



NiPN
National Information
Platforms for Nutrition



Dataset Aggregator

Milko Skofic

NIPN Technical Gathering
July 4th, Paris



Summary

- Issue
- Solution
- Audience
- Limitations

	C	D	E	F	G	H	I	J	K	L	M			
1	1133	1133	1133	1133	266	1132	997	983	1067	1051				
2	TEAM	ID	HH	SEX	BIRTHDAT	MONTHS	WEIGHT	HEIGHT	EDEMA	MUAC	WAZ-NCI			
3	2	1	1	f		53	15.7	106.2	n	151	-0.			
4	2	2	1	f		30								
5	2	1	2	m	26/06/2009	57.46	12.2	95.3	n	124	-2.			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Date	Grappe	Equipe	Identifiant	HH	Sexe	Naissance	Mois	Poids	Hauteur	EDEM			
2	05/05/2016	32	6	2	2	m	20/07/2012	45.5	16.1	99.6	n			
3	05/05/2016	32	6	1	2	m	10/01/2012	51.81	18.7	103.9	n			
4	05/05/2016	32	6	1	1	f	10/09/2015	7.82	6.7	69.2	n			
	Region	Numero District	District	Prevalence Malnutrition aiguë (SMART 2014*)			Burden 100% 2015 variante mediane			CASELOAD pe				
				Global 6-59 mois	Sévère 6-59 mois	Modérée 6-59 mois	Global 6-59 mois	Sévère 6-59 mois	Modérée 6-59 mois	Global 6-59 mois	Sévère 6-59 mois			
2	Kayes	1	Kayes	11.3	1.5	9.8	24,202	5,075	19,127	17,195	3,80			
4	Kayes	2	Bafoulabe	5.6	0.8	4.8	3,028	679	2,349	2,154	50			
5	Kayes	8	Oussoubigiana	11.3	1.5	9.8	4,960	1,040	3,920	3,524	78			
6	Kayes	3	Diema	12.2	2.1	10.1	11,079	2,935	8,144	7,902	2,20			
7	Kayes	4	Kenieba	11.3	1.5	9.8	9,154	1,919	7,235	6,504	1,43			
8	Kayes	5	Kita	11.3	1.5	9.8	20,478	4,293	16,185	14,549	3,22			
9	Kayes	6	Nioro	11.3	1.5	9.8	10,866	2,278	8,588	7,720	1,70			
10	Kayes	7	Yelimane	15.7	2.2	13.5	11,746	2,587	9,159	8,352	1,94			
11	Total Kayes						95,513	20,806	74,707	67,899	15,60			
13	Koulikoro	1	Koulikoro	13.8	2.4	11.4	11,926	3,188	8,738	8,508	2,39			
14	Koulikoro	2	Banamba	13.8	2.4	11.4	10,747	2,873	7,874	7,667	2,15			
15	Koulikoro	3	Diola	13.8	2.4	11.4	14,984	4,007	10,977	10,689	3,00			



Aggregating datasets

- Why aggregate datasets
 - Overview data spanning time
 - Overview from geographic locations
 - Other aggregations
 - Query the aggregate as a whole
- Which datasets can be aggregated
 - Same methods
 - Compatible formats
 - Share compatible set of variables



Aggregating datasets

- Process
 - Define a default data dictionary
 - Find all related datasets
 - Re-organise the variables in all datasets
 - Append one by one all datasets into a single one
 - Store and document the aggregated dataset
- Workflow
 - Repeat for all different types of datasets
 - Repeat each time the process is needed



Aggregating datasets

- Issues
 - Find the required files
 - Share the same variable definitions
 - Add metadata to individual datasets
 - Attach or update original copy
 - Organise aggregated datasets



Solution

- Network application
- Centralised data dictionary
- Dataset level metadata
- Study level metadata
- Archive for files and data
- Assisted variables harmonisation workflow
- Database for datasets and their data
- Query interface on datasets and their data

Usage

- Register study
 - Document study
 - Upload original files (datasets, reports, etc.)
 - Document individual files
 - Set dataset summary statistics

Usage

- Upload dataset data
 - Store in work area
 - Match with data dictionary variables
 - Data validation reports
 - Correct eventual errors
 - Load data in aggregated collection

Usage

- Search
 - Query dataset metadata
 - Query dataset data
 - Two sets of results
 - Studies
 - Data
 - Download study archives
 - Download data selections

Audience

- Archiving
- Working datasets
- Summary reports
- Harmonise historical data
- Share standards with data providers



Limitations

- Currently under development
- Initially targets Standardised Monitoring and Assessment of Relief and Transition (SMART) surveys
- Need feedback to enrich features

Questions?

Thank you!

