The Education Register

& ISOPED -
The system behind it

Aliza Krovi & Gustavo Schifris

Israel Central Bureau of Statistics

Preface

- **The Education Register** – a register that was developed by a team in the CBS
- A pilot study was made for one year
- To implement the register, a database of all education files is needed
- **ISOPED**-**I**ntegrative **S**ystem **O**f **P**rocessing **E**ducation **D**ata is a system that will produce the database

---

The presentation includes:

The Education Register
- What is it?
- How was it built?
- What do we need to make it an updating register?

ISOPED
- What is it ?
- Why do we need it?
- The components of the ISOPED
- The system that was developed

What is the Education Register?

- A statistical register that combines education data from several administrative sources

- Calculates 4 measures **for every person in the Population Register**

- The measures summarize the educational attainment of a person up to a certain year

The 4 measures are:

1. Is in education at present
2. Number of schooling years
3. Type of last school
4. Highest certificate attained

   These measures are available today from the Labor Force Survey for a sample of approximately 7000 persons

What are the uses of the Education Register?

- Production of current statistics on educational attainment
- Data for researchers
- International statistics such as OECD data collections
- Will be a part on the administrative census

Possible uses of the measures

	Is in education at present	Number of schooling years	Type of last school	Highest certificate
Characterize the population				v
International comparison	v		v	v
Sampling needs	v		v	v
Research		v	v	v

How to calculate these measures

Planning:
- Find the relevant files & variables for each measure
- Estimate the reliability of the variables
- Build an algorithm to calculate each measure while considering the time aspect

Calculating:
- Bring the variables to a common standard
- Match records from all files with the Population Register
- Write a SAS program
- Run!

Administrative files available

- The Population Register – the basis population

Education files such as:
- Students in primary & secondary education
- Matriculation examinations
- Universities and colleges – students & degree recipients
- Recognition of foreign degrees

**The average range of years covered: 15
Number of files per year: 20
20x15=300 files !**

Reliability of the data

1- Reported by an institution
2- Self reporting +presenting a document
3- Self reporting
4- Indirect reporting

The variables were classified into the 4
  levels of reliability

Reliability of data for every measure

Measures: / Files:	Is in education at present	Number of schooling years	Type of last school	Highest certificate
Recipients of academic degrees				1
University students	1		1	
Students in primary & secondary education	1	1(student) 3(parents)	1	
teachers				1
Recognition of foreign degrees		2		

A pilot study was made for one year

- ◻ An algorithm was built for a specific year
- ◻ Files were matched to the Population Registry
- ◻ The 4 measures were calculated from the data

Problems :
1. Matching 300 files to the Population Register
2. Standardization of the variables
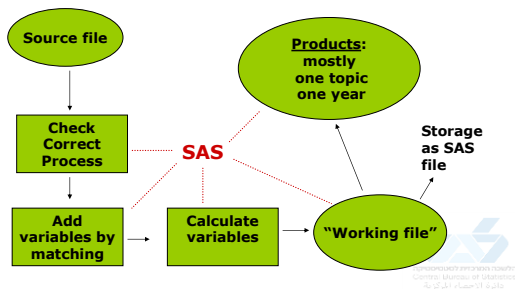3. Only good for one year… we want a register

The solution: ISOPED

Integrative
System
Of
Processing
Education
Data

A solution for a better work process, better storage and better use of the data

Working with Administrative files - current method:

```
Source file

Check
Correct          SAS
Process

Add              Calculate        "Working file"
variables by     variables
matching

Products:
mostly
one topic
one year

Storage
as SAS
file
```
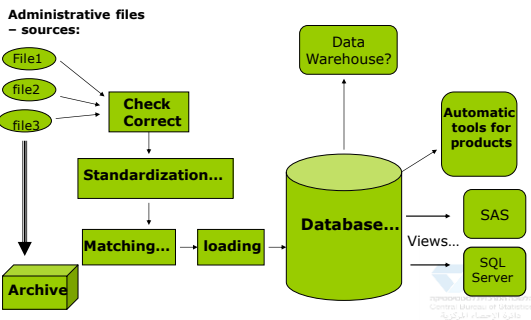
Problems with this method:

- Various working techniques
- File storage
- Files & variables Metadata
- Inefficient methods of producing products
- Not enough products based on data crossing years & topics

The ISOPED system – an alternative working method: THE VISION

Administrative files – sources:

File1
file2
file3

Check Correct

Standardization...

Matching... → loading

Archive

Data Warehouse?

Database...

Views...

Automatic tools for products

SAS
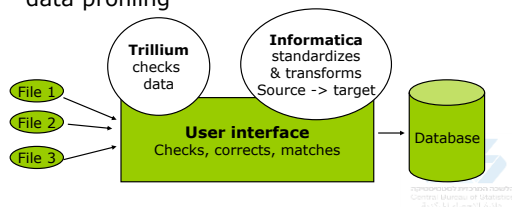
SQL Server

We need a tool that will be able to:

- Read files from various formats
- Manage metadata
- Check the data
- Do transformation from source to standard/target
- Match records from files with Database
- Load data into the Database
- Be friendly… 😊

The tool that was built

A user interface based on two tools:

1. Informatica: PowerCenter
2. Trillium Software System: Discovery, data profiling

File 1
File 2
File 3

Trillium checks data

Informatica standardizes & transforms Source -> target

User interface Checks, corrects, matches

Database

The project's current status

- A database was built, views were planned
- For every file (about 200) an Informatica map was created
- For every file a set of checks was designed
- A matching system was built (only exact match)
- The education files from two years were processed into standard tables
- We are about to do the matching and loading

Shortcomings of the system

- The system is not flexible - built for files of specific structure
- We lack a centralized metadata system
- The matching system is designed only for exact matching

It will take time until the system reaches its final shape

Thank You !!!

Aliza Krovi
alizak@cbs.gov.il