

Results of analysis of OJV data - a case study on data for Poland

(Jacek Maślankowski, Statistics Poland)

Table of contents

1. Online job vacancies - case studies in Statistics Poland	2
2. Basic characteristics of CEDEFOP OJV data for Poland	3
2.1. Number of online job vacancies by NUTS levels and reference time	3
2.2. Number of online job vacancies by ESCO level, date and NUTS	4
2.3. Number of online job vacancies by NACE, date and NUTS.....	4
2.4. Number of online job vacancies by date	5
3. Possible new indicators	6

1. Online job vacancies - case studies in Statistics Poland

We have made some investigations on the possible use of Online job vacancies in Poland to support labour market statistics. Several research papers were published regarding methodological issues as well as reports. The most recent paper was published in September 30th and is available online only in Polish:

Maślankowski J., *The collection and analysis of the data on job advertisements with the use of big data*, Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 60-74.

Bearing in mind the differences in definition of online job vacancy vs. job demand in surveys, we can find a comparison of number of vacancies based on traditional survey and online job vacancies. This information is included in the table presented in Figure 1.

Figure 1. Table presenting values for survey data vs. OJV data

TABL. 1. WOLNE MIEJSCA PRACY
WEDŁUG SPRAWOZDANIA Z-05 A OFERTY PRACY W ŹRÓDŁACH BIG DATA
W II KWARTALE 2017 R.

Wyszczególnienie	Z-05 (wakaty)	Big data (oferty pracy)
	w tys.	
Ogółem^a Total	Survey 122,0	OJV data 110,0
w tym:		
Woj. mazowieckie Voivodship	27,9	23,2
Pracownicy usług i sprzedawcy Sales industry	14,5	15,3

a Dane za II kwartał 2018 r. wynoszą odpowiednio: 164,7 i 131,0.
Źródło: opracowanie własne oraz GUS (2018).

Source: Maślankowski J., *The collection and analysis of the data on job advertisements with the use of big data*, Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 60-74.

We can see that the total number of online job vacancies is a little smaller comparing the number of vacancies based on traditional survey. The same result is when we investigate Mazowieckie voivodship (NUTS-2). However, making analysis of specific occupation can result in reverse dependence - e.g., in service and sales workers, the number of job vacancies by traditional survey is smaller than the result of OJV data.

The most important conclusion is that there are several duplicates in the data and one of the key issues in data processing is to deliver a dedicated de-duplication framework. Another issue is that number of OJV data in different websites is very shaky and can change significantly year-by-year. Average number of online job vacancies in 2nd/3rd quarter of 2017/2018 was presented in Figure 2 (by four country portals, two dedicated portals and one regional).

Figure 2. Average number of job vacancies data by year and type of OJV portal

TABL. 2. UŚREDNIONA LICZBA OFERT PRACY
W II I III KWARTALE

Portale		2017	2018
		w tys.	
Ogólnopolskie:	1	140	64
Country:	2	124	19
	3	46	86
	4	39	54
Branżowe: Dedicated:	1	1,2	0,6
	2	0,4	0,8
Regionalny Regional:		9,9	8,5

U w a g a. Cyfry w kolumnie Portale oznaczają kolejne badane portale.
Źródło: opracowanie własne.

Source: Maślankowski J., *The collection and analysis of the data on job advertisements with the use of big data*, Wiadomości Statystyczne. The Polish Statistician, 2019, vol. 64, 9, 60-74.

As it was demonstrated in Figure 2, we can observe very shaky data sources, in terms of number of job vacancies. It leads to the necessity of the use of various Internet portals with OJV without concentrating on specific ones.

2. Basic characteristics of CEDEFOP OJV data for Poland

To have a general overview of CEDEFOP data for Poland, four basic indicators were prepared and analysed. It includes:

- 1) Number of online job vacancies by NUTS levels (macro_region, region, province, city) and reference time (year_grab_date, month_grab_date, day_grab_date vs. year_expire_date, month_expire_date, day_expire_date)
- 2) Number of online job vacancies by ESCO code (esco_level_1 - 4), date, NUTS
- 3) Number of online job vacancies by NACE (macro_sector, sector), date, NUTS
- 4) Number of online job vacancies by day (to show seasonal effect e.g., total 1444290 for Poland in the last quarter of 2018)

Total number of OJV data in CEDEFOP database for Poland and published in Polish language is 2396651.

The strength of this dataset is a timeliness and cost of acquiring data. However there are some weaknesses, including different definitions of job demand/job vacancies, possible duplicates in the data and unknown target population.

In subchapters 2.1-4, indicators were shown with the value of 10 first columns having the biggest number of observations.

2.1. Number of online job vacancies by NUTS levels and reference time

The goal of this indicator is to show the cities in which there is the largest number of online job vacancies. It allows to check possible duplicates and shows if there are any issues with algorithms.

Table 1. Number of OJV by cities

	macro_region	region	province	city	year_grab_date	month_grab_date	day_grab_date	Number_of_OJV
1	REGION POŁUDNIOWY	Małopolskie	Miasto Kraków	Kraków	2018	9	4	3333
2	REGION CENTRALNY	Mazowieckie	Miasto Warszawa	Warszawa	2019	2	19	3244
3	REGION POŁUDNIOWO-ZACHODNI	Dolnośląskie	Miasto Wrocław	Wrocław	2018	9	4	2828
4	REGION CENTRALNY	Mazowieckie	Miasto Warszawa	Warszawa	2019	6	26	2642
5	REGION CENTRALNY	Łódzkie	Miasto Łódź	Łódź	2018	9	4	2465
6					2019	6	10	2389
7	REGION CENTRALNY	Mazowieckie	Miasto Warszawa	Warszawa	2018	7	12	2261
8	REGION CENTRALNY	Mazowieckie	Miasto Warszawa	Warszawa	2019	2	18	2233
9	REGION PÓŁNOCNO-ZACHODNI	Wielkopolskie	Miasto Poznań	Poznań	2018	9	4	2215

In the table we can see that there are some regions with blank names - it means that probably it was not possible to identify a region with algorithms already used.

2.2. Number of online job vacancies by ESCO level, date and NUTS

The second indicator is the analysis of job vacancies by ESCO level. It is important to know whether occupations are ranked the same way as in publications from traditional surveys. It was presented in Table 2.

Table 2. General overview of number of online job vacancies

	esco_level_1	esco_level_2	esco_level_3	esco_level_4	Number_of_OJV
0	Service and sales workers	Personal service workers	Waiters and bartenders	Waiters	178751
1	Professionals	Business administration and professionals	Administration professionals	Training and staff development professionals	141062
2	Clerical support workers	Customer services clerks	Client information workers	Enquiry clerks	121227
3	Elementary occupations	Labourers in mining, construction, manufacturi...	Transport and storage labourers	Freight handlers	71727
4	Elementary occupations	Labourers in mining, construction, manufacturi...	Manufacturing labourers	Hand packers	69882
5	Service and sales workers	Sales workers	Shop salespersons	Shop sales assistants	59800
6	Elementary occupations	Cleaners and helpers	Domestic, hotel and office cleaners and helpers	Cleaners and helpers in offices, hotels and ot...	56725
7	Clerical support workers	Numerical and material recording clerks	Material-recording and transport clerks	Stock clerks	55436
8	Craft and related trades workers	Handicraft and printing workers	Handicraft workers	Potters and related workers	47308
9	Service and sales workers	Sales workers	Other sales workers	Contact centre salespersons	45069

Data in table 2 shows that the largest number of job vacancies in the reference period is the largest for service and sales workers. This occupation is also identified as the most popular in official statistics data.

2.3. Number of online job vacancies by NACE, date and NUTS

Next indicator concerns the sector having the biggest number of online job vacancies data. It was shown in table 3.

Table 3. Number of OJV by region, sector and date

	macro_region	macro_sector	sector	year_grab_date	month_grab_date	day_grab_date	Number_of_OJV
0	REGION CENTRALNY	Human health and social work activities	Human health activities	2019	2	19	5558
1	REGION POLUDNIOWY	Human health and social work activities	Human health activities	2019	2	19	3679
2	REGION CENTRALNY	Human health and social work activities	Human health activities	2019	2	18	3346
3	REGION PÓŁNOCNO-ZACHODNI	Human health and social work activities	Human health activities	2019	2	19	3291
4	REGION CENTRALNY	Human health and social work activities	Human health activities	2019	6	26	3219
5	REGION CENTRALNY	Human health and social work activities	Human health activities	2019	3	1	3133
6	REGION CENTRALNY	Human health and social work activities	Human health activities	2019	4	20	3118
7	REGION CENTRALNY	Human health and social work activities	Human health activities	2019	4	19	2915
8		Accommodation and food service activities	Food and beverage service activities	2018	9	4	2902
9	REGION PÓŁNOCNY	Human health and social work activities	Human health activities	2019	2	19	2701

In every region one of macro sector is dominating - which is human health and social work activities. However, in row 8, we can see that not all sectors have the location assigned.

2.4. Number of online job vacancies by date

The aim of this indicator is to show seasonal effect e.g., total 1444290 for Poland in the last quarter of 2018. In Table 4, we can see the number of OJV by date.

Table 4. Number of OJV by date

	year_grab_date	month_grab_date	day_grab_date	Number_of_OJV
0	2018	9	4	73619
1	2019	2	19	22938
2	2019	2	6	21886
3	2018	7	12	21059
4	2019	4	18	20686
5	2019	6	26	20327
6	2019	2	18	19465
7	2019	3	20	18643
8	2019	4	14	18639
9	2018	7	17	18216

A quick analysis of the data presented in Table 4 led to the conclusion that there is a big disproportion in number of OJV by date, e.g., 4th of September 2018 - 73.6 thous., 19th of February 2019 - 22.9 thous. Please note that data in descending order by Number of OJV were presented in Table 4.

3. Possible new indicators

One of the possible indicators that is not covered in online tool Skills OVATE, is the analysis of soft skills needed by educational attainment. For example, the following tables shows what skills are the most demanded by different educational attainment.

Table 5. Examples of number of the most popular skills by educational attainment

	educational_level	escoskill_level_3	number
1	Post-secondary non-tertiary education	use a computer	358069
2	Post-secondary non-tertiary education	adapt to change	295066
3	Post-secondary non-tertiary education	work as a team	219734
4	Post-secondary non-tertiary education	teamwork principles	218418
5	Post-secondary non-tertiary education	Smalltalk	198743
6	Post-secondary non-tertiary education	use communication techniques	184131
7	Post-secondary non-tertiary education	think creatively	181362
8	Post-secondary non-tertiary education	public relations	180556
9	Post-secondary non-tertiary education	job market offers	147570
10	Post-secondary non-tertiary education	show responsibility	138571
1	Short-cycle tertiary education	adapt to change	14767
2	Short-cycle tertiary education	use a computer	11708
3	Short-cycle tertiary education	work as a team	11169
5	Short-cycle tertiary education	use microsoft office	8014
7	Short-cycle tertiary education	show responsibility	6632
8	Short-cycle tertiary education	teamwork principles	6382
9	Short-cycle tertiary education	create solutions to problems	4308
10	Short-cycle tertiary education	problem solving	4158
13	Short-cycle tertiary education	Smalltalk	2963
14	Short-cycle tertiary education	think creatively	2721
16	Short-cycle tertiary education	quality standards	2612
17	Short-cycle tertiary education	manage time	2605
0	Master or equivalent	use a computer	25585
1	Master or equivalent	adapt to change	25027
2	Master or equivalent	work as a team	22402
3	Master or equivalent	teamwork principles	21603
4	Master or equivalent	create solutions to problems	17214
5	Master or equivalent	think creatively	15996
7	Master or equivalent	problem solving	15881
8	Master or equivalent	project management	14788
9	Master or equivalent	use communication techniques	13831
10	Master or equivalent	develop strategy to solve problems	13600
11	Master or equivalent	show responsibility	12988
12	Master or equivalent	analyse problems for opportunities	12357
14	Master or equivalent	think analytically	11122
15	Master or equivalent	communication	10987
16	Master or equivalent	lead a team	9933

It is worth to note that almost each OJV have assigned educational attainment by ISCED code (International Standard Classification of Education). It means that we can make analysis in 8-levels scale, from “no education needed” to “doctoral educational attainment”.

A suggested approach is to calculate the following indicator that is not presented in the current Skills OVATE tool:

$$\text{Skills needed by educational attainment} = \frac{N_{sk}}{OJ_{ved}}$$

where:

N_{sk} - Number of OJV for a specific skill by educational attainment,

OJ_{ved} - Number of OJV for a specific skill by educational attainment

The results were presented in Table 6.

Table 6. Percentage of skills needed by educational attainment

ID	Educational attainment		Skills	Number of OJV	Percentage in total number of OJV
0	Bachelor or equivalent		adapt to change	44058	64,8%
1	Bachelor or equivalent		work as a team	39599	58,2%
2	Bachelor or equivalent		use a computer	39386	57,9%
3	Bachelor or equivalent		teamwork principles	35566	52,3%
4	Bachelor or equivalent		use microsoft office	29731	43,7%
5	Bachelor or equivalent		create solutions to problems	27814	40,9%
0	Master or equivalent		use a computer	25585	64,4%
1	Master or equivalent		adapt to change	25027	63,0%
2	Master or equivalent		work as a team	22402	56,4%
3	Master or equivalent		teamwork principles	21603	54,4%
4	Master or equivalent		create solutions to problems	17214	43,3%
5	Master or equivalent		think creatively	15996	40,2%
0	Post-secondary education	non-tertiary		826958	46,6%
1	Post-secondary education	non-tertiary	use a computer	358069	20,2%
2	Post-secondary education	non-tertiary	adapt to change	295066	16,6%
3	Post-secondary education	non-tertiary	work as a team	219734	12,4%
4	Post-secondary education	non-tertiary	teamwork principles	218418	12,3%

In Table 6 we can see that no skills are needed for ca. 46.6% of OJV for post-secondary non-tertiary education. For Master graduates, the most skill needed is to use a computer (64.4%). Employers expect from Bachelor graduates a skill “adapt to change” in 64.8% of job vacancies.

This indicator helps to develop the educational strategy for different schools.