Jacek Kowalewski

CENTRAL STATISTICAL OFFICE OF POLAND

STATISTICAL OFFICE IN POZNAŃ

USING A MAP OF SHORT-TERM STATISTICS IN THE PROCESS OF ORGANIZING BUSINESS SURVEYS

Inspiration

• disproportion between demand and supply on the information market

more, faster, higher quality, more details

• the need to change the functional model of statistics

cost reduction, respondent burden reduction, a change in the overall coordination of surveys, forms integration

• reluctance to change in statistical offices

complex and extensive surveys, internal connections between surveys, the need to ensure data coherence and continuity, poor knowledge of modern statistical techniques

• the need for detailed information about the whole survey system

# A map of a system of statistical surveys

What is a map of a system of statistical surveys?

Generalized, integrated and ordered description of statistical surveys including the following elements:

• Basic information about surveys containing: survey scope, target population, survey frequency, survey administration date, data collection methods, survey implementation.

• A description of input data (X) – sources and forms of data; input variables, quality assessment criteria (number of refusals, missing data, error estimates).

• A description of survey results (Y) i.e. the format (raw data, adjusted seasonally or on a working day basis, trend), form of presentation (indicator, absolute value), date and place of releasing data or making them available.

• A description of the transformation of input data into results, which is essential for further, more in-depth analyses.
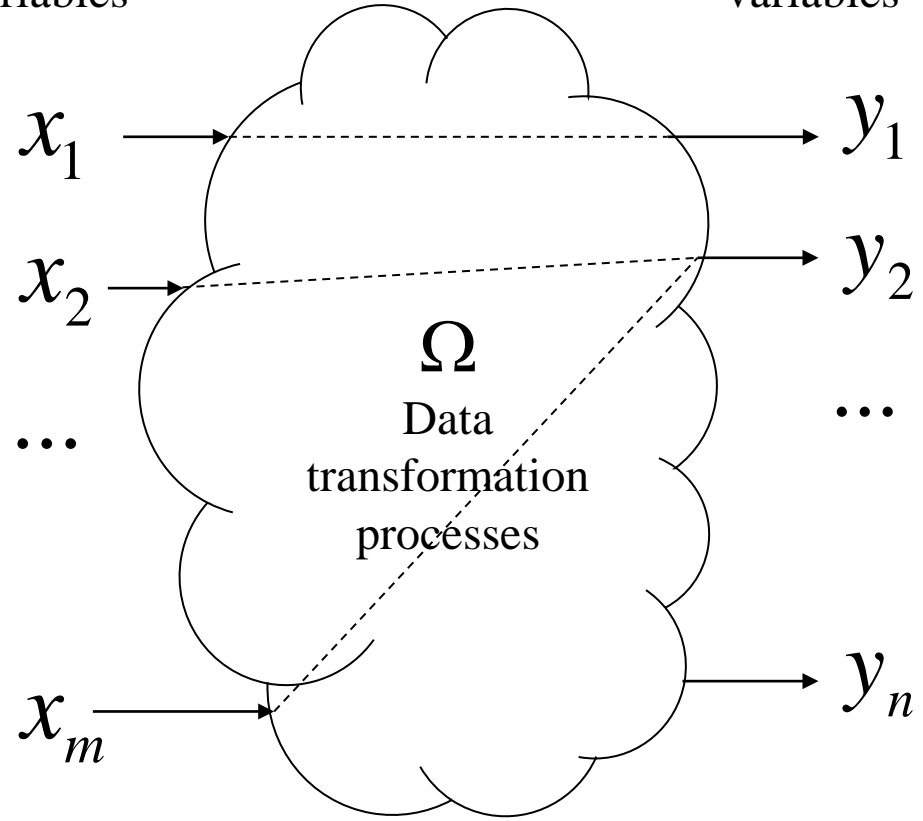
Data cube
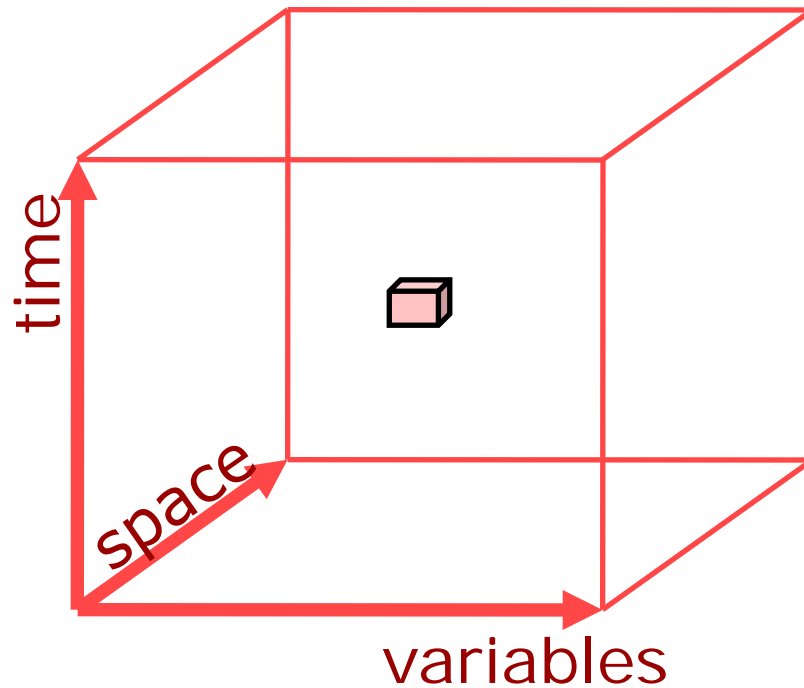
sections, divisions, classes…

Matrix of Transformation

$$M = \begin{bmatrix} m_{11} \ldots m_{1j} \ldots m_{1n} \\ \\ \ldots \qquad m_{ij} \qquad \ldots \\ \ldots \\ m_{m1} \ldots m_{mj} \ldots m_{mn} \end{bmatrix}$$

where

$$m_{ij} = \begin{cases} 1 & \text{, when input variable } x_i \text{ is used to estimate the value of variable } y_j \\ 0 & \text{, otherwise} \end{cases}$$

Implementation of short-term statistics in Poland

DG1 survey

- business activity report
- the basic source of short-term information about economic activity of businesses
- covers only enterprises with over 9 employees
- timeliness – 5th day after the end of each month
- current prices, constant prices
- digital reporting - over 90%
- efforts made to use administrative registers
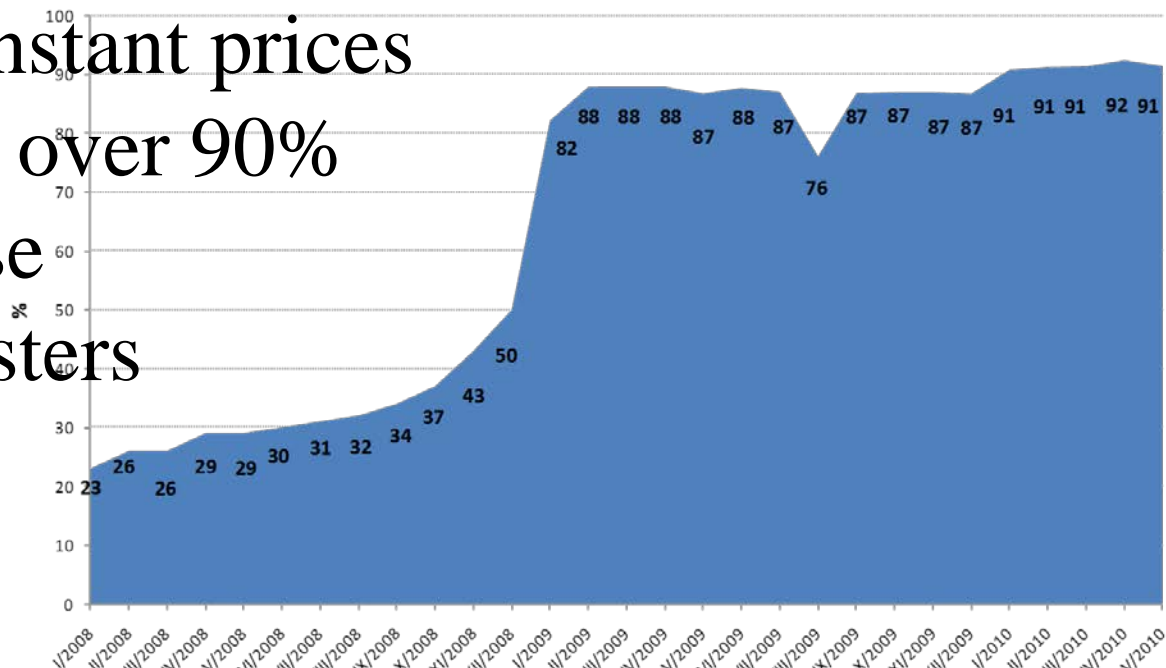
DG-1 SURVEY SCOPE

BY THEME

SALES

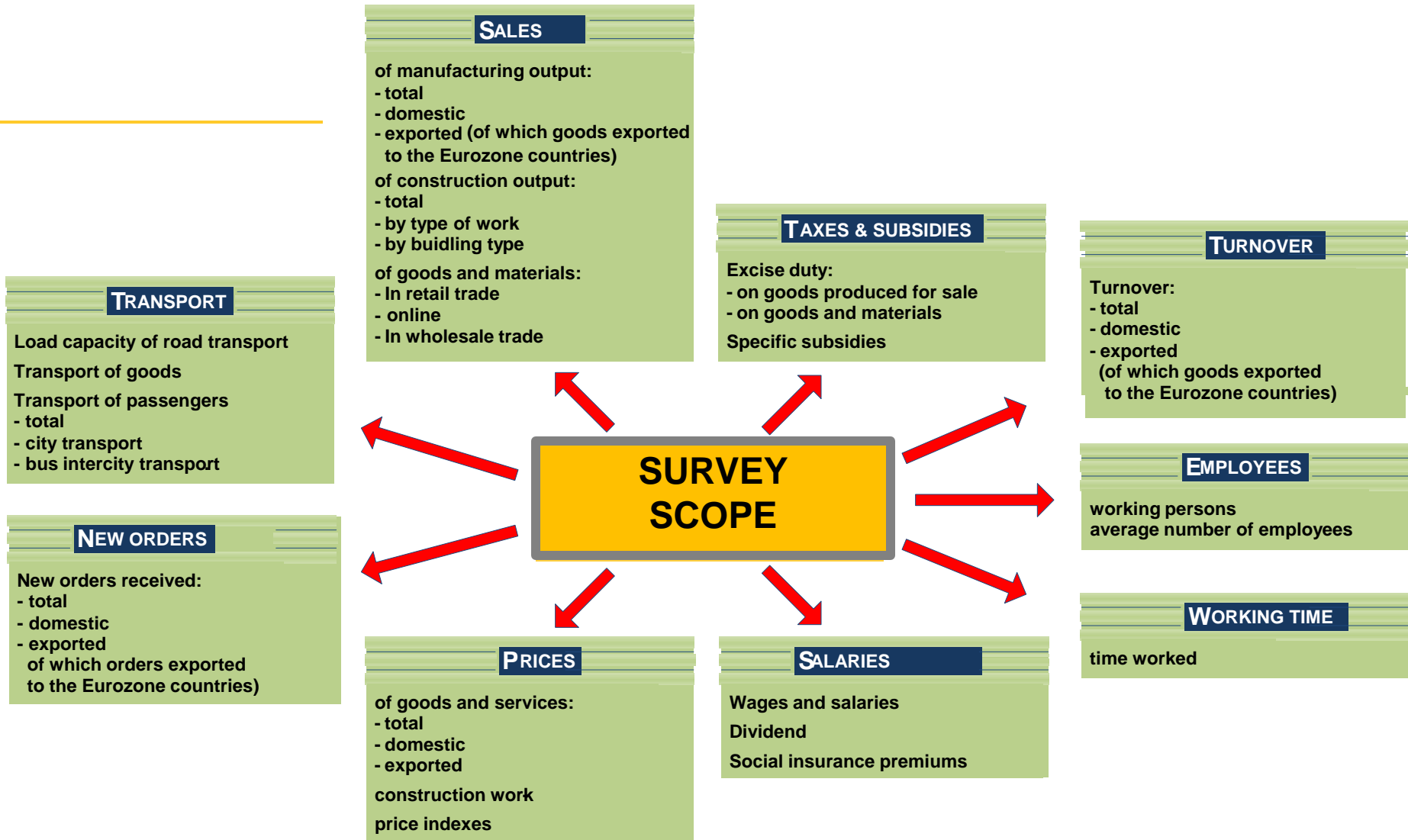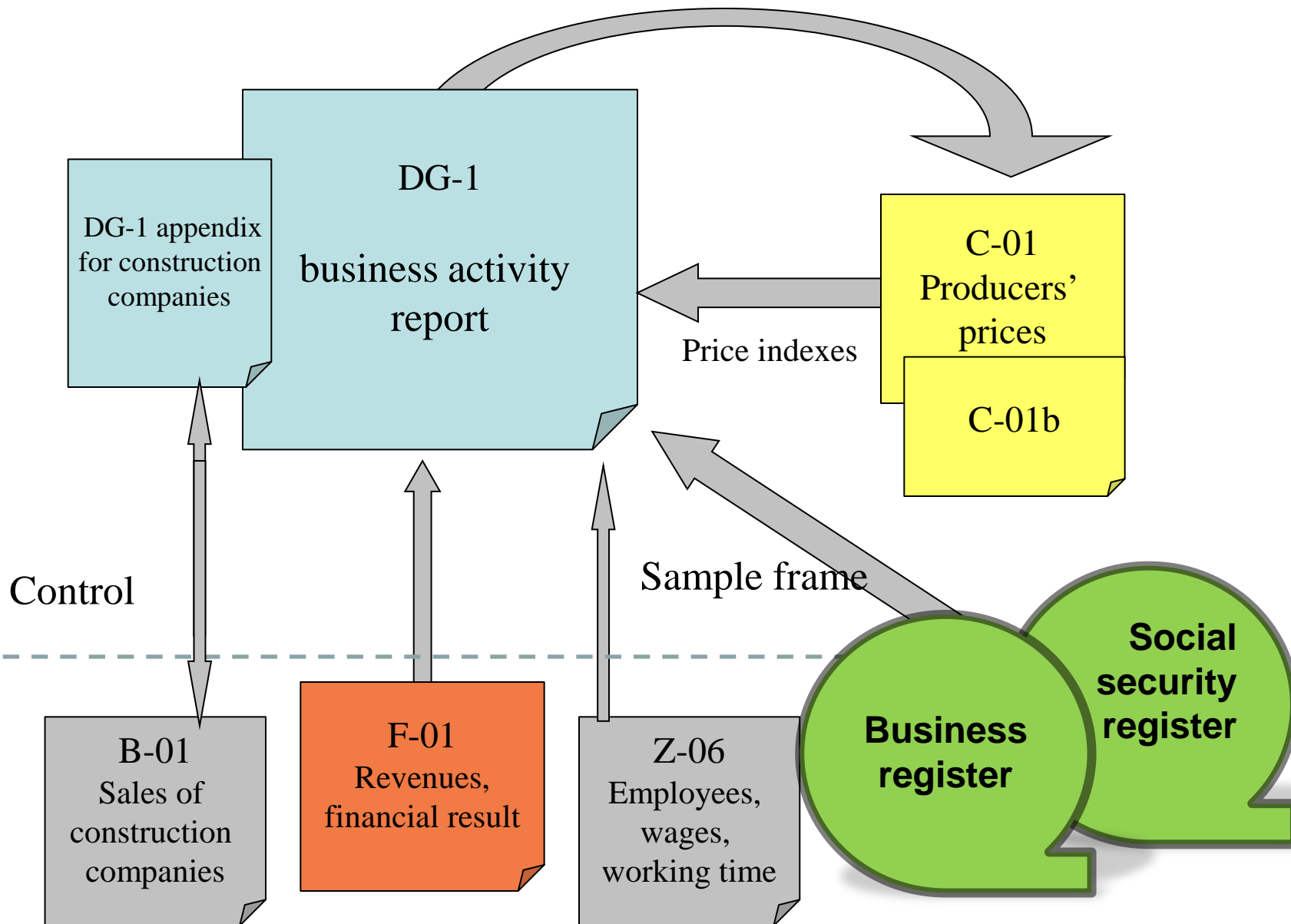of manufacturing output:
- total
- domestic
- exported (of which goods exported
  to the Eurozone countries)

of construction output:
- total
- by type of work
- by buidling type

of goods and materials:
- In retail trade
- online
- In wholesale trade

TAXES & SUBSIDIES

Excise duty:
- on goods produced for sale
- on goods and materials

Specific subsidies

TURNOVER

Turnover:
- total
- domestic
- exported
  (of which goods exported
   to the Eurozone countries)

TRANSPORT

Load capacity of road transport

Transport of goods

Transport of passengers
- total
- city transport
- bus intercity transport

SURVEY SCOPE

EMPLOYEES

working persons
average number of employees

NEW ORDERS

New orders received:
- total
- domestic
- exported
  of which orders exported
  to the Eurozone countries)

PRICES

of goods and services:
- total
- domestic
- exported

construction work

price indexes

SALARIES

Wages and salaries

Dividend

Social insurance premiums

WORKING TIME

time worked

A Map of Short-term Statistics

A MAP OF SHORT-TERM STATISTICS

Legal basis

Subjective scope

Number of entities

Objective scope

Frequency

Data collection

Organization

Regulation European Communities No 1165/98 concerning short-term statistics

(A)

DG-1 report
about 32,7 thousands
Annex to DG-1
about 2,5 thousands
C-01 questionnaire
about 3,8 thousands
C-0b questionnaire
about 0,5 thousands

(B)

Monthly:
DG-1,
C-01
C-01b
Quarterly:
Annex to DG-1

Electronic questionnaire

(C)

(A) **SUBJECTIVE SCOPE**

Legal persons, organizational entities without legal personality and natural persons conducting economic activity, employing more than 9 persons.
Self-financing units, not being budgetary entities or budgetary plants.
Entities classified according to Nace Rev.2:
DG-1 report:
- Section B, C, D, E, F, G, H, I, J, L, N, R
- Section M (excluding divisions: 72, 75)
- Division: 02, 95, 96
- Class: 03, 11
 Annex to DG-1
- Section F
 **C-01 q*uestionnaire*
- Section A (division: 02, 03)
- Section B, C, D, E, H
- Section J (division 61)
 C-01b questionnaire
- Section F

(B) **OBJECTIVE SCOPE**

Sale
Turnover
Taxes and subsidies
Number of persons employed
Hours worked
Gross wages and salaries
Prices
New orders
Transport /complementary data/

(C) **ORGANIZATION**

Vovoidship stage – Statistical Office in Poznań according to following organization:
*Short-term Statistics Centre in Poznań***:**
 vovoidships: wielkopolskie, mazowieckie, dolnośląskie
Branch in Kalisz:
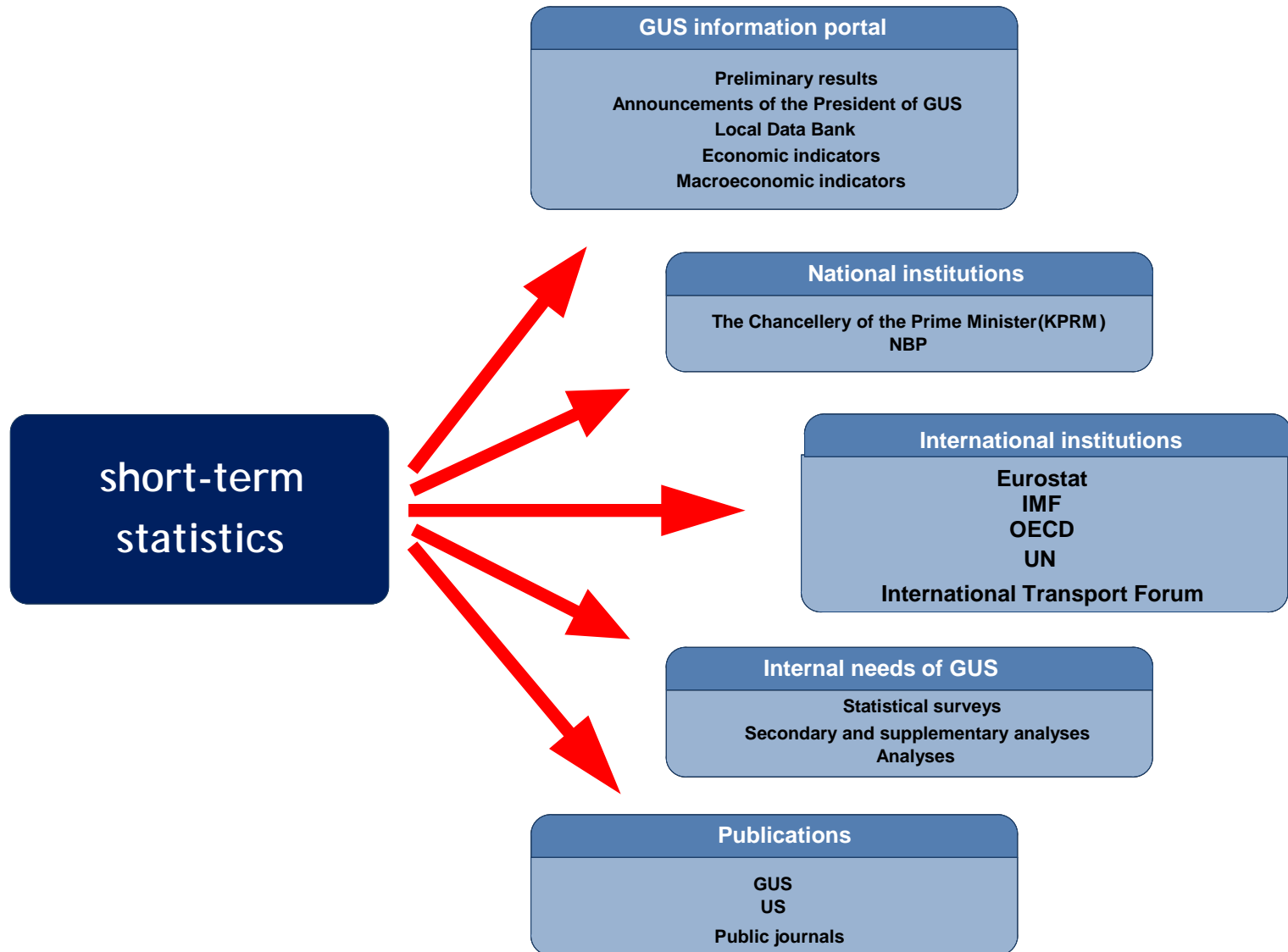 vovoidships : śląskie, lubuskie, warmińsko-mazurskie, opolskie
*Branch in Konin***:**
 vovoidships: lubelskie, świętokrzyskie, zachodnio-pomorskie,
 podlaskie
*Branch in Kościan***:**
 vovoidships: małopolskie, łódzkie
*Branch in Piła***:**
 vovoidships: podkarpackie, kujawsko-pomorskie, pomorskie

General stage – Central Statistical Office*:*
 Programming and Coordination of Statistical Surveys Department,
 Enterprises Department,
 Production Department,
 Social Surveys and Living Conditions Department,
 Demographic Surveys and Labour Market Department,
 Trade and Services Department.

Destinations for short-term data

short-term statistics

GUS information portal

Preliminary results
Announcements of the President of GUS
Local Data Bank
Economic indicators
Macroeconomic indicators

National institutions

The Chancellery of the Prime Minister(KPRM)
NBP

International institutions

Eurostat
IMF
OECD
UN

International Transport Forum

Internal needs of GUS

Statistical surveys
Secondary and supplementary analyses
Analyses

Publications

GUS
US

Public journals

INTERNATIONAL INSTITUTIONS

OUTPUT VARIABLES BY DOMAIN

Domain	Output variable	International institutions				
		ECMT	Eurostat	IMF	OECD	UN
Industry	*New orders*		✓			
	Domestic new orders		✓			
	Non-domestic new orders		✓			
	Turnover		✓			
	Domestic turnover		✓			
	Non-domestic turnover		✓			
	Sold production of industry	✓	✓	✓	✓	✓
	Price indices of sold production of industry		✓	✓	✓	✓
	Price indices of sold production of industry on domestic market		✓			✓
	Price indices of sold production of industry on non-domestic market		✓			
Construction	*Sale of construction and assembley production*		✓		✓	✓
	Sale of construction and assembley productionof buildings		✓			
	Sale of construction and assembley production of civil engineering		✓			
	Sale of construction and assembley production of new residential buildings		✓			
Retail trade	*Turnover in retail trade*		✓			
	Retail sales of goods				✓	
	Deflator of sales		✓			
Services	*Turnover in services*		✓			
	Price indices of services		✓			
Employment, wages and salaries	*Hours worked*		✓			
	Persons employed		✓			
	Persons employed in industry				✓	
	Average monthly wages and salaries in industry			✓		
	Average paid employment in industry			✓		
	Average monthly wages and salaries in enterprise sector				✓	
	Gross wages and salaries		✓			

STS INDICATORS FOR EUROSTAT

THE DATE OF TRANSFERRING DATA BY POLAND AFTER THE END OF THE REPORTING PERIOD

Domein	Output variable	Month				Quarter	
		30 days	35 days	45 days	60 days	60 days	90 days
Industry	*New orders*	✓					
	Domestic new orders	✓					
	Non-domestic new orders	✓					
	Turnover	✓					
	Domestic turnover	✓					
	Non-domestic turnover	✓					
	Sold production of industry	✓ [1]			✓ [2]		
	Price indices of sold production of industry			✓			
	Price indices of sold production of industry on domestic market		✓				
	Price indices of sold production of industry on non-domestic market		✓				
Construction	*Sale of construction and assembly production*	✓					
	Sale of construction and assembly production of buildings	✓					
	Sale of construction and assembly production of civil engineering	✓					
	Sale of construction and assembly production of new residential buildings				✓		
Retail trade	*Turnover in retail sales*	✓					
	Deflator of sales	✓					
Services	*Turnover in services*					✓	
	Price indices of services						✓
Employment, wages and salaries	*Hours worked*						✓
	Employed persons				✓		
	Gross wages and salaries				✓		

1) provisional data   2) final data

PUBLISHED BY CSO ON AN ONGOING BASIS

Blok tematyczny	Zmienna wynikowa	Dziennik Urzędowy RP "Monitor Polski"	Dziennik Urzędowy GUS	Informacja o sytuacji społeczno-gospodarczej kraju	Biuletyn Statystyczny	Ceny w gospodarce narodowej	Ceny robót budowlano-montażowych i obiektów budowlanych	komunikaty i obwieszczenia Prezesa GUS	Wyniki wstępne	Wybrane miesięczne wskaźniki makroekonomiczne
		Dzienniki Urzędowe		Publikacje GUS				Portal informacyjny GUS		
Ceny	Wskaźnik cen producentów w przemyśle na rynku krajowym				✓					
	Wskaźnik cen produkcji sprzedanej przemysłu			✓	✓	✓			✓	✓
	Wskaźnik cen produkcji sprzedanej wyrobów przemysłowych					✓				
	Wskaźnik cen obiektów budowlanych						✓			
	Wskaźnik cen produkcji budowlano-montażowej		✓	✓	✓	✓		✓	✓	✓
	Wskaźnik cen robót budowlano-montażowych						✓			
	Wskaźnik cen produkcji usług				✓	✓				✓
Sprzedaż	Produkcja sprzedana przemysłu			✓	✓				✓	✓
	Wydajność pracy w przemyśle			✓						
	Sprzedaż produkcji budowlano-montażowej			✓	✓				✓	✓
	Sprzedaż produkcji budowlano-montażowej wg charakteru robót			✓						
	Sprzedaż detaliczna towarów			✓	✓					✓
	Sprzedaż hurtowa towarów			✓	✓					
	Sprzedaż usług			✓						
Obroty	Obroty w handlu detalicznym				✓					✓
Nowe zamówienia	Nowe zamówienia				✓					✓
	Nowe zamówienia niekrajowe				✓					
Transport /dane uzupełn./	Przewozy ładunków			✓	✓					
	Przewozy pasażerów			✓	✓					
Pracujący	Pracujący w sektorze przedsiębiorstw				✓					
	Przeciętne zatrudnienie w sektorze przedsiębiorstw			✓	✓				✓	✓
Wynagrodzenia	Przeciętne miesięczne wynagrodzenie w sektorze przedsiębiorstw	✓	✓	✓	✓			✓	✓	✓
	Przeciętne miesięczne realne wynagrodzenie w sektorze przedsiębiorstw			✓	✓					✓
	Przeciętne miesięczne wynagrodzenie w sektorze przedsiębiorstw bez wypłat z zysku	✓	✓		✓			✓	✓	

TRANSFORMATION MATRIX

MANUFACTURING

Questionnaire	Input variable	New orders	Domestic new orders	Non-domestic new orders	New orders to the euro zone	Turnover	Domestic turnover	Non-domestic turnover	Domestic turnover to the euro zone	Sold production at constant prices - preliminary results	Sold production at constant prices - final results	Sold production at current prices	Labour productivity at current prices	Labour productivity at constant prices
DG-1	Net revenues from the sale of products					✓	✓			✓	✓	✓	✓	✓
	Net revenues from the sale of products to the euro zone								✓					
	Net revenues from the sale of products for the non-domestic market						✓	✓						
	Net revenues from the sale of goods and materials					✓	✓							
	Net revenues from the sale of goods and materials to the euro zone								✓					
	Net revenues from the sale of goods and materials for the non-domestic market						✓	✓						
	Value of manufactured products not classified as sales									✓	✓	✓	✓	✓
	Received new orders	✓	✓											
	Received non-domestic new orders		✓	✓										
	Received new orders to euro zone				✓									
	Excise tax on goods and materials					✓	✓							
	Excise tax on products of own production					✓	✓			✓	✓	✓	✓	✓
	Subsidies									✓	✓	✓	✓	✓
	Avarage paid employment												✓	✓
	Price indices of sold production (previous month=100)									✓				
C-01	Prices of the sale of representative products and services (excluding VAT and excise taxes) in the previous month									✓	✓			✓
	Prices of the sale of representative products and services (excluding VAT and excize taxes) in current month									✓	✓			✓
	The value of the sale of representative products and services (excluding VAT and excize taxes)									✓	✓			✓

Elements of analysis
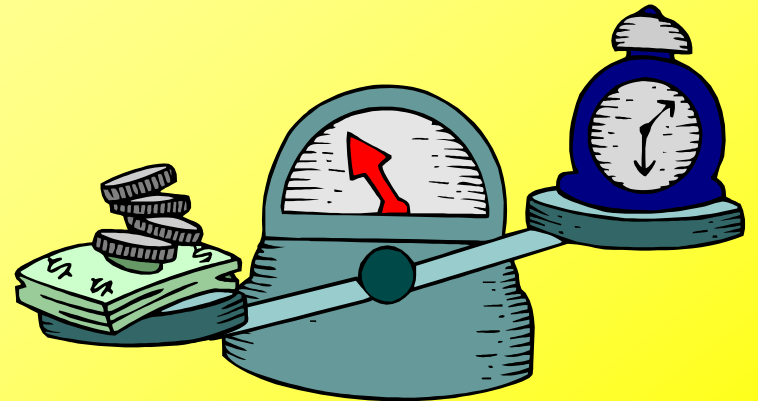
Short-term statistics

- annually over 450,000 reports
- 62 input variables
- 9 thematic areas

- 72 groups of output variables
- 104,000 figures released
- most figures released for the variable
Average monthly wage (19,500)

Matrix of Transformation

$$M = \begin{bmatrix} m_{11} \dots m_{1j} \dots m_{1n} \\ \dots \quad m_{ij} \quad \dots \\ \dots \\ m_{m1} \dots m_{mj} \dots m_{mn} \end{bmatrix}$$

																	Y						
																						Σ	
																				1			
							1		1						1		1						
				1		1					1		1		1		1						
																	1					...	
X				1		1					1		1				1		1			**11**	
											1											...	
			1										1										
											1												
																	1						
	Σ									...	**14**	...											

• To calculate *Sales of construction and assembly output* 14 input variables are required

• *Average number of employees* is used to estimate 11 different output variables

Matrix of Transformation

$$M = \begin{bmatrix} m_{11} \ldots m_{1j} \ldots m_{1n} \\ \ldots \quad m_{ij} \quad \ldots \\ \ldots \\ m_{m1} \ldots m_{mj} \ldots m_{mn} \end{bmatrix}$$

Y

Σ

															Σ
													1		
				1		1				1	1				
			1		1				1	1	1				
											1			...	
X						1								**1**	
														...	
		1							1						
											1				

Σ ...   **1**   ...

- Only 6 x and y variables are mutually and unequivocally matched.
- They can be analysed irrespective of the whole system.

Optimization criteria

Priority (weight)
W

Time
T

Cost
C

Quality
Q

Y_i
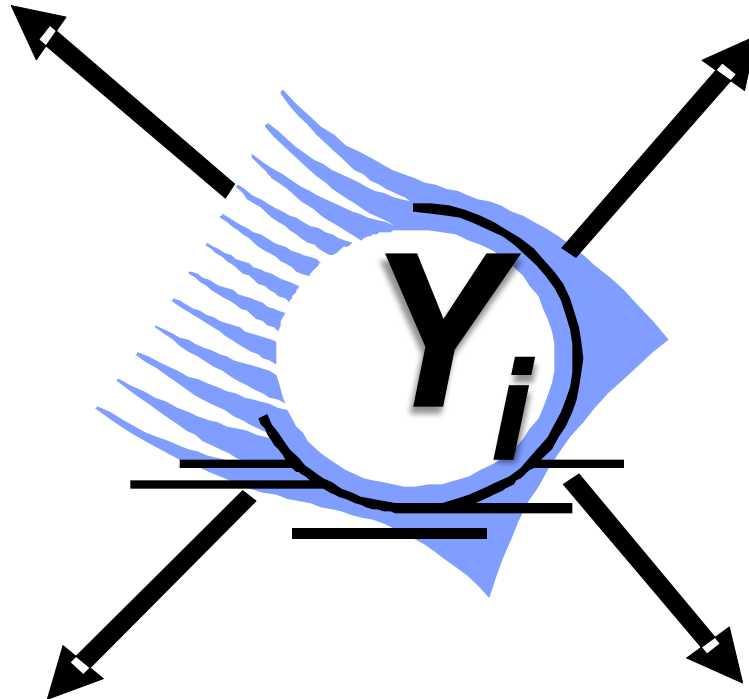
Assumptions for the criteria

- Priority (weight) –3 categories of output information have been distinguished. The largest weight has been assigned to data, which are to be estimated in order to comply with international obligations or national regulations. The lowest priority has been given to supplementary information.
- Release date – output values have been adopted on the basis of data release practices that have been in use so far, which are described in the Map.
- The cost of obtaining a variable – has been calculated by decomposing the total cost of a survey (as described in the Survey Programme of Public Statistics) by means of a matrix of transformation M.

Release date   ***T***

- 49% of variables are monthly ones
- They account for 96.5% of all estimates in those surveys
- 2/3 of output monthly variables have the highest priority
- 6 input variables are used only to estimate quarterly or annual data
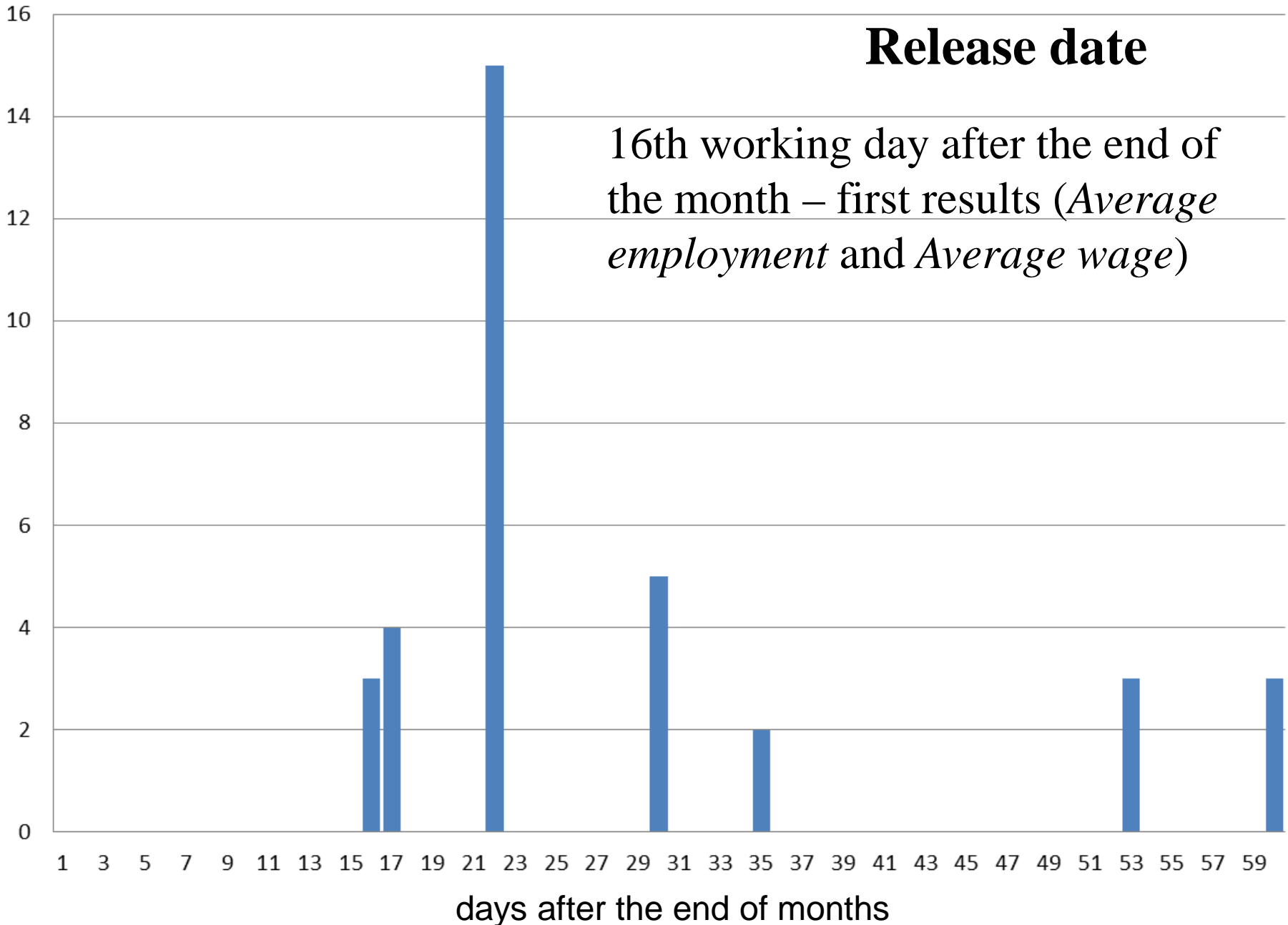
Costs

1. For each l-th survey s_l it is possible to determine:
- the total **cost of survey implementation** CB(s_l),

2. Based on values calculated for survey b_l it is possible to determine the cost value (***xc***) for **output variables** X

$$xc_j^l = f(CB(b_l))$$

3. **Costs of data transformation processes** (Ω) are minimal and therefore **negligible** in the model

Costs C

$xc_1^l = \frac{1}{3}CB$

$xc_5^l = \frac{1}{3}CB$

$xc_8^l = \frac{1}{3}CB$

CB

b_l

x1

x5

x8

y1 $\frac{1}{2}CB$

y2 $\frac{1}{6}CB$
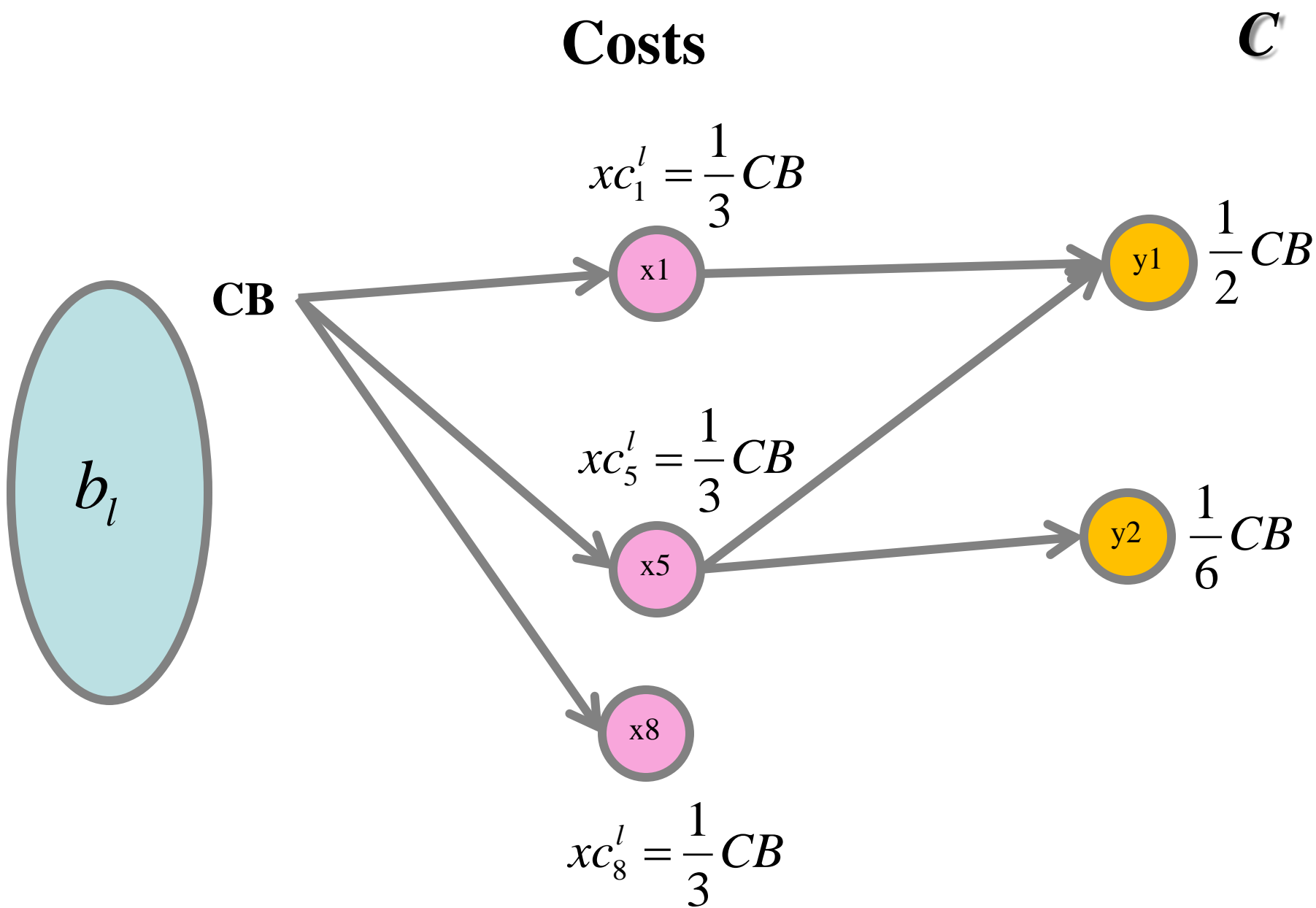
Costs

- PLN 42 (~ €10) – average cost per one report
- +PLN 35 (~ €8,5) – average cost of report's preparing
- PLN 185 (~ €45) – average cost of obtaining one estimate
- PLN 1.6 m (€380,000) – the maximum cost of one variable: *Indicator of producers' prices on the domestic market*
- PLN 400,000 (€95,000) – the maximum cost of estimating one value (*Base amount GN*)
- PLN 3 (€0.70) – the lowest cost of estimating one value (*Average employment*)

Conclusions

What for?

• The adjustment of the statistical output to changes in the surrounding environment should be treated as an ongoing process

• Different actions are necessary at various levels – looking for new data sources, using modern methodologies, changing survey organisation

• the need for advanced education (like EMOS)

• A map of a system of surveys provides essential, integrated information about implemented surveys, which helps to identify end users and   release dates, and describes relationships between different variables.   In other words, the map provides a quick answer to questions about what, where, when and how.

• The map has a motivational aspect

• Necessary to conduct in-depth analyses

• Can be useful for integrating statistical forms

• The map can help to optimize surveys. A matrix of transformations with precisely described and quantified criteria is a necessary prerequisite to   determine the optimal way of data processing in statistical surveys,   which can be obtained by solving a problem in mathematical modeling.

Model of optimization of statistical surveys
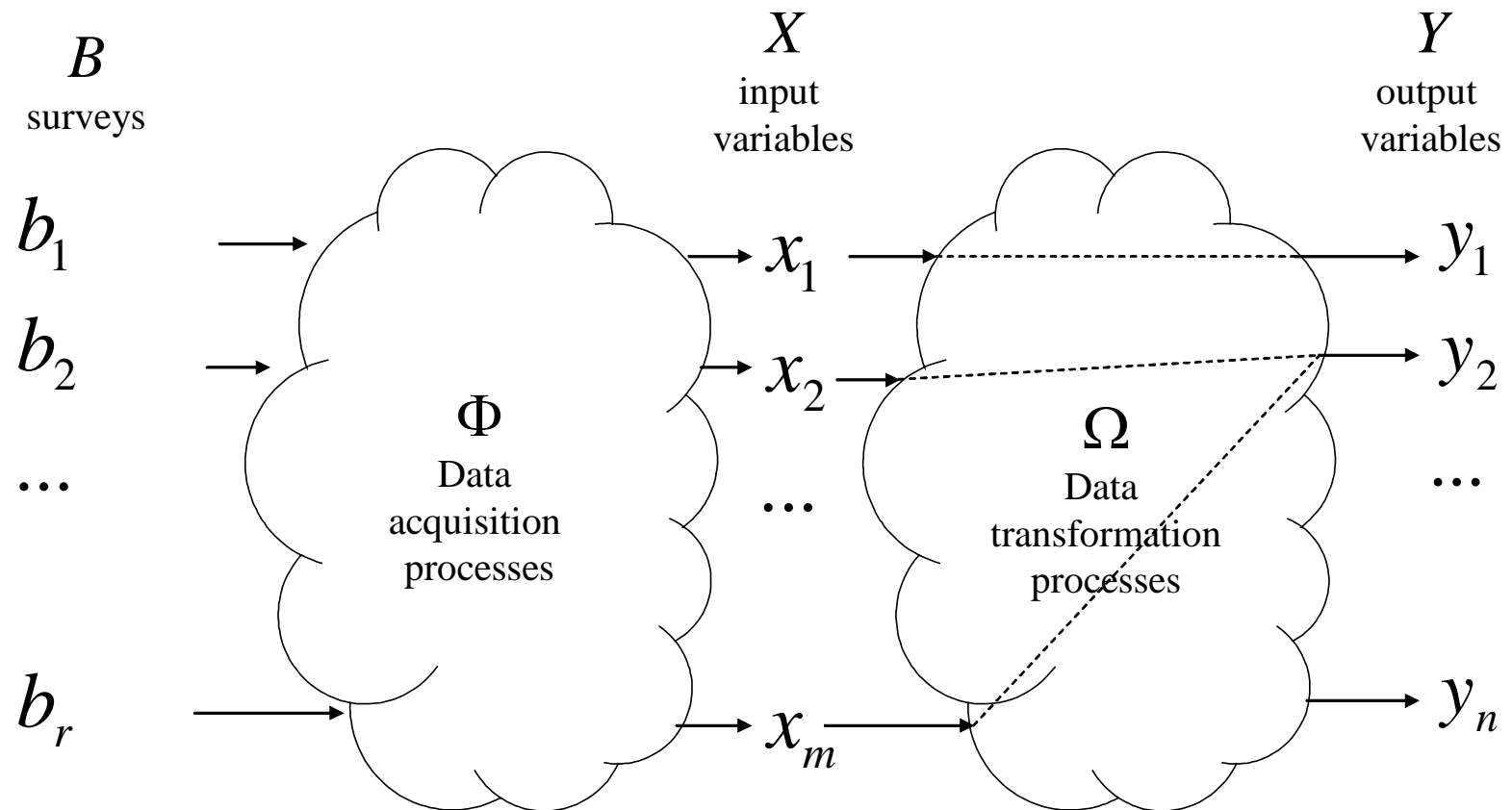
Two-stage multi-criterion problem
Φ data acquisition processes pose a binary problem,
Ω data transformation processes pose a combinatorial problem.

STUDIUM PRZYPADKU

➤ Platforma analityczno-raportująca DG-Analysis
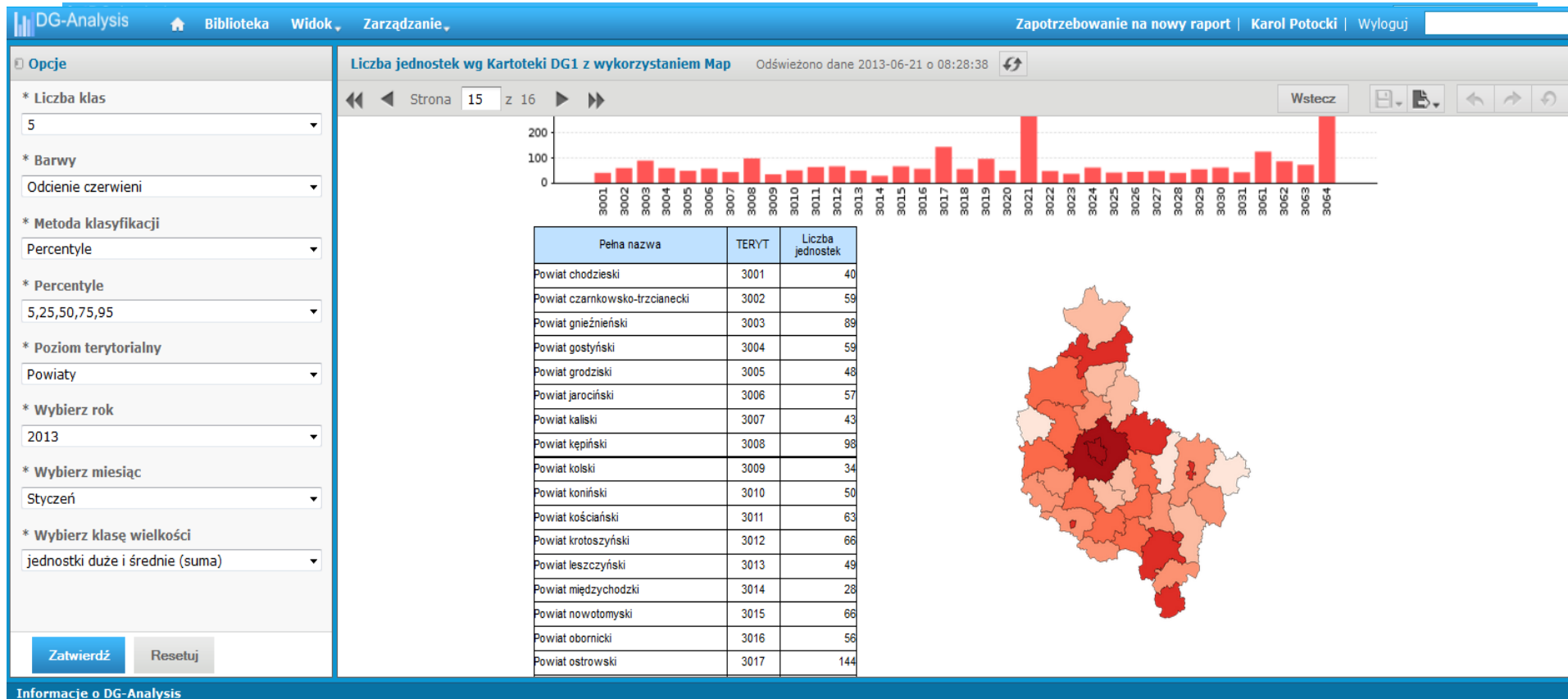
- wdrożenie intuicyjnego i elastycznego narzędzie do analizy i raportowania danych (w technologii Java)
- dostosowanie do specyfiki badania (DG-1) i środowiska (intranet – zintegrowane logowanie)
- liczne grono odbiorców - oddziały US Poznań, GUS, urzędy statystyczne
- parametryzowane zestawienia tworzone przez zespół projektantów
- ciągły rozwój platformy (ostatnio rozszerzono o wizualizację na mapach)
- w przypadku wizualizacji kartograficznej możliwe tworzenie zestawień zbiorczych

(np. odrębnie dla każdego województwa w podziale na powiaty)

Surveys

Information

b1

b2

b3

x1

x2

x3
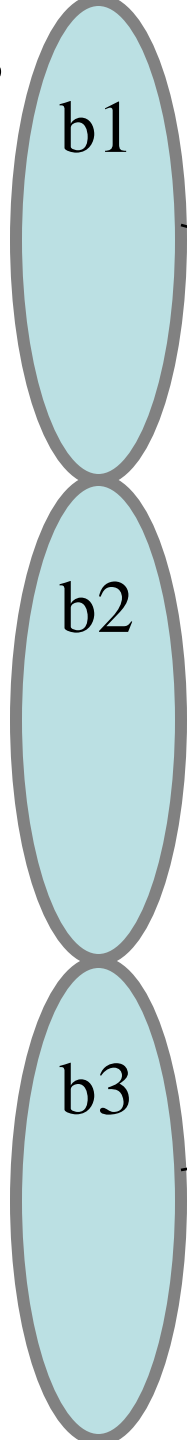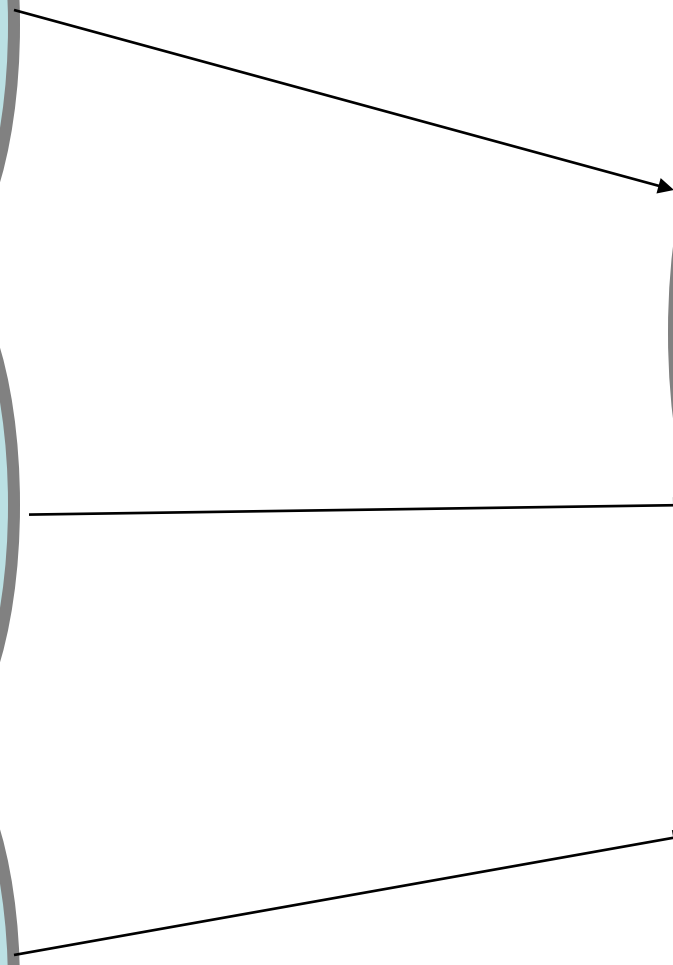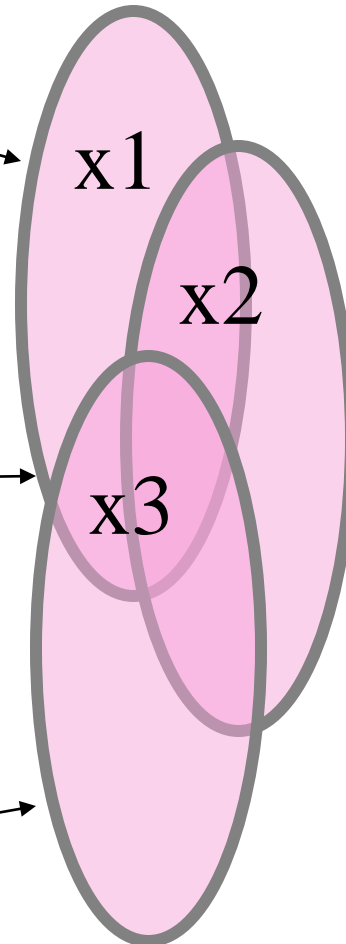
Statistical information

Sundgren (1973): information is defined as a message content:

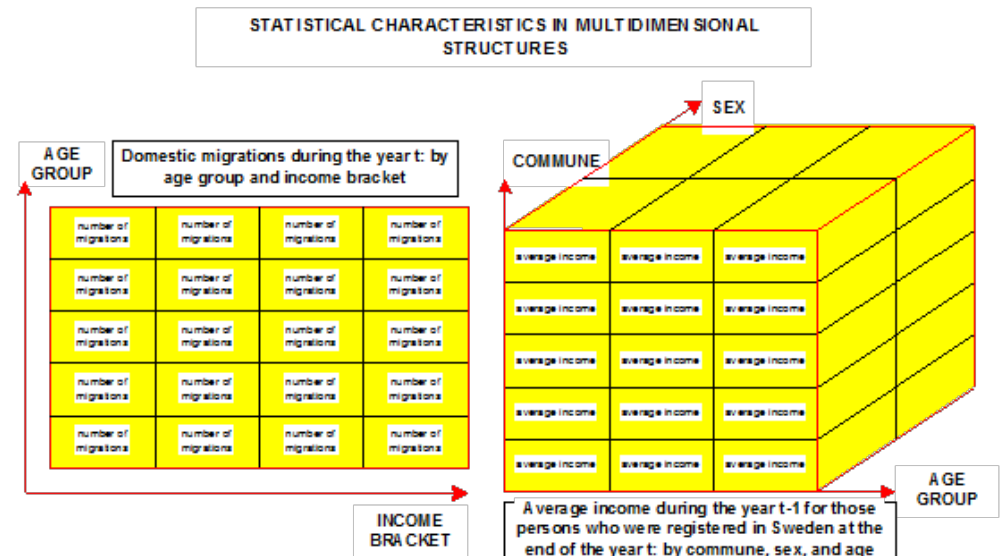$K := <O, X, x, t>$

where

K – infological message,

O – described object,

X – object attribute,

x – attribute value X,

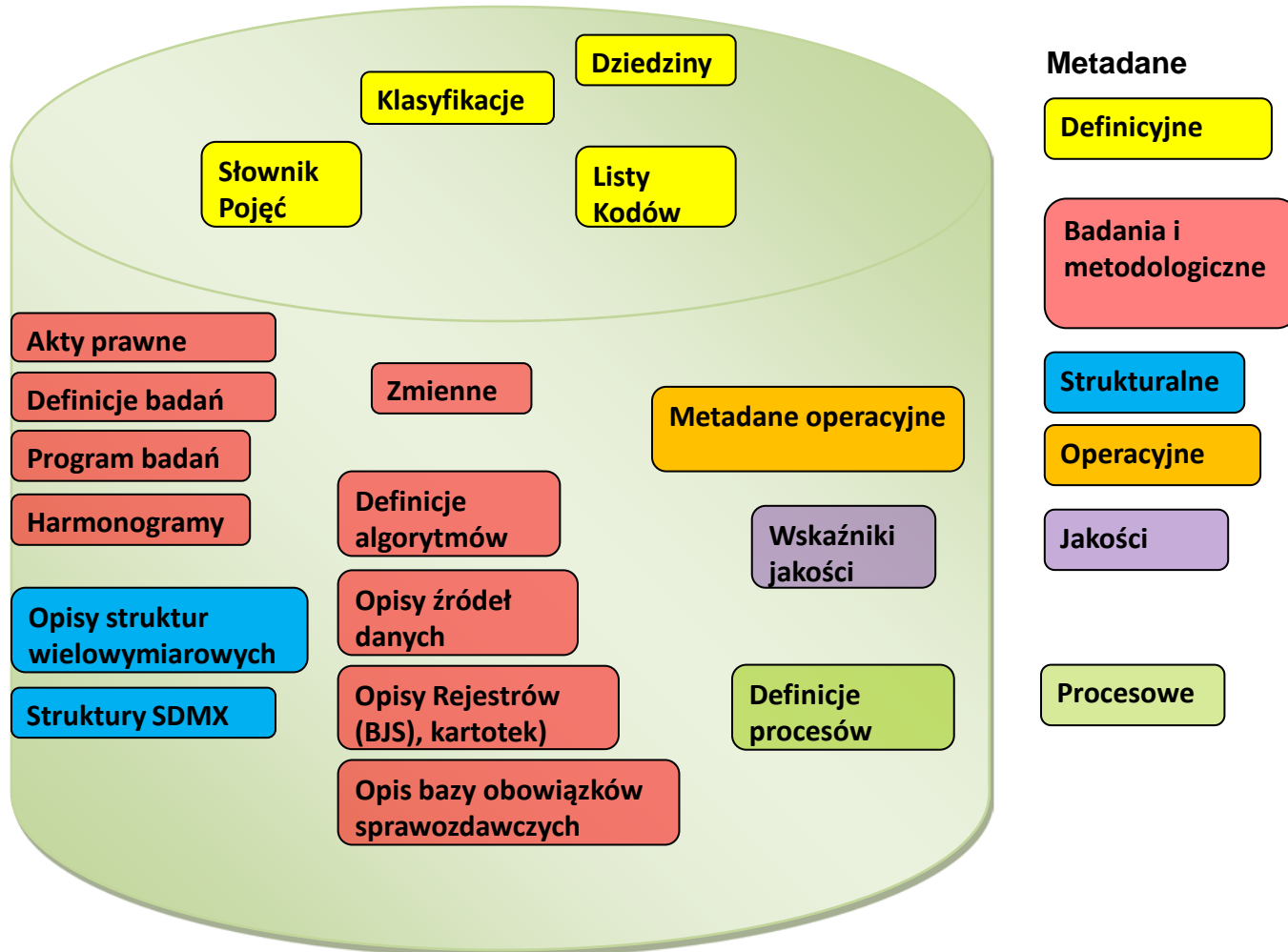t – time when attribute X of object O has value x.

Metadata

STATISTICAL CHARACTERISTICS IN MULTIDIMENSIONAL STRUCTURES

Repozytorium metadanych wg grup

18 255	81 393

17 921 (98,2%)	14 071 (17,3%)	Sample **31 992**

9,6% refusals

5,1% reminders

4,5% explanations

Analysis of the processing of the DG-1 survey

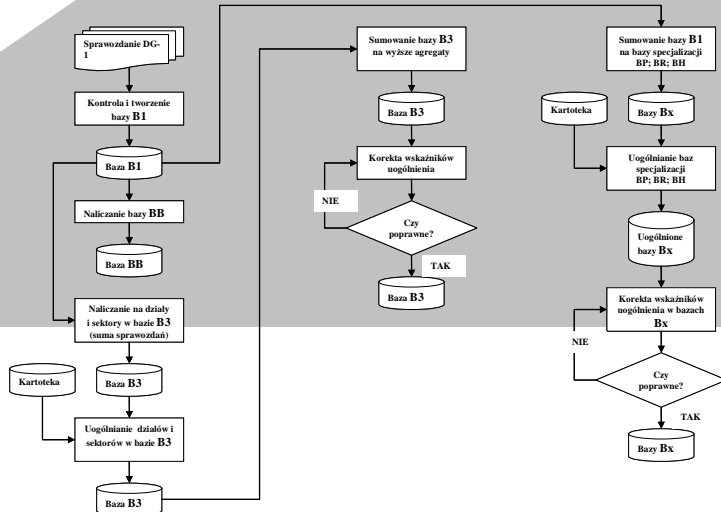18 255   81 393

17 921 (98,2%)   14 071 (17,3%)   sample

sections, divisions, classes

aggregation

generalization