



# ROMANIA'S EXPERIENCE with DATA MATCHING

**Nicolae Draganescu**, Agency for Social Benefits and Inspection, Romania  
**Sabin Chiricescu**, ICT Consultant, The World Bank

## Session 6, Romania

WORKSHOP ON  
REDUCING ERROR, FRAUD & CORRUPTION (EFC) IN SOCIAL  
PROTECTION PROGRAMS  
June 8-12, 2014, Opatija, Croatia



# OUTLINE

- Why
- How we started
- Detecting data mismatches
- What we did (2013)
- Lesson Learned
- 2014, Plans and Road ahead

# Why use data matching?



To identify cases where beneficiaries' information is recorded differently in different public databases

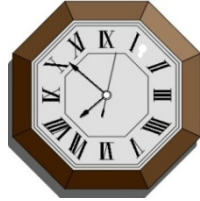
Use these data mismatches to target administrative checks or inspections

Compared to paper-based verification, these checks are cheap, comprehensive and cost-efficient

- You can check all beneficiary files at once
- Marginal cost of an additional check is close to zero
- High cost-benefit ratio: Instead of random inspection, focus the inspections on beneficiary files with data mismatches, especially those benefits with high level of potential overpayment
- Suggestions for improving data compatibility among different institutions for reducing error, fraud and corruption (EFC)
- Reducing administrative errors (on management information systems [MIS] of social assistance [SA] and IT systems– external data sources)

**Data mismatches → files with suspicion of EFC → ranked by risk level → high risk cases were verified**

# Context – Romania's Social Inspection



- **Social Inspection established in 2007**
- **Social Inspection focused on social services until 2011**
- **Starting 2012, the focus shifted on inspecting cash transfer programs using random inspections; limited data matching for the disability allowance program; poor data sources**
- **2013-2014:** cross-checking beneficiaries' data from the Social Assistance MIS (SAFIR) with information from other public databases: Civil Registry, tax authority, pension service, unemployment registries
  - First bulk data matching November 2012 – March 2013
  - Covered 4 large social assistance programs (4.5 million beneficiaries)
  - 11 million monthly records checked

# Main social assistance programs under data matching



Program	Key eligibility criteria	Number of beneficiaries
<b>Family allowance (FA)</b>	Means tested program: <b>income</b> and <b>asset</b> test Benefit level: fixed amount per child Assistance unit: family with children Conditionality: school attendance	~ 260.000
<b>Guaranteed Minimum Income (GMI)</b>	Means tested program: <b>income</b> and <b>asset</b> test Benefit level: difference between income guarantee and household income Assistance unit: household Conditionality: work search, workfare	~ 170.000
<b>Child raising benefit (CRB)</b>	Period of benefit could be one or two years; Benefit level: for the first year 85% of the monthly medium <b>income</b> computed for a period of one year before child birth; for the second year a fixed amount	~ 170.000
<b>State child allowance (SCA)</b>	Universal allocation for all children by 18 years and/or by the end of school (lyceum or equivalent)	~ 3.800.000

# How we started?

## Pre-requisites for data matching



- Each Romanian has a Personal ID with unique identifier (CNP)
- Most public databases are indexed on CNP
- Civil registry issues CNP and maintain the database of the population
- However, there is no database with family composition
- Different type of incomes are recorded in Tax, Pension, Unemployment and Social assistance databases



# How we started? Identification of the a-priory EFC risks of each program



Program	EFC risks	Availability and quality of external data
<b>Family Allowance (FA) and Guaranteed Minimum Income (GMI) programs (means tested)</b>	Family	No (accurate) electronic records about family composition
	Person identity; entry in civil registry (e.g. deceased)	Good quality data (Civil registry)
	<b>Income (high risk)</b>	Good quality data (Tax and Pension)
	<b>Assets (high risk)</b>	No national, electronic registries of key assets (land, houses, cars, livestock etc.)
	Other conditionalities: some of them have data sources, but not data-matching protocols; some of them checked through MIS (e.g. school absences)	
<b>Child Raising Benefit (CRB)</b>	Person identity	Good quality data (Civil registry)
	<b>Income (high risk)</b>	Good quality data (Tax and Pension)
	<b>Working during SA period (high risk)</b>	Good quality data (Tax) – if there is an income this means he/she works while he/she is claiming not to work
<b>State Child Allowance (SCA)</b>	Person identity	Good quality data (Civil registry)
	Age of child	Could be checked; it can generate false suspicions of EFC



# Detecting data mismatches: deceased beneficiary

Payments records from the Social assistance MIS (SAFIR) (order by person ID and month)

MONTH	PERSON ID (CNP)	AMOUNT
Jan 2014	1640403nnnn	100
Feb 2014	1640403nnnn	100
Jan 2014	1750131mmm	100
Feb 2014	1750131mmm	100
March 2014	1750131mmm	100
<b>April 2014</b>	<b>1750131mmm</b>	<b>120</b>
<b>May 2014</b>	<b>1750131mmm</b>	<b>120</b>



Civil registry information  
on birth and deaths

PERSON ID	PERSON NAME	BIRTH DATE	DECEASE DATE
1750131mmm		Jan 31, 1975	<b>March 10, 2014</b>
123456mmm	Sabin Chiricescu	April 3, 1964	





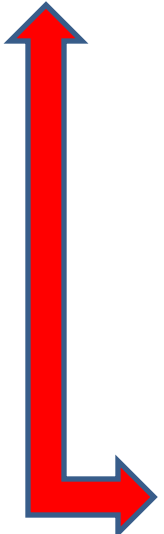
# Detecting data mismatches: Not reporting a type of income ?

For precise comparisons, you need incomes classified by type, earner and period (month)

MONTH	Family ID	PERSON ID (CNP)	income TYPE	income AMOUNT
Jan 2014	1	1750131mmm	Monthly Wage	200
Jan 2014	1	1750131mmm	Pension	100
Jan 2014	<b>1</b>	<b>2800131nnnn</b>	<b>House Rent</b>	<b>50</b>
<b>Total per family</b>				<b>350</b>

Income information from Tax administration

MONTH	PERSON ID	income TYPE	income AMOUNT
Jan 2014	1750131mmm	Monthly Wage	200
Jan 2014	1750131mmm	Pension	100
Jan 2014	<b>2800131nnnn</b>	<b>Copyright</b>	<b>50</b>
<b>Total per family</b>			<b>350</b>





# Detecting data mismatches: Compare aggregate data per family !

Sometimes it is difficult to compare records that are not identical or are recorded at different level of aggregation - Solution: compare total family income from Social assistance MIS with Tax authority MIS

MONTH	Family ID	income TYPE	income AMOUNT
Jan 2014	1	Monthly Wage	350
Total per family			350

Family details

Family ID	PERSON ID (CNP)
1	1750131mmm
1	2800131nnnn

Income information from Tax administration

MONTH	PERSON ID	income TYPE	income AMOUNT
Jan 2014	1750131mmm	Monthly Wage	200
Jan 2014	1750131mmm	Pension	100
Jan 2014	2800131nnnn	Copyright	50
Total per family			350

Conclusion: No suspicion in this family

# Step by step approach

## Preparation



1. **Identify EFC risks (e.g. rent income not reported or under-reported)**

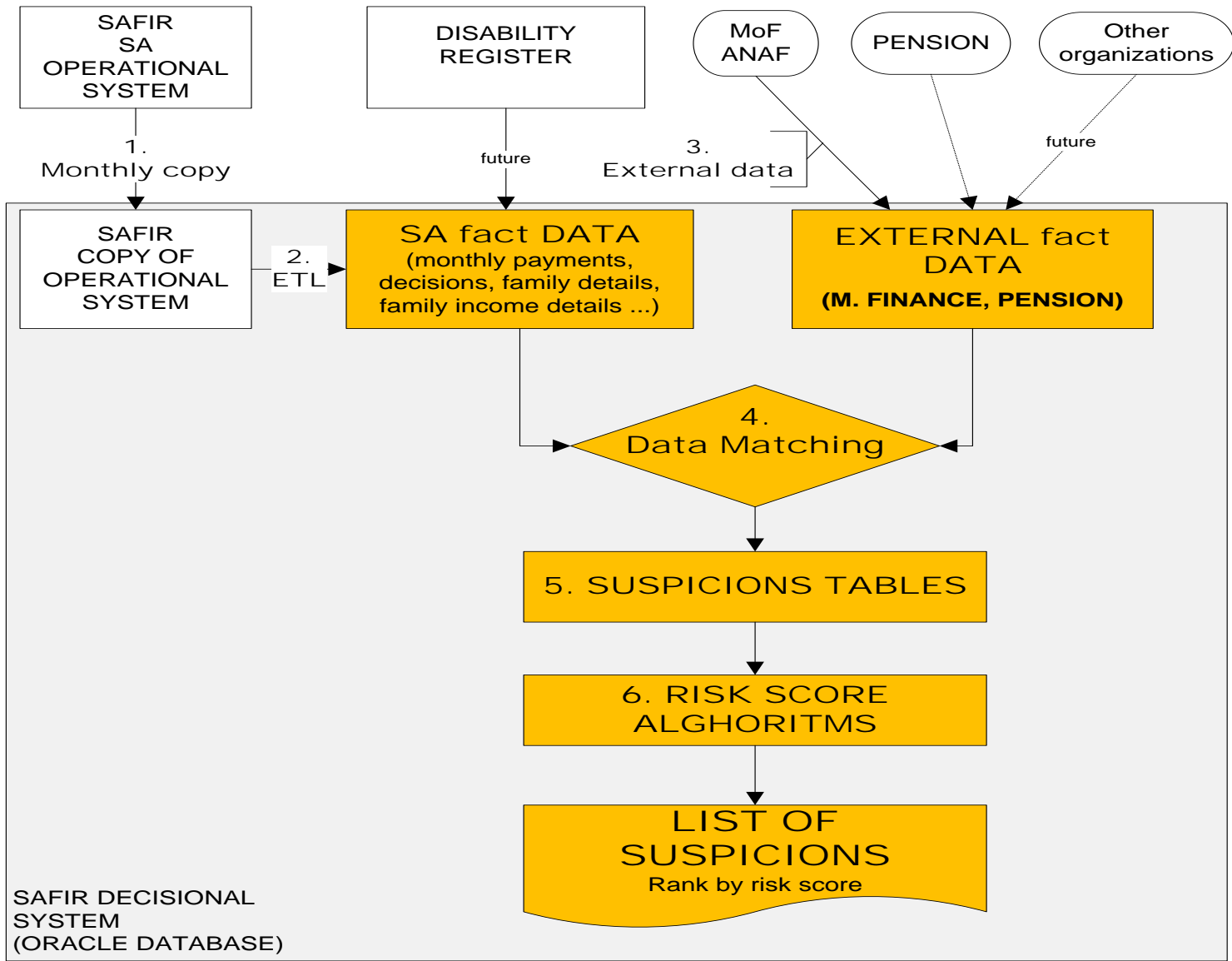
**There is not a rule for how you do it, depends on legislation, SA program, MIS (SAFIR) data, previous social inspection (SI) experience**

**Focus on high risks (e.g. income and assets)**

2. **Do we have an external data source to compare?**
3. **Conclude data exchange protocols with external institution**
4. **Asses the quality of external data (ex. Tax records on income from abroad are of poor quality – incomplete and inaccurate)**

# Step by step approach

## Data matching flow



# What we did (2013)

## Identification of risks - verification area



- **Verification of deceased persons** who are included for payment in the SAFIR system after their death
- **Verification of non-existent persons** who do not appear in external databases but are included for payment in the SAFIR system
- **Verification of income** (salaries and/or pensions and/or other social benefits) earned by persons/families while they receive a social assistance benefit

# What we did (2013)

## Data needed to be obtained



### ➤ **Personal** data (id, name, birth date, death date)

In Romania any person has a personal id (CNP) used on any major national databases

### ➤ **Data regarding the income** earned by person (wages, pensions, allowances, amounts from leases, royalties, copyrights etc.)

## Providing institutions of external data

- Ministry of Public Finance, National Tax Administration Agency,
- National Employment Agency,
- Civil Persons Register - Ministry of Interior
- National Public Pension House
- **MIS (SAFIR) itself for SA programs income (child allowance, child rising allowances should be added on family income for FA, GMI)**

# What we did (2013)

## Data sources



DATA TYPE	PROVIDING INSTITUTION
Personal identification data	Fiscal Hub (instead of Civil Register)
Monthly <b>income</b> earned by person (wage, unemployment allowance ... around 20 types)	Fiscal Hub
Yearly <b>income</b> ( <i>amounts from leases, royalties, copyrights</i> )	<i>Fiscal Hub</i>
Monthly pension <b>income</b>	Pension House
Monthly SA allowances - <b>income</b>	NASBSI - SAFIR itself

**Conclusion: It's a matter of protocols and data exchange format**

# What we did (2013)

## Suspicion – what we are looking for

(programs: FA, GMI, RCB, SCA)



### ➤ Suspicions related to “personal data”:

- Non-existent persons
- Deceased persons
- Persons’ age (children) – SCA benefit only

### ➤ Suspicions related to “income of persons”

- **Over-reporting in CRB** to achieve a higher benefit (for the first year the benefit is 85% of reported medium monthly income)
- **Not reporting all incomes (or under-reporting) in GMI**



# What we did (2013)

## Use data based on SA needs and constraints



- Compared non-identical data, even at risk of many false positives - compared **estimated** net wage (based on gross wages in Tax records and tax rules) with **actual** net wages recorded in the social assistance MIS

Use an algorithm for computing net income as requested by SA legislation

- **Filter data - cross-checking cases with most likelihood of irregularities**

Lesson:

- To avoid many false positives, you need IDENTICAL data records in the two databases;
- But you can still get accurate indications of cases with EFC risk even with non-identical data

Better to start data matching than expect the perfect data

# What we did (2013)

## Ranking suspicions based on severity



- **Generated more suspicions than SI had to capacity to investigate (b/c of comparison of estimated with actual records)**
- **Assigned different risk level for each suspicion**
  - **From 1 to 10 (the highest)**
  - **Higher risks for longer duration of irregularity, level of overpayments or for larger differences (e.g. in reported wage incomes in the two databases)**
  - **The Social Inspection team focused capacity on the cases with highest risk**
  - **Then checked whether our risk scoring algorithm worked or not after inspection**

### **Examples of parameters for CRB benefit y (risk)**

- **Over payment months**
- **Total SA amount during over payment period**
- **External average monthly gross income**

Higher risk level (9,10) computed for longer over payments period (more than 6 months) and large overpayments amounts and gross income

Assigning risk is important in prioritizing cases for verification



# What we did (2013)

## Example of a of cross-check (FA)

FAMILY ID	SUSPICION TYPE	No OVER PAYMENTS MONTHS	RISK LEVEL	PERSON ID	PERSON NAME	EMPLOYER ID	SA BENEFIT	FISCAL income	PENSION income	SA income
nnn	FA limit	6	7	1	Parent 1	10953691		600	-	-
						13341930		191	-	-
						16892658		2,471	-	-
						14075542		594	-	-
				2	Child		SCA	-	-	630
				3	Parent 1			-	5,625	-
<b>TOTAL INCOME for OVER PAYMENT MONTHS</b>								<b>3,856</b>	<b>5,625</b>	<b>630</b>
<b>TOTAL INCOME PER FAMILIY</b>								<b>10,111</b>		
<b>AVERAGE INCOME PER MONTH PER FAMILY</b>								<b>1,685</b>		
<b>AVERAGE INCOME PER MONTH PER FAMILY MEMBER</b>								<b>562</b>		
<b>INCOME LIMIT for means tested FAMILY ALLOWANCE (per member)</b>								<b>370</b>		

# What we did (2013)

## Data matching summary results, 2013



Program	Area of investigation	Number of verified files	The number of verified persons	Number of payments subject to review	Number of cases with suspicions of major EFC	Number of cases with suspicions of minor EFC	Number of total cases identified where EFC is suspected
<b>CRB</b> 26 months: Jan 11 - Feb 13	Person	435,178	785,801	4,713,983	544	5,719	<b>6,263</b>
	Income	435,178	785,801	4,713,983	23.612	6,943	<b>30,555</b>
<b>GMI</b> 15 months: Dec 11 - Feb 13	Person	282,874	544,096	2,978,448	3.048	3,137	<b>6,185</b>
	Income	143,739	168,043	1,379,948	2.080	7,699	<b>9,779</b>
<b>FA</b> 15 months: Dec 11 - Feb 13	Person	425,232	1,370,455	4,503,198	7.426	4,899	<b>12,325</b>
	Income	295,074	933,863	3,074,041	14.405	24,934	<b>39,339</b>
<b>SCA</b> 26 months: Jan 11 - Feb 13	Person	4,463,455	7,890,364	100,057,521	9.544	2,404	<b>11,948</b>
	Age	4,463,455	4,411,719	100,057,521	50.970	478,584	<b>529,554</b> <i>High rate of cases with potential EFC may also be caused by incomplete data</i>

# What we did (2013)

## Overall results of data matching



Program	SA MIS (SAFIR)	Number of cases with suspicions - data matching	Number of cases investigated 2013	Number of cases with EFC, 2013	Total over payments (debts) 2013 (Lei)	Recovered debts by Mar 31, 2014 (Lei)	Costs Dec 31, 2013 (Lei)
CRB	Yes	36,818	27,720	5,650	11,569,170	5,117,824	115,692
FA	Yes	51,664	34,123	15,151	2,643,421	1,690,999	1,129,939
GMI	Yes	15,964	9,676	2,959	1,730,223	877,075	
SCA	Yes	541,502	27,308	2,549	1,189,902	692,476	74,964
Heating		35,851	24,030	4,874	755,368	241,046	490,989
Disabilities		5,457	5,457	2,724	1,301,069	551,914	143,118
<b>Total</b>			<b>128,314</b>	<b>33,907</b>	<b>19,189,153</b>	<b>9,171,334</b>	<b>1,954,701</b>

Cost-benefit ratio on overpayments (estimated debts)	9.82
--	------

Cost-benefit ratio on recovered debts	4.69
---------------------------------------	------

<i>SA MIS (SAFIR) Cost-benefit ratio on recovered debts</i>	<i>6.34</i>
---	-------------

# Lessons learned after the first data matching exercise



- Prior to 2013, there was neither a concept nor a capacity to match data covering a long time period between SAFIR and other institutional/external data
- Controls focused on EFC suspicions → saving significant material, financial and human resources
- Change in the manner of control → from administrative to verifications where there are suspicions

**SA datamatching exercise was the first on this scale (tens of million rows; three data sources)**

# Lessons learned after the first data matching exercise



- No harmonised data among institutions (ex. gross against net; different income types)  
  
On 2014 the fiscal legislation was improved for asking net income requested by SA, Statistics, etc.
- Good data matching and risk profiles starts from program requests (if income per person and per type is not asked at the outset of a claim, it becomes more difficult to verify income later on)
- Understanding data is a must → if not, lots of suspicious cases
- Assigning a risk level is important → SI should prioritize most suspicious cases

# Data matching 2013

## Few resources – Great team - Great results



Half of data matching team (third of IT team)  
Not doing only data matching

Data matching desk  
Without laptops





# Data matching 2013

## Few resources – Great Team - Great results



Data matching & SI

Data matching rules and calendars



# 2014: Data matching becomes standard business practice



- 2012: First data matching exercises, not bulk, not standardised
- 2013: First bulk data matching for 6 programs
- **2014: Data matching twice per year, regular business function of the National Agency for Social Benefit**
- Improvements:
  - New cross-checking criteria : CRB medium income for computing SA amount
  - New approach: verify families instead of persons) → then a file could be checked once for all possible irregularities (till 2014 the SI was focused on programs)
  - Improved risk level computing algorithm (criteria and their importance )

# Plans to expand data-matching



- Major programs (FA, GMI, CRB, SCA) – checked twice per year
- Disability allowance and heating programs– checked once per year
- Identification of new data sources for additional cross-checks
  - Annual income (amounts from leases, royalties, copyrights)
  - Unemployment – quarterly eligibility document (“adeverinta ANOFM”)
- Identification of new risk groups/irregularities
- Develop a dedicated MIS social inspection module

# MIS social inspection module

## End-to-end monitoring system



- Inspection planning
  - Collects and record suspected cases of EFC from different channels (public, staff, data matching, risk profiling)
  - Documents how the sample of cases to inspect has been selected
- Inspection implementation
  - Records the results of each review
  - Records the action taken, sanctions applied and the amount of over- and under-payments
- Reporting on results and continuous monitoring
  - Produces tables with key results of the inspection for monitoring purposes
  - Follow up after a number of months

# Road ahead

## Future MIS and Social Inspection flow

