sti_symptom, sti_treatment,	included	included	included
Pregnancy		age at first sex, contraceptive		
Discrimination		verbal discrimination due to sex work, denial of health services, denial of police assistance/services	included	included
Internalized Stigma		feels of shame, guilty, self-blame,	included	included
COVID			experience during lockdown,	

Harmonization Lesson 3:

A Data Catalogue, an inventory of data sources and variables is needed to capture essential details about each data source

We reviewed the existing JHU Data Ccatalogue

The Boloka Codebook was developed to describe the variable naming system

	A	B	C	D
1	**VARIABLE / FIELD NAME**	**DESCRIPTION**		
2	ever_hiv_test	ever had an HIV test		
3	time_recent_hiv_test	how long ago was the last HIV test		
4	art_use	currently on ART		
5	current_art_adherence	current adherence to ART		
6	place_hiv_test	Where was this test done?		
7	outcome_hiv_test	What was your test result?		
8	informed_hiv_results	Have you been told/informed of the result of test?		
9	current_hiv_status	As far as you know today, what is your HIV status?		
10	first diagnosed_hiv_test	If positive, when were you first diagnosed with HIV?		
11	disclosure_partner_hiv_status	Have you told your partner that you are living with HIV		
12	when_disclose_partner_hiv_status	When did you tell your partner that you were living with HIV?		
13	aware_partner_hiv_status	Do you know your partner's HIV status?		
14	partner_art_status	Is your partner on treatment?		
15	past3months_disclosed_hiv_status_paying_clients	In the past 3 months, have you disclosed to any payingclient what your HIV status is?		
16	past3months_disclosed_hiv_status_other_womenwh(	In the past 3 months, have you disclosed to any otherwomen who sell sex, what your status is?		
17	start_time_art_use	When was the first time that you ever started taking treatment?		

Harmonization tool   **CODEBOOK**   (+)

Operational definitions developed for coding purposes

- **Indicator:** Broader category of HIV-related interest areas (e.g. HIV testing, ART)
- **Variable:** Element of an indicator (e.g. Ever HIV tested)
- **Variable name:** Specific term for capturing the variable (e.g. ever_hiv_test)
- **Question:** Phrase posed in the questionnaire/program tool that captures the variable
- **Response categories:** Possible answer options provided on a questionnaire/program tool
- **Dataset:** Collection of variables on a specific population (e.g. Sex workers survey, Male clients survey)
- **Data source:** Group of related datasets (Klerksdorp)
- **Codebook:** Reference table for describing the variable naming system
- **Data catalogue:** Inventory of data sources and variables
- **Harmonisation:** The bringing together of different data sources with various data structures
- **Data partner:** Organisational origin of data (e.g. HSRC, NACOSA)
- **Population:** Group of individuals for which the data captures (e.g. FSW, MSM)
- **Harmonisation Tool:** Shared excel sheet designed to support harmonisation process

Harmonization Lesson 4:

A coding table describing the variable naming system and operational definitions is needed to guide the harmonization process

This prevents duplicates, errors, contradictions, discrepancies, and inconsistencies

Questionnaire Reviews – Assess similar variables that can potentially be harmonized under each indicator –PHRU HIV Testing variables

50	pre_result_hiv Show the result field ONLY if: [previous_hiv]= '1'	What was your test result?	radio, Required 0 — I do not know 1 — Negative 2 — Positive 3 — Indeterminate Custom alignment: RH
51	result_status	As far as you know today, what is your HIV status?	radio, Required 0 — I do not know 1 — Negative (no HIV) 2 — Positive (living with HIV) Custom alignment: RH
52	year_diagnosed_hiv Show the result field ONLY if: [previous_hiv]= '1'	If positive, when were you first diagnosed with HIV?	text (date_dmy), Required Custom alignment: RH
53	partner_hiv Show the result field ONLY if: [previous_hiv]= '1'	Have you told your partner that you are living with HIV?	radio, Required 0 — No 1 — Yes 2 — I do not have a partner Custom alignment: RH
54	part_when Show the result field ONLY if: [previous_hiv]= '1'	When did you tell your partner that you were living with HIV?	text (date_dmy), Required Custom alignment: RH
55	partners_HIV	Do you know your partner's HIV status?	radio, Required 0 — Unknown 1 — Negative 2 — Positive 99 — No partner Custom alignment: RH

SABSSM Survey 2017 HIV Testing Variables

SECTION 9	HIV COUNSELLING AND TESTING
	I am now going you ask you a few questions about HIV testing

9.1	Do you know of a place nearby where you can get an HIV test?	Yes	No
		1	2

9.2	Have you <u>ever</u> had an HIV test?	Yes	No	No response
		1	2	3
			GO TO 9.24	GO TO SECTION 10

9.3	How long ago did you have your most recent HIV test?	
0 to 3 months		1
4 to 6 months		2
7 to 11 months		3
Less than a year ago		4
Between 1-2 years ago		5
Between 2-3 years ago		6
Three or more years ago		7

9.9	You indicated that you were previously tested for HIV. Are you willing to tell me the last HIV test result you received?	Yes	No	Never received result
		1	2	3
			GO TO 9.18	

9.10a	What was the result of that HIV test?	Positive	Negative	Indeterminate
		1	2	3
			GO TO 9.18	

INSTRUCTION	*Please note that you should not tell me about the actual result. I am only interested whether you have been told/informed of the result of the test.*

9.4	Have you been told/informed of the result of your most recent test?	Yes	No
		1	2

9.5	Where did you get your most recent HIV test?	
Public hospital		1
Private hospital		2
Public clinic or doctor		3
Private clinic or doctor		4
Mine hospital		5
Traditional healer		6
NGO		7
Pharmacy/chemist		8
HIV testing centre		9
Workplace		10
Health Jamboree event		11
Other		

9.6	During your most recent HIV test, did you have counselling <u>before</u> the HIV test?	Yes	No
		1	2

9.7	During your most recent HIV test, did you have counselling <u>after</u> the HIV test?	Yes	No
		1	2

9.8	What was the <u>main</u> reason for going for your last HIV test?	
I wanted to know my HIV status		1
My partner asked me to go for testing		2
I wanted to start a new sexual relationship		3
I wanted to get married		4
I applied for an insurance policy		5
I applied for a loan		6
My employer requested it		7
I was feeling sick		8
I was instructed by a health worker (nurse/doctor)		9

Beyond Zero HIV Testing Variables

Information on HIV Status of Clients

1. Have you had an HIV test before? ☐ Yes ☐ No (*If no, skip to question 7)

2. If yes, when was your last HIV test? ☐ Less than 3 months ago ☐ 3 to 6 moths ☐ more than 6 months
Please provide the date of your last test (DDMMYYYY)_____

3. If yes, what was your status at your last HIV Test? ☐ Known Positive ☐ Known Negative (*If known Negative, please skip to question 7)

4. If you are known HIV +ve, are you linked to care? ☐ Yes ☐ No 5. If yes, which clinic? _____

6. If no, please allow our outreach programme to assist you access care (outreach staff member to provide the client with a referral letter to the of their preference and document the name of the clinic here_____

7. Would you like to have an HIV test today? ☐ Yes ☐ No If no, what is your reason? _____

8. If yes, proceed to the consent process with the outreach staff member (client to complete consent process and sign the consent form)

National Department of Health HIV Testing Variables

HIV TEST 15 YEARS AND OLDER (EXCL ANC)	DEFINITION	HIV test client 15 years and older (excluding antenatal clients)	INCLUSIONS	None	COLLECTED BY	Clinician
	EXTENDED DEFINITION	None	EXCLUSIONS	» EXCLUDE antenatal clients » EXCLUDE repeat tests on the same client to confirm diagnosis	COLLECTION POINTS	All facilities
	USE AND CONTEXT	None	FREQUENCY	Monthly	TOOLS	PHC Comprehensive Tick Register; HTS Register (HIV Testing Services) or HCT module in TIER.Net
HIV TEST POSITIVE 15 YEARS AND OLDER (EXCL ANC)	DEFINITION	HIV test positive client 15 years and older (excluding antenatal clients)	INCLUSIONS	None	COLLECTED BY	Clinician
	EXTENDED DEFINITION	None	EXCLUSIONS	» EXCLUDE antenatal clients » EXCLUDE repeat tests on the same client to confirm diagnosis	COLLECTION POINTS	All facilities
	USE AND CONTEXT	None	FREQUENCY	Monthly	TOOLS	PHC Comprehensive Tick Register; HTS Register (HIV Testing Services) or HCT module in TIER.Net
HIV-POSITIVE CLIENT SCREENED FOR TB	DEFINITION	HIV-positive clients who have been screened for TB after positive HIV test	INCLUSIONS	» INCLUDE antenatal clients who tested HIV-positive	COLLECTED BY	Clinician
	EXTENDED DEFINITION	None	EXCLUSIONS	None	COLLECTION POINTS	All facilities
	USE AND CONTEXT	None	FREQUENCY	Monthly	TOOLS	PHC Comprehensive Tick Register; Tick register Hospital
HIV POSITIVE CLIENT ELIGIBLE FOR IPT	DEFINITION	All HIV-positive clients eligible for Isoniazid preventive therapy (IPT). A client is eligible for IPT when they have screened negative for TB using the TB screening tool.	INCLUSIONS	» INCLUDE HIV-positive antenatal clients » INCLUDE asymptomatic HIV-positive TB contact children older than 5 years and younger than 15 years.	COLLECTED BY	Clinician
	EXTENDED DEFINITION	Count clients who are eligible for IPT prevention therapy only once.	EXCLUSIONS	» EXCLUDE asymptomatic TB contact children younger than 5 years	COLLECTION POINTS	All facilities
	USE AND CONTEXT	Criteria for excluding active TB refer to the 4 questions on the TB screening tool: » Cough of any duration in HIV-positive; » Unintentional weight loss (e.g. > 1.5kg in last month) » Fever > 2 weeks; » Drenching night sweats	FREQUENCY	Monthly	TOOLS	PHC Comprehensive Tick Register

HIV

A ELEMENTS

Example: Variable Matching for harmonization across datasets

Variables considered*	SABSSM 2017 (cross-sectional)	Klerksdorp dataset - PHRU (cross-sectional)	Variable matching	Harmonization success
1. HIV testing	Construct: have you previously tested for HIV? Type of Data: categorical 1 Yes 2 No	Construct: Were you tested for HIV at the TB clinic? Type of Data: categorical 1 Yes 2 No 3 I don't know	• Complete matching of construct • Matching of response or data type and coding, except missing value coding	• Record if harmonization was successful or not • Was data pooled into one variable?
2. Sexual history	Construct: "How old were you when you had sex for the first time?" Type of Data: continuous Missing: are there any?	Construct: "How old were you when you first had sex?" Type of Data: continuous Missing: are there any?	• Complete matching of construct • Matching of response or data type and coding, except missing value coding	• Record if harmonization was successful or not • Was data pooled into one variable?

After identifying variables, the next step is to match them and merge variables from the different datasets into a new variable - merging requires at least one variable in common (Adhikari et al, 2021).

We matched selected variables across categories of indicators in multiple datasets that elicited comparable information

Harmonization Lesson 5:

The harmonization process can be complex, e.g. differences in scales of measurement, variable coding schemes, and definitions of concepts.

Metadata quality is VERY important in harmonization precision *- how the variable for a particular indicator was collected, recoding of variables, developing new variable names, and Deciding which variables to merge*

.

A Boloka Data Harmonization Tool was created to support the harmonization process

Variable names

- Developed by the PACER/JHU team
- When adding a new row to the Harmonization Tool, check the codebook to see if the variable has been previously defined
- If the variable had not been previously defined, create a new variable with the following steps
- Not too long (4 words max)
- Words separated by underscores (_), *do not use spaces or –*
- Variable name must be lowercase, no capitalization
- Variable names must be kept general, specifics can be mentioned under the "QUESTION (copy exactly)" and "RESPONSE CATEGORIES" sections. E.g., if a dataset variable covers ART adherence in the past 3 months, use a variable name such as art_adherence, *do not use a name such as art_adherence_past_3_months*. This way, the variable name can also be used for dataset items covering ART adherence for different periods (e.g., past 6 months, year, etc.)

Variable name lengths

- The variable name should be kept short and easy to read (not more xxx characters)

Open text fields - case by case

- Open-ended questions should be excluded from the harmonisation tool
- Researchers/students can use the open-ended questions if needed

An extract of The Boloka Data Harmonization Tool

	A	B	C	D	E
1	INDICATOR	POPULATION	VARIABLE NAME	QUESTION (copy exactly)	RESPONSE CATEGORIES
2	HIV testing	General population	ever_hiv_test	Have you ever had an HIV test?	yes, no, no response
3	HIV testing	General population	ever_hiv_test	Have you ever had an HIV test?	yes, no, no response
4	HIV testing	MSM	ever_hiv_test	Have you ever had an HIV test before?	yes, no
5	HIV testing	Male clients	ever_hiv_test	Have you previously tested for HIV?	yes, no
6	HIV testing	FSW	ever_hiv_test	Have you previously tested for HIV?	yes, no
7	HIV testing	General population	time_first_hiv_positive_test	What was the month and year of your first HIV positive test?	IF "DON'T KNOW" MONTH, THEN
8	HIV testing	General population	time_recent_hiv_test	How long ago did you have your last HIV test?	year, 1-2 years, 2-3 years, 3 or more years
9	HIV testing	General population	time_recent_hiv_test	How long ago did you have your last HIV test?	year, 1-2 years, 2-3 years, 3 or more years
10	HIV testing	MSM and TG	time_recent_hiv_test	When was your last HIV test?	<3 months, 3-6 months, >6 months
11	HIV testing	FSW	ever_hiv_test	Have you previously tested for HIV?	yes, no
12	HIV testing	MSM	ever_hiv_test	Have you previously tested for HIV?	yes, no
13	ART	General population	current_art_use	medications?	yes, no
14	ART	General population	current_art_adherence	medications, daily?	yes, no
15	HIV testing	General population	place_hiv_test	test?	yes, no, no response
16	HIV testing	General population	place_hiv_test	test?	yes, no, no response
17	HIV testing	General population	informed_hiv_results	recent test?	yes, no

Harmonization tool | CODEBOOK | (+)

Harmonization Lesson 6:

A standard operating procedure containing processes for creation and definitions of harmonized variables is needed to serve as a reference point

In case errors occur or are identified by users at a later time, all data conversions should be reversible.

Harmonization Lesson 7:

It is important to establish an institutional mechanism for managing data harmonization

For The Boloka Project we worked through the UJ Information and Communication Services (ICS), who secured a REDCap license from the REDCap consortium. The license grants UJ permission to utilize the REDCap software, along with access to the consortium's support tools and resources. Currently, ICS is in the process of implementing the REDCap system in compliance with ICS security standards and adhering to the best practices recommended for REDCap.

Closing Remarks

- There are **nuances to be considered in the harmonization of data**; Once such hurdles are addressed the potential for high-level integrated analysis is formidable.
- The harmonization of multiple data elements requires **manual labor, time, continuous engagement -** you have to
- The Data harmonization Tools are informed by several factors - **types of data, volume of data sources, objectives of the study, structure of data, etc.**
- A Data processing plan should be in place' that includes descriptions of **how the producer(s) of the harmonized data** deal with the following:
  Differences in study design, differences in measures and scales of measurement, differences in operational definitions of concepts, routing of questions, respondents asked), differences in how estimates are generated (imputation, weighting, or nonresponse adjustments)
- It is important to have a **diverse trained harmonization team with interdisciplinary skills** to design and implement the harmonization process

Selected References

Adhikari, K et al. International Journal of Population Data Science (2021) 6:1:1680
Bello B, Ndagurwa P, Omogiate S, Luwaca B, Motsieloa L. Republic of South Africa: 2020 Global AIDS Monitoring Report (2021). Johannesburg: CESAR; 2021.
Johns Hopkins Data Catalogue

Phaswana-Mafuya, RN, Edith Phalane, Katharine S. Journeay, Haley I. Sisel, Claris Siyamayambo, Betty Sebati, Francois Wolmarans, Katherine Rucinski, Amrita Rao, Kalai Willis, Xiaoming LI, Olatosi Banky, Stefan D. Bara. 2023. Harnessing big heterogeneous data to evaluate the potential impact of HIV responses among key populations in generalized epidemic settings in Sub Saharan Africa: The Boloka Data Repository. Proceedings of the 4th National Big Data Health Science Conference. *BMC Procedings* **17** (Suppl 19), 32 (2023). https://doi.org/10.1186/s12919-023-00281-y

South African National AIDS Council (SANAC). National Strategic Plan for HIV, TB, STIs 2023-2028. Available from: https://sanac.org.za/wp-content/uploads/2023/05/NSP-Document.pdf (accessed on 09 May 2023)

UNAIDS. (2021). *UNAIDS Data 2021.*
https://www.unaids.org/sites/default/files/media_asset/JC3032_AIDS_Data_book_2021_En.pdf

Acknowledgements

Acknowledgements