Strategy-wide Harmonization for Reporting: Creating Public Data and Documentation

2015 USAID Global Education Summit

Leveraging Goal One Data: New Opportunities and Resources

Roger Stanton
Optimal Solutions Group, LLC

November 4 , 2015

Agenda

- Overview of Secondary Analysis for Results Tracking (SART)
- Harmonizing datasets and documents
- Standardized data and descriptions

Overview of SART

Standardizing data to tell a unified story

Challenge

- Not all Goals 1 and 3 data are the same
- Data coded and documented inconsistently across implementers
- Lack of sufficient technical documentation

Solution

- Systematically harmonize & standardize Goals 1 and 3 data
  - Develop process to harmonize disparate data sets
  - Have the standardized, open source data available to all stakeholders

USAID FROM THE AMERICAN PEOPLE

SART Ed | USAID Ed Strategy | Map | Data | Reports

WAKANDA

"USAID works in partnership with the Government of Wakanda to improve learning outcomes in early-grade reading, provide pathways to education for bright but disadvantaged students, and empower youth with skills and opportunities to lead the way to a peaceful and prosperous Wakanda."

Download Data

Programs

Goal 1

- Wakanda Reading Program
- Wakanda Early Grade Reading Project
- Wakanda Reading Initiative

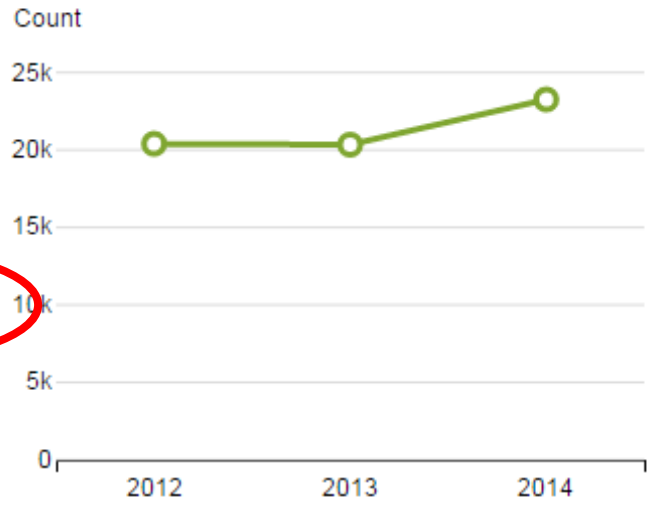
Goal 3

- Wakanda Reading for All
- Wakanda Project on Educational Access
- Wakanda Education Unites

Goal 1

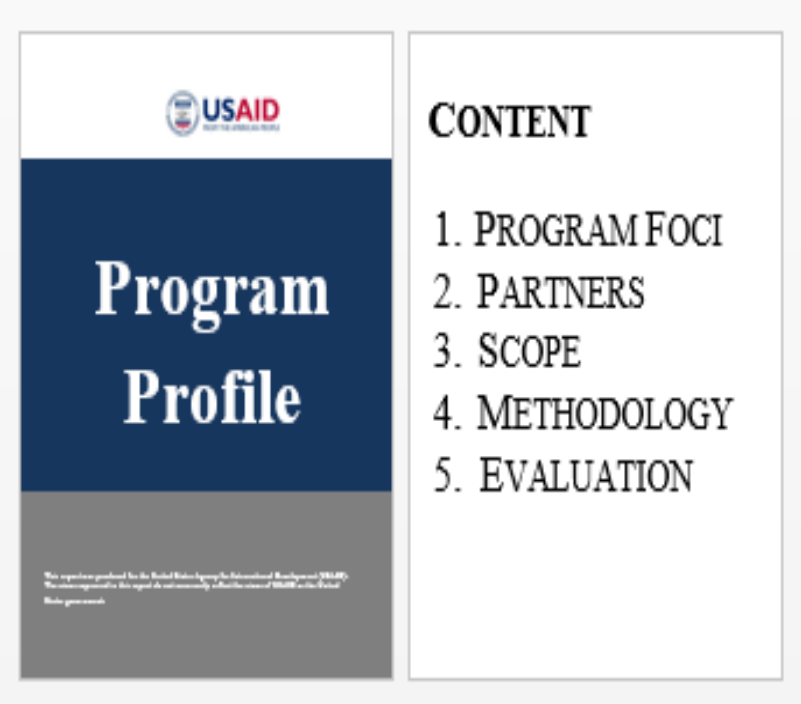
Summary

Beneficiaries



View Details

Harmonizing Data and Meta-Data



+

Sex	Language	School type	Grade	ORF
M	English	Urban	1	35
F	French	Urban	2	45
F	French	Rural	4	50
M	English	Rural	6	48
M	English	Rural	5	62
M	Bangla	Urban	3	50
F	Bangla	Urban	1	35

USAID
FROM THE AMERICAN PEOPLE

Normalized Structure

- Normalized structure
  - Cross-sectional data for a phase
  - Columns representing unique variables, with each row corresponding to a unique record (long)

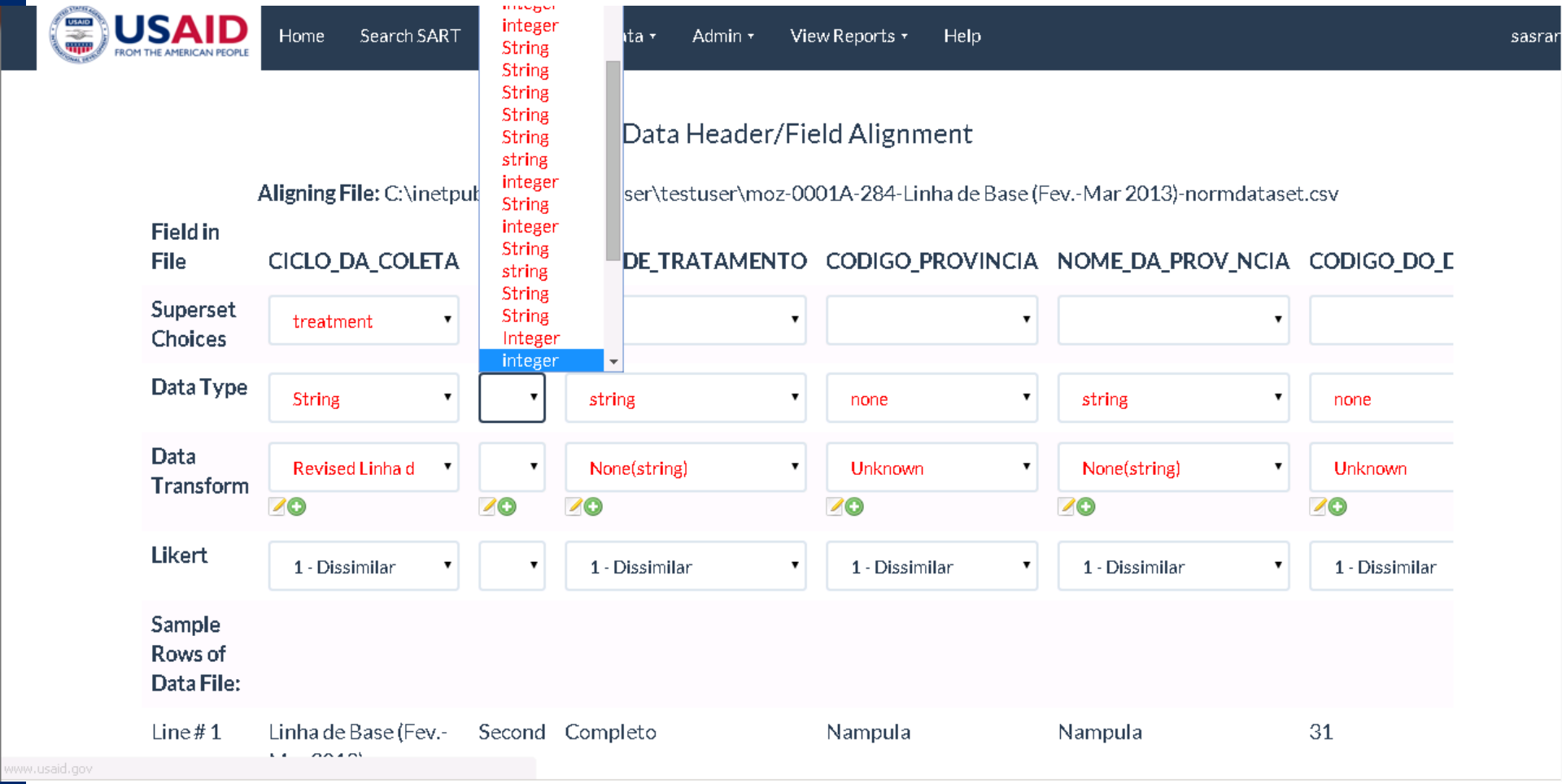
- If in-bound not in this structure, we normalize it!
  - Split files containing multiple phases
  - Reshape to long
  - Capture embedded documents including codebook

In-bound data						
student_id	language	k_language	orf	k_orf	cwpm	k_cwpm
1	English		2.8		9.7	
2		Kiswahili		3.4		15
3		Kiswahili		4.6		22
4	English		9.5		12	

Normalized data			
student_id	language	orf	cwpm
1	English	2.8	9.7
2	Kiswahili	3.4	15
3	Kiswahili	4.6	22
4	English	9.5	12

Transformation

- Typical “transformations” include
  - Create missing variables e.g., country, language, etc.
  - Create missing calculated variables from other variables (e.g., oral reading fluency)
  - Alter in-bound to meet ingestion limitations
    - Special characters
    - Limit character width to 50
    - All null values*


USAID
FROM THE AMERICAN PEOPLE
Home
Search SART
Admin
View Reports
Help
Data Header/Field Alignment
Aligning File: C:\inetpub
ser\testuser\moz-0001A-284-Linha de Base (Fev.-Mar 2013)-normdataset.csv
integer
String
String
String
String
String
string
integer
String
integer
String
string
String
String
Integer
integer
Field in File
CICLO_DA_COLETA
DE_TRATAMENTO
CODIGO_PROVINCIA
NOME_DA_PROV_NCIA
CODIGO_DO_D
Superset Choices
treatment
Data Type
String
string
none
string
none
Data Transform
Revised Linha d
None(string)
Unknown
None(string)
Unknown
Likert
1 - Dissimilar
1 - Dissimilar
1 - Dissimilar
1 - Dissimilar
1 - Dissimilar
Sample Rows of Data File:
Line # 1
Linha de Base (Fev.-
Second
Completo
Nampula
Nampula
31

Purpose: Accurate, Standardized, Documented

- Process has accuracy checks throughout
- All changes are documented
- Produces a standardized dataset and associated documentation

The Ingested Public Use Microdata Series (IPUMS)

- IPUMS is the standardized data ingested into the SART database

- Can be downloaded as a .CSV file

Dataset Profile

- Document accompanies IPUMS data download
- One profile for each country | project | phase
- Overview and details

Key Variables

Variable name	Variable definition	Available in IPUMS dataset?
country	Country name	Yes
project	Name of project within the country?	Yes
treat_phase	In which phase of the study is this treatment school student?	Yes
grade	What is the student's grade level?	Yes
female	Is the student female?	Yes
language	Language the test was conducted in	Yes
year	In what year was the student tested?	Yes
treatment	What level of treatment is the school receiving?	Yes
clpm	Correct Letters Per Minute	Yes
cwpm	Correct Familiar Words Per Minute	Yes
cnonwpm	Correct Invented Words Per Minute	No
orf	Oral Reading Fluency	Yes
read_comp_score	Total correct Reading Comprehension questions	Yes

Complex Survey Specifications

Complex survey specifications	Available
Probability weight variable	final_weight
No. of Stages of Stratification	2
Variance Estimate Method	Linearized
Single-unit Handling	Scaled
Stage 1	
Strata: Variable (strata1)	strata1
Strata: Description	Division and year
Sampling Unit	masked_school_code
Sampling Unit: Description (if applicable)	School
Finite Population Correction	fpc1
Stage 2	
Strata: Variable (strata2)	strata2
Strata: Description	Grade and sex
Sampling Unit	student_id

IPUMS Dataset Codebook

Variable name	Variable definition	Response category labels	Variable type	Notes	Available in IPUMS dataset?	Constructed by SART?
country	Country name		string		Yes	Yes
region	The largest geographic breakdown of the program		string		No	Not reported
state	The second largest geographic breakdown of the program		string		Yes	No
district	Third largest geographic breakdown, which may reflect school districts		string		No	Not reported
project	Name of project within the country?		string		Yes	Yes
year	In what year was the student tested?		integer		Yes	Yes
month	In what month was the student tested?	January; February; March; April; May; June; July; August; September; October; November; December	string		No	Not reported
date	On what date was the student tested?		datetime	Date format: MMDDYYYY	No	Not reported
masked_school_code	Masked code to identify school's within country.		string		No	Not reported

Dataset Supplementary Materials

- Describes changes made during normalizing and transforming data
  - Changes to structure (cross-sectional/longitudinal; long/wide)
  - Changes to variables
    - Alignment of variable names
    - Alignment of categorical variable options

Program Profiles

- In addition to the IPUMS data and documentation
  - Program profiles based on review of program documentation (e.g., annual reports, evaluation reports)
  - The SART Ed website
    - Submit data
    - View information for Goal 1 and 3 programs

Summary

- SART's objective is to collect, harmonize & standardize data and meta-data, and so that *ANY methodology* can be used to track and report on the progress of Goals 1 and 3 of the USAID Education Strategy.
- For more information see SART Ed at https://sartdatacollection.org