Decade of Roma Inclusion progress monitoring indicators

Results from the pilot in Bulgaria

Summary

1. **Types of indicators**
2. **General principles and approaches to data disaggregation by ethnicity**
3. **Possible data sources**
4. **Examples of their strengths and weaknesses**
5. **Sequence of the next steps**

Types of indicators

- Input indicators

- Output indicators

- Outcome indicators

- Impact indicators
    - …as well as sustainability indicators
    - …positive/negative externalities indicators
- All these indicators should be present in the NAPs and all need require kind of data

Types of indicators: one example

Hypothetical project aiming to boost employment trough the requalification of unemployed persons

- **Input** indicators: number of trainings per unemployed, number of lectures per unemployed, unit cost of training
- **Output** indicators: number of unemployed who passed a requalification training as share of unemployed
- **Outcome** indicators: percentage of those who found a job out of the total number of those who passed a requalification training
- **Impact** indicators: registered changes in the household income of those who have passed a requalification training (with sub-group "of those who found a job)
  - ☐ Sustainability indicators: duration of the job, found after the requalification
  - ☐ Externalities indicators: decrease of the rate of the drop-outs from school, decrease of the social fragmentation

General principles for the design of Decade indicators

- It is neither possible nor reasonable to invent and implement specific "**Roma indicators**".
- The **targets** – and not the instruments that measure the progress toward the targets – can be specific, reflecting the specificity of the challenges
- For the monitoring of the Decade **standard** socio-economic and human development indicators should be applied
- Standard indicators must be fed with **ethnically disaggregated data** to achieve ethnically disaggregated indicators
- Anything solution that is too simple in that regard is inevitably simplistic and hence **misleading**

Main challenges

- How to identify the universe under study (answering the question "**who is Roma?**")
- What kind of **ethnic markers** can be used for disaggregation of socio-economic data by ethnicity and thus compute ethnically disaggregated indicators?
- Which of the existing instruments and ongoing statistical data collection exercises can be used?
- What type of data concerning past periods can be disaggregated retrospectively for comparative purposes and trends monitoring?

Possible approaches to ethnically disaggregated data

1. Disaggregating hard statistics using **personal identification numbers as a common link** between mutually complementing data sets

2. Disaggregating hard statistics using **territorial tags** as ethnic markers

3. Extending the samples of regular sample based surveys with **ethnic boosters**

4. Conducting custom **"on the spot" surveys** among recipients of different social services

5. Collecting data at a community level by **community-based data collectors and monitors**

UN
D P

Data sources

- Regular population censuses
- Sample based surveys (household budget surveys, labor force surveys, LSMS, MICS, sociological surveys, etc.)
- Administrative registries
- Line ministries registries (in particular, Ministry of Education and Ministry of Health)
- Special agencies registries (Health insurance institute, National social insurance institute)
- Anonymous surveys conducted on the spot by service providers (labor offices, hospitals)
- Data collected at community level

Links between the different indicators and sources

Time frame

Long term

Short term

- Different types of indicators:
  - □ address different phases of the process
  - □ require different type of information that can be obtained from different sources

- Impact indicators ➔ data from regular population census

- Outcome indicators ➔ data from HBS, LFS and other similar instruments

- Output and input indicators ➔ data from individual institutions reporting systems.

Using personal identification numbers as common link between different data sets

Assumptions of the approach:

- Administrative and other registries do not maintain data on ethnicity
- So does Personal Identification Number
- Ethnicity however is registered during census and so is PIN
- Most of administrative registries use PIN as well
- Using PIN as common link between ethnic attributes from census and different data sets, various administrative registries can de disaggregated by ethnicity and ethnic-sensitive indicators can be computed
- This should be done on aggregate level (not revealing individual ethnic identity)
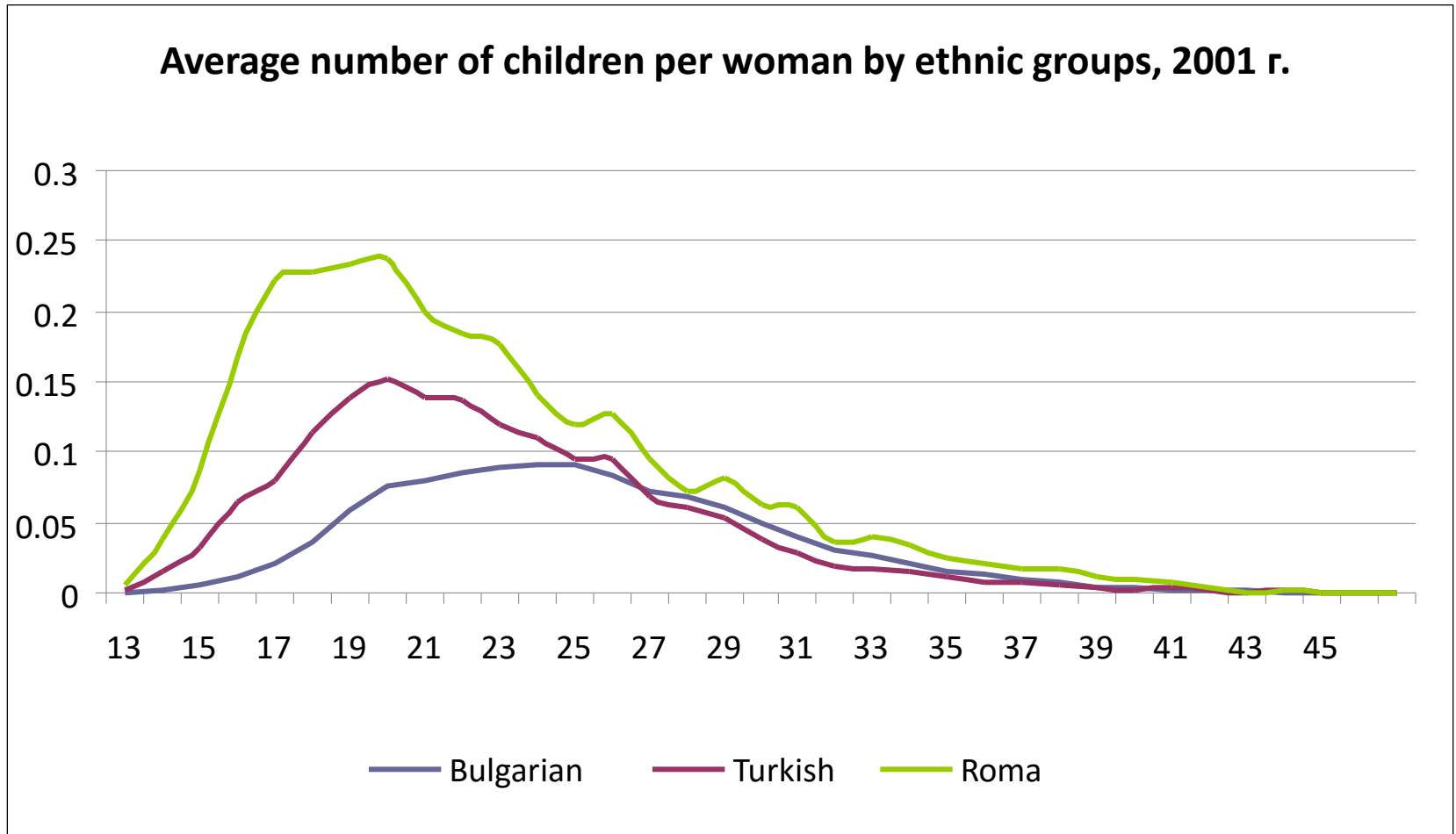
Using personal identification numbers as a common link – the logic of the approach
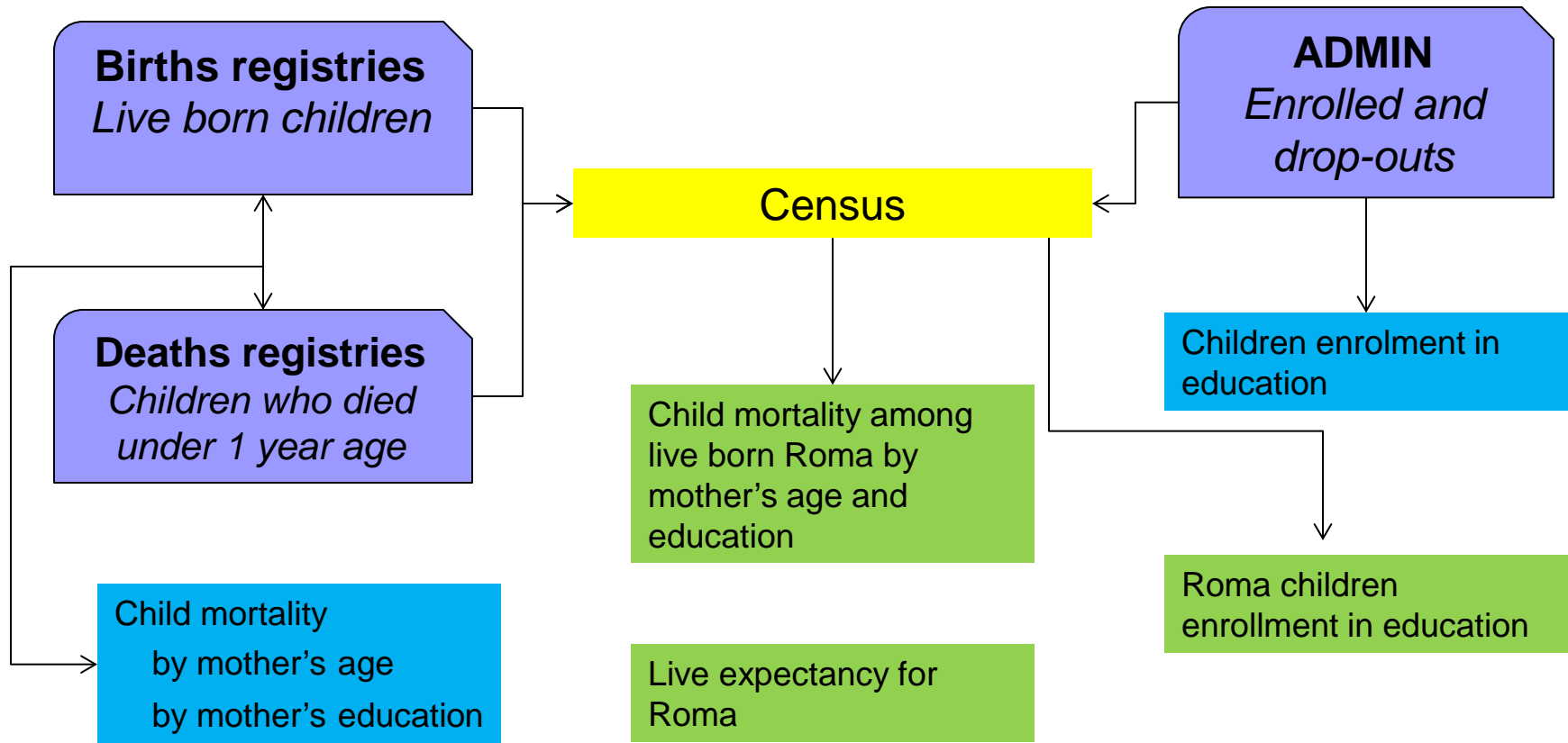
# Using personal identification numbers as a common link – examples

Ethnic group	1992	2001
Average number of children per woman		
Bulgarian	1,41	1,16
Turkish	1,92	1,64
Roma	2,93	2,77
Early (juvenile) birth rate (births per 1000 of age below 18)		
Bulgarian	66,2	41,3
Turkish	283,1	179,6
Roma	690,3	508,8
Extremely young birth rate (births per 1000 of age below 15)		
Bulgarian	3,1	2,4
Turkish	20,3	21,5
Roma	70,1	35,6
Child mortality by ethnic group (deaths per 1000)		
Bulgarian		9.9
Turkish		17
Roma		28

Using personal identification numbers as a common link – examples

Average number of children per woman by ethnic groups, 2001 r.

Using personal identification numbers as a common link – examples

Life expectancy
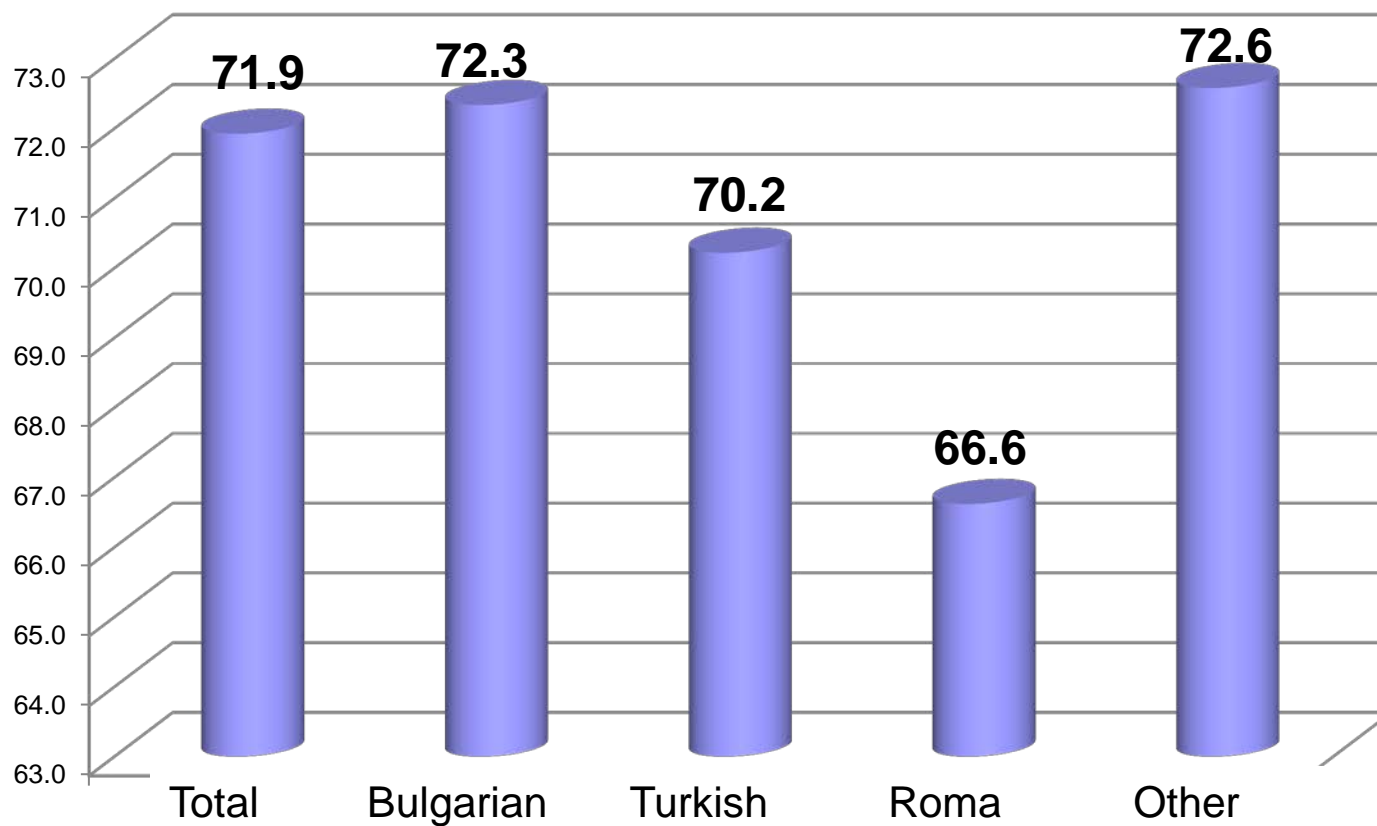
Health indicators that are possible to compute using PIN as a common link

- Prenatal, neonatal and postnatal mortality
- Number of not hospitalized births out of the total number of births
- Child mortality by mothers' age
- Roma morbidity (most common illnesses)
- Percentage of Roma with health insurance
- Percentage of Roma covered by screening surveys
- Number of Roma who passed a regular medical check-up
- Number of Roma registered in the system of social service's primary health care

Territorial tags as ethnic markers

Assumptions of the approach:

- Most of the vulnerable Roma are isolated and excluded territorially in separate (often segregated) communities
- Territorial mapping of those communities is possible
- Once a detailed map of Roma-dominated communities is available, it will be possible to correlate ethnic characteristics with territorial tags (individual's address)
- This will allow to monitor a standard set of statistical indicators for a population **leaving in an area with ***% of Roma**

# Ethically-disaggregated data based on territorial tags

- Allows to estimate **the absolute number of the population** more precisely that census (the absolute number is crucial to determine both the resources needed and the target indicators)
- It can be more **reliable** solving the problem with the refusal to declare real ethnicity in the census or to declare different one
- It is less susceptible to fluctuations due to changes in political environment
- Can be combined with GIS mapping
  BUT
- It grasps the marginalized, visually excluded segment of the Roma population
- It is complementary to the PIN-based approach and it does not replace it
- To be meaningful, the approach should be used on a level lower than "municipality"

Territorial mapping of Roma (concentration)

Share of Roma population by municipalities, 2001
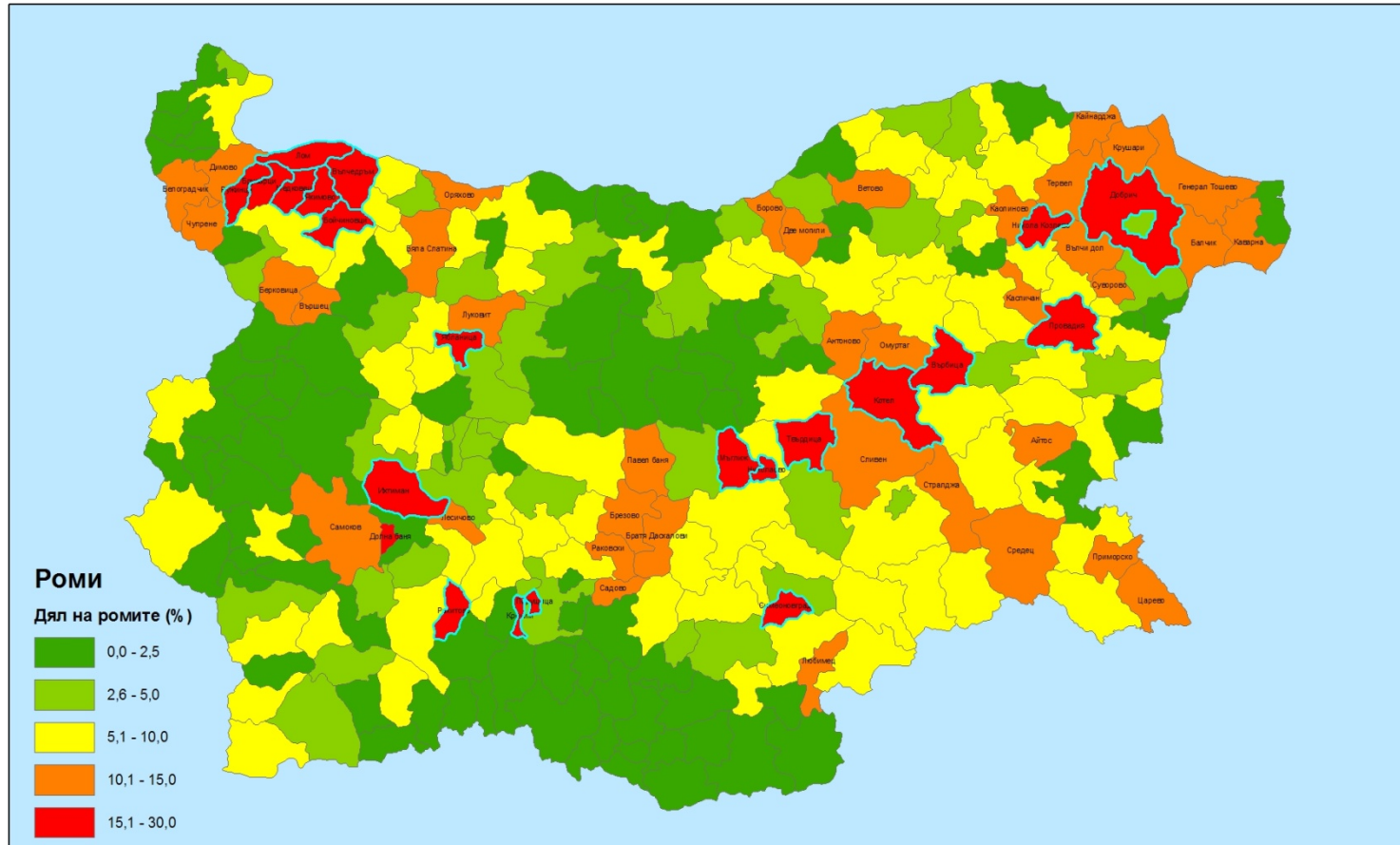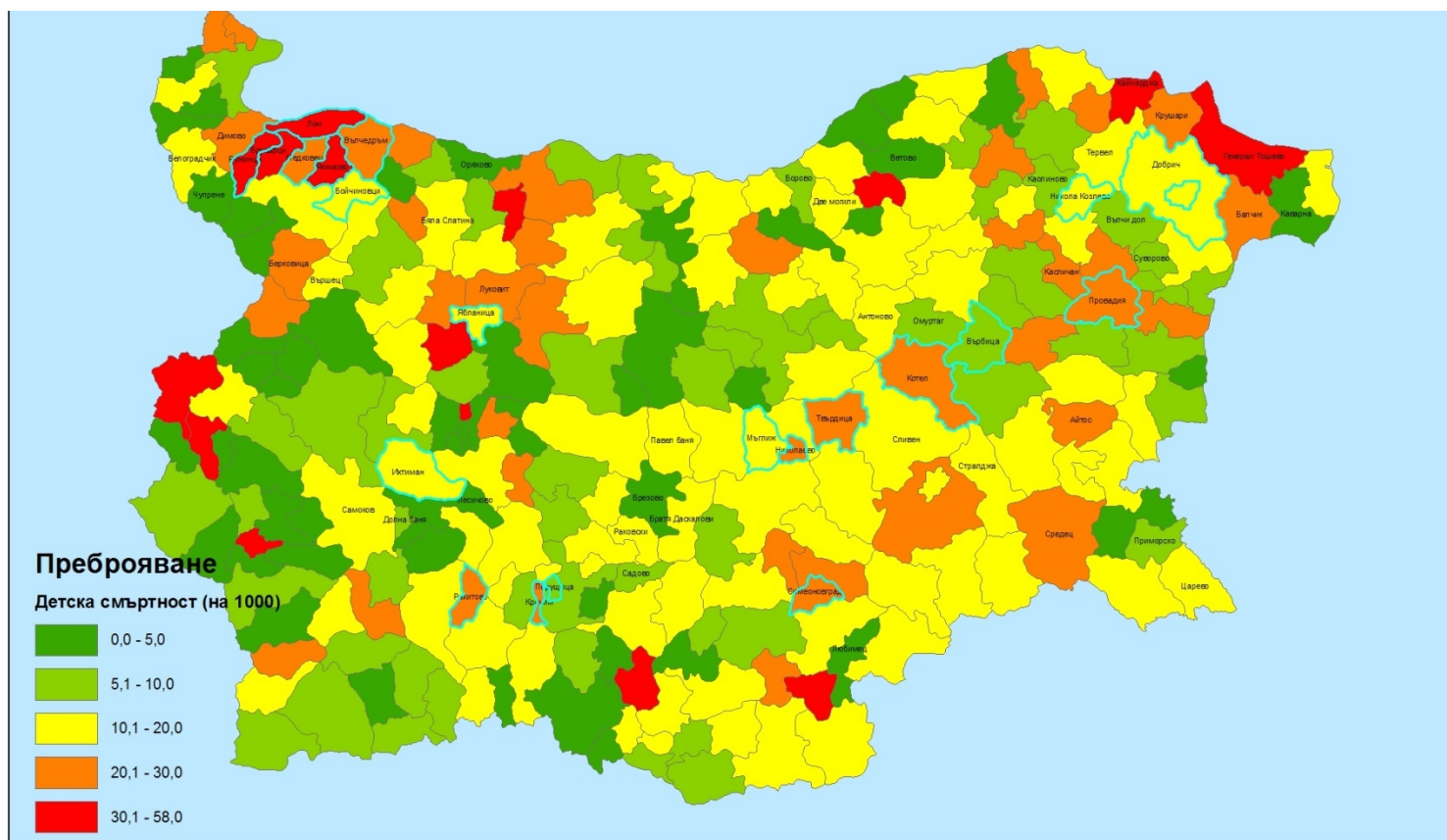
Illiteracy (%) by municipalities, 2001
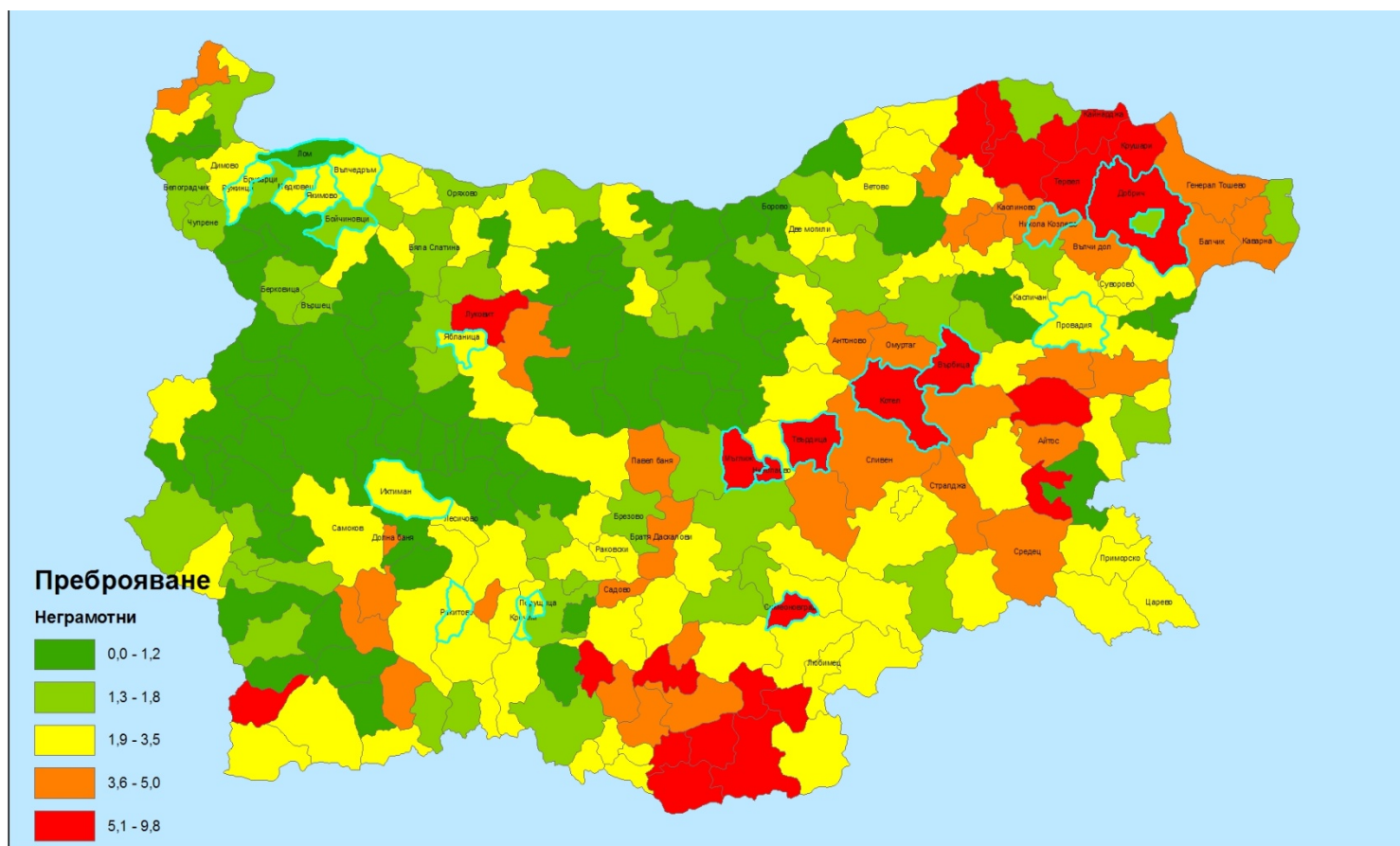
Correlation between share of Roma population and % of illiteracy is valid for some municipalities but not for all…

Child mortality (under 1) by municipalities, 2001

…the same for child mortality

Territorial mapping on a lower level – Sofia

Share of Roma – census 2001 data
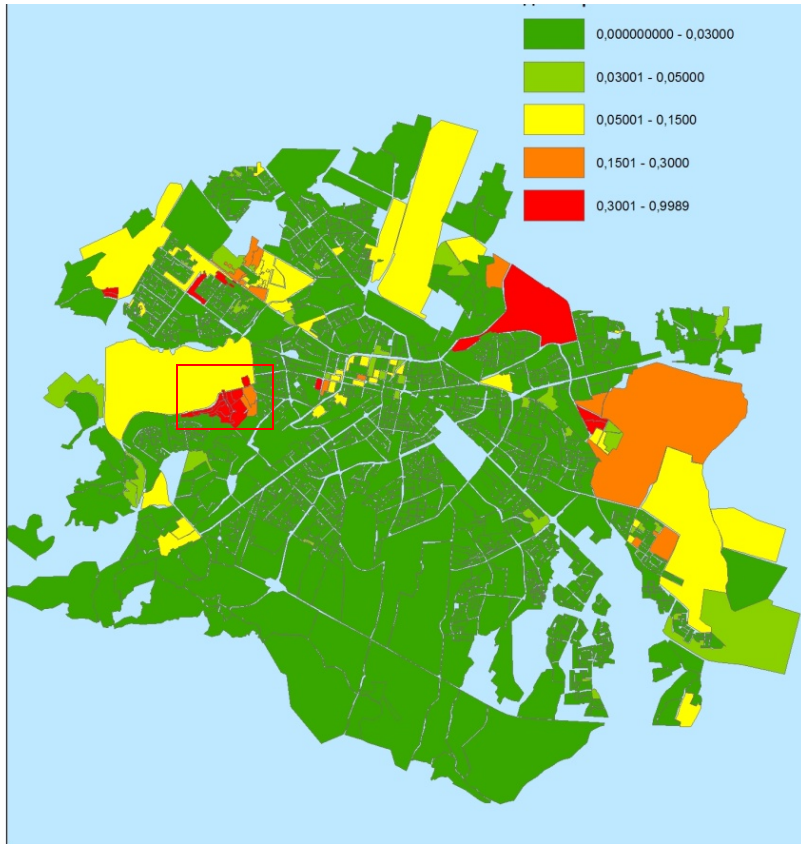
Share of illiterate – census 2001 data

Legend (Share of Roma):
- 0,000000000 - 0,03000
- 0,03001 - 0,05000
- 0,05001 - 0,1500
- 0,1501 - 0,3000
- 0,3001 - 0,9989

Legend (Share of illiterate):
- 0,000000000 - 0,007
- 0,008 - 0,02
- 0,03 - 0,04
- 0,05
- 0,06 - 0,4
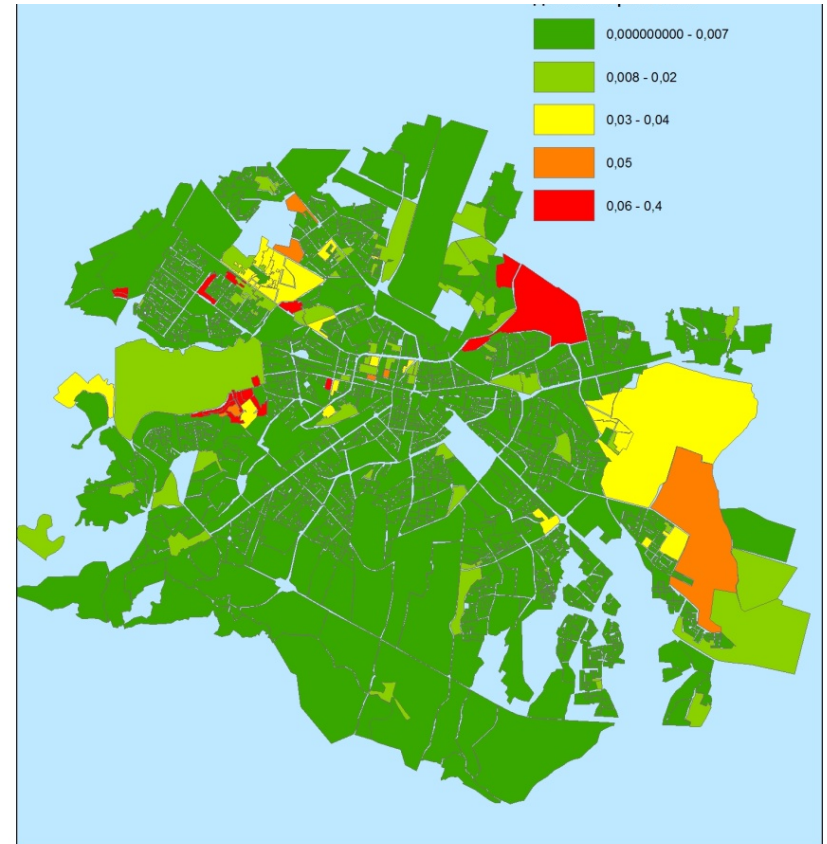
But is perfectly valid at district level

Territorial mapping at statistical control units in *Fakulteta* district
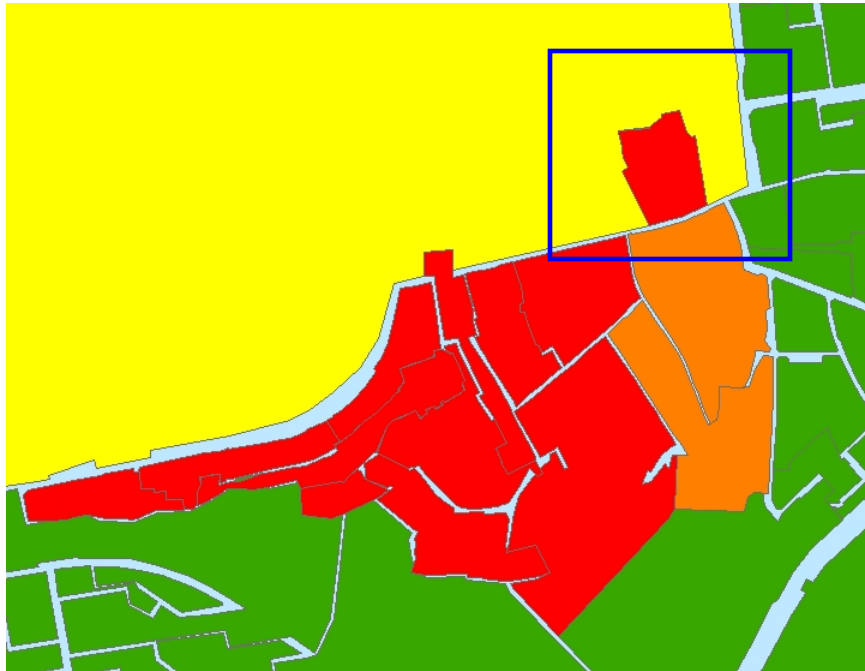
Census data – a snapshot once every 10 years (last in 2001)

GIS (Google Earth image) makes possible updates between censuses

Correlations between territorial concentration of Roma population and selected indicators

Concentration		Share of Roma	Share of illiterate	Share of people with primary education	Child mortality	Density
		2001			2004-2005	2005
Under 5%	1	1,9	1,3	16,8	8,9	98,5
5-10%	2	7,0	2,3	24,3	12,5	52,8
10-15%	3	11,9	3,4	29,1	16,0	36,3
15-20%	4	17,9	2,9	29,7	22,7	36,7
20-25%	5	21,7	5,2	33,1	19,4	35,1
Over 25%	6	27,0	7,2	41,7	27,5	27,4
Total		4,7	1,8	20,3	11,0	69,4
Correlation ratio			**0,94**	**0,97**	**0,95**	**-0,83**

Roma-sensitive indicators
(Indicators that strongly distinguish areas populated by Roma communities)

Indicator	Correlation with the presence of a Roma community
Natural increase	0,82
Unemployment	0,81
Population with water supply restrictions	0,66
Employment in the agricultural sector	-0,81
Average wage	-0,83
Employment in the industrial sector	-0,87
Companies net sales revenues per person	-0,93

Possible ethnic-sensitive indicators based on territorial tags

- Types of dwellings
- Size of the dwelling; m2 per household member
- Average number of members per household
- Average number of households per dwelling
- Child mortality under 1
- Frequency of mother mortality by age and by main death causes
- Frequency of hereditary diseases
- Frequency of sexually transmitted diseases
- Percentage of the children under school age covered by health services
- Percentage of family/mothers who renounce to have basic health cares for their children
- Progress/regress in school desegregation

All these indicators are **"indicators for a population living in certain area with certain parameters"** and they are not directly "indicators for this or that ethnic group"

Roma boosters in sample based surveys

- Theoretically, they would provide comprehensive information on income, expenditures, consumption patterns, employment status and qualification of the labor force; this data would be important input for monitoring progress under Priority 3 (housing) and Priority 4 (employment)

- Data about the educational aspects and children and youth status will be poorer; MICS – not in all countries and not done on regular basis

- But constructing the sample boosters may be a problem because the number of Roma population is not clearly determined ("who's Roma?" question)

- Samples can be also constructed on the base of the territorial distribution of the ethnic groups – provided a map of their distribution exists

- GIS sampling can complement mapping of Roma neighborhoods

Sample surveys based data for indicators to monitor NAPs' targets

1. **Status of the household**
   - **Electricity, clean water, sewage, major HH items**
2. **Education profiles of its members**
   - **Enrollment rates, literacy rates, attainment, reasons for non-attainment**
3. **Incomes**
   - **Total HH incomes and by HH members, by source (type of contract, sector)**
4. **Expenditures**
   - **Total and by type, consumption patterns**
5. **Employment and unemployment status**
   - **By sex, qualification, duration, enrollment in employment programs**
6. **Perception of different threats**

Example: "healthy life expectancy" in Bulgaria based on data from sample surveys

Age category	Men				Women			
	Live expectancy		Live expectancy на in good health status		Live expectancy		Live expectancy на in good health status	
	1996	2001	1996	2001	1996	2001	1996	2001
15-19	53,84	54,30	46,12	45,38	60,88	61,11	48,21	47,24
20-24	49,07	49,51	41,50	40,71	56,02	56,23	43,51	42,50
25-29	44,37	44,80	36,93	36,18	51,16	51,36	38,87	37,83
30-34	39,68	40,08	32,40	31,65	46,33	46,52	34,19	33,35
35-39	35,05	35,43	27,98	27,13	41,51	41,71	29,62	28,75
40-44	30,58	30,91	23,66	22,84	36,76	36,96	25,13	24,46
45-49	26,35	26,61	19,64	18,78	32,11	32,31	20,91	20,24
50-54	22,40	22,65	15,87	14,93	27,58	27,77	16,86	16,26
55-59	18,74	19,01	12,42	11,86	23,16	23,37	12,96	12,54
60-64	15,38	15,63	9,27	8,83	18,98	19,17	9,59	9,06
65-69	12,32	12,61	6,56	6,05	15,05	15,20	6,57	6,20
70-74	9,48	9,80	4,28	4,03	11,46	11,58	3,90	3,72
75-79	7,05	7,40	2,57	2,37	8,45	8,44	2,29	1,95
80+	5,16	5,49	1,15	1,35	5,95	5,89	1,02	0,97

Individual "on the spot" surveys

- Anonymous thematic questionnaire that must be filled by the social service users voluntarily

- They can have a "ethnicity" field

- They can be source of information about the ethnic profile of the user of the respective service, as well as about the way the service providers work (for example, show if there are some ethnically motivated prejudices).

  But:

- These data are not representative of the population itself

- Representativity of the respective provider's clients is limited

Examples of survey forms in the field of health care

Possible questionnaire:

- How do you evaluate your health status as a whole? – on a 5 grade scale

- Do you have a chronic disease or a health problem? Yes/No

- Do you have a health insurance? Yes/No   Is it important for your health status? Yes/No

- How many times and when for the last time have you asked for medical help (a GP, emergency doctor, pediatrician – for children under 17, a specialist, I have not asked) ?

- What was the reason that made you ask for medical help (disease, trauma or injury, regular check-up, prescribe medicines, administrative procedures – medical certificate and other, other reasons)

- In the last 2 years have you ever visited a gynecologist?

- Do you think that young age pregnancy and birth (under 16) are dangerous for the mother and child's health?

- Have you encountered problems in access to health services related to your ethnicity?

Community-based monitoring

It is a system to collect data about a certain community by members of this same community. This system would provide:

■ Quantitative information on the community status - number of households, their housing conditions, number of children attending school, their age and grade, number of drop-outs, number of new-born, number of vaccinated children etc.

■ Quantitative information on occurrence of certain events relevant from Decade monitoring perspective (power cuts and their duration, accidents, conflicts with majority or other Roma groups, NGOs activities etc.)

■It will give the possibility for a real (and not only declarative and formal) involvement of Roma

Community-based monitoring – probable problems

- The communities are "interested party" and data collected by communities members may be biased
- Local monitors can be under pressure from local leaders, who may have veiled interests
- Necessary qualifications may be insufficient
- Lack of "common interest" spirit ("*us* versus *them*" phenomenon)
- Incentives for scrupulous periodicity reporting may be insufficient (certain issues may receive higher priority than data collection)
- Linguistic and semantic problems may exist

Comparing different approaches do data disaggregation

	Statistical relevance of data collected	Anticipated costs	Methodological difficulties
PIN as a link	High	Low	Low
Territorial tags	High	High but only for initial mapping	Medium
Extended samples	High	Medium but on regular basis (every quarter)	Medium (related to sampling)
Custom surveys	Low	Low	Low

Comparing different approaches do data disaggregation

	Opportunities for Roma involvement	Legal framework amendment	Feasible in:
PIN as a link	Low	Yes	Short term perspective given legal framework in place
Territorial tags	High	No	Mid-term perspective
Extended samples	Low	No	Short term given legal framework in place
Custom surveys	Low	No	Short term perspective

Conclusions

- Disaggregating statistical data by ethnicity is possible even when exact number of Roma population is unclear

- Constructing ethnically sensitive indicators is possible – both national and internationally comparable

- Problems exist, however they are not methodological, technical or financial but rather of political nature

- Given the concerns regarding individual data integrity, such disaggregations and construction of indicators should be done by specially appointed agency operating in line within clear legislation on the matter

- The NAP needs revision – to be amended by sets of relevant input-output-outcome and impact indicators and to become M&E consistent tool

- The roles and responsibilities of institutions involved in Decade implementation and monitoring should be clearly specified and streamlined to avoid duplication and internal rivalry

Sequence of the steps in case of replication of the pilot elsewhere

- Inventory of the necessary components
    - □ PIN as element of the census data is it available, registered?
    - □ What standard statistical sample surveys exist (HBS, LFS, LSMS, MICS), what is their periodicity and do they use Roma samples?
    - □ Which of the available administrative and other data bases can be matched?
    - □ Has a mapping of Roma community been conducted?
    - □ Legal framework overview (existing legislation on personal data protection)
    - □ Existing administrative structures (who does what and is responsible for what in regards to Decade monitoring and NAP implementation)
- Discussion with Roma organizations and agreeing on joint actions in the area of data collection
- Pilot test of the methodology
    - □ Computation of major indicators
    - □ Extending the samples
    - □ Training Roma data collectors if community level data collection is implemented
- Institutionalizing the system (making it part of the administrative structures)
- Updating the NAP