RAISE

Inventory of data sources for analysing intersectional inequalities quantitatively

Deliverable 3.1 (WP3) – July 2024

Gábor Hajdu, Vera Messing

Central European University

Disclaimer:

Dissemination Level: Public

Contents

**Inventory of data sources
for analysing intersectional inequalities quantitatively**

1. Introduction

The RAISE project focuses on structural racism stemming from ethnic, racial and religious identities. Its aim is to contribute to the understanding of processes that result in structural racism and challenge these by providing agency to people with a minority background and to contribute to raising awareness of the problem. For this purpose, the consortium includes research and awareness raising elements and partners with both academic research institutions and NGOs. WP3 is a workpackage that aims to contribute to building more knowledge on the subject matter.

This paper - part of Work Package 3 (WP3) - aims to show the role of intersectional identities in structural inequalities on the labour market. The objective of this paper is to provide an overview of the quantitative data sources that are suitable for studying the intersections of racial, ethnic, religious and gender identities and inequalities in the labour market comparatively across countries as well as nationally. This overview constitutes an essential input for the subsequent work in WP3 that aims to analyse intersectional identities and their roles in the (re)production of inequalities related to employment and the wider labour market and consequently to social status.

First and foremost, we list datasets that are suitable for comparative, cross-national analysis. However, given the limitations of these datasets, we also list national datasets that can provide a deeper understanding of the relationship between labour market inequalities and racial, ethnic, religious and gender identities in the national contexts.

The next section presents the selection criteria for identifying data sources, and the following section lists the limitations and challenges of quantitative analysis of racial, ethnic, religious and gender identities and labour market inequalities. Section 4 lists the data sources, first cross-national surveys and then national surveys. Section 5 presents the research strategy for the subsequent quantitative analysis in WP3.

2. Selection criteria

To identify suitable data sources, we searched the internet, databases and data repositories of surveys, and sought information from our project partners. When identifying suitable data sources, we had six criteria that the datasets had to meet.

1. Geographical coverage

As a basic requirement, the database must include Hungary, Poland, Belgium, the Netherlands, Germany and Turkey as these are the countries are partners in the RAISE project. In addition, we aimed to find datasets covering more European countries.

2. Indicators

Since the aim of the analysis is to assess the role that intersecting identities (race, ethnicity, religion, gender) play in the formation and operation of labour market inequalities, we need datasets that include as many of these variables as possible (preferably all of them).

According to the project proposal, our outcome variables to measure labour market inequalities are unemployment and social status.

In addition to these variables, we are considering analysing how other possible identities (e.g. age, parenthood) are related to differences in unemployment and social status.

3. Sample

Although the focus of the proposed analysis is on labour market disadvantage based on ethnic, racial, religious and immigrant background, it is important to emphasise that appropriate datasets should not only include minorities, immigrants or people of a particular religious denomination, as labour market inequalities are relative and can therefore be evaluated relative to the general population. This means that we need datasets which's target groups include the general population as well as immigrants, descendants of immigrants and ethnic minorities. In other words, surveys on members of ethnic, religious and other minorities may be an important and useful source of information on individual experiences of labour market discrimination, but they are less appropriate for the specific focus of this analysis.

4. Number of observations (sample size)

The size of the dataset is an important selection criterion when examining the intersections of racial, ethnic, religious and gender identities and inequalities in the labour market. A relatively large number of observations ensures that minority groups, including smaller minority groups, are adequately represented, allowing for more reliable analyses. This is particularly important in countries where the proportion of ethnic, racial or religious minorities is small, as a smaller dataset may not contain enough observations from these groups to draw meaningful conclusions. Therefore, datasets such as larger labour force surveys, large longitudinal studies or large cross-national surveys are preferred as they can provide a robust sample size.

5. Time coverage

A relevant, but not necessarily the most important, factor in the selection of datasets is time coverage. Ideally, recent surveys are preferred, but relatively older surveys are acceptable if all other criteria are met.

6. Availability

Finally, the dataset must be freely available for scientific purposes, accessible online and accompanied by documentation in English.

3. Limitations and challenges

There are a number of limitations that need to be taken into account when analysing the relationship between racial, ethnic, religious and gender identities and labour market inequalities, and when listing the quantitative data sources that are appropriate for this analysis.

Ethnicity refers to a shared cultural heritage, ancestry, nationality or historical experience, often including a shared language and traditions. While ethnicity is more about objective cultural markers, ethnic identity is subjective and reflects an individual's sense of belonging and identification within their ethnic group.[1] Most surveys include questions about country of birth, but often do not ask about ethnic identity, despite European Commission guidelines that stress the importance of asking people how they identify (self-identified ethnicity) and the importance of allowing people to indicate multiple ethnic/group affiliations (EC, 2021). Moreover, when considering racism and xenophobia, it is probably even more important to understand how people are perceived by those who discriminate: "To measure (in)equality, it may be as important to identify the perceived racial and ethnic origin as the one self-identified. (Perceived) racial origin is an important element of the discrimination experiences of Afro-Europeans and European Muslims, as well as the Roma." (Farkas, 2017, p. 6) However, it is almost impossible in surveys to collect information on the respondent's ethnicity as perceived by others. It is however, to a limited extent, possible to collect information in surveys based on which perceptions are formed. Such information includes language used in private contexts, membership of minority organizations or experience of being discriminated due to race, ethnicity.

A further challenge for analysing the relationship between labour market inequalities and intersecting identities is that surveys using the official or main (or perhaps second) language of the country de facto exclude parts of the population who tend to have limited knowledge of the main language spoken in the country (Font & Méndez, 2013a).[2]

Cross-national surveys usually do not include all the variables that are the focus of the project (racial, ethnic, religious and gender identities). This makes it difficult to examine the relationship between all identities - racial, ethnic, religious and gender - and labour market inequalities. It is possible that national surveys include all relevant variables, but in this case

---

[1] However, unlike other minorities, there are no objective criteria for determining Roma ethnicity, as Roma are a culturally heterogeneous population, even within a country and even more so between countries (Messing, 2014).

[2] For other potential methodological challenges, see Font and Méndez (2013b) or Sumption (2020).

the comparability of the analysis may be limited due to possible differences in questions, concepts and wording.

It is important to note that surveys are of limited use for measuring labour market inequalities, especially when it comes to precarious forms of employment since a significant part of labour market inequalities can be linked to informal employment or to characteristics of job quality that are not or rarely asked about in questionnaires. Preliminary results from the ethnographic approach of Work Package 1 (WP1) confirmed in local settings that boundary making processes related to intersecting identities are complex and difficult to measure in surveys. Therefore, qualitative analysis in WP3 (focus groups) will provide valuable insights for understanding the relationship between intersecting identities and labour market positions and inequalities in specific localities or related to specific social/ minority groups.

In terms of methodological challenges, recent initiatives to address intersecting identities from quantitative perspective have been critically addressed (Bauer, 2014; Bauer et al., 2021). A significant challenge in analysing the relationship between multiple identities and labour market inequalities is the additional complexity introduced by modelling and interpretation. The large number of potential interaction effects complicates the examination of the results and the sample size in most social strata becomes a significant issue (Evans et al., 2018). However, recent literature has proposed the use of multilevel regression applications for large numbers of intersections using individual-level data, which shows promise as a statistically efficient method (Evans et al., 2018, 2024).

4. List of data sources

This section presents a list of data sources that are most suitable for studying the intersections of racial, ethnic, religious and gender identities and inequalities in the labour market. We came to this selection after an in-depth mapping and inventory of available data sources in Europe, and identified their characteristics described in section 2 (selection criteria).

4.1. Cross-national surveys

4.1.1. European Social Survey

The European Social Survey (ESS) is a scientific cross-national survey that has been carried out since 2002. The survey is conducted every two years through face-to-face interviews. Due to COVID related restrictions of personal contacts, it has been permitted to conduct the survey using self-completion mode in some countries in its recent round (Round 10, 2021/22).

A major advantage of the ESS is that its core modules are asked in every round, but it also includes 2-3 rotating modules each round. This allows the survey to cover a wide range of topics over time. The most important questions on labour market status and work, country of

birth, and religion are included in the core modules. The rigorous methodology[3] and consistent data collection over time make the ESS a reliable source for studying how different social identities influence labour market inequalities in a European context.

The ESS includes specific questions on the country of birth of respondents. One question asks for the respondent's country of birth, and the country of birth of both the respondent's father and mother is asked if either was born in a country other than the country of the survey. The questionnaire also asks about the respondent's religion, whether the respondent belongs to a discriminated group, native language and whether the respondent is a citizen of the country.

Table 1: Information on the European Social Survey

URL	https://www.europeansocialsurvey.org/
Geographical coverage	European countries. 39 countries have participated in at least one round since 2002/03; 19 countries participated in at least 8 of the first 10 waves, 15 countries in all 10 waves.
Time coverage	10 waves between 2002 and 2022. The first edition of wave 11 was published in June 2024 for 13 countries.
Sample size	All countries must aim for a minimum "effective achieved sample size" of 1,500, or 800 in countries with an ESS population of less than 2 million after adjusting for design effects. The total sample size in the first 10 waves is
Labour force variables	Unemployment, several variables on social status (including household income, feeling about household income)
Identity variables	Country of birth, mother's and father's country of birth, citizenship, belonging to a discriminated group, religion, language used in home setting. No questions on ethnic background except for belonging to a minority ethnic group in the country.
Availability	Free of charge for non-commercial use

4.1.2. European Union Labour Force Survey (EU-LFS)

The EU-LFS is the largest household sample survey in Europe, providing quarterly and annual data on labour participation of people aged 15 and over, as well as information on the economically inactive. It is conducted by national statistical institutes across Europe and processed centrally by Eurostat, with the current coverage extending to 34 European countries.

---

[3] https://www.europeansocialsurvey.org/methodology/methodology-overview

The questionnaire includes questions on the country of birth of respondents and the country of birth of both parents. Based on these answers the dataset contains a derived variable on migration status (native-born, native-born with one parent born abroad, native-born with both parents born abroad, foreign-born). This makes the survey suitable for analysing the relationship between labour market inequalities and gender and immigrant background. However, data on ethnic, racial and religious identities are not collected.

In addition, an "ad hoc module" is included in the questionnaire each year. In 2021 (and in 2014) the topic of the ad hoc module was "Labour market situation of migrants and their immediate descendants", which provides additional information on the respondents' skills in the main language of the host country.

The main advantage of the EU-LFS is its large sample size, which results in a relatively large number of immigrants in the sample in most countries. Another important advantage of the survey is that it uses the same concepts and definitions, uses common international classifications (e.g., ISCO, ISCED), and records the same set of variables in each country, which ensures the comparability of the data across countries. The main disadvantage of the survey is the lack of questions on ethnic, racial and religious background. This makes the dataset a potentially useful but limited source of data for examining the intersections of racial, ethnic, religious and gender identities and inequalities in the labour market.

Table 2: Information on the EU Labour Force Survey

URL	https://ec.europa.eu/eurostat/web/lfs/overview
	https://www.gesis.org/en/missy/metadata/EU-LFS/
Geographical coverage	34 European countries
Time coverage	Quarterly and annual data from 1998. (The data collection covers the years from 1983 onwards.)
Sample size	The achieved sample is approximately 0.29 % of the total population.
Labour force variables	Unemployment (ILO employment/unemployment status), variables on social status (including income in decile, permanency of main job). The ad hoc module of "Labour market situation of migrants and their immediate descendants" includes a question on feeling discriminated against at work in the current job.
Identity variables	Country of birth, citizenship, country of birth of both the father and the mother. The ad hoc module of "Labour market situation of migrants and their immediate descendants" includes a question on skills in the main host country language. No questions on ethnic background except for belonging to a minority ethnic group in the country. No questions on religion.
Availability	Access to microdata is granted for scientific purposes only. Access can be requested from Eurostat.

4.1.3. European Union (EU) Statistics on Income and Living Conditions (EU-SILC)

Similar in many aspects to the EU-LFS, the EU-SILC is an individual and household data collection based on nationally representative probability samples. The sample sizes are smaller than those of the EU-LFS. The questionnaire includes questions on the respondents' country of birth and the country of birth of both parents, but does not collect data on ethnic, racial or religious identity. Access to the microdata is identical to that of the EU-LFS.

The EU-LFS is therefore the preferred data source.

Table 3: Information on the EU-SILC

URL	https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions/
Geographical coverage	32 European countries
Time coverage	Annual data from 2003
Sample size	The minimum effective sample sizes ranges from 3750 to 14500.[4]
Labour force variables	Unemployment (self-defined current economic status), variables on social status (including income, permanency of main job).
Identity variables	Country of birth, citizenship, country of birth of both the father and the mother. No questions on ethnic background and religion.
Availability	Access to microdata is granted for scientific purposes only. Access can be requested from Eurostat.

4.1.4. Eurobarometer surveys

The Eurobarometer is a series of surveys on a wide range of topics conducted regularly on behalf of the European Commission and the European Parliament.

One of the special topics of the survey series is cultural identity, more specifically ethnic groups, minorities and discrimination. In recent years, in 2019 (EB 91.4), 2018 (EB 90.4), 2015 (EB 83.5), 2012 (EB 77.4), and 2009 (EB 71.2), surveys have been conducted that include specific questions on belonging to a minority group and on religion. These surveys are therefore suitable for analysing the relationship between racial, ethnic, religious and gender identities and labour market inequalities.

These surveys ask respondents whether they consider themselves to be a member of a minority group, including ethnic minorities and religious minorities, among other categories. They also ask respondents to identify their religious affiliation from a list, including options such as various Christian and Muslim denominations. Respondents are also asked whether they have personally felt discriminated against on a number of grounds and are also asked about the circumstances (e.g., when looking for a job or at work).

The regular sample size is n=1000 respondents per country (except for small countries, e.g., Malta, Luxembourg), which is smaller than for the other surveys mentioned above. In order to achieve a sufficient number of respondents with a minority background, it is necessary to

---

[4] https://ec.europa.eu/eurostat/statistics-explained/SEPDF/cache/39791.pdf

pool a few rounds of Eurobarometer surveys, although this results in a relatively long time frame.

Table 4: Information on the Eurobarometer surveys

URL	https://europa.eu/eurobarometer/about/eurobarometer https://www.gesis.org/en/eurobarometer-data-service
Geographical coverage	EU member countries. Accession and Candidate Countries are included intermittently.
Time coverage	Special topic on ethnic groups / minorities: 2019 (EB 91.4), 2018 (EB 90.4), 2015 (EB 83.5), 2012 (EB 77.4), 2009 (EB 71.2)
Sample size	The regular sample size is n=1000 respondents per country, except small countries
Labour force variables	Unemployment (self-defined current economic status), variables on social status (including subjective social class, subjective financial situation).
Identity variables	Being member of a minority group (including an ethnic minority, a religious minority). Nationality. Religion (including different Christian and Muslim denominations). No questions on country of birth (in recent years).
Availability	The dataset, questionnaires, codebook and other documents can be downloaded from the GESIS Data Archive.

4.1.5. Second European Union Minorities and Discrimination Survey (EU-MIDIS II)

The aim of EU-MIDIS II is to collect comparable data in all EU Member States on people with a migrant or ethnic minority background, including Roma, in order to help EU institutions and policymakers develop evidence-based policies to respect, protect and fulfil their fundamental rights (FRA, 2017).

The survey interviewed 25,500 respondents from different ethnic minority and migrant backgrounds, including Roma, in all EU Member States in 2015-2016. The minimum sample size was 500 interviews per country, with a minimum of 400 interviews per target group. The number of target groups is one or two per country (except in Italy, where there are three). The survey aimed to obtain a random probability sample of the selected target groups that would be representative of the target populations with a migrant or ethnic background in each country, but the lack of reliable strategies for obtaining probability samples for hard-to-reach populations, such as immigrants and ethnic minorities, makes the representativeness of the samples questionable (FRA, 2017).

The survey covers a range of issues relating to discrimination on a number of grounds, including colour, ethnic origin and religion. The survey also collected a range of demographic characteristics and information on labour market attachment (including experiences of discrimination at work).

The survey can be a useful source of data for analysing the relationship between labour market attachment, experiences of discrimination and racial, ethnic, religious and gender identities, as it covers a relatively large number of respondents from different ethnic minority and migrant backgrounds that are relevant in each country. However, it also has some important limitations. First, it covers only one or two target groups per country (e.g., immigrants and descendants of immigrants from Turkey and Sub-Saharan Africa in Germany, Roma in Hungary, recent immigrants in Poland). Secondly, different target groups are covered in different countries, making it difficult (or even impossible) to compare data across countries. The heterogeneity of the sampling methods used in the countries makes the data incomparable between countries.

In conclusion, the survey can be used for descriptive analysis of the labour market attachment of people with different racial, ethnic, religious and gender identities, and as benchmark data for the general population, indicators from Eurostat (or other data sources) can be used, but it is less suitable for detailed analysis using several control variables or for more sophisticated comparisons.

Table 5: Information on the EU Minorities and Discrimination Survey

URL	https://search.gesis.org/research_data/ZA6703
Geographical coverage	27 European countries (EU member states)
Time coverage	2015-2016 (varies by country)
Sample size	25,515 respondents from different ethnic minority and migrant backgrounds, including Roma, in all EU Member States.
Labour force variables	Unemployment (employment/unemployment status), experience of discrimination at work, variables on social status (including income).
Identity variables	Country of birth, citizenship, country of birth of both the father and the mother. Religion, use of traditional or religious clothes different from those typically worn in country, extent of religious self-identification.
Availability	Registered users can download the dataset, questionnaires, codebook, and other documents from the GESIS Data Archive.

The primary aim of the quantitative analysis in WP3 is to understand the relationship between labour market outcomes and intersecting identities through a comparative analysis. However, as national surveys can provide larger samples, a wider range of labour market variables and variables on ethnic minority background, immigrant background and religion that are more relevant in the specific national context, we list national datasets with a brief description (from the countries participating in the RAISE project) that are suitable for the analysis. We arrived to this selection after consulting our project partners and the ETHMIGSURVEYDATA project outputs.

4.2.1. TÁRKI Household Monitor – Hungary

The TÁRKI Household Monitor is one of the largest Hungarian household surveys. It focuses on labour market attachment, income, economic expectations and economic behaviour. The multi-stage, proportionally stratified national probability sample is representative of the Hungarian adult population.

A major advantage of the survey is that it includes a rich set of questions on labour market activity and income, as well as questions on ethnic identity, religion and attendance at religious services. It does not include questions on country of birth and immigrant status, but these are less relevant in the Hungarian context. Importantly, the questionnaire includes a question asking the interviewer whether he/she thinks the respondent is of Roma origin, which provides a proxy variable for how the respondents are perceived by others.

The relatively large sample size ensures a sufficient number of respondents with an ethnic minority background.

Table 6: Information on the TÁRKI Household Monitor

URL	https://www.tarki.hu/eng/haztartas-monitor
Geographical coverage	Hungary
Time coverage	1998-2015 (11 waves)
Sample size	5053 respondents in the last wave
Labour force variables	Unemployment (employment/unemployment status), rich set of other variables on labour market activity, variables on social status (including income, subjective social class, subjective financial situation).
Identity variables	Ethnic identity (self-assessed), religion. Roma origin as perceived by others.
Availability	The dataset is available from the TÁRKI Data Archive

4.2.2. Longitudinal Internet studies for the Social Sciences (LISS) – The Netherlands

The LISS panel is a representative panel survey of the Dutch population. It consists of 5000 households and about 7500 individuals aged 16 and over. The LISS Core Study is an annual longitudinal survey consisting of several questionnaires covering a wide range of topics. (There are also externally funded surveys and experiments that can be merged with the Cory Study.)

A major advantage of the dataset is that the questionnaires include a rich set of relevant questions on ethnic origin, ethnic identity, religion and self-identified gender.

The project included a special immigrant panel between 2010 and 2014 (n=2499 participating panel members). This sample included the four major non-Western immigrant groups and two major Western immigrant groups in the Netherlands, as well as other people of Western European, Western non-European and non-Western origin.

Table 7: Information on the LISS panel

URL	https://www.lissdata.nl/ https://www.dataarchive.lissdata.nl/study-units/view/1
Geographical coverage	The Netherlands
Time coverage	Monthly surveys since 2007, yearly LISS Core Study
Sample size	Approximately 7500 individuals (from 5000 households), the average individual response rate varies between 71% and 80% since 2019.
Labour force variables	Unemployment (employment/unemployment status), rich set of other variables on labour market activity and work. Variables on social status (including income, subjective financial situation).
Identity variables	Country of birth, immigrant background, ethnic identity, language generally spoken at home and a wide range of other questions relevant to ethnic identity and immigrant background. Religion, self-identified gender.
Availability	The published data can be accessed free of charge for scientific research.

4.2.3. Dutch Parliamentary Election Study 2021 (DPES) and Dutch Ethnic Minority Election Study 2021 (DEMES)

Both the DPES and the DEMES consist of two waves: a pre-election survey and a post-election survey around the 2021 Dutch elections. The first wave of the DPES sample consisted of 2396

Dutch citizens aged 18 and over. The first wave of the DEMES sample consisted of 765 people aged 18 and over with Dutch citizenship and a first- or second-generation non-Western immigrant background. The questions in DEMES are based on the DPES questionnaire.

The questionnaires include questions on having an immigrant background and on religion (religious denomination and frequency of attending religious services). The DEMES questionnaire includes more detailed questions on migration history, parents' migration history and the country or region of origin of the partner.

The two datasets can be used together to analyse the intersections of racial, ethnic, religious and gender identities and inequalities in the labour market.

Table 8: Information on the DPES and the DEMES

URL	https://ssh.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/dans-xcy-ac9q https://ssh.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/dans-26b-xrqu
Geographical coverage	The Netherlands
Time coverage	2021
Sample size	2396 in DPES, 765 in DEMES
Labour force variables	Unemployment (self-defined labour force status). Variables on social status (including income, subjective social class).
Identity variables	Immigrant background, religion (religious denomination, attendance of religious services). Only in DEMES: country or region of origin.
Availability	The datasets are available on the DANS Data Station.

4.2.4. German Family Panel ('Panel Analysis of Intimate Relationships and Family Dynamics' – PAIRFAM)

The German Family Panel is a longitudinal study of partnership and family dynamics in Germany. Data are collected annually from a random sample of initially more than 12,000 individuals from three birth cohorts and their partners, parents and children. A fourth birth cohort was added to the sample in wave 11 as a replacement and refreshment sample. The sample size has varied between 5000 and 8800 in recent waves.

Although the main focus of the survey is on partnership and family dynamics, the questionnaires also include questions on country of birth, country of birth of partner and parents, citizenship, German language skills and religion.

Table 9: Information on the PAIRFAM survey

URL	https://www.pairfam.de/en/
	https://paneldata.org/pairfam/
Geographical coverage	Germany
Time coverage	14 waves, annually since 2008
Sample size	8800 individuals in wave 11 (2018/2019), 5000 individuals in wave 14 (2021/2022)
Labour force variables	Unemployment (employment status), rich set of other variables on labour market activity and work (including temporary employment contract, work schedule). Variables on social status (including income, subjective financial situation).
Identity variables	Country of birth, citizenship, country of birth of partner and parents, ethnicity and migration status (generated variable based on parents' country of birth), German language skills, religion.
Availability	The data is available free of charge for scientific research.

4.2.5. German Socio-Economic Panel (SOEP)

The German Socio-Economic Panel (SOEP) is a longitudinal survey launched in 1984 with a sample size of about 30,000 individuals from 15,000 households. It includes data from both West and East Germany, with an East German enlargement sample added after reunification in 1990. Other enlargement samples are migrant samples, typically integrated after periods of increased inflows (e.g., immigration of EU citizens from Central and Eastern Europe after EU accession, immigration of refugees from the Middle East in 2016-2017).

The survey covers a wide range of topics, including labour force status, income, education, health, ethnicity and religion.

An advantage of SOEP over PAIRFAM (described above) is that it is a representative sample of the general population, rather than a sample of four cohorts.

Table 10: Information on the SOEP

URL	https://www.diw.de/en/diw_01.c.678568.en/research_data_center_soep.html https://paneldata.org/soep-core/
Geographical coverage	Germany
Time coverage	Annually since 1984
Sample size	About 30,000 individuals (15,000 households)
Labour force variables	Unemployment (ILO definition, registered unemployed status), rich set of other variables on labour market activity and work. Variables on social status (including income, subjective financial situation).
Identity variables	Country of birth, citizenship, parents' country of origin, German language skills, religion.
Availability	The anonymous microdata can only be used for scientific purposes.

4.2.6. Polish Panel Survey (POLPAN)

The Polish Panel Survey (POLPAN) has been conducted every five years since 1988. It focuses on changes in the social structure of Poland over the last 30 years. The survey pays particular attention to the employment history of respondents, labour market inequality and how respondents adapt to changing living conditions in Poland.

The questionnaire includes questions on religion, occupational history and labour force status (including unemployment history). There is no question on ethnicity and country of birth, but as Poland is one of the most ethnically (and religiously) homogeneous countries in Europe, with more than 97% of the population identifying themselves as ethnically Polish, this may be of less concern.

Table 11: Information on the POLPAN survey

URL	https://polpan.org/en/
Geographical coverage	Poland
Time coverage	Every five years since 1988
Sample size	N=2161 in 2018, n = 2780 in 2013.
Labour force variables	Unemployment, rich set of other variables on labour market activity and work (including being employed by a temporary work agency, job loss likelihood). Variables on social status (including income, subjective financial situation of the household).
Identity variables	Religion (including religious affiliation, frequency of church attendance, religious commitment). No questions on ethnic background and country of birth.
Availability	The integrated data file (1988–2018) is available on the Harvard Dataverse.

4.2.7. Belgian electoral surveys

The RepResent Longitudinal Survey 2019-2021 and the PartiRep II Longitudinal Electoral Survey, 2014 are two longitudinal surveys conducted around the Belgian elections of 2014 and 2019.[5]

The target population of the RepResent Longitudinal Survey was the population eligible to vote in the elections of 26 May 2019. The aim was to obtain a net sample that matched the gender, age and education distribution of the electorate in their respective regions. The target population of the PartiRep II Longitudinal Electoral Survey was the population eligible to vote in the 25 May 2014 elections.

Both questionnaires include a question about the respondent's religion, but do not ask about country of birth or ethnicity. In the PartiRep II survey, respondents were asked about the language they usually speak with their family.

---

[5] Similar to the 2021 Dutch Ethnic Minority Election Study, the 2019 Belgian Ethnic Minorities Election Study (BEMES) could be used as a specific dataset for people with an immigrant background. However, unlike the Dutch representative survey, the BEMES focuses on Belgian citizens with a specific immigrant background (Turkish or Moroccan origin) living in two cities (Antwerp, Liège). (https://soc.kuleuven.be/ceso/ispo/projects/bemes)

Table 12: Information on the Belgian electoral surveys

URL	https://ssh.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/dans-xf5-djem https://ssh.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/dans-xwm-mq42
Geographical coverage	Belgium
Time coverage	RepResent: Four waves between 2019 and 2021; PartiRep II: two waves in 2014
Sample size	RepResent: n=7351 in wave 1, smaller sample sizes in later waves; PartiRep II: n=2019 in wave 1, n=1532 in wave 2.
Labour force variables	Unemployment. Subjective social status in the RepResent survey (satisfaction with household income).
Identity variables	Religion. Language spoken at home in the PartiRep II survey. No questions on ethnic origin and country of birth.
Availability	The datasets are available on the DANS Data Station.

5. Summary

The aim of this paper has been to provide an overview of the quantitative data sources that are suitable for studying the intersections of racial, ethnic, religious and gender identities and inequalities in the labour market. As the focus of the research is comparative, the cross-national datasets serve as the main data sources. As these datasets have their own limitations (e.g., no or no specific information on ethnic minority status, relatively small sample sizes), the national surveys listed in Section 4.2. can provide additional insights, although they are not directly comparable. It should be noted that identities (and related boundary making processes on the labour market) are complex and may be difficult to measure in surveys. Therefore, qualitative research element (focus group research) in WP3 will provide a further, deeper understanding of the relationship between intersecting identities and labour market inequalities in local settings and with reference to certain minority groups.

A major challenge for the subsequent quantitative analysis is that the large number of possible intersecting identities may raise issues of model parsimony, sample size, and interpretability. We plan to address these challenges by using a limited number of variables (e.g. immigrant status, religion and gender) or by using statistically efficient methods proposed in recent literature (Evans et al., 2018, 2024). However, a balance needs to be struck between interpretability and complexity.

References

Bauer, G. R. (2014). Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. *Social Science & Medicine*, *110*, 10–17. https://doi.org/10.1016/j.socscimed.2014.03.022

Bauer, G. R., Churchill, S. M., Mahendran, M., Walwyn, C., Lizotte, D., & Villa-Rueda, A. A. (2021). Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM - Population Health*, *14*, 100798. https://doi.org/10.1016/j.ssmph.2021.100798

EC. (2021). *Guidance note on the collection and use of equality data based on racial or ethnic origin*. European Commission.

Evans, C. R., Leckie, G., Subramanian, S. V., Bell, A., & Merlo, J. (2024). A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). *SSM - Population Health*, *26*, 101664. https://doi.org/10.1016/j.ssmph.2024.101664

Evans, C. R., Williams, D. R., Onnela, J.-P., & Subramanian, S. V. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Social Science & Medicine*, *203*, 64–73. https://doi.org/10.1016/j.socscimed.2017.11.011

EthmigSurveyData: The International Ethnic and Immigrant Minorities' Survey Data Network. https://ethmigsurveydatahub.eu/

Farkas, L. (2017). *Data collection in the field of ethnicity*. European Commission.

Font, J., & Méndez, M. (2013a). Introduction: The methodological challenges of surveying populations of immigrant origin. In J. Font & M. Méndez (Eds.), *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies* (pp. 11–41). Amsterdam University Press. https://library.oapen.org/handle/20.500.12657/33868

Font, J., & Méndez, M. (Eds.). (2013b). *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies*. Amsterdam University Press. https://doi.org/10.26530/OAPEN_450851

FRA. (2017). *Second European Union Minorities and Discrimination Survey. Technical report*. Publications Office of the European Union. https://doi.org/10.2811/326094

Messing, V. (2014). Methodological puzzles of surveying Roma/Gypsy populations. *Ethnicities*, *14*(6), 811–829. https://doi.org/10.1177/1468796814542180

Sumption, M. (2020). How useful are survey data for analyzing immigration policy? *Data & Policy*, *2*, e19. https://doi.org/10.1017/dap.2020.20