# Crowd-Sourcing: Integrated Data Collection Systems
# Palestinian Statistical Business Register, Obstacles and Methodologies

**Osaid Ismail**
oalarda@pcbs.gov.ps
Member in the team work of the Palestinian Statistical Business Register
**Palestinian Central Bureau of Statistics**

**PCBS**

## Introduction

"Business Register" (BR) is considered as one of the most important examples of Integrating data from several sources, that contains different forms in terms of coverage and quality, this difference, due to the difference in registration methodologies and main purpose of registration of each source.

"Business Register" (BR) as a concept has two forms, "Administrative Business Register (ABR)" and "Statistical Business Register (SBR)"; the main difference between them is that the census records which serves statistical purposes, belong to SBR and not to ABR. Moreover, ABR is the starting point to build the SBR.

## SBR Objective

### Cost
Very low cost system compared to implementing a comprehensive census
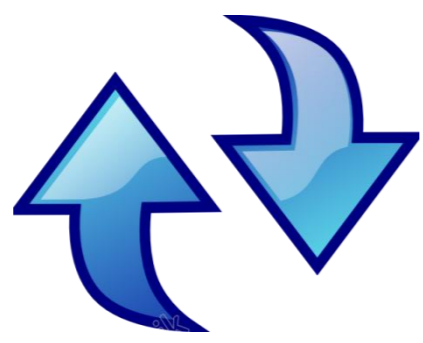
### Quality
High quality due to the interconnection of several sources compared to data quality from only one source
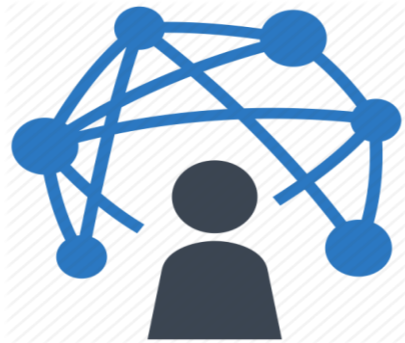
### Update
The Update of the system by partners is continuously compared to a census executed every 5 years

### Coverage
Broader coverage of institutions, benefiting from coverage of several sources and several methodologies in registration

## Life Cycle of SBR and Data Flow

### Data Collection
The files of data should be collected from each party, computerize the non-computerized ones, and keep them in the original and excel/csv format.
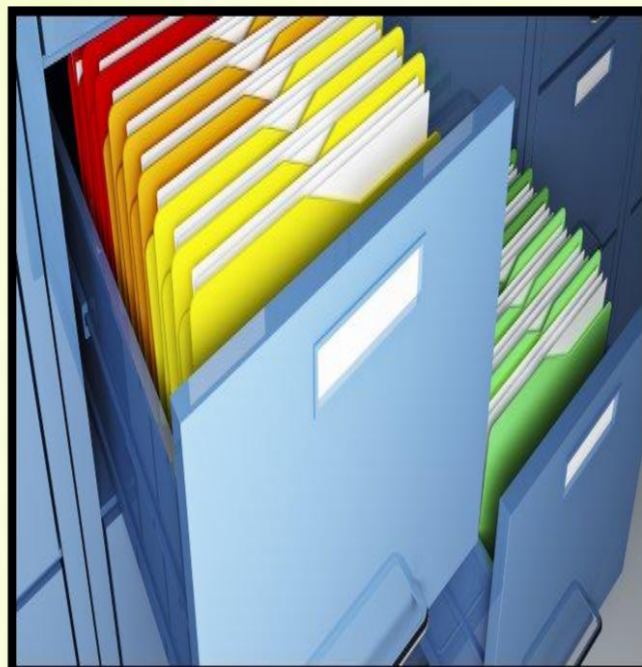
### Data Description
The data should be described in a special form to analyze the current situation, to facilitate the programming work in building the matching program and to study the problems of each file.

### Data Cleaning
A set of rules are implemented to be applied on different files to make the data more consistent as much as possible to improve the quality and coverage of matching.

### Data Matching
The matching process is based on comparing the matching variables together in a set of priorities in a certain score through Duke tool in JAVA.

### Manual Validation
Several variables are used to take the decision if the automatic matching is true or not, such as the ID, commercial name, owner name, phone numbers, activity, and the address.

### Values of the Variables
After the validation, the best value of each variable is chosen according to a set of priorities (depending on the quality of the files)

## General Description of SBR-V1

V1

The BR-V1 covers enterprises and establishments (local units) in the West Bank that are registered in the partners files which intersect between at least two partners. Almost **40,000** establishments (to compare with the **100,000** establishments recorded in establishments Census 2017) are already included in first version.
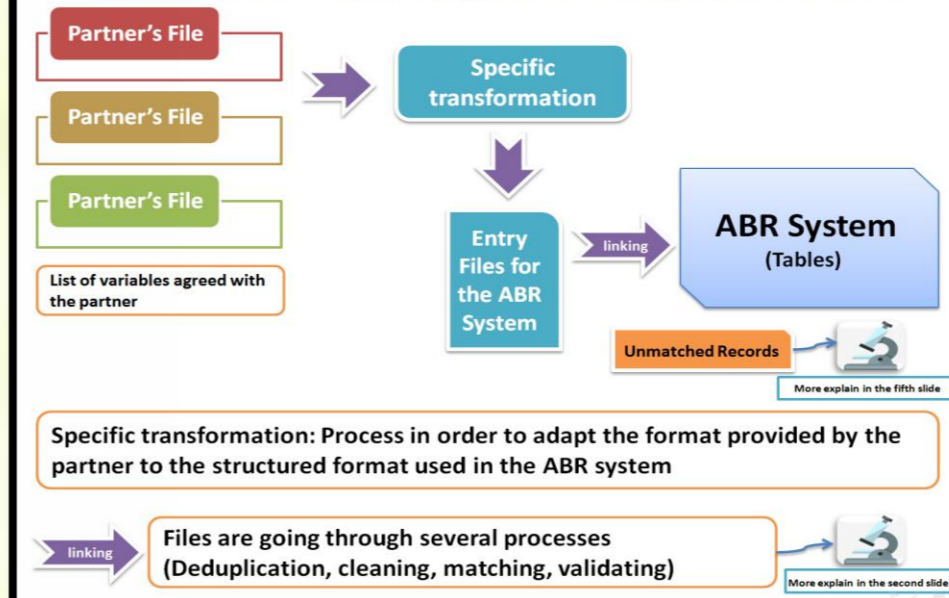
- Ministry of Finance
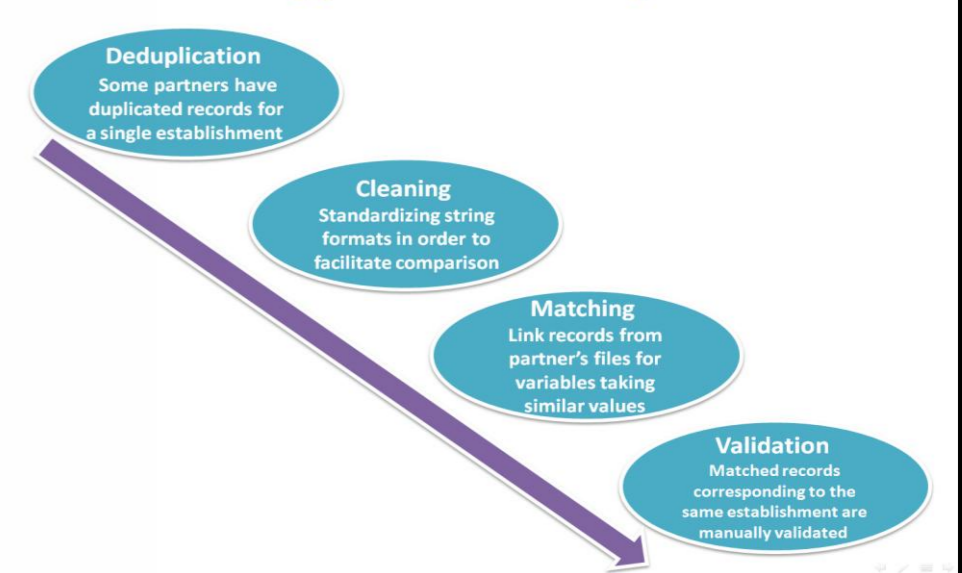- Ministry of National Economy
- Census
- 10 Municipalities
- Chambers of Commerce

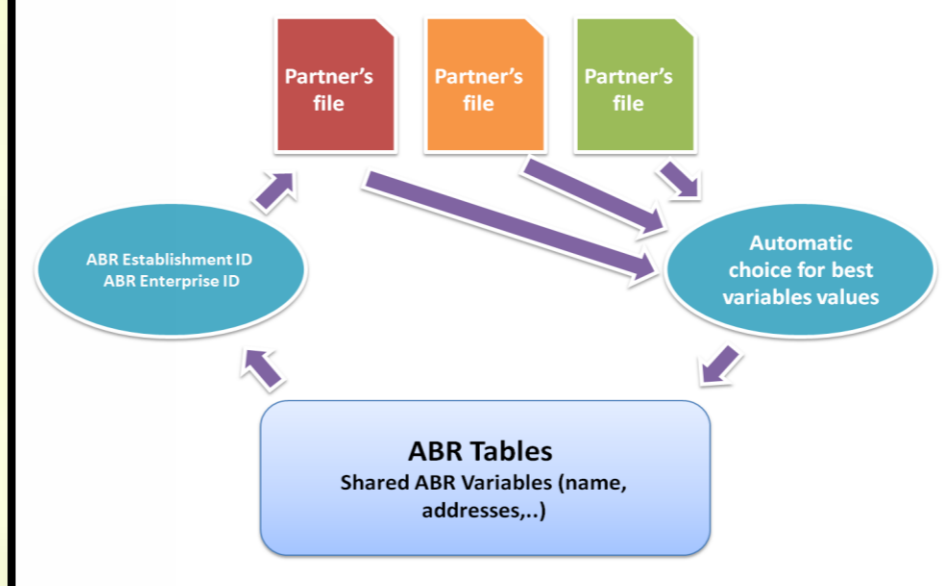## Building and Updating the Registry Processes
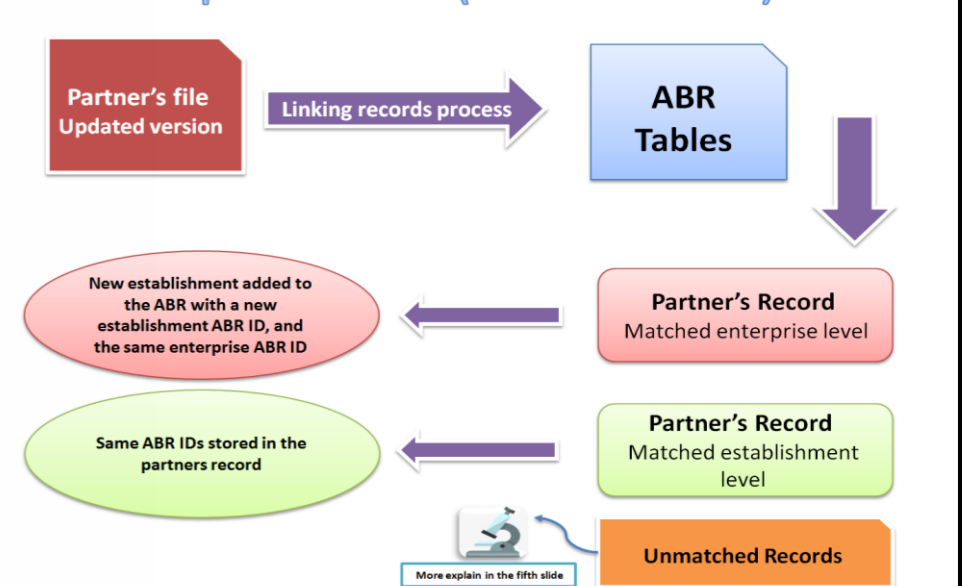
### Initial process for import of the partner's files



Specific transformation: Process in order to adapt the format provided by the partner to the structured format used in the ABR system

Files are going through several processes (Deduplication, cleaning, matching, validating)

### Linking partner's records process



- **Deduplication** Some partners have duplicated records for a single establishment
- **Cleaning** Standardizing string formats in order to facilitate comparison
- **Matching** Link records from partner's files for variables taking similar values
- **Validation** Matched records corresponding to the same establishment are manually validated
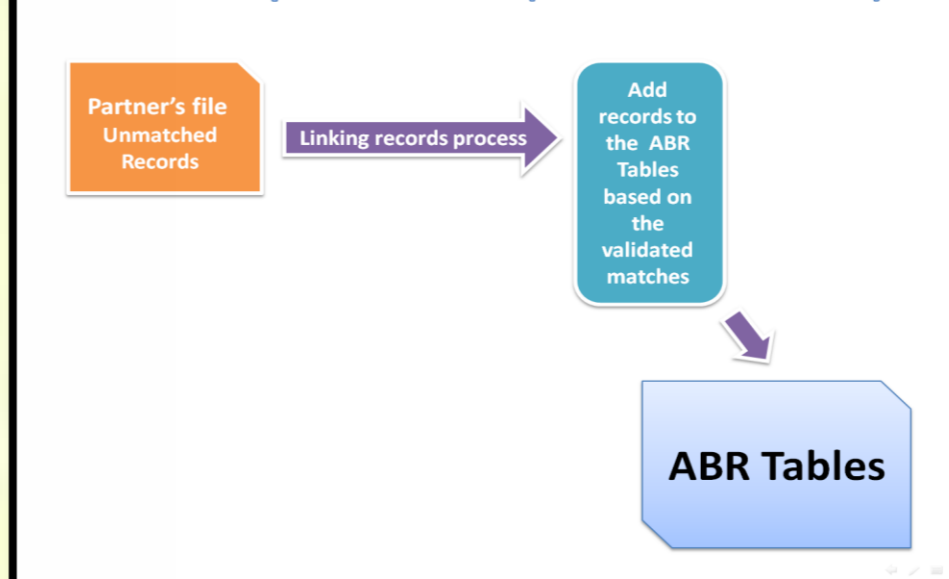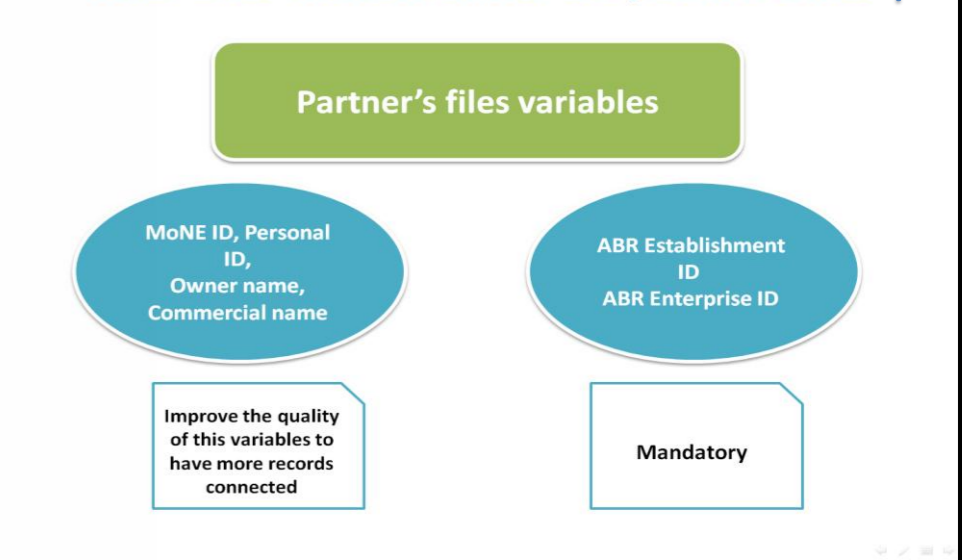
### Flow data into the ABR & Partner's feedback



### Update Process (Matched Record)



### Manual update Process (Unmatched Record)



### What does the ABR enrich the partner's files ?

Partner's files variables

- MoNE ID, Personal ID, Owner name, Commercial name
- ABR Establishment ID, ABR Enterprise ID
- Improve the quality of this variables to have more records connected
- Mandatory

## SBR System Basic Characteristics

- No duplication of establishments or enterprises records
- "ABR shared variables" that allow contacting economic units
- Storage all the Data in efficient, reliable, and secured way
- Connects the establishments registered to the enterprise they belong to
- Sharing the database among partner in order to use it for updating its own register

## The Risks and Difficulties (Challenges)

- Poorly the coverage of partner register on whish the ABR coverage is based
- Partner files do not record "matched" variables correctly , these variables is the most crucial information to create secure links
- The duplication of the records in the data of the partners, one of the responsibilities of the ABR management team
- The data collected by partners is poorly recorded, not updated, and not comprehensive
- The partner does not provide sufficient information to have an effective process of determining the links between records
- The follow-up of institutions to modernize their "active" statute, this follow-up should be undertaken by local administrations

**References :**
"Building ABR System Description" Document Developed by the SBR team in PCBS under the supervision of a French team

PCBS  Insee  EXPERTISE FRANCE

### Contact Info

PCBS Email:
Diwan@pcbs.gov.ps
P.O Box: 1647
Tel. Office: 00970 2982700