Levels and Stages of Integration

Fourth RAP Regional Workshop on Building Training Resources for Improving Agricultural and Rural Statistics: Survey Methods for Agricultural Statistics- Current Practices and International Recommendations

14-18 December 2014, Tehran, Iran.

Alick Nyasulu
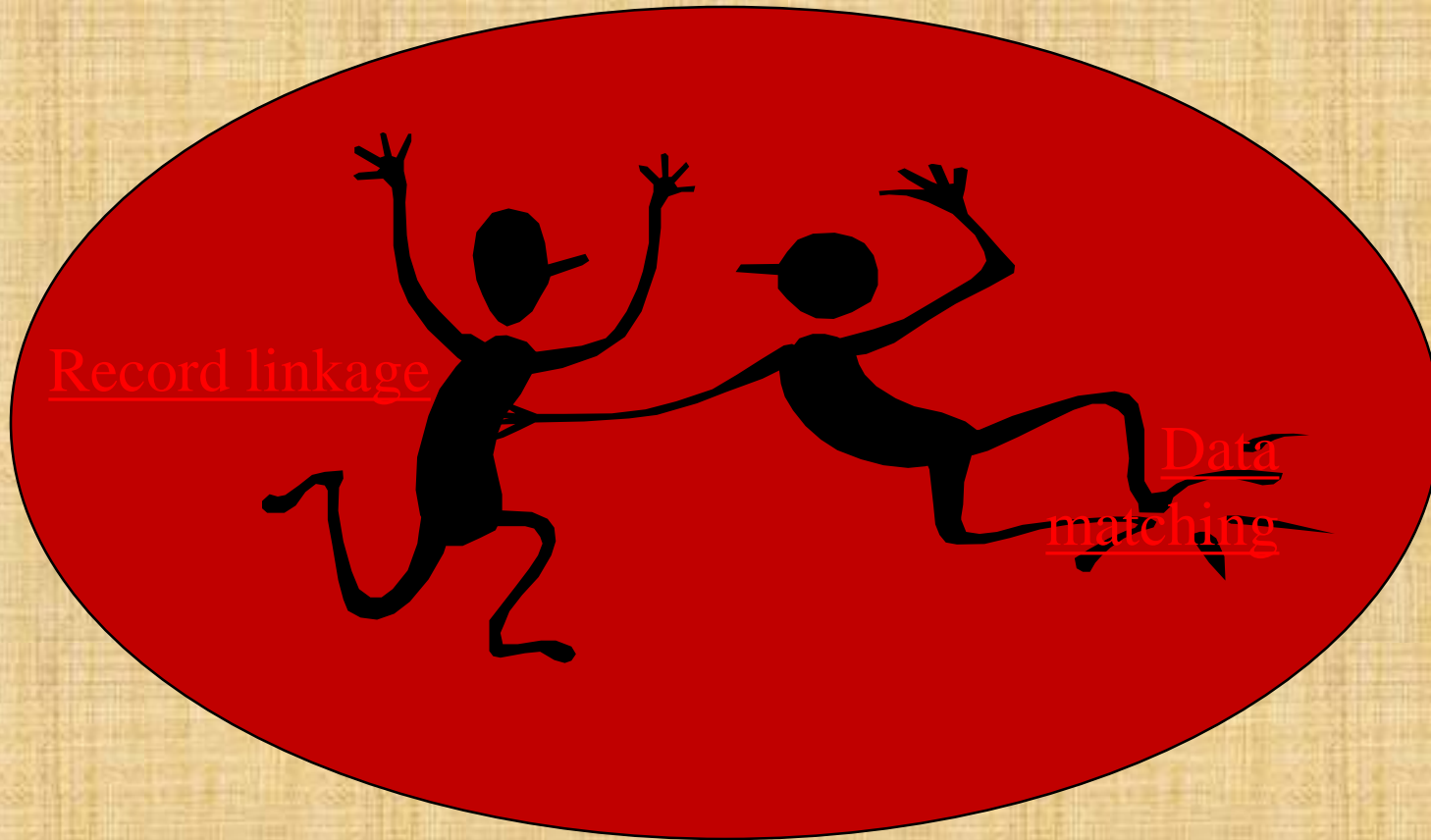## Statistical Institute for Asia and the Pacific (SIAP)

Content

- Data Integration
- Stages of Integration
- Case Studies

Introduction

- Data integration is broadly defined as the *combination of data from different sources on the same or a similar individual or unit.*
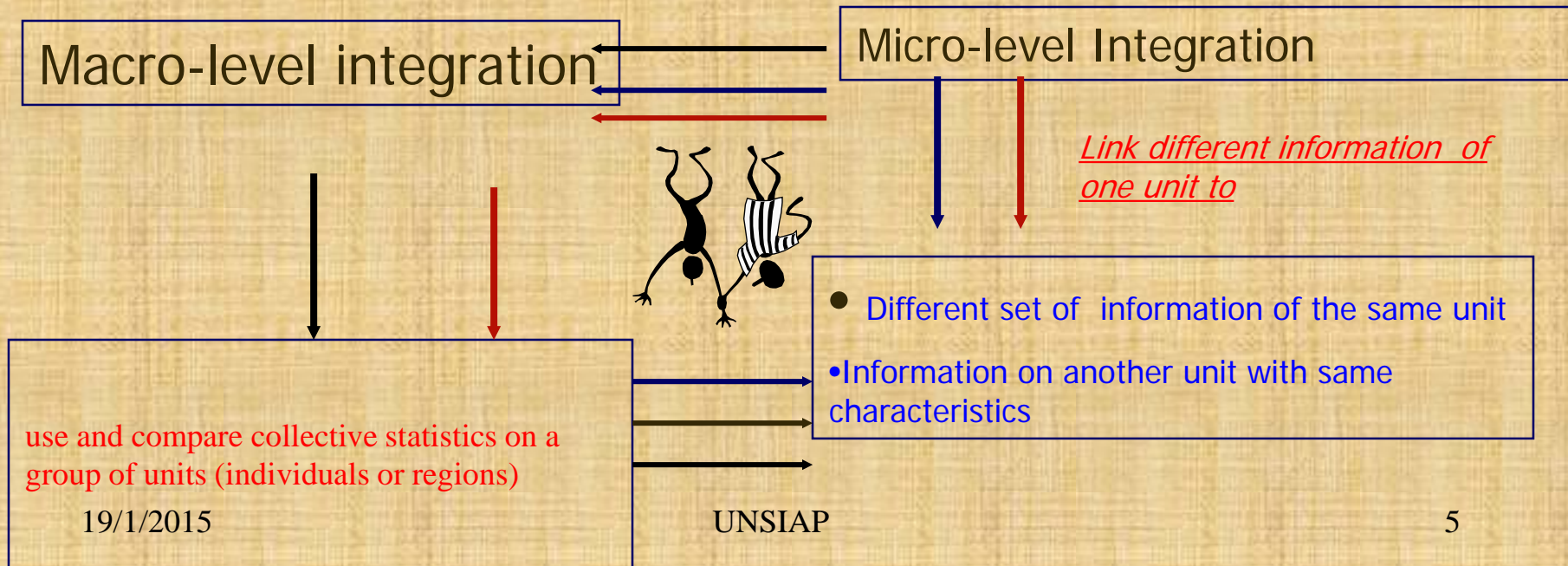
Record linkage

Data matching

Other terms used to describe the process of data integration

Data Integration

Data Integration

Macro-level integration ← Micro-level Integration

Link different information of one unit to

- Different set of information of the same unit

- Information on another unit with same characteristics

use and compare collective statistics on a group of units (individuals or regions)

Statistical Units

- Dwelling Units
- Household

- Persons
- Land parcels

- Agriculture holding
- Farm Household

Stages in integration

- Integration can occur at different stages of the statistical process

Stages in Data Integration

Before data collection

International standards

Sample designs for data integration

Definition of variables

Common frames of reference

Stages in Data Integration

During field operations or data processing

International protocol for interviews

Sharing Field Materials

Sound methodologies

Methodologies of estimation

Stages in Data Integration

During field operations or data processing

International protocol for interviews

Sharing Field Materials

Sound methodologies

Methodologies of estimation

Stages in Data Integration

During Data Dissemination

Domains and Indicators

Common Interpretive schemes

Information systems

Links to other sources

Case Studies

Cases of different methods: Household survey data: different income/consumption modules

Example of Ethiopia 2000 surveys:

Reason: different consumption modules

	Sample size	Mean exp./p	Headcount (%)	Gini (%)
Welfare Monitoring survey 2000	25016	46.0	81.3	49.0
HH income & exp. survey 2000	16672	92.5	21.9	30.0

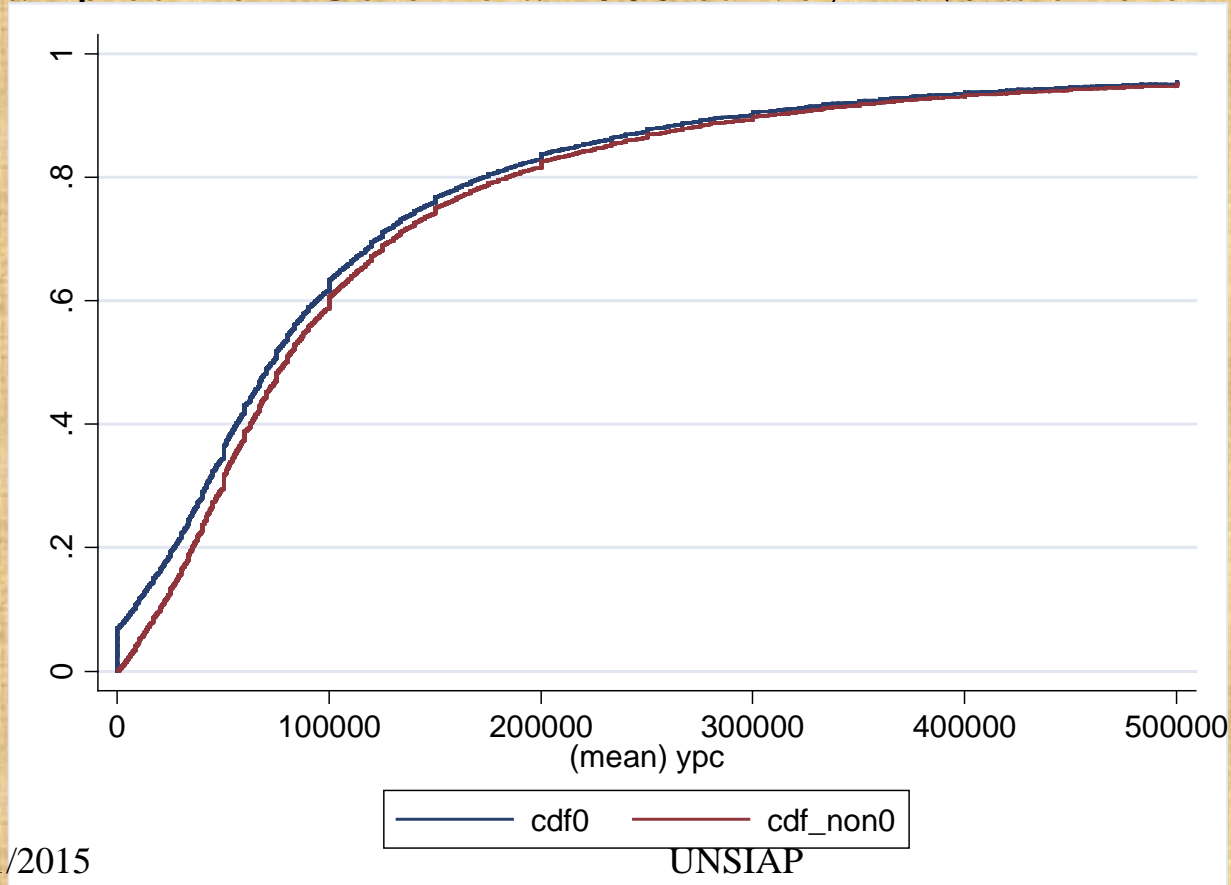# Common problems on data processing

1. Income/consumption aggregates
2. Valuing income in kind
3. Missing value
4. Outliers

Missing, zero and outliers

- Never mix missing value and zero;
- Examples from Latin America Caribbean labor force surveys
- Outliers: check carefully and always keep original records
  - Income by sources
  - Sub components of consumptions

Missing and outliers

Examples from Colombia 2000 survey – 7% are zero income

UNSIAP

Thank You