Central European
Labour Studies
Institute

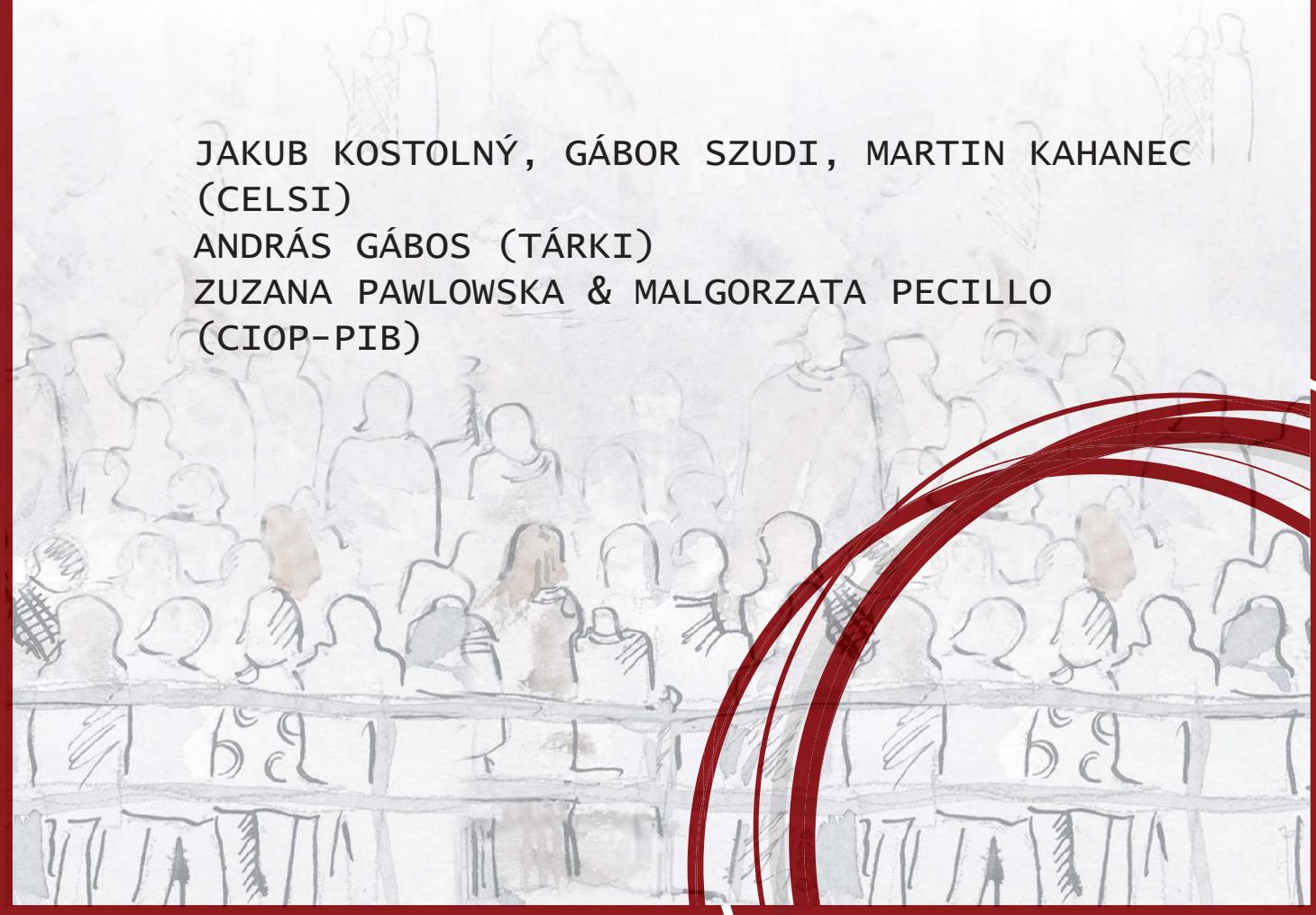
www.celsi.sk

CELSI Research Report No. 31

RESEARCH INFRASTRUCTURE GAPS REPORT CENTRAL AND EASTERN EUROPE

FEBRUARY 2020

JAKUB KOSTOLNÝ, GÁBOR SZUDI, MARTIN KAHANEC (CELSI)
ANDRÁS GÁBOS (TÁRKI)
ZUZANA PAWLOWSKA & MALGORZATA PECILLO (CIOP-PIB)



Research
Infrastructure gaps report Central and Eastern Europe

CELSI Reserach Report No. 31
February 2020

Jakub Kostolný, Gábor Szudi & Martin Kahanec (CELSI)
András Gábos (TARKI)
Zofia Pawlowska & Malgorzata Pecillo (CIOP-PIB)

Central European Labour Studies Institute (CELSI)

Zvolenská 29
821 09 Bratislava
Slovak Republic

Tel/Fax: +421-2-207 357 67
E-mail: info@celsi.sk
Web: www.celsi.sk

CELSI Reserach Report No.31

The aim of this report is to map the gaps in data sets and research infrastructures in Central and Eastern Europe and to analyse the main obstacles and challenges that underpin these gaps. The report looks at the gaps from the perspective of four types of data: (i) international and national survey data, (ii) data provided within the InGRID project (iii) administrative data, and (iv) data on vulnerable groups. In order to study the gaps an expert questionnaire has been constructed following a roundtable discussion and pilot interviews. Two case studies with focus on gaps in data on vulnerable groups in Hungary and working conditions in Poland are presented as well.

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under Grant Agreement No 730998

Corresponding Author:
Jakub Kostolný
Central European Labour Studies Institute
Zvolenská 29, 821 09 Bratislava, Slovakia
E-mail: jakub.kostolny@celsi.sk

Contents

1. Introduction

Gaps in access to data inhibit, and often preclude important research and policy analysis. Even though a large variety of data is collected at the European, national and sub-national levels, including publicly financed data sets and standardised European survey data sets, access to much of the existing data is often limited by financial constraints, lacking administrative capacity of the data owner, or lacking political will to provide a framework for data access. The lack of access to reliable and comparable data is even more evident when it comes to studying research questions and policy issues related to hard-to reach vulnerable groups. Further barriers to effective access to data arise when studying sensitive topics such as poverty or social and labour market exclusion of such groups.

The lack of access to data is a pressing issue especially in the regions with weaker tradition of providing data for social science research, such as Balkan or Central and Eastern Europe (CEE). The first InGRID project (Hamon-Cholet et al., 2017) raised these concerns and concluded that research infrastructures in CEE should be more integrated into European research infrastructures and that research should be more promoted in order to overcome such barriers.

The aim of this report is to map the gaps in data in Central and Eastern Europe, to identify the main drivers of these gaps and possibilities on how to overcome them.

2. Methodology

The analyses in this report are based on four main data sources: (1) pilot interviews and roundtable debate on Central and Eastern European research gaps, (2) the InGRID-2 CEE expert questionnaire on data gaps, (3) the InGRID-2 user survey, and (4) country case studies on gaps in data sets and research infrastructures in Hungary and Poland by InGRID-2 partners TÁRKI Social Research Institute and CIOP-PIB, respectively.

The pilot interviews and the roundtable were organised by CELSI on 7-8 March 2019 within the InGRID-2 project. One of the main purposes of the event was to inform the design of the InGRID-2 CEE expert questionnaire. The roundtable was conducted with four experts in the panel and 40 representatives of government analytical institutions, statistical offices and researchers. During the event five experts were approached for pilot interviews in order to inform and better formulate the InGRID-2 CEE expert questionnaire. In particular, the InGRID-2 CEE expert questionnaire was constructed with regards to the main findings and discussions from the CEE research infrastructure gaps roundtable,[1] a pilot interview. To ensure cross-comparability, CELSI considered the InGRID-2 user survey and the questionnaire on gaps in Balkan countries by PANTEION.[2] Based on these inputs, the InGRID-2 CEE expert questionnaire was designed to cover four main themes: (1) gaps in international and national data, (2) gaps in data provided in the InGRID-2 project (3) use of administrative data and (4) gaps in data on vulnerable and hard-to-reach groups.

The InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe was disseminated among more than 200 stakeholders in the studied Central-Eastern European countries (Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia) between 25 September and 25 November 2019. The database of experts that were approached was constructed based on contacts, research institutions, research fellows and experts from the region and also shared within the network of experts who attended the CEE roundtable and data forum events organised within InGRID-2. The questionnaire was disseminated primarily by email with a personalised message, but also via CELSI website and social media, InGRID-2 website and newsletter. To boost response from the Baltic states, on November 15th the survey was sent to 15 additional experts in Estonia, Latvia and Lithuania

2.1 The sample

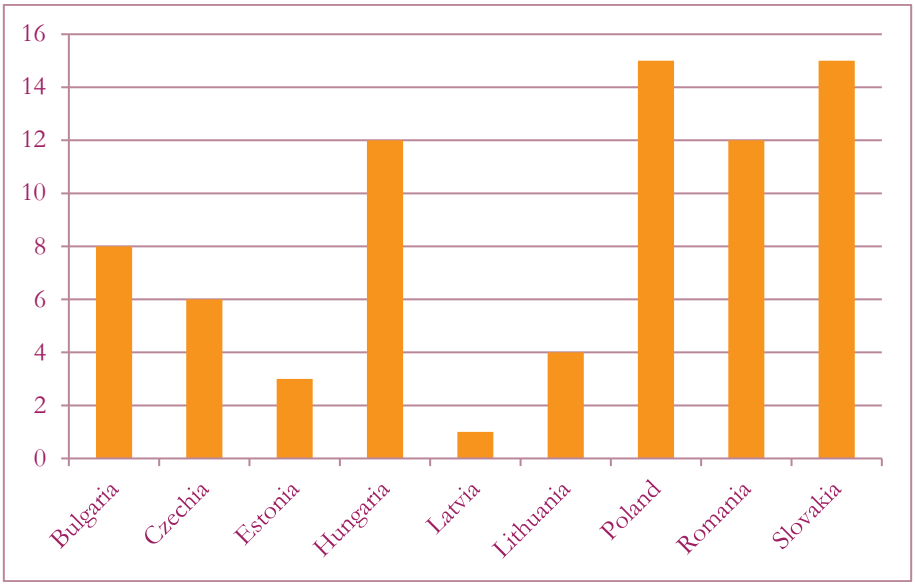
The InGRID-2 CEE expert survey focused specifically on the gaps in research data sets, data infrastructure, and data about vulnerable groups (see Annex 1). The total number of experts from the CEE target countries was 18 (the total, including countries outside the CEE, was 35; 17 experts were based outside of the CEE region but referred to the region in their relevant answers). The InGRID-2 CEE expert respondents were complemented with 57 responses received for the target CEE countries through the INGRID-2 user survey developed by HIVA-KU Leuven in July 2019. Thus, the total number of 75 responses gathered from all CEE target countries are used in this report. Due to some differences between the research questions in the InGRID-2 CEE expert questionnaire and

1 http://www.inclusivegrowth.eu/round-tables/round-table-on-main-gaps-in-research-infrastructures-in-the-balkan-countries-panteion-29-october-2018

2 http://www.inclusivegrowth.eu/round-tables/round-table-on-main-gaps-in-research-infrastructures-in-central-eastern-europe-celsi-7-8-march-2019)

the InGRID-2 user survey, we treat the gathered data separately. In particular, while some general questions cover all 75 responses, only the relevant responses are covered in the analysis of specific question. Figure 1 shows country-composition of the sample.

Figure 1. Country composition of the sample



There are 63 female experts and 31 male experts. Most experts, 51, are from the age group of 35-49 years old, 23 are aged 18-34 and 16 experts are between 50 and 64 years old. 19 experts have up to 5 years of tenure, 16 have 6 to 10 years of tenure and 14 have more than 11 years of experience. Among the experts, there are 25 junior researchers, 33 senior researchers, 10 professors 7 directors or research managers.

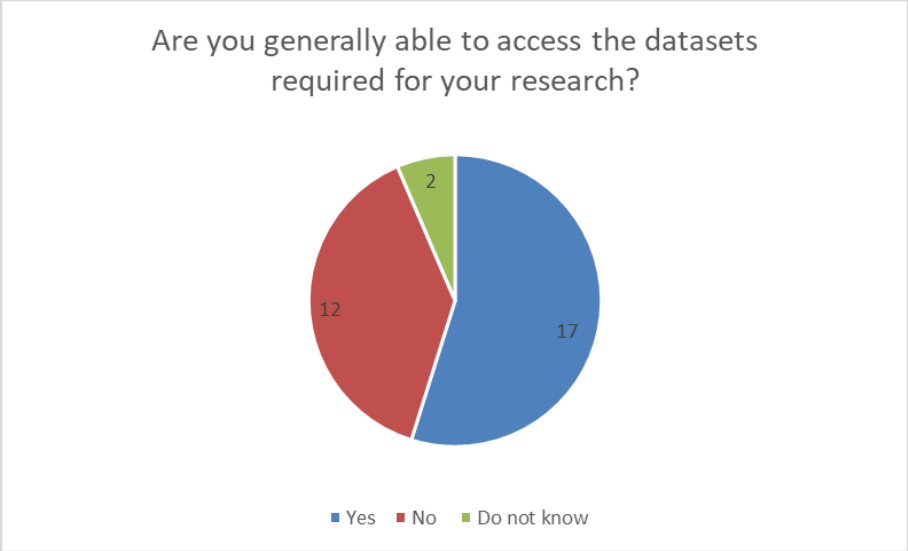
The main areas of expertise are: labour market and vulnerability (20), inequality and social exclusion (18), social policy and welfare state (13), poverty and living conditions (11), working conditions and employment (10) and industrial relations and collective bargaining (6).

3. Analysis

3.1 Satisfaction with access and use of data

Out of the 31 experts answering the question 'Are you generally able to access the data sets required for your research', 17 confirmed that they are generally able to access the data sets required for their research work, and 12 experts answered that they cannot access such data sets (Figure 2). Two experts did not know or preferred not to give an answer.

Figure 2. Access to data sets



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe, Question 9

Based on *experience*, we observe that experts with more years of experience have fewer problems with access to the relevant research data sets relevant to CEE region. Around 50% of experts in each category with fewer years of experience (people with max 5 years of experience, with experience between 6-10 years and experience between 11-15 years) could generally access the relevant databases: altogether 12 experts were generally able to access relevant databases, while 11 could not do so or could not specifically indicate it. However only 1 out of the 5 experts with the highest level of experience (between 16-20 years and with more than 20 years of experience) indicated that he/she cannot generally access the relevant databases which may suggest that experts with more years in their relevant study fields will also gain more experience on the ways and methods to reach these databases.

Similarly, we found that experts aged 50 and above may have fewer problems with accessing such databases. Half of younger experts (in the age cohorts of 18-34 years and 35-49 years) could generally access the relevant databases: altogether 11 experts were generally able to access relevant databases, while 11 could not do so or could not specifically indicate it. However, only 1 out of the 6 older experts (between 50-64 and above 64) indicated that he/she cannot generally access the relevant databases.

It is interesting to note that while the majority of experts who work in a research institute, think tank or governmental organisation indicated they are generally able to access the databases (10 out of 13), most experts employed at non-governmental or civil society organisations and at higher education institutes or universities claimed that they cannot generally access such data sets (9 out of 14), which may suggest that knowledge about or accessibility to such research databases is more limited in these types of institutions.

There are no detectable differences in the reported ability to access databases in the CEE region with regard to *gender* or *area of expertise* of experts. Among those answering this question, 3 out of 5 men and 7 out of 10 women could not generally access the relevant databases (with 2 women not giving a specific answer). When checking the pre-defined expertise fields (Q4: industrial relations, collective bargaining and representation; inequality and social inclusion; labour market vulnerability and precariousness; poverty and living conditions; social policy, welfare states, social services; working conditions, employment conditions, personnel management; other) we observe that overall more than half of the experts in each field claimed accessibility to relevant data sets (16 out of 28 experts answering this question) with around half of experts having access to databases in each field with the exception of industrial relations and social policy where all 5 experts (respectively 3 and 2) claimed to have a general access to databases.

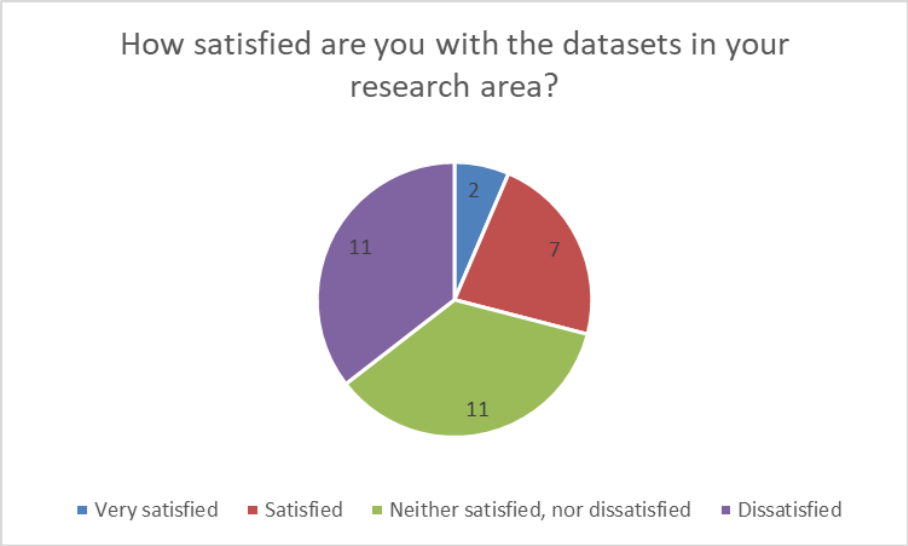
When asked in the InGRID-2 user questionnaire whether the currently available data, indicators, methods and tools are sufficient to support research and policymaking on the topics of vulnerability and working conditions 21 experts say no while 20 says the data are sufficient.

We observe some variation across countries. For Slovakia, 6 experts report that data is generally available and 1 claims that 'effective criteria for representativeness and reliable data on collective bargaining coverage are needed.' In Poland, 5 out of 10 experts claim that the data is generally not available with reasons such as: 'more site-specific data is required', 'the data about firms are not freely available' and that there is 'lack of data on the level of single households'. In Bulgaria 2 out of 5 experts claim data is not accessible, with one expert claiming 'there is a need for additional ad-hoc surveys and longitudinal studies'. In Romania 5 claiming the data is insufficient and 1 saying it's enough. 'It's partly sufficient. Some of the topics need more attention from ethical point of view, other topics (like environment) need the attention of the social sciences.', 'data protection policy', 'statistics on mobile individuals need to be improved'. In Estonia 1 out of 2 experts is saying there are: 'uncovered topics, low comparability, data access problems'.

In Latvia there is one positive answer and in Lithuania 2 out of 3 experts claim that: 'Some data available, but limited access and possibility to use merged admin data', 'EU-wide data on employment of persons with disabilities is based on 2011 data; decision-makers can make informed decisions on improving the situation of this group on this basis. Also longitudinal data of any sort is missing' (InGRID-2 user survey). Based on the InGRID-2 CEE expert survey we were able to get one more response from Lithuania saying: 'We wanted to include an overview of labour market segmentation in the country as there's really nothing on the topic to date, but having tried to access administrative data I came to understand why. Social Security data was limited – for example no indicators on education, and also not accessible. Moreover there was no way even in theory to get linked data by means of an individual identifier between e.g. social security and education.'

With regards to satisfaction with the data in the InGRID-2 CEE expert questionnaire out of the 31 experts, 11 said that they are neither satisfied nor dissatisfied with the data sets in their research area, while 9 of them confirmed their satisfaction (satisfied or vary satisfied) and 11 experts voiced their dissatisfaction. Overall, if neutral opinions are not taken into consideration, more experts were generally dissatisfied with the data sets than satisfied but while 2 persons indicated a high level of satisfaction, no experts had an extremely negative stance on the databases (very dissatisfied).

Figure 3. Satisfaction with data sets



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe, Question 10

Concerning satisfaction based on *seniority* of experts, we can observe that senior researchers (having at least 10 years of experience in a given field) were more likely to be less satisfied (2 out of 9 indicated concrete satisfaction) than average or junior researchers, who were likely to be more satisfied (6 out of 19 indicated concrete satisfaction). The difference is not statistically significant even if we take into account that younger experts were also more likely to give neutral answers (neither satisfied nor dissatisfied). This small difference may indicate that the available data sets are appropriate until a specific expertise level but not for more 'sophisticated' data looked for by senior experts or they do not know where to look for relevant data (which is in line with our findings on the relation between experience and access to data sets). This difference corresponds to the findings related to satisfaction pursuant to the *age* of experts since researchers between 35 and 64 were more critical (4 out of 17 satisfied) than the younger (18-34 years old) age group (3 out of 10 dissatisfied).

As in the case of accessibility, those experts replying to the question and working in governmental organisations (9) indicated that they are satisfied (4) or at least neutral (2) towards data sets, while each expert from non-governmental or civil society organisations (4) stated their dissatisfaction with databases. This distinction may again hint that information available within these data sets is appropriate for governmental *working environments*, but not for the civil sector.

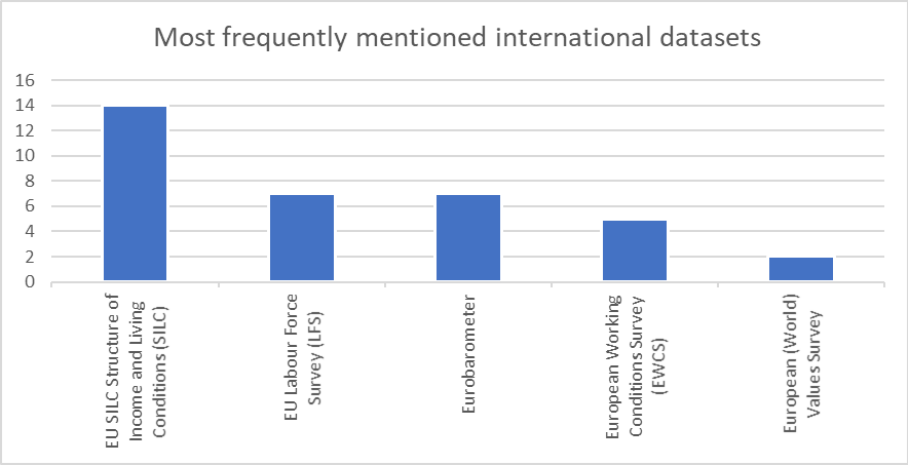
Regarding the *expertise fields* of experts, those working in areas related to labour markets and inequality and social exclusion are the most dissatisfied with current databases (only 2 out of 13 experts were satisfied, 1 and 1 expert respectively). Experts (2 out 3) in the field of industrial relations remained neutral, i.e. neither satisfied nor dissatisfied. Satisfaction with databases was generally stated by experts in economic fields (statistics, economics and business) citing a general content with data sets on microsimulations of tax and transfer systems, as well as fiscal policy.

The *concrete gaps in research topics* mentioned by the experts were more numerous. A social scientist mentioned that data on trade union density, working conditions and workplace employment relations are not systematically collected, and a senior economist expressed the view that there is no comparative cross-country data set available for labour mobility or return migration in Central-Eastern Europe; two experts (social scientist and economist) underlined the lack of longitudal studies using administrative data are missing in their respective countries.

3.2 International, national and administrative data sets, as well as data provided in INGRID-2 project used in this analysis

The expert questionnaires also aimed to find out the most *frequently used data sets* in the context of Central-Eastern Europe. The databases were divided into three categories: international, INGRID and administrative data sets. In the first two categories, experts could choose from a set of pre-defined databases, while they could freely provide any administrative data sets used.

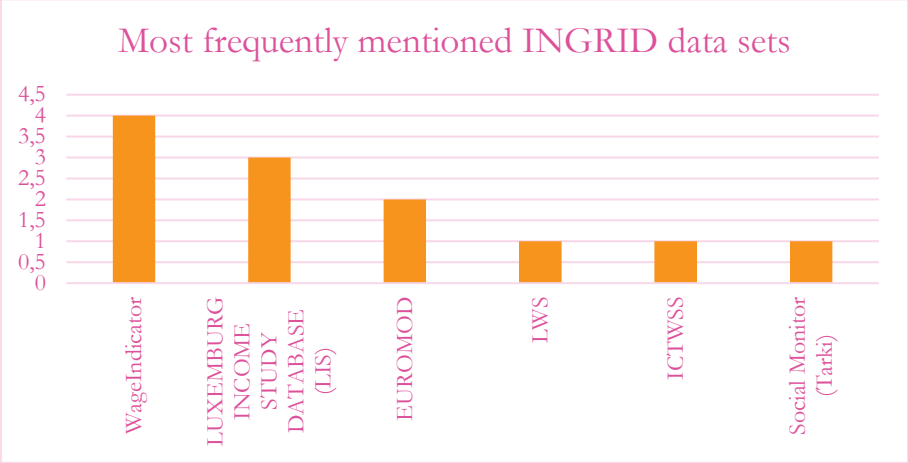
Figure 4. Most frequently mentioned international data sets



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

The most frequently used *international and national data sets* were the Structure of Income and Living Conditions data set (EU SILC) (14 mentions), followed by the EU Labour Force Survey (LFS) and Eurobarometer (7-7 mentions), European Working Conditions Survey (EWCS) (5 mentions) and European World Values Survey (2 mentions). The other surveys were mentioned less than two times.

Figure 5. Most frequently mentioned InGRID data sets



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

The *databases used in the INGRID project* were much less selected by the experts, indicating a lower level of knowledge about these data sets. 4 experts stated the frequent usage of the WageIndicator database, while 3 experts said that they use the Luxembourg Income Study (LIS) database and 2 experts mentioned the frequent use of EUROMOD. 1 expert each indicated the use of the Luxembourg

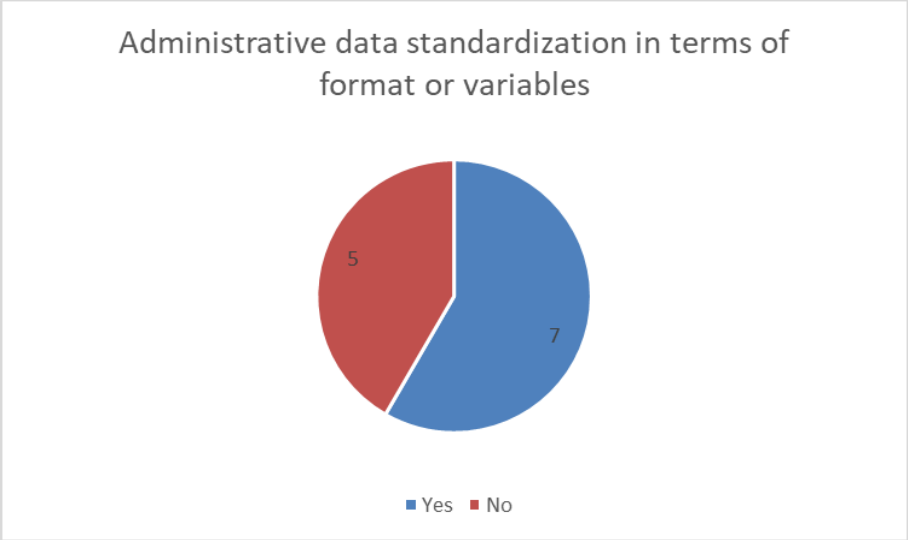
Wealth Study Database, Tárki's Social Monitoring database or the University of Amsterdam's database on international characteristics of social dialogue as frequently used in his/her daily work. It should be noted that more options were given to experts, but no mentions were given to other databases, such as the Social Assistance Explorer, the University of Stockholm's Comparative Social Policy databases, the University of Antwerp's Minimum Income Protection database or Tárki's IPOLIS (Integrated Poverty and Living Conditions Indicator System.

In case of *administrative databases* used, experts could freely express their choices. Therefore, there is a higher variety of such data sets mentioned.

The following 12 administrative data sets were explicitly mentioned by experts in CEE region:

1. PES administrative data about jobseekers in Slovakia.
2. Data of Slovak Social Security Agency.
3. Data of Czech Ministry of Labour and Social Affairs.
4. Health insurance company data in Czech Republic.
5. Labour Ministry data on social transfer beneficiaries in Czech Republic.
6. Average Earnings Information System in Czech Republic.
7. Hungarian Migration Office statistics.
8. The foreign workers' registry in Czech Republic.
9. School performance statistics by the Czech Ministry of Education.
10. Annual data on social in/exclusion in Hungary.
11. The Lithuanian Social Insurance data (and outside of the region).
12. Norwegian microdata on the labour market (used in Czech Republic).

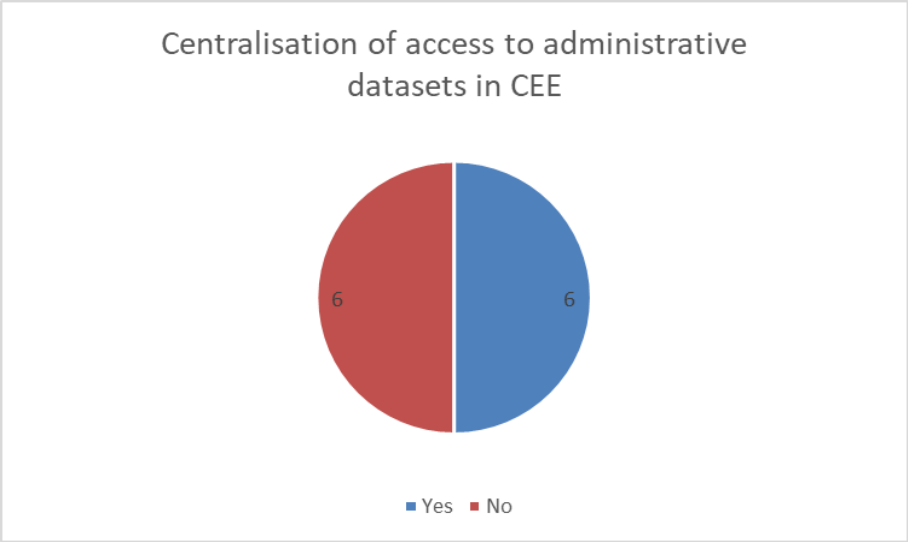
Figure 6. Standardised procedures for obtaining of administrative data



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

In four cases, experts mentioned that there is a *standardised process of obtaining the data* for such an administrative data set: the Norwegian data set outside of the CEE region, as well as the Lithuanian data with a standard data gathering by the Social Insurance Institution. In case of the Czech Labour Ministry's data on social transfer beneficiaries and the school performance statistics provided by the Czech Ministry of Education, the sharing of data between administrative bodies is generally described by law (formal letter asking for the data and formal answer which includes the data which exist and can be shared).

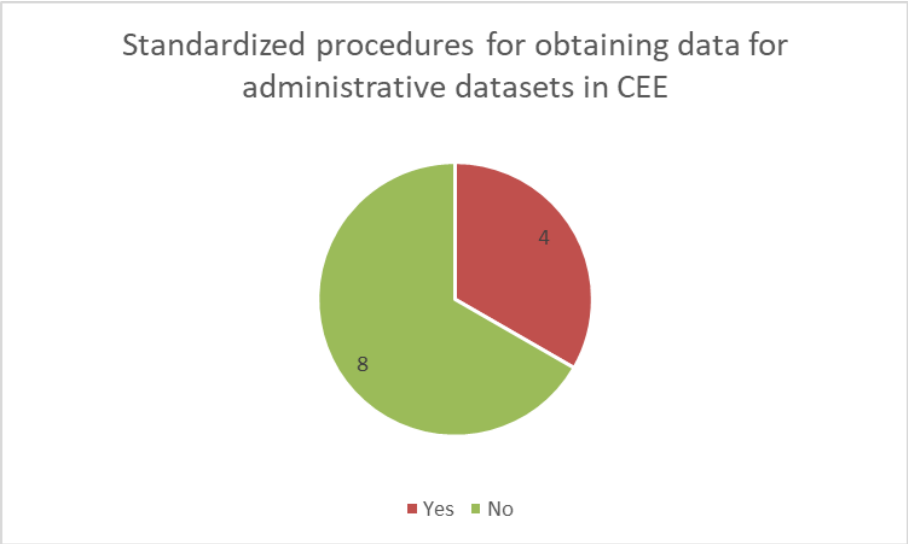
Figure 7. Administrative data standardisation



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

In case of data *standardisation in terms of format or variables* used for different administrative data sets in the CEE region, the values indicated a higher level of standardisation than for data (obtaining) procedures. 7 such databases are standardised either in format or variables, and only 5 are left not standardised. Most commonly the variables and the units of observation (5-5 databases) are standardised, followed by sources of data and format (3-3). The Norwegian microdata and the School performance statistics by the Czech Ministry of Education are the databases standardised in each four aspects.

Figure 8. Centralisation of access to administrative data sets

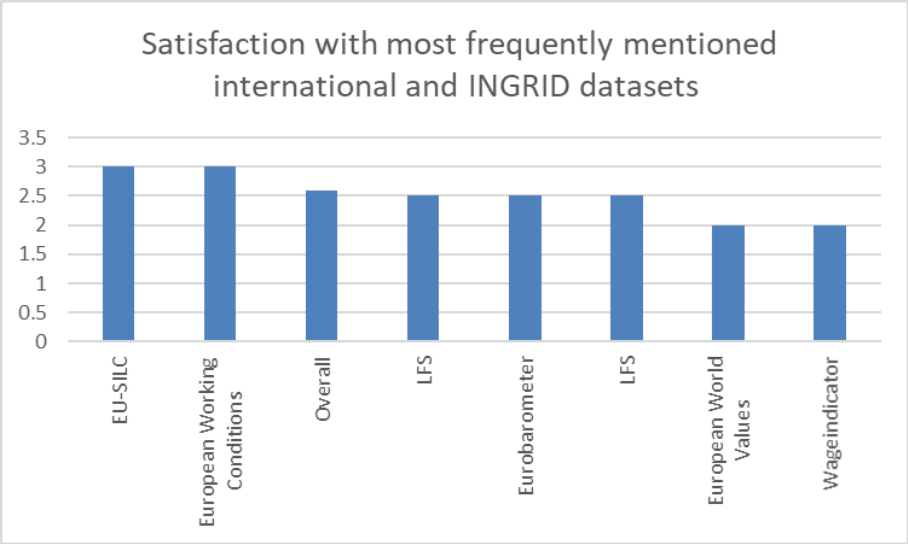


Source INGRID expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

Regarding the *centralisation of access to the regional administrative data sets,* the expert questionnaires revealed that half of these are centralised and thus can be accessed at one organisation, and half are decentralised and can be accessed for instance at different units of ministries or branches of national authorities. Some experts detailed the central organisations for respective data sets: the Central Labour Office of Labour, Social Affairs and Family may access administrative data about jobseekers in Slovakia, the Social Security Agency may access the social security data in Slovakia, the Lithuanian

Social Insurance Institution may access the social insurance data, the Ministry of Social Affairs may access the foreign workers' registry in Czech Republic.

Figure 9. Satisfaction with International and InGRID data sets



* 0: very dissatisfied – 4: very satisfied.
Source INGRID expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

If we check the *general level of satisfaction* with the most frequently mentioned international and INGRID data sets, we can observe a generally positive picture. In a scale from 0 to 4 where 0 stands for total dissatisfaction and 4 for complete satisfaction, the overall group of data sets are valued at 2.6 with the most frequently used databases closely around this average. The WageIndicator data set received the lowest average value with 2 (indicating a neutral stance of experts), while the EU-SILC and European Working Conditions databases received a higher than average (3) satisfaction rate.

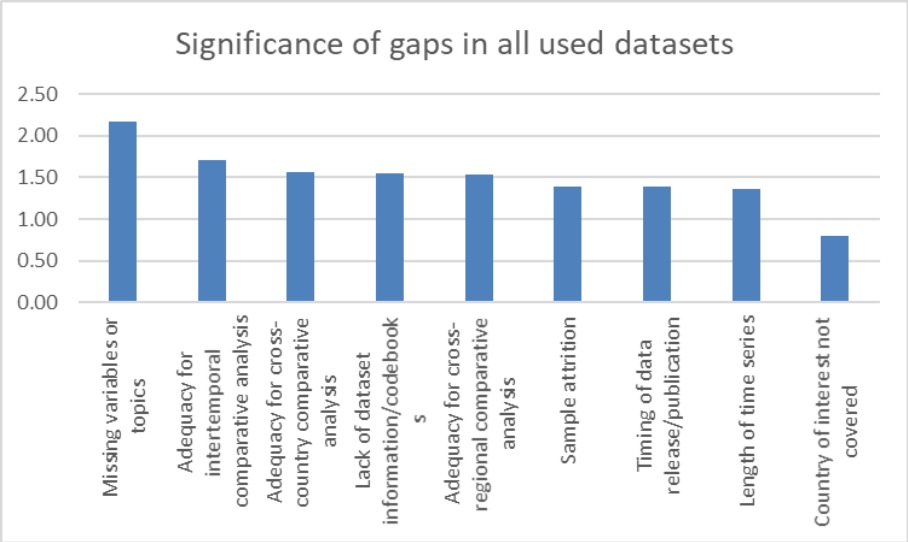
3.3 Significance of gaps and obstacles

We aimed to delve into problematic issue deeper in each category of data sets therefore specific research gaps in the international, INGRID and administrative data sets were also inquired. The experts were asked to rate the significance of the following gaps:

- missing variables or topics;
- sample attrition;
- country of interest not covered;
- adequacy for cross-country comparative analysis;
- adequacy for cross-region comparative analysis;
- adequacy for intertemporal comparative analysis;
- timing of data release/publication;
- length of time series;
- lack of data set information/codebooks.

Based on their evaluation, the following can be concluded in a scale where 0 is insignificant (i.e. there is no such gap in a data set) and 4 is very significant (i.e. the data set has serious problems in terms of the category):

Figure 10. Significance of gaps in all data sets



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

In case the *research gaps* encountered in all three types of databases (international, INGRID and administrative) are evaluated, experts stated the most significant gaps in terms of missing variables and topics and also experienced serious gaps in all used data sets in terms of adequacy for inter-temporal, cross-country and cross-regional comparative analysis, indicating the overall problems with such data sets when used for comparative analytical purposes. We must however underline that only one category (missing variables or topics) received more than 2 points (showing rather an overall dissatisfaction among experts). It thus can be concluded that majority of experts was generally satisfied with data sets used despite obvious shortcomings. The least gaps were encountered with regard to potential lack of coverage of country of interest (country and regional coverage can be deemed appropriate).

Based on the findings from the InGRID-2 user questionnaire, which studied the gaps and obstacles in general, not by the categories of data as InGRID-2 CEE questionnaire, experts seem to specify five main issues (using 1-8 ranking) when using data and indicators in the following order (most important issues, responses ranked 1-3): (1) accessibility, (2) missing data and data availability including sample size, completeness or unit of analysis, (3) coherence and comparability including harmonisation of the data, (4) timeliness and punctuality and coverage of topical issues and (5) data quality, accuracy and reliability.

In a follow-up multiple choice question, the recurring drivers which underpin these issues seem to be: (1) lack of comparability of data sources, difficulties to link data of multiple source, (2) limited coverage of specific groups (e.g. sectors, companies, hard-to-reach individuals), (3) limited availability of microdata, difficulties in getting access to microdata and (4) lack of comparability of data across time.

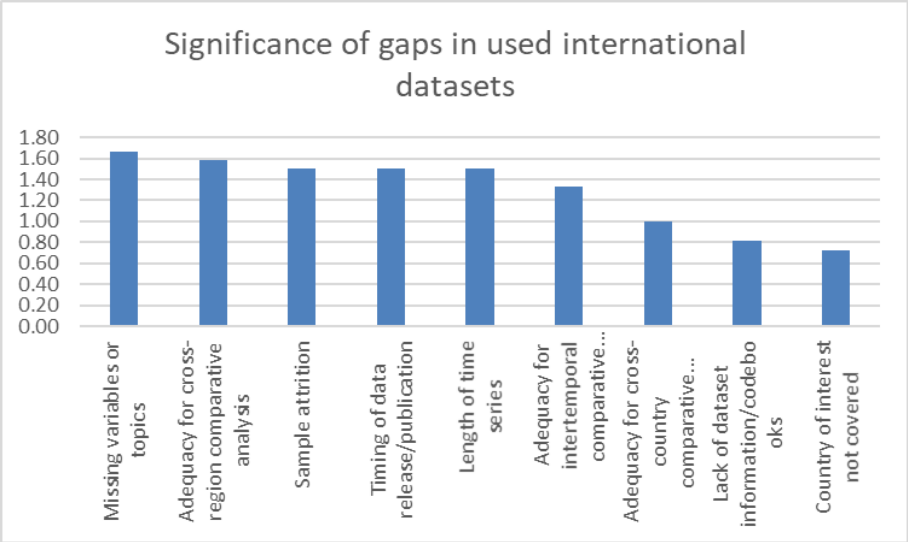
The user questionnaire also focuses on challenges which prevent bridging the gap. The experts focus the most on lack of communication expressed by: (1) limited cooperation between researchers and policymakers, (2) lack of awareness of available research outcomes and their usability among policymakers, (3) communication issues, lack of mutual understanding, (4) mismatch between policy and research cycles and (5) lack of skills to interpret research outcomes.

When checking the specific research gaps mentioned in separate types of databases we get a slightly different overview. If only the *gaps within international databases* are checked, the first issue to be observed is the lowest general scores given to each gaps, indicating that experts are the most satisfied with such databases. If expressing dissatisfaction with certain aspects, they most frequently view the missing variables and topics, as well as adequacy for cross-regional comparative analysis as the most

relevant gaps (similarly when overall results on gaps are checked). However, more significant gaps are perceived in this category of data sets in terms of sample attrition, timing of data release or publication and the length of the series. These issues should be better considered by respective international and national organisations dealing with such databases. Again the lack of country coverage gains the lowest scores, along with a perceived lack of data set information or codebooks meaning that such databases are well-developed by their respective organisations with adequate countries covered and necessary information included for research purposes.

On a country level we can observe that Czech experts were the least critical and gave lower-than-average scores, while Polish and Hungarian experts were less satisfied and indicated more significant gaps, in particular in terms of time series length and adequacy for cross-country and intertemporal comparative analyses. One Slovak expert in particular mentioned that cross-country data sets are missing (making these type of analyses more difficult) on labour mobility, return and Roma migration in Central-Eastern Europe. One Czech expert working at a research institute mentioned the need for more data on trade union density, working conditions and workplace employment relations, but expressed general satisfaction with the European Working Conditions Survey (EWCS) and the European Company Survey (ECS).

Figure 11. Significance of gaps in international data sets

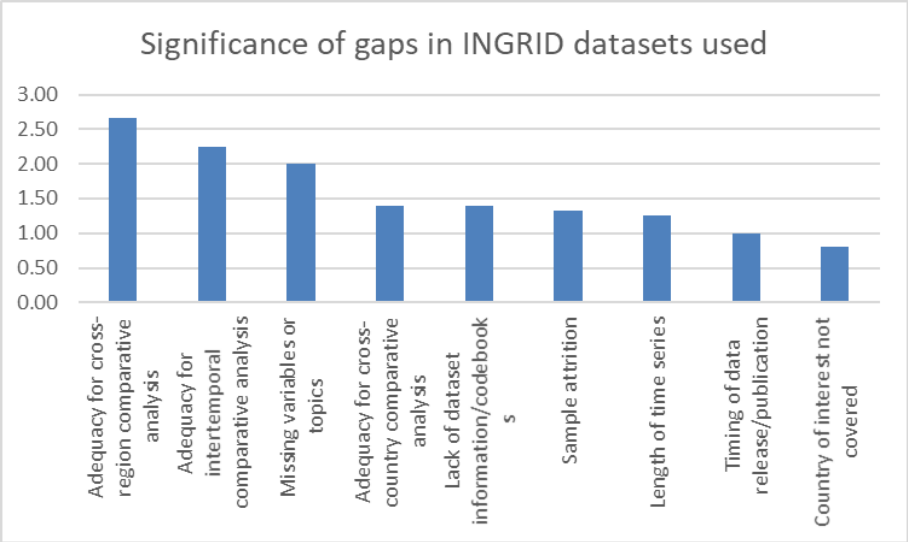


Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

If checking the gaps mentioned by experts for INGRID databases most frequently used, we observe higher average scores. Experts had the most serious problems (with a rating above 2) in terms of adequacy for cross-regional or intertemporal comparative analysis and missing variables or topics. It seems so that not missing variables, but rather their usability for comparative research purposes is the biggest problem concerning INGRID databases. The least significant gaps in the INGRID data sets were specified with regard to lack of country coverage (as in all categories), but experts did not perceive relevant gaps for the length of time series or the timing of data release either – in contrast to international data sets.

On a country level, Czech experts indicated more serious research database gaps, while Slovaks voiced the most favourable opinions towards INGRID databases. A Czech expert working at a higher education institution in the field of labour market and precariousness formulated dissatisfaction with current data sets in terms of missing variables and a lack of data set information.

Figure 12. Significance of gaps in InGRID data sets

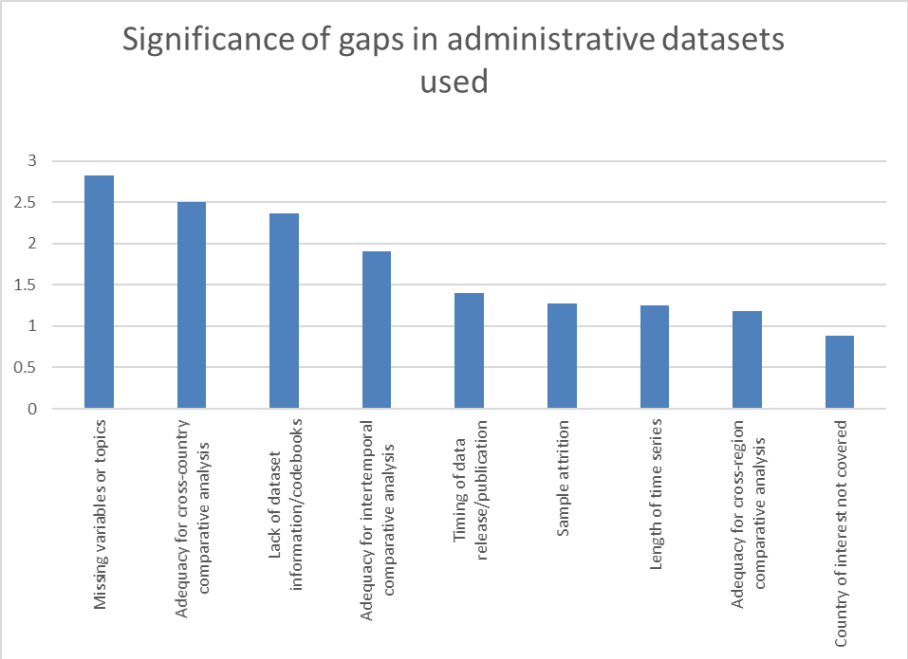


Source INGRID expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

The most serious gaps (with the highest overall scores and biggest variations) were mentioned by experts regarding the various administrative databases most frequently used in CEE region. Again, the missing variables or topics caused the biggest issues, but also significant gaps were generally perceived in terms of potentially missing variables or topics and adequacy for cross-country comparative analysis. The greater variety in administrative data sets also meant more significant gaps encountered with the lack of data set information or codebooks, potentially indicating more difficult usability of such data sets. The country coverage was deemed appropriate in case of administrative data sets too, and no significant gaps were observed concerning the length of time series or sample attrition either.

On a country level, we can observe that Czech experts were the least critical and gave lower-than-average scores, while Polish and Hungarian experts were less satisfied and indicated more significant gaps, in particular in terms of missing variables or topics or lack of data set information and codebooks. One Czech expert working at a governmental institution said that there are fragmented data and a lot of 'expert estimates' concerning homelessness and social housing provision where data on ethnicity and discrimination is missing (with small exception of education), and therefore intersectionality is practically impossible (however sometimes data on employment and social benefits are interlinked).

Figure 13. Significance of gaps in administrative data sets



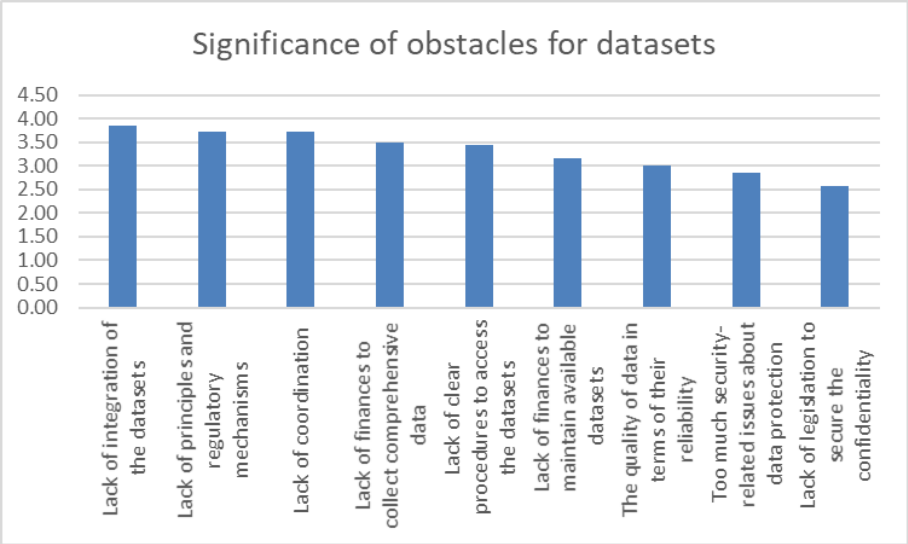
Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

Finally, we asked the opinion of experts about the *significance of possible obstacles* particularly concerning the collection and maintenance of data sets on poverty, living conditions, working conditions and vulnerability. Based on their own experience they rated the following obstacles:

- lack of finances to collect comprehensive data;
- lack of finances to maintain available data sets;
- lack of clear procedures to request/access the data sets;
- lack of legislation to secure the confidentiality of sensitive data;
- too much security-related issues with regard to data protection;
- lack of principles and regulatory mechanisms governing the establishment and use of large-scale, linked data sets;
- lack of coordination of what is collected/available and what is missing;
- the quality of data in terms of their reliability;
- lack of integration of the data sets – lack of multiple government agency data sets.

The experts evaluated the significance of the above obstacles as follows (in a scale of 4 to 0 where 4 indicates a very significant obstacle and 0 a non-existent, insignificant obstacle):

Figure 14. Significance of obstacles for data sets on poverty, living conditions, working conditions and vulnerability



Source InGRID-2 CEE expert questionnaire on the data gaps in research infrastructures in Central-Eastern Europe

As is visible in the figure, compared to the research gaps in various data sets used, the experts were much less positive regarding the obstacles about collection and maintenance of data sets on poverty, living conditions, working conditions and vulnerability. The scores ranked between 3.86 and 2.57 (where a score of 4 signalled the most significant obstacle) meaning that serious gaps should be overtaken in this area by relevant stakeholders. This is particularly true for the lack of integration of such data sets, the lack of principles and regulatory mechanism, the lack of coordination and the lack of finances used for collection of comprehensive data sets in these highly relevant research topics. The least serious obstacles were both concerning security issues: experts mentioned least issues with data protection or confidentiality (regulatory) mechanisms. On a country level, Czech experts seemed to observe the least significant obstacles compared to other countries, in particular to lack of finances to maintain or collect comprehensive databases and the data quality in terms of reliability.

As concrete gaps regarding data for difficult-to-reach groups Roma, people with disabilities, as well as people neither employed nor registered as unemployed (thus not receiving any social transfers if in working age), people who have dropped out of school, people released from prisons or young adults leaving institutional care were mentioned. More efforts should be taken by relevant stakeholders to collect and analyse data for such population cohorts in Central-Eastern Europe. One expert also mentioned that - even if such data are collected - information collected is rarely used in policy debates and policymaking on national level. The data is either not evaluated at all or the results are not communicated to relevant national stakeholders in order to focus the relevant policy discussion.

4. Case study: gaps in data and research infrastructures for vulnerable groups in Hungary

András Gábos (TÁRKI Social Research Institute)

4.1 Data sets available to monitor the situation of vulnerable groups in Hungary

General surveys and administrative data

Being an EU member, Hungary takes part in all Eurostat coordinated surveys. These surveys provide the main source of research related to inclusive growth. They cover either the whole population in private households (EU-SILC, HBS) or the adult population (e.g. EU-LFS, EHIS). For the moment, no administrative data are part of these data sources. The Hungarian EU-SILC provides the micro-level data input for EUROMOD.

As an alternative household income survey to EU-SILC, the TÁRKI Household Monitor Survey is also part of the Hungarian data infrastructure in income inequality, poverty and social inclusion (1992-1997: Hungarian Household Panel, 1998-2015: cross-sectional survey). Until recently, it provided input for LIS (Luxembourg Income Study) and OECD distributional analysis. External identification of the Roma was part of the survey. It made possible cross-check with Hungarian EU-SILC. Additional major surveys are: EQLS, ESS, HFCS (ECB survey on consumption and wealth).
Administrative data in several fields of taxation, social insurance, labour market participation, public work schemes and education are part of the data infrastructure.

Vulnerable group-specific data sources available for Hungary

Hungary is part of all major survey that provides infrastructure to monitor the situation of vulnerable groups within a cross-country comparative frame.
Children: PISA, TIMSS, PIRLS, ESPAD, HBSC.
Elderly: SHARE.
Roma: UNDP and FRA regional comparative surveys, EU-MIDIS II.

Ethnicity variables on major HCSO run surveys in Hungary[3]

The ethnicity variable is part of Hungarian Census since 1941, although its wording has been subject to change time to time. For the last Census (2011) and Microcensus (2016), a double question on ethnicity was asked, allowing therefore for an expression of a double identity. Answers to questions on ethnicity are not compulsory.

After a one-year pilot on LFS (2013) and EHSIS (2012), HCSO includes these double identity questions on all major Eurostat co-ordinated, non-mandatory surveys (EU-SILC, AES, EHSIS).

This initiative makes it possible to cover the most important indicators of poverty, living conditions and quality of life for the Roma under the same methodological background as for the overall population.

3 This short section is based on the presentation of Natalie Jamalia (HCSO) entitled 'Ethnicity variable in the social surveys of the Hungarian Central Statistical Office' at the InGRID-2 expert workshop on *Methods and data infrastructure to measure the quality of life of various vulnerable groups: extending IPOLIS*, held in Budapest, on 25-27 April 2018.

4.2 Gaps in the Hungarian data infrastructure

General problems

1. Important segments of the population are still missing from surveys conducted on private households. Individual or household-level non-response affects in the first place the bottom and the top of the income distribution and, as a result, vulnerable groups are under-surveyed. This is a growing concern and may (strongly) affect the estimates of poverty and quality of life indicators and as a consequence, policy effectiveness.
2. People living in institutions or institutional households (children, disabled people, older people) are missing from surveys. Again, this may lead to bias in indicator estimates and in public policies.

Country-specific problems

1. In accessing data for research, a few obstacles still exist. The micro data of HCSO run surveys can be access only through the researchers' room. The pre-harmonised national microdata sets that belong to the Eurostat coordinated surveys (e.g. to the EU-SILC), as well as other national survey data are accessible for researchers' community only at site, through the HCSO-based researchers' room. According to the HCCSO, this is necessary to due to data protection reasons. While the access to the data is ensured in this way, a few restrictions exist: e.g. a user fee must be paid, it involves additional administrative burden, it allows some discretionality for the HCSO to evaluate the access to the data against the aim of the research for which the data are requested, only printed outputs can be taken out from the room, the use of the data sets is limited in time and space.
2. The identification problem for Roma still exists. The estimated number of the Roma based on the double identity questions on HCSO run surveys is around 300,000, while alternative estimates set this number at 6-800,000 people (*Bernát and Messing, 2016*). This means that about half of the Roma population is missing from these surveys. Further research is needed to assess to what extent the two groups (the one self-identified in surveys and the others) are similar or not, and how this affects estimates for poverty and quality of life indicators.

Although general and country-specific drawbacks are present, overall, the Hungarian data infrastructure is adequate for monitoring poverty, living conditions and other domains of quality of life.

5. Case study: data sets for working conditions in Poland

Zofia Pawłowska, Małgorzata Pęcitło (CIOP-PIB)

5.1 General information

The main source of information on working conditions in Poland is the national survey of working conditions (SWC), conducted annually by the Central Statistical Office (CSO) in accordance with the programs of statistical surveys on public statistics. Entities covered by the survey are enterprises employing more than 9 persons and conducting economic activity in the 14 sectors indicated in the program of statistical surveys. They are obliged to provide to CSO their reports on working conditions, which are defined as a group of factors in the work environment. Information collected in the survey include, among others:

- number of persons employed, exposed to harmful factors whose concentration or intensity exceeds MAC (the maximum allowed concentration) or/and MAI (the maximum allowed intensity);
- number of persons employed exposed to strenuous factors (e.g. an awkward working posture or excessive physical exertion);
- number of persons employed exposed to mechanical factors, associated with particularly dangerous machinery (listed in legal regulations).

An important source of information on working conditions is also the ad-hoc LFS module survey 'Accidents at work and work-related health problems', conducted by the CSO according to the European Commission Regulation.

The information on working conditions in Poland, as in other EU countries, is also provided by European Working Conditions Survey, conducted by European Foundation for the Improvement of Living and Working Conditions (Eurofound) and by European Survey of Enterprises on New and Emerging Risks (ESENER), conducted by European Agency for Safety and Health at Work (EU-OSHA).

Although all these surveys relate to working conditions, their objectives are formulated in different ways (Table 1).

Table 1. Aims and periodicity of the surveys providing information on working conditions in Poland

Name of the survey	Institution	Aim (in relation to working conditions)	Periodicity
The survey on working conditions (SWC)	Central Statistical Office	Assessment and monitoring workplace hazards in work environment (physical, chemical, mechanical) and hazards related to strenuousness of work	Yearly
ad-hoc LFS module survey "Accidents at work and work-related health problems"	Central Statistical Office (conducted as a part of the Programme of ad hoc modules for the Labour Force Survey)	Determination of the scope of occurrence of factors at work that adversely affect physical health or mental well-being	Every 4 years
European Working Conditions Survey	European Foundation for the Improvement of Living and Working Conditions	Assessing and quantifying working conditions of both employees and the self-employed across Europe on a harmonised basis. Analysing relationships between different aspects of working conditions Identifying groups at risk and issues of concern as well as of progress	Every 5 years
European Survey of Enterprises on New and Emerging Risks (ESENER)	European Agency for Safety and Health at Work	Providing information on how workplace risks, and especially new and emerging risks, are being managed across Europe	Every 5 years

The basic characteristics of these surveys (including their methodology and informational potential) are presented in the next subsections.

5.2 General characteristics of the methodologies of the surveys providing information on working conditions in Poland

General characteristics of the methodologies of the surveys providing information on working conditions in Poland, including mode of data collection, topics covered and target population are presented in the Table 2.

Table 2. Mode of data collection, topics covered and target population in the surveys providing information on working conditions in Poland

Name of the survey	Mode of data collection	Topics covered	Target population
The survey on working conditions (SWC)	Companies' reports, prepared on the basis of obligatory measurements of harmful factors in working environment	The number of persons employed in hazardous conditions, in which exposure exceeds MAC - the maximum allowed concentration and/or MAI - the maximum allowed intensity: In total by risk factors (including physical, chemical, mechanical factors and strenuousness of work)	Enterprises from 14 sections, with the employment of 10 and more persons
Ad-hoc LFS module survey "Accidents at work and work-related health problems"	Interviews carried out with the use of additional (to LFS) questionnaire	Factors at work that adversely affect mental wellbeing or physical health: physical factors (e.g. lighting, noise, micro-climate), chemical factors (e.g. toxic substances) and biological factors (e.g. bacteria), occurring within the area of the work place (e.g. factory room, work position) as well as within the area surrounding the establishment	The population of the currently employed persons
European Working Conditions Survey	Face-to-face interviews, at the expert's home; average duration of 45 minutes	The scope of the survey questionnaire has widened substantially since the first edition. Themes covered today include employment status, working time duration and organisation, work organisation, learning and training, physical and psychosocial risk factors, health and safety, work-life balance, worker participation, earnings and financial security, as well as work and health	All residents of the surveyed countries and in employment at the time of the survey (according to the ILO definition).
European Survey of Enterprises on New and Emerging Risks (ESENER)	Interviews with the person who knows best about the way safety and health risks are managed at their workplace in each establishment surveyed	General safety and health risks in the workplace and how they are managed Psychosocial risks, such as stress, bullying and harassment Drivers of and barriers to OSH management Worker participation in safety and health practices	All establishments that have five or more employees, covering all sectors of economic activity except for private households

In two of the presented surveys (SWC and ESENER), the source of data is information from enterprises, but it is collected in different ways. In the SWC, mandatory reports containing data from enterprises' registers are the source of information, while in the ESENER interviews with representatives of enterprises are performed. In two other studies (ad-hoc LFS module survey and EWCS) information from working people is the source of information.

5.3 Informative potential of the working conditions surveys

Working conditions refer to the working environment and aspects of an employee's terms and conditions of employment. This covers matters such as the organisation of work and work activities, training, skills and employability, health, safety and well-being, working time and work-life balance. The following can be taken into account when comparing the informative potential of the various surveys on working conditions:

- aspects of working conditions covered by the survey (comparison is presented on the Figure 15);
- information on workers covered by the survey, such as sex, age, education, employment status, occupation (comparison is presented on the Figure 16);
- information on size and sector of companies covered by the survey.

Figure 15. Aspects of working conditions covered by different surveys

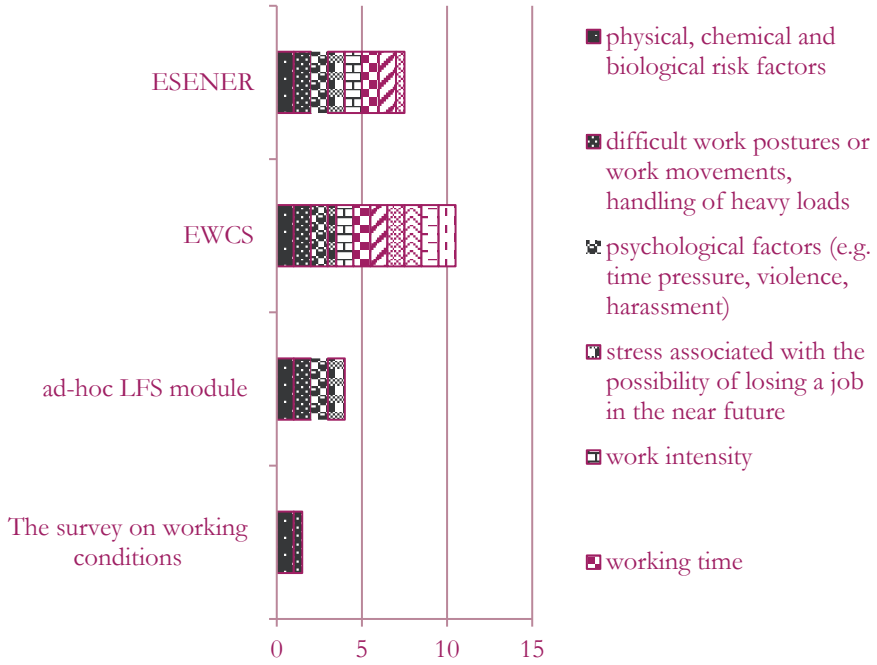
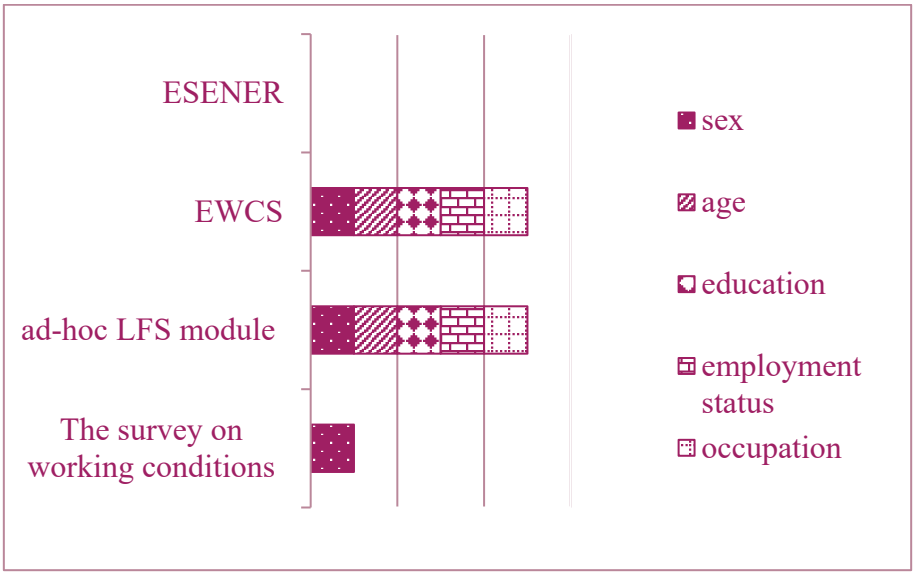


Figure 16. Information on workers covered by different surveys



The national survey of working conditions carried out every year covers such risk factors in the working environment, whose concentrations and intensities exceed the limit values established by

law. It does not provide information on other aspects of working conditions or on exposure to psychosocial factors and on factors occurring in the working environment in concentrations below the limit values. The survey also collects information on the activities and size of the undertaking and the gender of exposed persons. The strength of this survey is the quality of the data and the frequency with which they are collected. However, the scope of information on working conditions is limited.

The scope of the European survey of enterprises ESENER is broader as it covers all typical risk factors identified in enterprises, including psychosocial factors. Additionally, information on risk factors in small enterprises (employing more than 5 persons) is also available.

The scope of information on working conditions collected in the ad-hoc LFS module is broader than in WCS. In particular, it covers also psychosocial factors and selected factors in the working environment which workers consider to be harmful to their health, whether or not exposure exceeds the limit values. However, the number of factors covered by the survey is lower than the number of factors covered by the SWC survey. Information on sex, age, education, employment status and occupation of the surveyed persons is also collected.

The scope of information collected in EWCS survey is the widest. The survey covers a lot of aspects characterising working conditions (such as employment status, working time duration and organisation, work organisation, learning and training, physical and psychosocial risk factors, health and safety, work-life balance, worker participation, earnings and financial security, as well as work and health). Information on sex, age, education, employment status and occupation of the surveyed persons is also collected.

5.4 Basic issues related to the use of existing data sets on working conditions

Data on working conditions collected in various surveys provide information on various aspects of working conditions in Poland. The use of data collected in these surveys depends on the needs of users, who, however, must remember that data from different databases cannot be compared. Still a problem is the analysis of working conditions in enterprises employing less than 5 people, as well as working conditions of individual farmers. It is also not possible to determine on the basis of collected data the working conditions of some particularly sensitive groups (e.g. migrant workers). However, it is noteworthy to point out that surveys of working conditions are adapted to the needs of the changing world of work and cover more and more issues.

6. Discussion and conclusions

The aim of this report was to map the gaps in data sets and research infrastructures in Central and Eastern Europe and to explore the main obstacles and challenges which underpin these gaps. In order to map the gaps we organised a roundtable discussion with 40 regional experts and researchers to inform the research design of the InGRID-2 CEE expert questionnaire. Two case studies with focus on gaps in data on vulnerable groups in Hungary and working conditions in Poland were presented as well. Based on the expert input we categorised the focus of the questionnaire on four types of data: (i) international and national survey data, (ii) data provided within the InGRID project (iii) administrative data, and (iv) data on vulnerable groups. The results of the expert questionnaire along with the responses from the InGRID-2 user questionnaire offer some interesting findings and exploratory mapping of the issue of infrastructure gaps in Central and Eastern Europe. However the aim is not statistical inference or comparative conclusions.

We find that experts are generally able to access the databases, however the majority of them are rather dissatisfied with the available databases. Among the main issues are: missing data and topics, sample size or unit of analysis, coherence, timeliness and quality of data. Country coverage does not seem to be an issue. The expert opinions suggest that work experience could play a role, with the more experienced and older experts able to access the databases better and being more satisfied. Moreover experts working at a research institute or governmental organisation seem to be more satisfied and better able to access the data than experts employed at non-governmental or civil society organisations.

When talking specifically about the data sets on poverty, living conditions, working conditions and vulnerability, the experts gave quite disfavourable scores in the questionnaire. The most serious obstacles mentioned were the lack of integration of such data sets, difficulties to link data of multiple sources and lack of comparability across time, the lack of principles and regulatory mechanism, the lack of coordination and the lack of finances used for collection of comprehensive data sets. Communication issues, limited cooperation between researchers and policymakers, issues with presentation of research outcomes were mentioned as some challenges preventing the overcoming of gaps.

appendix 1 **Survey CELSI**

Expert Questionnaire Data Gaps in CEE

The information will be processed by CELSI for research purposes strictly. In particular, it will not be disclosed to third parties, respecting anonymity of the respondents.

The survey is conducted within InGRID-2 research project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 730998. For more information please visit the project's website: http://www.inclusivegrowth.eu/ .

Please let us know about yourself:

* 1. In which country do you work?

- ◯ Bulgaria
- ◯ Czechia
- ◯ Estonia
- ◯ Hungary
- ◯ Latvia
- ◯ Lithuania
- ◯ Poland
- ◯ Romania
- ◯ Slovakia
- ◯ other

(please specify)

* 2. Which of the categories below best describes your current or last primary affiliation?

- ◯ government, governmental/municipal organization, or public organziation
- ◯ university or higher-education organization
- ◯ research institute or think tank
- ◯ non-governmental or civil society organization
- ◯ business company or organization
- ◯ other

(please specify)

* 3. Please choose which of the following best describes your professional status:

- ◯ student
- ◯ postdoc
- ◯ research assistant
- ◯ junior researcher (less than 10 years of experience)
- ◯ senior researcher (10 or more years of experience)
- ◯ assistant professor
- ◯ associate professor
- ◯ professor
- ◯ director or manager
- ◯ other

(please specify)

* 4. What is your main area of expertise?
Choose one of the following answers.

- ◯ poverty, living conditions
- ◯ social policy, welfare states, social services
- ◯ inequality, social in/exclusion
- ◯ labour market, vulnerability, precariousness
- ◯ working conditions, employment conditions, personnel management, HRM
- ◯ industrial relations, collective bargaining, representation
- ◯ other

(please specify)

* 5. How many years of experience do you have in your main area of expertise?

- ◯ 0-5 years
- ◯ 6-10 years
- ◯ 11-15 years
- ◯ 16-20 years
- ◯ More than 20 years

* 6. What is your main scientific field?

Choose one of the following answers.

- ◯ Economics and business
- ◯ Sociology
- ◯ Political science
- ◯ Educational science
- ◯ Psychology
- ◯ Law
- ◯ Demography
- ◯ Mathematics
- ◯ Statistics
- ◯ Computer sciences / ICT
- ◯ Health sciences
- ◯ (Research) methodology
- ◯ Not applicable
- ◯ Other

(please specify)

* 7. What is your age?

Choose one of the following answers.

- ◯ 18-34 years old
- ◯ 35-49 years old
- ◯ 50-64 years old
- ◯ 65 years old or older

* 8. What is your gender?

- ◯ female
- ◯ male
- ◯ gender-variant/non-conforming
- ◯ prefer not to answer
- ◯ other

(please specify if you pefer)

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

General questions

* 9. Are you generally able to access the datasets you require for your research work?

- ◯ Yes
- ◯ No
- ◯ Don't know

* 10. How satisfied are you in general with the datasets in your research area(s)?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
◯	◯	◯	◯	◯

11. Is there a specific research topic in your field of expertise that is currently not covered by any dataset but should be researched? Please explain below.

Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

International and national data

Dataset 1

* 12. During the last five years, have you made use of any of the following datasets?

Please select the **one**, which you have used most frequently. You will be able to select two more of the datasets listed above further in the questionnaire.

If you have **not** used any of the listed datasets, please select "None of the datasets listed above".

- ◯ EU Labour Force Survey (LFS)
- ◯ EU SILC Structure of Income and Living Conditions (SILC)
- ◯ Structure of Earnings Survey (SES)
- ◯ Eurobarometer
- ◯ European (World) Values Survey
- ◯ European Quality of Life Survey (EQLS)
- ◯ European Working Conditions Survey (EWCS)
- ◯ European Company Survey (ECS)
- ◯ None of the datasets listed above

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

International and national data

Dataset 1

* 13. How satisfied are you with this dataset?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
◯	◯	◯	◯	◯

* 14. Please rate the significance of the following gaps in this dataset

Please rank all the options

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

International and national data

Dataset 2

* 15. Have you made use of any other of the following datasets during the last five years?

If so, please select the **one**, which you have used next most frequently (after those you have indicated on the previous pages).

If **not**, please select "None of the datasets listed above".

- ◯ EU Labour Force Survey (LFS)
- ◯ EU SILC Structure of Income and Living Conditions (SILC)
- ◯ Structure of Earnings Survey (SES)
- ◯ Eurobarometer
- ◯ European (World) Values Survey
- ◯ European Quality of Life Survey (EQLS)
- ◯ European Working Conditions Survey (EWCS)
- ◯ European Company Survey (ECS)
- ◯ None of the datasets listed above

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

International and national data

Dataset 2

* 16. How satisfied are you with this dataset?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
◯	◯	◯	◯	◯

* 17. Please rate the significance of the following gaps in this dataset

Please rank all the options

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

International and national data

Dataset 3

* 18. Have you made use of any other of the following datasets during the last five years?

If so, please select the **one**, which you have used next most frequently (after those you have indicated on the previous pages).

If **not**, please select "None of the datasets listed above".

- ◯ EU Labour Force Survey (LFS)
- ◯ EU SILC Structure of Income and Living Conditions (SILC)
- ◯ Structure of Earnings Survey (SES)
- ◯ Eurobarometer
- ◯ European (World) Values Survey
- ◯ European Quality of Life Survey (EQLS)
- ◯ European Working Conditions Survey (EWCS)
- ◯ European Company Survey (ECS)
- ◯ None of the datasets listed above

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

International and national data

Dataset 3

* 19. How satisfied are you with this dataset?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
○	○	○	○	○

* 20. Please rate the significance of the following gaps in this dataset

Please rank all the options

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	○	○	○	○	○	○
Sample attrition	○	○	○	○	○	○
Country of interest not covered	○	○	○	○	○	○
Adequacy for cross-country comparative analysis	○	○	○	○	○	○
Adequacy for cross-region comparative analysis	○	○	○	○	○	○
Adequacy for intertemporal comparative analysis	○	○	○	○	○	○
Timing of data release/publication	○	○	○	○	○	○
Length of time series	○	○	○	○	○	○
Lack of dataset information/codebooks	○	○	○	○	○	○

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Ingrid data

Dataset 1

* 21. During the last five years, have you made use of any of the dataset/tools that are included in the InGrid-1 or InGrid-2 EU Project?

Please select the **one**, which you have used most frequently. You will be able to select two more of the datasets listed above further in the questionnaire.

If you have **not** used any of the listed datasets, please select "None of the datasets listed above"

- ◯ EUROMOD
- ◯ LUXEMBURG INCOME STUDY DATABASE (LIS)
- ◯ LUXEMBOURG WEALTH STUDY DATABASE (LWS)
- ◯ WageIndicator
- ◯ Social Monitor (Tarki)
- ◯ Social Assistance Explorer
- ◯ Comparative Social Policy databases SOFI - University of Stockholm
- ◯ CSB-MIPI: Minimum Income Protection (University of Antwerp)
- ◯ ICTWSS (AIAS – University of Amsterdam): institutional characteristics of social dialogue
- ◯ IPOLIS (Integrated Poverty and Living Conditions Indicator System, Tarki)
- ◯ None of the datasets listed above

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Ingrid data

Dataset 1

* 22. How satisfied are you with this dataset?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
◯	◯	◯	◯	◯

* 23. Please rate the significance of the following gaps in this Ingrid dataset

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Ingrid data

Dataset 2

* 24. During the last five years, have you made use of any other of the dataset/tools that are included in the InGrid-1 or InGrid-2 EU Project?

Please select the **one**, which you have used next most frequently. (after those you have indicated on the previous pages)

If you have **not** used any of the listed datasets, please select "None of the datasets listed above"

- ◯ EUROMOD
- ◯ LUXEMBURG INCOME STUDY DATABASE (LIS)
- ◯ LUXEMBOURG WEALTH STUDY DATABASE (LWS)
- ◯ WageIndicator
- ◯ Social Monitor (Tarki)
- ◯ Social Assistance Explorer
- ◯ Comparative Social Policy databases SOFI - University of Stockholm
- ◯ CSB-MIPI: Minimum Income Protection (University of Antwerp)
- ◯ ICTWSS (AIAS – University of Amsterdam): institutional characteristics of social dialogue
- ◯ IPOLIS (Integrated Poverty and Living Conditions Indicator System, Tarki)
- ◯ None of the datasets listed above

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Ingrid data

Dataset 2

* 25. How satisfied are you with this dataset?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
◯	◯	◯	◯	◯

* 26. Please rate the significance of the following gaps in this Ingrid dataset

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Ingrid data

Dataset 3

* 27. During the last five years, have you made use of any other of the dataset/tools that are included in the InGrid-1 or InGrid-2 EU Project?

Please select the **one**, which you have used next most frequently. (after those you have indicated on the previous pages)

If you have **not** used any of the listed datasets, please select "None of the datasets listed above"

- ◯ EUROMOD
- ◯ LUXEMBURG INCOME STUDY DATABASE (LIS)
- ◯ LUXEMBOURG WEALTH STUDY DATABASE (LWS)
- ◯ WageIndicator
- ◯ Social Monitor (Tarki)
- ◯ Social Assistance Explorer
- ◯ Comparative Social Policy databases SOFI - University of Stockholm
- ◯ CSB-MIPI: Minimum Income Protection (University of Antwerp)
- ◯ ICTWSS (AIAS – University of Amsterdam): institutional characteristics of social dialogue
- ◯ IPOLIS (Integrated Poverty and Living Conditions Indicator System, Tarki)
- ◯ None of the datasets listed above

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Ingrid data

Dataset 3

* 28. How satisfied are you with this dataset?

Very satisfied	Satisfied	Neither satisfied, nor dissatisfied	Dissatisfied	Very dissatisfied
○	○	○	○	○

* 29. Please rate the significance of the following gaps in this administrative dataset

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	○	○	○	○	○	○
Sample attrition	○	○	○	○	○	○
Country of interest not covered	○	○	○	○	○	○
Adequacy for cross-country comparative analysis	○	○	○	○	○	○
Adequacy for cross-region comparative analysis	○	○	○	○	○	○
Adequacy for intertemporal comparative analysis	○	○	○	○	○	○
Timing of data release/publication	○	○	○	○	○	○
Length of time series	○	○	○	○	○	○
Lack of dataset information/codebooks	○	○	○	○	○	○

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Administrative data 1

* 30. During the last five years, have you made use of any administrative data for research or analytical work? (e.g. records from the social insurance company, tax authority, health insurance company, or similar)

- ◯ Yes
- ◯ No
- ◯ Don't know

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Administrative data 1

* 31. If yes, please let us know which one you have used most frequently. You will be able to select two more data sets further in the questionnaire.

* 32. Is there a standardized procedure of obtaining this data? (e.g. publicly available policy of access, an access point, data service center)?

- ◯ Yes
- ◯ No
- ◯ Don't know

33. If yes, please explain

* 34. Is the data standardized in terms of the format or variables?

- ◯ Yes
- ◯ No
- ◯ Don't know

35. If yes, how is the data standardized? (multiple choice)

- ☐ Format
- ☐ Variables
- ☐ Source of the data
- ☐ Unit of observation
- ☐ Other

* 36. Is access to this dataset centralized at one organization (e.g. the statistical office), or is it decentralized (e.g. at different units of ministries or branches of national authorities)?

- ◯ Centralized
- ◯ Decentralized
- ◯ Some centralized, other decentralized

37. If centralized, what institution is responsible for the centralization?

* 38. Please rate the significance of the following gaps in this administrative dataset

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Administrative data 2

* 39. During the last five years, have you made use of any other administrative dataset?

- ◯ Yes
- ◯ No

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Administrative data 2

* 40. If yes, please let us know which is the other frequently used administrative dataset. (after those you have indicated on the previous pages)

* 41. Is there a standardized procedure of obtaining this data? (e.g. publicly available policy of access, an access point, data service center)?

- ◯ Yes
- ◯ No
- ◯ Don't know

42. If yes, please explain

* 43. Is the data standardized in terms of the format or variables?

- ◯ Yes
- ◯ No
- ◯ Don't know

44. If yes, how is the data standardized? (multiple choice)

- ☐ Format
- ☐ Variables
- ☐ Source of the data
- ☐ Unit of observation
- ☐ Other

* 45. Is access to this dataset centralized at one organization (e.g. the statistical office), or is it decentralized (e.g. at different units of ministries or branches of national authorities)?

- ◯ Centralized
- ◯ Decentralized
- ◯ Some centralized, other decentralized

46. If centralized, what institution is responsible for the centralization?

* 47. Please rate the significance of the following gaps in this administrative dataset

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Administrative data 3

* 48. During the last five years, have you made use of any other administrative dataset?

- ◯ Yes
- ◯ No

InGRID
Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

Administrative data 3

* 49. If yes, please let us know which is the other frequently used administrative dataset. (after those you have indicated on the previous pages)

* 50. Is there a standardized procedure of obtaining this data? (e.g. publicly available policy of access, an access point, data service center)?

- ◯ Yes
- ◯ No
- ◯ Don't know

51. If yes, please explain

* 52. Is the data standardized in terms of the format or variables?

- ◯ Yes
- ◯ No
- ◯ Don't know

53. If yes, how is the data standardized? (multiple choice)

- ☐ Format
- ☐ Variables
- ☐ Source of the data
- ☐ Unit of observation
- ☐ Other

* 54. Is access to this dataset centralized at one organization (e.g. the statistical office), or is it decentralized (e.g. at different units of ministries or branches of national authorities)?

- ◯ Centralized
- ◯ Decentralized
- ◯ Some centralized, other decentralized

55. If centralized, what institution is responsible for the centralization?

* 56. Please rate the significance of the following gaps in this administrative dataset

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Missing variables or topics	◯	◯	◯	◯	◯	◯
Sample attrition	◯	◯	◯	◯	◯	◯
Country of interest not covered	◯	◯	◯	◯	◯	◯
Adequacy for cross-country comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for cross-region comparative analysis	◯	◯	◯	◯	◯	◯
Adequacy for intertemporal comparative analysis	◯	◯	◯	◯	◯	◯
Timing of data release/publication	◯	◯	◯	◯	◯	◯
Length of time series	◯	◯	◯	◯	◯	◯
Lack of dataset information/codebooks	◯	◯	◯	◯	◯	◯

Expert Questionnaire Data Gaps in CEE

Data on poverty and vulnerable groups

* 57. Below is a list of possible obstacles concerning the collection and maintenance of datasets on poverty, living conditions, working conditions and vulnerability. Please rate how significant they are in your country from your own experience.

	Very significant	Significant	Neither significant, nor insignificant	Rather not significant	Insignificant	Don't know
Lack of finances to collect comprehensive data	○	○	○	○	○	○
Lack of finances to maintain available datasets	○	○	○	○	○	○
Lack of clear procedures (lack of transparent engagement process) to request/access the datasets, e.g. administrative/government microdata	○	○	○	○	○	○
Lack of legislation to secure the confidentiality of sensitive data (e.g. on health)	○	○	○	○	○	○
Too much security-related issues with regard to data protection	○	○	○	○	○	○
Lack of principles and regulatory mechanisms governing the establishment and use of large-scale, linked datasets	○	○	○	○	○	○
Lack of coordination of what is collected/available and what is missing	○	○	○	○	○	○
The quality of data in terms of their reliability	○	○	○	○	○	○
Lack of integration of the datasets – lack of multiple government agency datasets	○	○	○	○	○	○

58. Are there any difficult-to-reach groups for which there are no data available to support research about them?

59. Regarding data gaps in the areas of poverty, living conditions, working conditions and vulnerability, are there any other remarks/comments/suggestions that you would like to share?

InGRID

Supporting expertise in inclusive growth

Expert Questionnaire Data Gaps in CEE

60. Please enter your contact details below if you would like to be involved in subsequent discussions, such as workshops, conferences, seminars etc, based on the findings of this research

Email Address

Phone Number

61. In case you would like to share a more detailed experience in a short semi-structured online interview, please indicate your choice below and we will contact you.

- ◯ Yes
- ◯ No

Thank you for your input!

References

CELSI, (2019). Round table on 'Main gaps in research infrastructures in Central-Eastern Europe', CELSI, 7-8 March 2019. More information http://www.inclusivegrowth.eu/round-tables/round-table-on-main-gaps-in-research-infrastructures-in-central-eastern-europe-celsi-7-8-march-2019

Dimoulas, C., Belegri-Roboli, A., Michaelides, P. G., Papadopoulou, D., Karlaganis, P., & Konstantantakis, K.N. (2019). GAP ANALYSIS ON THE USE OF RESEARCH INFRASTRUCTURES IN SOUTHEAST EUROPEAN (BALKAN) COUNTRIES, Leuven, H2020 InGRID-2 project.

Hamon-Cholet, S., Salibekyan, Z., Van Gyes, G., & Mottier, L. (2017). InGRID and the Central and Eastern European countries: scoping the challenges. Futuring briefing note 3 (Deliverable 2.3). Leuven: InGRID.


CELSI