Final Report

Improving equality data collection in Belgium

UNIA

unia.be

SERVICE PUBLIC FÉDÉRAL
JUSTICE
FEDERAL PUBLIC SERVICE
JUSTICE

FEDERALE OVERHEIDSDIENST
JUSTITIE
FÖDERALER ÖFFENTLICHER DIENST
JUSTIZ

Gelijke Kansen
Egalité des Chances
Chancengleichheit
Equal Opportunities
.be

Content

1 Introduction

This report is the result of the project *'Improving equality data collection in Belgium'* (IEDCB). This is a one-year Belgian project, co-funded by the Rights, Equality and Citizenship (REC) programme of the Directorate-General for Justice of the European Commission. It was realised thanks to a partnership between Unia and the Equal Opportunities Team of the Federal Public Service for Justice.

This project is based on the *Guidelines on improving the collection and use of equality data*[1], developed by the *Subgroup on equality data*. This Subgroup was set up by the *High Level Group on Non-Discrimination*, *Equality and Diversity* of the European Commission and is facilitated by the European Union Agency for Fundamental Rights (FRA). In addition to these *Guidelines*, we also used the European Commission's *Handbook on Equality Data*[2] as a reference during this project. We have adapted these guidelines to the Belgian context and to the unique characteristics of the Belgian data landscape.

It's important to note, however, that the present report is a shortened and translated version of the original report that was written in French and Dutch. The full length report can be found on the website of Unia in either Dutch or French.

1.1   What is equality data?

According to the definition of the *Subgroup on equality data*, equality data is any piece of information that allows us to describe and analyse[3] the state of equality or inequality in society. 'Data' is used as a general term and can refer to both quantitative and qualitative data. Such data can be collected from a number of data sources such as official data (census, administrative databases), household and individual surveys, victimisation surveys, complaints, discrimination testing, diversity monitoring and qualitative research.

It may also concern data that was not collected for the purpose of measuring equality or discrimination, such as the administrative databases for social security. The data from these sources is mainly collected for a reason other than measuring or demonstrating inequality, but it can still be used as equality data. In the chapter 'Methodology', we discuss the data that was the focus of this project in more detail.

1.2   Why is equality data important?

There are several reasons why collecting and using equality data is important. The adage 'to measure is to know' applies here. In the context of combating discrimination and inequality, it is necessary to collect equality data to gain a more precise picture of reality. Accurate and complete data provides the foundation for identifying problems and addressing them through a policy based on facts, rather than

---

[1] High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2018), *Guidelines on improving the collection and use of equality data*
[2] European Commission (2016), *European Handbook on equality data*
[3] High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2018), *Guidelines on improving the collection and use of equality data*, p. 2

intuition. Below we list the main reasons:[4]

- Equality data enables us to monitor inequalities across different domains of society such as employment, housing and education. This enables policy making based on facts. In addition, equality data can be used to evaluate the impact of policy aimed at tackling inequalities.

- Equality data can support policy work and awareness-raising in the field of equality and non-discrimination.

- Equality data can make hidden forms of discrimination visible, for example by using situation testing (discrimination testing) in domains such as housing and work, or by conducting victimisation surveys.

- Equality data can be applied in the fight against discrimination and inequality, such as positive actions[5],data mining and situation testing (discrimination testing)[6] to detect possible discrimination.

- Equality data is indispensable to uncover and understand the structural mechanisms of inequality and discrimination.

- Data demonstrating direct or indirect discrimination can be used as reliable evidence in legal discrimination proceedings.[7]

- Equality data is important for reporting in the context of international human rights obligations.

The **purpose of the data** is the most important principle that emerges when using equality data. It is not enough to produce or collect data simply to 'know'. It is important to 'know why' we want to 'know'. The collection, processing or analysis of equality data must serve a precise, legitimate and proportionate purpose and must meet the requirements for the protection of privacy and human rights.

1.3 The Belgian context: one of the reasons for this project

Belgium has a fragmented landscape when it comes to equality data. There are many data sources, but there are major differences in the collection methods and definitions used. For example, there are at least three different versions of the 'origin' variable based on administrative data, depending on the definition used by the institutions or organisations concerned (different geographical descriptions, different ways of processing missing data, etc.).

While the living conditions of certain groups has been well documented in certain domains, there is a conspicuous lack of information for other groups or domains. Even when data sources exist, they are not always used for actions or policy work that promotes equality. The partners involved may not be aware that this data is available. Some data is not well known or is not easily accessible; this leads to a lack of

---

[4] High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2018), *Guidelines on improving the collection and use of equality data*, p. 4-5

[5] Unia, Note cadre sur les actions positives.

[6] Unia, Tests de situation 2.0 : quelques nouvelles propositions d'Unia.

[7] In Belgium, Article 33 of the Anti-Discrimination Act allows the use of "general statistics on the situation of the group to which the victim of discrimination belongs", such as facts that give rise to a presumption of discrimination. (see: Act of 10 May 2007 to combat certain forms of discrimination (Anti-Discrimination Act).

awareness about these resources and their use by stakeholders.

Since there is no consultation and no central coordinating actor, efforts in the field of equality data are very fragmented and based on ad-hoc initiatives. This results in a lack of coordination and strategic direction for the design and implementation of policy. Equality and non-discrimination are cross-cutting themes of relevance to a range of public institutions and structures. This is certainly the case in Belgium, where certain areas fall within the competence of the federal government (police, justice, criminal law), while others fall within the competence of the communities (education, culture) or of the regions (housing, mobility).

In addition, there is usually very little consultation with the rights holders and civil society organisations involved. This goes against the principles of the human rights-based approach advocated by the European Commission for producing, collecting and analysing equality data. The **human rights-based approach to data** comes from the Office of the High Commissioner for Human Rights (OHCHR). It encompasses a set of principles[8] to improve the quality, relevance and use of data and statistics, following international human rights standards and principles. In this report, we review these different principles and how they are – and can be – applied in the context of equality data in Belgium.

Finally, this project is a response to a series of recommendations from international institutions.[9] These recommendations call for a better development of equality data for Belgium. There are also specific requests for data from European monitoring mechanisms[10] for non-discrimination or equality, requests to which Belgium is not always able to respond. This project therefore aims to be a first step towards formulating answers to these various recommendations and requests.

To respond to these different needs and issues, this project positions itself as a first step on the road to a coordinated strategy for collecting and using reliable equality data in Belgium, with respect for human rights-based principles.

1.4   Project objectives

This project aims to strengthen and promote the collection and use of equality data in Belgium. On the one hand, this means working with the existing data, by centralising it in order to make it more

---

[8] These principles are explained in detail in the Guidance Note on the collection and dissemination of data from the Office of High Commissioner for Human Rights. This note is available on their website.

[9] In February 2019, at the conclusion of their official visit to Belgium, the United Nations Working Group of Experts on People of African Descent explicitly recommended "that the Government collects, compiles, analyses, disseminates and publishes reliable statistical data disaggregated by race and on the basis of voluntary self-identification…" (United Nations Working Group of Experts on People of African Descent (2019), *Statement to the media by the United Nations Working Group of Experts on People of African Descent, on the conclusion of its official visit to Belgium,* 4-11 February 2019).

[10] In 2014, in the fifth report on Belgium, ECRI reiterated its recommendation "that the authorities consider collecting data broken down according to categories such as citizenship, ethnic origin, language and religion and to ensure that this is done in all cases with due respect for the principles of confidentiality, informed consent and the voluntary self-identification of persons as belonging to a particular group. Such a system should be drawn up in close co-operation with all those concerned, including civil society organisations, and should take into consideration the possible existence of multiple discrimination" (European Commission against Racism and Intolerance (2014), *ECRI Report on Belgium (fifth monitoring cycle),* CRI(2014)1).

accessible and transparent; and on the other hand, that the problems associated with this data in Belgium must be analysed and the associated gaps identified.

This general objective translates into **three concrete** results:

1) **A mapping of equality data in Belgium.** In order to improve the development and use of equality data, it was first and foremost necessary to outline a clear and complete picture of existing equality data. This search also led to the discovery of useful data sources that would otherwise have remained under the radar because they are not specifically aimed at measuring equality.

2) **To carry out a needs assessment based on findings of the mapping exercise and produce a series of recommendations on ways to improve the collection and use of equality data in Belgium.** This is the subject of this report.

3) **A data hub that centralises equality data sources to make data more accessible and transparent.** This data hub is accessible free of charge at **www.equalitydata.unia.be**. On this online platform, data sources can be searched by criterion and by domain. Website visitors can search various data sources that Unia has collated between June 2020 and March 2021. These sources are either immediately accessible online or available upon request. This data hub is kept up-to-date and users can also suggest new data sources or changes to information about existing data sources to Unia.

In addition, the project also aims to achieve the **following objectives**:

- **Improving the use of equality data via the data hub** so that equality data can be used more by policy makers and civil society.

- **Raising awareness of the importance of the collection and use of equality data by public institutions and civil society**, through their participation in the advisory committees that guided the project. We were able to explain to them why it is necessary to improve equality data and to start a discussion about the ways in which we can do this in Belgium.

- **Developing a human rights-based approach to data in the field of equality data in Belgium** thanks to the participation of the advisory committees that guided the entire project. This was complemented by a very broad information gathering (crowdsourcing) among actors working in the fields of anti-discrimination, diversity and equality.

- **Giving Belgium the opportunity to more effectively respond to international recommendations.** This project will give Belgium the opportunity to create recommendations on equality data collection, so that we can also comply with international standards.

The beneficiaries of this project are numerous: governments, civil society, social partners, academics, rights holders, etc. Public authorities, as processors of equality data in Belgium, are also the ones who stand to benefit most from better coordination and increased awareness about the collection and use of equality data.

1.5 Which discrimination criteria and why?

This project is a first step in a strategy for the development of equality data. As this first phase of the

project was limited to one year, we focused on three initial groups of discrimination criteria. We want this project to be the first step in a gradual approach: we have laid a solid foundation for improving the collection and use of equality data. We hope that this will later include the other discrimination criteria (disability, age, etc.) and possibly other relevant characteristics.

The three groups of discrimination criteria[11] we have taken into account are:

1) The 'racial' criteria, encompassing so-called 'race', skin colour, nationality, descent and national or ethnic origin;

2) Philosophical or religious beliefs;

3) Sexual orientation, gender identity, gender expression and sex characteristics.[12]

The methodology contains a definition of these groups of criteria.

Why did we choose these three groups of discrimination criteria? We focus on the 'racial' criteria because the existing data on this group of criteria is inconsistent, because the definitions and forms of operationalisation differ and change over time, and there are various gaps in the data. It is also the group of criteria for which Unia receives the most discrimination complaints and questions. The choice for the discrimination criterion 'religion or philosophical belief' is motivated by the fact that discrimination based on 'racial' criteria in Belgium often intersects with religion and belief. Therefore, a focus on 'racial' criteria, without considering religion or belief, could lead to blind spots in the data sources on 'racial' criteria and their possible intersections.

Finally, we also selected 'sexual orientation' and 'gender identity' as criteria, as these are addressed in intersecting policy initiatives (such as the *Interfederal Plan of Action against discrimination and violence against persons based on their sexual orientation, gender identity, gender expression or intersex/DSD condition 2018-2019*), but for which there is little data and research available to inform policy. After the recommendations from the first meetings of the advisory committees, we added the criteria of sex characteristics and gender expression, criteria for which there are data gaps and which often co-occur with data on gender identity.

1.6 Content and limitations of this report

This report is the conclusion of our work on the IEDCB project. We realise that – sometimes, albeit in a brief way – it raises many issues inherent to the topic of equality data. One of the most important issues is undoubtedly recognising the need for information and data to measure and combat the phenomena of discrimination and inequality, while avoiding reproducing these discriminations or inequalities by creating categories that divide. Added to this are the mandatory requirements of privacy legislation and a human rights-based approach.

This report is not intended to cover all these topics in-depth. It is a first step, on the basis of which

---

[11] The discrimination criteria used for this project are based on the Belgian anti-discrimination legislation. For more information see https://www.unia.be/en/grounds-of-discrimination/grounds-of-discrimination

[12] These last three criteria are part of the Gender Act (Law of 10 May 2007 combating discrimination between women and men) and not of the Anti-Racism Act (Law of 30 July 1981 punishing certain acts motivated by racism or xenophobia, amended by the Act of 10 May 2007 and Act of 17 August 2013) and the Anti-Discrimination Act (Act of 10 May 2007 combating certain forms of discrimination), as is the case for the other criteria listed here.

further work can be done. We narrowed our analysis down to the elements that emerged from our research in the field (survey, interviews, etc.), the advisory committees and our mapping. Thus, we have remained focused on the Belgian reality in terms of the existing data and the concerns raised by our respondents.

2 Methodology

In this chapter we present the methodology of the project in three parts. In the first part, we explain more about how we defined and delineated the three groups of discrimination criteria. In the second part, we take a closer look at the methodology for inventorying equality data and for analysing this inventory. In the third part, we explain more about the method of collecting information for the mapping and the present report.

2.1 Definition of the three groups of discrimination criteria

As mentioned in the introduction, for the project we made a mapping of existing equality data in Belgium for three groups of discrimination criteria:

- The 'racial' criteria, encompassing: so-called race, skin colour, nationality, descent and national or ethnic origin;

- Philosophical or religious beliefs;

- Sexual orientation, gender identity, gender expression and sex characteristics.

In this section, we define more precisely these three groups of criteria.

Since the theme of equality data is closely linked to anti-discrimination, we have used the so-called 'protected' criteria in anti-discrimination legislation to define the framework for this project. However, attitude polls, research reports, administrative databases – just to give a few examples of the type of data sources we included in our data mapping – tend to apply a sociological angle, with conceptual demarcations that deviate from the discrimination criteria as defined by legislation. The concrete translation of these concepts into studies or databases is also very diverse. Therefore, in order to match the reality of the data sources, we had to deviate from the legal definition of the discrimination criteria. In addition, we also had to deviate from the distinction between different discrimination criteria by clustering them. Here are the different categories we used for the mapping:

1) Regarding the first group of criteria – **the 'racial' criteria** – our mapping distinguishes between the data sources that contain information about nationality, skin colour and national/ethnic descent or origin of persons.

   - **Nationality:** for the purposes of this report, we define nationality as the current legal link between a person and a particular State. Many data sources contain this information.

   - **Skin colour:** include data sources that explicitly deal with skin colour (for example: a survey with a question about the prevention of discrimination based on skin colour or diversity monitoring in the media, which asks about the perceived skin colour).

   - **National or ethnic origin and descent**:

     o With regard to **national origin**, we are talking about data sources with information about persons whose national origins are linked to countries other than Belgium. This means that data with information about the nationality at birth falls under this

category and not under 'nationality', since it does not concern the current nationality and therefore refers to the national origin. In the context of this project, links to the **migration history** of individuals or their families also fall into this category for our purposes (for example using terms such as 'first generation', 'second generation' or 'immigrants').

o   The term **ethnic origin** refers to groups based on the concept of ethnicity. This concept is the subject of many discussions among social scientists. We will not delve deeper into the conceptual debate here as it goes beyond the scope of this project, but we would like to emphasise that actors in the social sciences argue that ethnicity is a social construct and a relational reality, rather than a substantial or an essential one.[13] For the purposes of this report, we are defining ethnicity as an aspect of social relations between social actors who see themselves and/or are perceived by others as culturally different.[14] Ethnicity can thus refer to bonds that unite a group of people with the same cultural characteristics, a sense of belonging or a belief in a common origin.

o   **Descent** refers to the genealogical origin of a person. It allows one to grasp the situation of an individual who is treated differently because of the real or supposed identity of his or her ancestors. In the context of anti-discrimination legislation, this term is mainly used to indicate the situation of Jewish persons or persons of Jewish origin, who do not necessarily profess the Jewish religion, and who can be victims of anti-Semitism, regardless of whether they consider themselves Jewish. In certain cases, this criterion may also apply to Travellers.

2) The second group includes the criterion of **philosophical or religious beliefs**: those beliefs that touch on the question of the existence or non-existence of a deity. Think of the different religions such as Judaism, Christianity or Islam. Other philosophies of life where the existence of the divine is not central are also included (e.g., atheism, liberalism, Buddhism, etc.).[15]

3) The third group encompasses the discrimination criteria of **sexual orientation, gender identity, gender expression, and sex characteristics.**[16]

▪   **Sexual orientation or preference:**[17] determines who a person is sexually and/or romantically attracted to, based on gender, gender identity and/or gender expression,

---

[13] M. Bessone, "Race et ethnicité", P. Savidan (dir.), *Dictionnaire des inégalités et de la justice sociale*, Paris, P.U.F., 2018, p. 1379-1385, p. 1383.

[14] M. Martiniello, *L'ethnicité dans les sciences sociales contemporaines*, Paris, P.U.F., p. 18.

[15] The concept of 'philosophical beliefs' is defined in a limited way in the context of this project. For example, we are not talking here about all the beliefs that are accepted in the context of the jurisprudence of the European Court of Human Rights such as pacifism, veganism, etc.

[16] The last three criteria are part of the Gender Act (Act of 10 May 2007 to combat discrimination between women and men).

[17] When we talk about the discrimination criteria as defined by law, we use the term "sexual inclination" because this is the term used in the law. However, when talking about the sociological concept used in data and research, we will use the term "sexual orientation" throughout the report. That term is now seen as more sociologically correct, as indicated by, for example, çavaria on their website: "In the past, the word sexual inclination was used. Nowadays we use orientation to indicate that who you are sexually and/or romantically attracted to is not always clearly defined and who you are attracted to can vary over time and in intensity (=sexual and/or romantic fluidity) over time."

for example, heterosexuality, homosexuality and bisexuality.

- **Gender identity:**[18] is the inner conviction and individual experience of being either a woman and/or/nor a man. This may or may not match the gender assigned at birth. For this project, we searched for data that looks at gender identity outside of the binary definition. Examples are cis and transgender, non-binary and gender fluid.

- **Gender expression:**[19] refers to the way people express their gender identity and how it is perceived by others. It can be expressed through language, clothing, behaviour, hairdo, and so on. For this project, we searched for data that explicitly used gender expression as a variable.

- **Sex characteristics:**[20] refers to physical, biological characteristics and the variations in these characteristics. These variations can relate to the following biological characteristics: chromosomes, gonads, hormones, reproductive organs, and pubertal characteristics. The term 'intersex' describes people whose sex characteristics do not fall within the traditional gender divide. For this project we therefore searched for data sources that have explicit information about variations in sex characteristics and intersex people.

These concepts can be applied in many different ways in data and research (see Chapter 5). When drawing up the mapping and completing the corresponding *mapping tool*[21] we had to decide which forms of operationalisation[22] fall under which category of criteria. For completing the *mapping tool* and categorising the data sources, we have adhered to the following guideline: we follow the operationalisation used by the producer of the data source. For example, in studies based on educational data, the variable 'home language not Dutch' is often used to capture students with a migration background. We have therefore categorised these data sources as data on the concept of migration background, without further discussing the choice to operationalise 'home language not Dutch' as a marker for a migration background.

2.2 Methodology of the mapping

2.2.1   Selection of the sources

When drawing up a mapping of equality data, we have tried to be as exhaustive as possible for the three groups of criteria involved in this project. A first necessary task was therefore to delineate the mapping and thus to define selection criteria: which data sources do we include in the mapping and why? After

---

[18] Institute for the Equality of Women and Men, Legislation.

[19]   Institute for the Equality of Women and Men, Legislation .

[20]Callens, N & Motmans, J (2020), *Brochure d'information à l'intention des parents d'enfants présentant des variations des caractéristiques sexuelles* (published by the Institute for the Equality of Women and Men).

[21] The 'mapping tool' is an instrument for drawing up a mapping of equality data made by the *Subgroup on equality data* of the FRA. In concrete terms, this is an Excel file that is intended to facilitate the inventory and categorization of the metadata about the data sources (European Union Agency for Fundamental Rights (2019), *Compendium of Practices on Equality Data: Equality Data Diagnostic Mapping Tool*).

[22] By operationalization, in this context we mean measuring abstract concepts by using indicators representing the abstract concept.

all, as mentioned in the introduction, the term 'equality data' is a very broad concept. Many different types of sources are therefore eligible: databases, research reports, annual reports of institutions and organisations, newspaper articles, web pages, books, etc. This resulted in a huge amount of data sources that could be inventoried. However, we wanted to create a mapping that is usable and relevant for the users of the data hub. The feasibility of the project in a short period of time[23] was also an important consideration. It was therefore necessary to identify certain selection criteria which we explain below.

Firstly, we limit ourselves to data sources that were **published between 2010 and 2020**. Subsequently, it must also concern data sources that are **publicly accessible**. By publicly accessible we mean data that can be consulted without a fee and/or data that can be obtained upon request. As stated above, the definition of equality data is very broad, so our next criterion makes a selection **based on the type of data**. For this we followed the classification in the *Guidelines on Equality data*[24] of the *Subgroup on Equality Data*. We have therefore only included data in the form of household and individual surveys, census data, administrative data, complaint data, situation testing (discrimination testing), diversity monitoring and qualitative research. For reasons of feasibility and pursuing clarity, we have made the following exemptions:

- We only included qualitative data sources that contain a clear social science methodology. We also follow the guidelines of the *Subgroup on Equality Data*, where qualitative research is defined as 'case studies, in-depth and expert interviews'. Therefore, we have not included publications with a legal methodology in the project.

- Some data sources contain information on the number of people having a foreign nationality in Belgium without link with a specific theme. This is not sufficient to be included in the mapping. In the data source, there needs to be an intersection between data about nationality and a specific theme. For example, we do not include data sources about the number of people having foreign nationalities in Belgium but we include data sources about the nationality of people working in the healthcare sector in Belgium.

- It was impossible, in practical terms, to include all academic research output. An additional argument here is the fact that much academic output is not freely available. Academic research that has been explicitly funded/requested/published by administrations or governments or by civil society is included, as are research reports published by the Support Centres for Policy Relevant Research in Flanders. Later in the report, we refer to these academic sources that were inventoried as 'academic sources published by a third party'.

2.2.2   Mapping tool

The *Subgroup on Equality Data* and FRA provided us with a 'mapping tool' in the form of an Excel file to make an inventory of the data sources. In the Excel file, each line represents a data source. For each source we filled in a series of information, more specifically about the criteria and the domains in the data source. The rest of the information collected is listed in the table below.

---

[23] The IEDCB project is a one-year project.
[24] High Level Group on Non-discrimination, Equality and Diversity, Subgroup on Equality Data (2018), *Guidelines on improving the collection and use of equality data*, p 4-5.

Table 1. Information categories included in the mapping

Data Sources	Title
	Author/publisher
	Internet link
	Language
Data characteristics	Type of data
	Geographical coverage
	Date of publication/last update
Domains	Work
	Education
	Health
	Housing
	Justice and Police
	Media
	Other[25]
Criteria	Nationality
	Skin colour
	National or ethnic origin/migration background/descent/other 'racial' criteria
	Religion or philosophical belief
	Sexual Orientation
	Sex Characteristics
	Gender identity
	Gender expression
Data specific to:	Experiences of discrimination
	Hate crime or hate speech
Remarks	Specify the 'other' category in the domains

We have adapted the mapping tool to the Belgian situation, partly on the basis of the feedback from the advisory committees (see section 2.3.1). For example, we have not included many categories of information related to technical properties of data (sample size, representativeness, etc.), because this only applies to quantitative data from polls, and they do not represent the majority of equality data in Belgium. We have also added the discrimination criteria for sex characteristics and gender expression, as these are also characteristics for which there is little (known) data; the latter are also often contained in the same data sources as those on gender identity.

We have also been pragmatic in the application of certain decisions. The mapping tool proposed by the FRA and the *Subgroup on Equality Data* contains all discrimination criteria. Our project is limited to three groups of criteria. However, in order to try and provide as much information as possible in the mapping, when we included a data source (see selection criteria above), we were also willing to include

---

[25] Because it was not possible to categorize all possible domains separately, we named a residual category 'other'. This includes data relating to a wide variety of domains including culture, participation, asylum and migration, poverty, security, sport, mobility, leisure, youth work, etc. In this category we have also indicated if the data does not relate to a certain specific domain but a pure 'count', such as Statbel's statistics on diversity by origin. Finally, we have also indicated in this category that data relates to an attitude survey with regard to a certain criterion and therefore does not relate to a specific domain. An example of this is data on attitudes towards people of non-Belgian origin in the workplace in Unia's Work Diversity Barometer.

information about other discrimination criteria (gender, disability, age, etc.). But the mapping tool was not well designed for this. Due to time constraints, we decided not to include this information. We have adapted the mapping tool during the project. This was a lengthy process, characterised by many discussions about the merits of each category of information in the mapping tool, and staff resources and capacity. We believe that the tool can now be reused in future projects on equality data, regarding other criteria and characteristics.

2.3 Information gathering: fieldwork

In the spirit of the guidelines of FRA and the Subgroup, but also of the Cooperation Agreement[26] for the establishment of Unia, this project takes a **participatory approach**. For the collection of information, we have conducted (1) meetings with advisory committees, (2) surveys and (3) interviews with experts. In this way we involved as many actors as possible.

2.3.1   Meetings with advisory committees

Via advisory committees, during the various phases of the project, we received input and feedback from relevant external parties. We set up two type of advisory committees: a committee with public institutions and academics, and a committee with members of civil society. We identified key actors and invited them to join the advisory committees and thus participate in the project.

By involving public institutions and civil society, we also aim to increase awareness of equality data among these actors. This has allowed us to emphasise the importance of equality data and the need for improvements. We therefore hope that the involvement of the members of advisory committees helped to ensure that the recommendations of this report are followed. We organised three different meetings:

- June 2020: kick-off meeting where we introduced the project and received feedback on the mapping tool.

- December 2020: we presented the draft of the mapping and gathered additional sources for the mapping and general comments.

- April and May 2021: we presented the online datahub and the report. We received (written) feedback on the datahub and gathered comments on the content of the report, in particular on the recommendations.

2.3.2   Surveys

Based on a short literature review, we administered a survey to gain insight into the use of equality data across Belgian civil society, public institutions and academia. We distributed the survey very widely in order to involve as many actors as possible in our project who could share relevant information about equality data with us.

We made two versions of the survey: a short and a long one. The long survey was sent to actors who, in our estimation, make greater use of equality data. The short survey was sent to actors with less expertise

---

[26] For more information, see the page *Accord de coopération Centre interfédéral pour l'égalité des chances* on the Unia website.

in the field of equality data. In both versions of the survey (long and short) we asked questions about which equality data is used by the person or organisation, about gaps in equality data, about the perceived need to collect more equality data and about self-identification. In the long version, we included additional questions about equality data that the actors process themselves, which equality data they plan to process in the future, and questions about intersectionality.

However, getting responses to our call to complete the survey proved challenging. Despite reminders (at least two) which were sent out, we recorded a response rate of approximately 14%.[27] This response rate is in itself a first result of our surveys: the subject of 'equality data' is after all very specific and we realise that a large part of the actors to whom we sent the survey may have never heard of equality data or were perhaps not actively involved with it and therefore did not complete the survey. We thus observed that in the first place there is a need for more context and awareness about the subject of equality data. A pedagogical approach to the issue is necessary.

2.3.3   Interviews

Based on the literature on equality data, we identified the themes that we wanted to explore more deeply with relevant stakeholders for the purpose of the report. This concerns the following themes: gaps in equality data, intersectionality, self-identification, conceptualisation and categorisation of groups, mainstreaming of equality data and instrumentalisation and sensitivity of data. We prepared a topic guide to use for interviews.

We conducted seven interviews with relevant stakeholders from academia in the field of equality data: two with French-speaking and five with Dutch-speaking experts.

2.3.4   Analysis of the surveys and interviews

The information obtained through the surveys and interviews served not only as input for the mapping, but also for this report. For the purposes of the report, we analysed both the surveys and the interviews in a systematic way. The results of the analysis have partly shaped the themes discussed in this report. For the writing of this report, however, we primarily used the mapping and its analysis. We supplemented the conclusions we had made on this basis, drawing on the results of the analysis of the surveys and interviews.

---

[27] We sent the long survey to 194 people and received 27 responses. This equates to a response rate of 13.9%. We sent the short survey to 212 actors and it was completed by 31 people. This equates to a response rate of 14.6%.

3   The legal framework: data protection

3.1   The GDPR and the Belgian law regarding the protection of personal data

There is a clear demand in Belgium for the collection and processing of equality data. Since equality data is largely personal data, its processing is subject to a legal framework at the Belgian and European level. This legal framework guarantees fundamental rights such as the right to privacy. One of the most important data protection instruments at the EU level is the **General Data Protection Regulation (GDPR)**. The GDPR[28] is the EU Regulation on the protection of the personal data of natural persons which entered into force in May 2018. As it is a Regulation, it has direct application in all EU Member States, including Belgium. In principle, a Belgian law can only further supplement or clarify the GDPR in those areas it allows.

The GDPR lays down a number of principles regarding the processing of personal data in general (Article 5), and defines the conditions that the processing of personal data in general must meet in order to be considered lawful (Article 6). In addition, the GDPR contains a ban on the processing of sensitive personal data, but also provides detail on the exceptions that make processing possible.

The Belgian Law of 30 July 2018 regarding the processing of personal data definitively cancels the privacy law of 8 December 1992 and integrates the principles of the GDPR into the Belgian legal order. This law must therefore be read in conjunction with the GDPR. The law supplements or clarifies the GDPR in those areas where the GDPR allows.

From this we can conclude that 1) there is a legislative framework that protects personal data; and 2) that there are possibilities to collect equality data within this framework.

3.2   The Belgian law on public statistics

In Belgium there is also **the Belgian law of 4 July 1962 on public statistics.** This regulates the functioning of the National Institute of Statistics, today known as **Statbel**. The law contains the standards for statistical production and imposes restrictions on the type of data with which Statbel may work. Article 24-quinquies makes explicit that:

"Under no circumstances shall the statistical investigations and studies of the National Statistical Institute relate to private life, including sexual conduct, belief or activity in political, philosophical or religious fields, racial or ethnic origin."

Today, private life and sensitive personal data are already protected by the GDPR and the Belgian law of 30 July 2018. We can therefore question the desirability of the very restrictive positioning of Article 24-quinquies in the law on public statistics. This positioning certainly offers Statbel and other institutions opportunities, but also creates important limitations and confusion. Our experience is that **this law is**

---

[28] Regulation (EU) 2016/679 of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union 04 May 2016, L119/1.

often interpreted in a highly restrictive way, even by institutions or organisations that are not subject to this law. It therefore makes it more difficult to determine what is possible and what is not possible in Belgium and thus creates uncertainty. It should be possible to open a discussion about the advantages and disadvantages of this Article 24-quinquies, as well as about the room for manoeuvre it offers. This discussion should provide better information about what is and is not allowed and should clarify the scope and relationship with Article 9 of the GDPR. In this way, the unclarity can no longer hinder the development of certain types of studies on inequalities or discrimination for fear of violating this article.

3.3 Ambiguity and concerns about the legislative framework

Our survey and interviews reveal an understandable fear and uncertainty regarding equality data. We also found that the limitations imposed by the legal framework are interpreted in different ways. On one hand, there is a concern with regard to the processing and protection of personal data and on the other - possibly related to this - a lack of knowledge about the legislative framework and **uncertainty about its interpretation**. It is understandable that people, institutions and organisations find it difficult to find their way in this fragmented and technical landscape and have a fear of the unknown. That is why it is first and foremost important to **develop, make accessible and disseminate clear knowledge about what is and is not possible within the legislative framework**. The legislative framework exists and considers both the point of view of the individual rights holders and that of the data controller and processer. For example, there is a fundamental prohibition on the processing of sensitive personal data, but the exceptions to this prohibition are also framed and weighed against the protection of other individual fundamental rights. The processing of data related to the discrimination criteria should not be subject to a general prohibition, and the possibilities provided for in the legislation should be used to process sensitive personal data.

In other words, existing standards do not preclude the collection of sensitive personal data,[29] but lay down fundamental safeguards and limitations on the extent and manner in which this type of information can be collected and processed. What is and is not possible will therefore also vary depending on the situation and the safeguards built into it. Raising awareness about these existing standards is therefore important and can form the basis for the debate about equality data. A public debate about why, how and which data is collected is desirable here.

---

[29] More specifically on this theme, see chapter 5 of Ringelheim, J. & De Schutter O. (2010). *Ethnic monitoring. The processing of racial and ethnic data in anti-discrimination policies: reconciling the promotion of equality with privacy rights.*

4 Gaps in Belgian equality data

4.1 Analysis of the mapping

This section provides an overview of the numerical analysis of the mapping regarding the criteria, domains, data collection methods and the language of the data sources.

At the time of writing this report, **335 sources** had been inventoried. They all contain data on at least one of the three groups of criteria covered by this project. Table 2 shows the number of sources that contain information about the different **criteria**. The focus is clearly on 'racial' criteria, while 'sex characteristics', 'gender identity' and 'gender expression' appear to be less prominent. 148 sources contain information about 'nationality' and 225 sources about 'origin/migration background/descent/other racial criteria'. The latter result is probably due, among other things, to the fact that the category is very broadly defined (see Chapter 2). 'Religious/philosophical beliefs' (65 sources) and 'sexual orientation' (51 sources) appear to be less prevalent criteria. In addition, only 38 sources contain information about gender identity and 24 about skin colour. Finally, only 8 sources contain information about 'sex characteristics' and 13 sources contain information about 'gender expression'.

Table 2. Number of sources with information about the criteria

Nationality	Skin colour	Origin/ migration background / descent / other 'racial' criteria	Religion / philosophy	Sexual orientation	Sex characteristics	Gender identity	Gender expression
148	24	225	65	51	8	38	13

In terms of the **domains** (see table 3), 'work' and the 'other' category appear to be the most important, with 129 and 181 sources respectively. The domain 'other' covers very diverse domains, ranging from 'asylum and migration' to 'social participation' and 'youth work'. In addition, 'education' (86 sources), 'health' (74 sources) and 'housing' (69 sources) also appear to be important. The categories 'media' (26 sources), and 'justice and police' (37 sources) appear to be less important.

Table 3. Number of sources with information about the domains

Work	Education	Health	Housing	Justice and police	Media	Other
129	86	74	69	37	26	181

On a closer analysis of the **intersections between the domains and the criteria** (see Table 4), it is noticeable that certain criteria clearly intersect more often with certain domains. For example, we find that 'sex characteristics' and 'gender expression', which are generally rare, occur largely in studies dealing with the domains 'health' and 'other'. Finally, and more generally, an important part of the 'racial' criteria can be found in studies dealing with 'work'.

Table 4. Breakdown of information on criteria by domains

%	Work	Education	Health	Housing	Justice /Police	Media	Other	Total
Nationality	26%	11%	11%	14%	5%	2%	29%	100%
Skin colour	18%	15%	9%	15%	18%	4%	22%	100%
Origin / migration background / descent / other 'racial' criteria	23%	16%	11%	13%	6%	5%	26%	100%
Religion / ideology	17%	17%	8%	10%	9%	7%	32%	100%
Sexual orientation	16%	16%	13%	10%	11%	4%	28%	100%
Sex characteristics	15%	15%	19%	4%	11%	8%	27%	100%
Gender identity	12%	12%	19%	5%	13%	5%	33%	100%
Gender expression	15%	15%	20%	5%	12%	5%	27%	100%

In terms of **data collection methods**, surveys (172 sources) and administrative data (102 sources) are clearly the most common methods. In contrast, only a small number of sources use 'discrimination testing' (10 sources) and 'complaints data' (15 sources). Sources with information about sex characteristics only use 'surveys' (7 sources), 'qualitative research' (2 sources) and 'administrative data' (1 source).

Finally, the mapping contains sources in four languages – Dutch, French, English and German – some of which exist in different languages. However, there is a clear majority of sources (152) that only appear in Dutch compared to the French-language sources (39). Nevertheless, many sources are available in both French and Dutch (110).

4.2   Gaps based on surveys and interviews

In the previous section, we described the mapping based on a numerical analysis. However, it is not self-evident to determine, on the basis of a numerical analysis alone, which areas can also be regarded as gaps and what exactly is still needed in order to improve equality data in Belgium. Based on the input we collected through our fieldwork and the literature, we identified the gaps in the Belgian equality data. After all, organisations and experts that have close contact with the field and groups involved are well placed to indicate what is needed. A summary of the gaps can be found in Chapter 6 'Recommendations'. The full list of the gaps identified can be found in the French or Dutch version of the report.

5 Operationalisation

5.1 Introduction

In this report we talk about measuring concepts such as religious beliefs, gender identity and national and ethnic origin. These are abstract sociological concepts that do not have a single, self-evident meaning, and whose meaning is still evolving in some cases. Although these concepts are used in everyday language, their exact meanings are rarely defined and **there is no single way to operationalise them.**

'Operationalisation' can be defined as making abstract concepts measurable by using indicators that represent the abstract concept. This is important because data is highly dependent on the method of operationalisation. The fact that there is no unambiguous definition and standardised operationalisation, of course makes data collection based on these concepts more difficult. When concepts do not have a clear definition and different indicators are used to group people, this has a direct impact on the reliability, robustness and comparability of the data.

Before going deeper into the different modes of operationalisation for the different concepts, we first discuss why it is so difficult to define and operationalise these concepts.

5.1.1 Identities: fluidity and heterogeneity

The concepts covered in this report are used by people in everyday life to identify differences or similarities, to place themselves and others in groups, and to adopt an identity. They are constructed continuously through social interactions. In other words, although they have a real impact on people's lives, they are social constructs whose meaning is not fixed.

The complexity of social reality and people's personal identities are difficult to capture through standard or fixed categories. The identities and characteristics of people and groups are not fixed and clear, but rather fluid and complex. Similarly, the definitions of the concepts covered in this report are dynamic and can change according to context and situation. This fluidity and complexity naturally complicate the operationalisation of the criteria.

In addition, the groups are internally heterogeneous. People belonging to a particular religion, for example, can vary in nature and intensity with which they experience their faith. In addition, the intersectional approach also requires taking into account that people belonging to a group, for example based on a specific origin, can differ from each other on other criteria such as gender identity, religion, sexual orientation, etc.

5.1.2 From reality to categories: a difficult transition

Trying to capture these fluid identities and heterogeneous groups in fixed concepts and categories means that **there is a risk of reducing people to variables that do not reflect real complexity and heterogeneity**. The information collected in this way is therefore less reliable as it does not correctly reflect the object it is intended to measure.

Moreover, this possibly also reinforces the myth of homogeneous identities and categories. Classifying people according to their origin seems to suggest that 'ethnicity' and so-called 'race' are real and fixed, regardless of how people interact with them. This risks reproducing the idea that there are groups of people who are fundamentally different from each other (essentialisation): *"There is a risk that the unintended difference in treatment towards minorities would be attributed to a characteristic or a culture that is their own, which negates any political and legal effort in the fight against discrimination".*[30] It is therefore important to keep in mind that these characteristics are social constructs whose meaning is not fixed, but which at the same time can have a real impact in human interactions and in society. As long as that impact is there, it is important to measure these characteristics by means of analytical categories in order to map social reality, without amplifying them, let alone generating them. Moreover, these analytical categories should not be used to reduce individuals to the (constructed) group.

However, to tackle discrimination and inequality, it is important to identify the groups who are discriminated against. But we should not lose sight of the risk of stereotyping. Stereotyping is mainly a result of political and social discourse and interpretations of data. Good communication about data is therefore an important means of combating stereotyping (see the recommendations in Chapter 6).

The above mentioned difficulties make it important for researchers to devote sufficient attention and thought to the operationalisation of personal characteristics. It is important to be aware of the possible consequences that a certain operationalisation can have on the outcome of the data collection. In addition, a researcher should also consider the influence of their own position on the data collection process. After all, dividing a population into concepts and categories is not free of norms or values and data collectors must be aware of the potential biases they have when making decisions with regard to such sensitive personal characteristics.[31]

Some concepts still require a great deal of research and debate to know how best to measure them. There is no unequivocal and timeless answer to this question. The relevance of the operationalisation of categories evolves over time and depends on the context and the purposes for which it is done. However, minimum conditions can be developed to ensure the quality and reliability of equality data. We discuss these in chapter 6. To frame these minimum conditions and without giving an exhaustive description[32] here, we outline a brief overview of the two types of operationalisation used in Belgium: self-identification and external definition.[33] We do not choose one over the other as an ideal operationalisation method. Both measurement methods have advantages and disadvantages that we will describe below. An important principle that we emphasise when choosing the operationalisation is that it depends on the objective, the executor, and the target group of the data or the research. In that

---

[30] Center for Intersectional Justice and Actiris (2018), *Intersectionaliteit: Analyse van de transpositie van het concept van intersectionaliteit in het kader van de hervorming van de instrumenten ter bevordering van diversiteit en bestrijding van discriminatie*, p. 4

[31] OHCHR (2018), *A Human Rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development. United Nations*, p. 11

[32] On this subject, see for example le rapport du Comité pour la mesure de la diversité et l'évaluation des discriminations en France : Héran, F (2010), *Inégalités et discriminations. Pour un usage critique et responsable de l'outil statistique* (2010) ; or chapter 4 of the book: Ringelheim, J. & De Schutter, O. (2010), *Ethnic monitoring. The processing of racial and ethnic data in anti-discrimination policies: reconciling the promotion of equality with privacy rights*

[33] The literature on equality data refers to *'proxy information'* instead of *'external definition'* (see, for example, *Guidelines on improving the collection and use of equality data*). However, we consciously choose not to use the term *'proxy'* because this seems to imply a certain value judgment. Moreover, the term *'external definition'* better expresses the intention.

regard, different approaches can also be used together.

5.2 The use of self-identification in equality data

5.2.1 Principle

Self-identification is the basic principle within the human rights-based approach to equality data. It means that data about personal characteristics is provided by the persons to whom the data relates and that they can do this at their own discretion. In addition, respondents should be free to choose whether or not to disclose such personal information. All questions about personal characteristics should also include the option of an open answer and of indicating multiple categories of identity.[34] Furthermore, the categories should be developed in a participative way, to ensure that respondents who fall within these discrimination criteria can contribute to the data collection in an optimal way.

The idea behind this is the *do no harm* principle. According to this principle, everyone should be aware of the risk that people might be exposed to further discrimination or stereotyping as a result of an intervention – such as collecting equality data, for example. Any objection from these groups in this regard must therefore be heard and acted upon by the data producers.[35]

5.2.2 Challenges

A key challenge concerns **defining the categories of a personal characteristic and the terminology of these categories.** When asking people to position themselves in terms of identity, it is important that the categories proposed correspond to social representations that are effectively prevalent in society at a given moment. These social representations are specific to the Belgian context and the history of the country (economic migration, colonisation…) and depend on the way in which certain themes are approached in the public debate. They must therefore not only be created in collaboration with the respondents, but can also evolve over time to continue to correspond to the prevailing social representations in the different groups.

If not, the risk of underreporting or data gaps grows. After all, it is possible that people do not recognise themselves in the concept as it has been operationalised. For example, intersex people who do not recognise themselves in the definition of intersex that is used.[36] The consequence of this is that these persons are not included in the data through self-identification, even though they may experience discrimination.

The underreporting also threatens to occur among people who do not want to identify themselves as belonging to a certain group for fear of being discriminated against or stigmatised. This is not unimportant, since in Belgium there is a perceived lack of trust in government. This is the case, for example, when data must be collected on the Roma. Anecdotally, some Roma in Belgium are very

---

[34] European Commission (2016), *European Handbook on equality data*, p. 56
[35] OHCHR (2018), *A Human Rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development. United Nations*, p. 11
[36] Many people don't use terms like 'intersex' or 'DSD,' but simply prefer a more descriptive way of expressing what's going on, such as 'I don't have a uterus' or 'I have fused labia'.

suspicious of the government and certainly of data collection from the government, and are therefore reluctant to participate in data collection. However, there is a great need for reliable data on inequality and discrimination against the Roma population in Belgium. Before data can be collected, it is important to invest in informing and involving the Roma population about the importance of data for understanding inequality and discrimination and to foster trust in that way. This is discussed further in the next chapter.

Nevertheless, complex and changing lists of categories can pose problems for the analysis. This is particularly problematic in statistical analyses, which often rely on the stability of the variables over time and which require clearly defined and mutually exclusive categories in a reasonable number.

Furthermore, when measuring discrimination and inequality, it is necessary to distinguish between how individuals see themselves and how others or the wider society view these individuals. People with a migration background who feel 'simply Belgian' can still be perceived as 'different' by society at large. It is important to know how certain groups are perceived in society, as people are discriminated against based on how others view them. One way to address this is to explicitly ask about experiences of discrimination and the perceived reasons for this discrimination. A disadvantage here, however, is that it concerns a perception of discrimination and it is often difficult to know what the motives of the discriminating party were.

Auto-hetero perception is a variant of self-identification in which persons can indicate how they think society would define them.[37] This is another way of mapping out how a person is perceived by others and therefore to what extent this person is potentially discriminated against. It takes into account the discrepancy between how a person sees themselves versus how society sees the person. Thus, questions can also be included such as 'I understand that others see me as a transgender person/Muslim, etc.'

Finally, there is also the danger that self-identification unwittingly shifts the responsibility for an anti-discrimination policy from the government to the discriminated groups. Self-identification can then have a perverse effect: if someone does not identify themselves as belonging to a certain discriminated group, the idea could arise that the inequality problem is smaller than it actually is.

When we talk about difficulties in applying the principle of self-identification, we should also mention a particular aspect in Belgium. A lot of equality data is created on the basis of administrative data (external definition). *A priori*, this goes against the principle of self-identification issued at European level and included in the *Guidelines on Equality Data* or in the *European Handbook on Equality Data*. We can only acknowledge this situation and ensure that it can be taken into account in the upcoming debates and questions about equality data in Belgium. Indeed, this **issue of the operationalisation of self-identification by public institutions** must be the subject of a debate and be discussed in all transparency. From this debate a position can emerge that protects individuals from potential abuse, while respecting their identities by taking into account the principles of the human rights-based approach to data.

5.3 The use of external definition in equality data

The other form of operationalisation that we want to consider is that of 'external definition'. The *Subgroup on Equality Data* recognises the value of this for equality data and indicates that where

---

[37] European Commission (2016), *European Handbook on equality data*, p. 38-39

possible the use of a combination of data based on self-identification and external definition can provide a complete picture of the target groups.

5.3.1   Principles

When using external definition to operationalise the personal characteristics, the person is assigned a certain category based on available information about this person. In this way, groups are often captured by means of an *a posteriori* reclassification. People cannot therefore indicate to which category they belong.

5.3.2   Challenges

Many of the drawbacks associated with self-identification disappear when using external definition. It is in a sense more objective, because it uses data that does not vary or varies less according to individual perception.[38] But this way of working also presents difficulties.

It effectively goes against the human rights-based approach that takes precedence in the production, collection and use of equality data. This approach is based on the participation of the groups involved in the data. The above use of the word 'objective' is only justified because the data concerned is 'factual information' (e.g., in the context of origin data: nationality at birth; in the context of sexual orientation data: official sex of the partner, etc.). However, the reclassification of this data into categories is not at all 'objective' and requires many decisions and trade-offs about the definitions and delimitations of the categories. Moreover, the reclassification, as a social construct, is influenced by the social and historical context in which it takes place. This work is usually done today without consulting the groups concerned. The only objective aspect is the method of collection - 'facts' are collected - but what is done with them next requires choices that are not 'objective'.

We therefore risk working with categories that do not correspond to the social reality of the groups. In many cases, the external definitions and categories used cannot capture the full range of inequalities and nuances within these groups. This can cause serious reliability problems in the field of anti-discrimination.

An example of the problematic use of an external definition is the use of a person's legal sex to determine gender identity. The legal sex is defined by administrative data. It is assumed that the legal sex of a person is the same as the gender identity. The person's legal sex, however, is the legal record of what was determined physically at birth. In many cases, the legal sex of a person matches the gender identity, and there is no problem. But this is obviously a problem for transgender people.

Another problematic example concerns the use of nationality or nationality at birth to determine whether or not someone can be regarded as a 'Roma'. 'Roma' is a term that includes various groups[39] and the use of nationality to determine this does not give an accurate picture of the Roma population. These various groups differ from each other in terms of nationality, migration history, religion, language, etc. They also have different histories of exclusion and discrimination. Some groups will have Belgian

---

[38] Comité pour la mesure de la diversité et l'évaluation des discriminations (COMEDD) (2010), *Inégalités et discriminations Pour un usage critique et responsable de l'outil statistique*
[39] H. Tocquet en J. Weyts (2013), *Context, drijfveren en opportuniteiten van Midden en Oost-Europese immigratie. Een exploratief onderzoek met focus op Roma*

nationality (for several generations, already) and others will not. (Birth) nationality is, in this case, not a very good way to determine origin. After all, Belgian Roma are also discriminated against. The solution to this could be the self-identification mentioned above.

6   Recommendations

The analysis of the mapping and the information collected through our surveys, interviews and meetings with the advisory committees form the basis for our recommendations on the minimum conditions to be respected for the development of equality data with respect for human rights and concerning the development of a coordinated strategy and a structural framework to collect and use equality data in Belgium.

6.1 Use the data hub and help keep it up to date

Our first recommendation is for all stakeholders working on the themes of discrimination and inequality. to use the data hub. The data hub is rich in quantitative and qualitative information pertaining to the three groups of criteria covered by this project, across many domains. Many of the data sources are little known, while they can be used to paint a better picture of the discrimination that certain groups experience and of the inequality in certain domains. Our data hub aims to be a first step towards bringing all data sources together and making them easier to access and use. We also have the ambition to further develop the data hub by adding new criteria. The impact of this has yet to be seen.

Given the expansion of equality data in Belgium, we will ensure to keep the hub up-to-date by adding the most recently published data sources. We therefore call on the users of the data hub to provide us with new data, communicate updates to data sources or any necessary corrections. We have provided a contact option in the data hub itself.

6.2 Filling gaps

This project has identified several gaps in equality data for our three groups of criteria. We shall briefly review some of these gaps here (see the Dutch or French version of the report for the full list). We invite everyone to make an effort to close the various gaps identified by this report with respect to the criteria, the target groups or the domains that concern them. These involve both the production of data itself and a plea to facilitate the production of this data.

Gaps regarding the content and type of data sources:

- There is a need for more data across the spectrum of **sexual orientation** from an intersectional perspective.

- There is a need for more data on the **'racial' criteria** in certain domains, such as in the **domain of 'justice and police'**. This data can map and tackle phenomena such as ethnic profiling or other forms of selectivity by the police. There are also no figures on the diversity in the judicial/police system.

- Within the domain of **'health'** there is a need for data on the **'racial' criteria** and more specifically on inequalities in health, unequal access to and discrimination in health care and the differential impact of health care.

- Here is also little high-quality data available in Belgium about **Roma and Travellers**. This is

reflected in most (international and national) reports: there is a need for complete and reliable data on which to base good policy.

- In Belgium, there are only a few data sources with data on inequality, discrimination, hate speech or hate crime based on **religious or philosophical beliefs** and **descent**. However, this is important data for tackling discrimination against Muslims,[40] for example, but also for combating phenomena such as Islamophobia or anti-Semitism.

- There is a need for data and research on the criterion of **sex characteristics**. More specifically, there is a need for data collection on discrimination and inequality with regard to intersex people, in which they themselves are actively involved and can participate. A structural policy on data collection on sex characteristics could fill this gap.

- There is a need for more data on the criteria of **gender identity and gender expression** in the general population, as well as on discrimination and inequality against **transgender people**.

- Since **discrimination tests** are a good way to identify and combat discrimination based on different criteria, there is a need for more discrimination tests in different domains and for different cities and regions.

- There is a need for more **intersectional data and research**. We will elaborate on this recommendation in section 6.4.

Gaps with regard to data quality and analysis:

- We identify several difficulties and problems related to the **operationalisation** of criteria (see Chapter 5). These problems and difficulties make it difficult to compare certain data, for example between different Belgian regions. For example, there are many different ways to define and measure the 'racial' criteria, and broad and different categories of origin are also often used. With regard to the operationalisation of gender identity, surveys often leave the option of indicating 'X' or 'other' in addition to the choice of 'man' or 'woman' (i.e., M/F/X), but this is not the most appropriate way to map[41] people's gender identity. There is a **need for coordination, research and harmonisation on the operationalisation of the various discrimination criteria**.

- Existing data is often difficult to compare over time. Data should therefore be **collected more regularly and longitudinal analyses should be performed** more frequently, where possible.

Gaps with regard to accessibility, transparency and legislation on the protection of personal data:

- It is not clear to many users who collects which data and for what purpose. There is a clear demand for a **central mapping with transparent information about data**. On the one hand, our project and the associated data hub partly meet this requirement. On the other hand, we have not included all academic sources and we warmly appeal to the academic community to consider accessibility of academic data sources and to work on an accessible mapping of these sources.

- There is a **lack of knowledge of the Belgian and European data protection legislation** on the protection of personal data and it is often assumed that the legislation allows nothing or very little. In practice, under certain important conditions, a lot is possible in the field of equality data

---

[40] More than 85% of the cases opened by Unia on the basis of religious or philosophical belief concern Muslims.
[41] Motmans, J., Burgwal, A. and Dierickx, M. (2020), *Adviesnota: Het meten van genderidentiteit in kwantitatief onderzoek*, Transgender Infopunt

(see Chapter 3). With this project, we are attempting to take a first step towards greater clarity about the existing legal framework, but we are also calling for **broader awareness** about this legal framework and the relevant safeguards it contains.

▪ There is a **need for a sound and broad public debate about the collection and processing of equality data** and to raise awareness about the value of equality data within the context of the current (legal) framework. This debate can help build trust (see the recommendations on minimum conditions in section 6.3) and can be initiated and led by a consultative body on equality data (see the recommendations in section 6.5). In the context of this debate, it is necessary in particular to question the advantages and disadvantages of Article 24-quinquies of the Law of 4 July 1962 on public statistics and to clarify its scope, in the light of Article 9 of the GDPR.

▪ We are calling for **the existing and high-quality data sources to also be used effectively to inform policy**. At the moment, equality data are still too often pushed aside when policy is made.

Of course, the production of data or of analyses must be done according to the principles that we list below.

6.3 Minimum conditions for high-quality equality data

This report does not advocate the constant development of data about everything and everyone, without a purpose and without a framework. Data is important, but equally important is its quality and respect for the context in which the data is collected, processed and analysed. Here, we list the minimum conditions that must be respected in order to achieve this goal.

The collection and processing of data is very dependent on the purpose that is being pursued. It is therefore difficult to define the concrete decisions about the data (method of collection, categories used, analysis, etc.) in a standardized way. Despite this ad hoc characterisation, we have identified basic principles that can be applied in all cases of data collection, processing and analysis. We have based this on the principles stated by the respondents in our surveys and interviews.

We noted with interest that most of the principles stated by our respondents overlapped with those of the human rights-based approach to data. There is thus a broad shared consensus that the principles in the human rights-based approach are the foundation of all work on equality data. For us, these basic principles are the minimum conditions that must be respected in order to make the production, collection and analysis of equality data a success.

6.3.1   Knowledge of and respect for the legislation on the protection of personal data at all stages of data processing

The respondents of the survey most often emphasised the importance of protecting personal data. They are generally in favour of the development of equality data for the three groups of criteria, with the *conditio sine qua non* that privacy is guaranteed, at all stages (collection, analysis, preservation, etc.). This principle is also part of the principles defined in the human rights-based approach to equality data.

We also believe that a better knowledge of the legal framework for the processing of personal data can benefit everyone, whether they are a user of data or not, in the institutions and in civil society. There is,

indeed, a tendency to think that certain data should not be touched. That idea needs to be nuanced, because according to the existing legal framework, it is possible to process certain data under specific conditions (see Chapter 3). This nuance must therefore be applied through clear and easily-accessible information about what is and what is not allowed in the context of the legislation on the protection of personal data. What can and cannot be done according to the legislation is only mentioned in general terms in this report, as this depends on the situation and differs according to the research, the data and the built-in guarantees for data protection. For more information and advice, citizens can visit the website of the Data Protection Authority.[42]

6.3.2   Participation of the groups involved: 'nothing about us, without us'

One of the most important principles of the human rights-based approach is that of participation. This principle is applied by ensuring the participation of the groups concerned in all stages of the data processing: data planning, method selection, data collection, analysis, storage, and dissemination. An important process is also the operationalisation of each criterion (religion, origin, gender identity, etc.) that must be developed in consultation with the groups concerned.[43] This is certainly the case when it comes to data based on self-identification, but it is also the case for other types of data; the groups concerned must be able to guide the debate and participate in the decisions on the categories used for the data, as well as the collection and analysis method used.

Two concrete fields of action are emerging:

- Setting up different research processes to work with the groups to determine how to collect data, which categories to use, etc.[44] Such a research approach should also be accompanied by a pedagogical process on the importance of equality data and its use.

- The participation of the groups involved must be organised in a structural manner. This can be done via a consultative body, which must be sustainable, to monitor possible evolutions in the field of discrimination or inequality and the sensitivities of groups. We will discuss this recommendation in more detail in section 6.5 below. There is a need to be sufficiently transparent and public, so that everyone has the opportunity to be informed and to participate. The challenge is also to strive for the participation of sufficient rights holders or persons directly affected to optimally represent the diversity in existing sensitivities and positioning.

6.3.3   Purpose of the data

An essential principle is that of the purpose of data, which incidentally, is central to the legislation on the protection of personal data. The collection, processing or analysis of equality data must have a well-defined, legitimate and proportionate purpose and this purpose must be justified. There can be no producing or collecting information just for its own sake. You have to know why you want to 'know'. Data is therefore not an end in itself, but a means.

---

[42] Website v of the Data Protection Authority : It is also possible to request information via this website.
[43] OHCHR (2018), *A Human Rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development. United Nations*, p. 3
[44] An example of such a study: Simon P. and Clement M. (2006), *Rapport de l'enquête « Mesure de la diversité » Une enquête expérimentale pour caractériser l'origine*, INED

This purpose must have a legitimate character in the context of the protection of human rights and the fight against inequality and discrimination. Under no circumstances should this data be used to stigmatise groups or to collect information about specific individuals. There must therefore be strict and transparent safeguards to ensure individual privacy is maintained on this data (see section 6.3.1 and section 6.3.5).

6.3.4   A specific method for each purpose

A simple basic principle follows from the previous points: a specific methodology fits each purpose and each social context. As we have pointed out, despite the established principles listed in this chapter, there is no *one size fits all*.

Thus, there are several ways to collect data, depending on the question being asked, the groups involved and the status of the actor collecting the data. A combination of methods may be a suitable solution. It can be useful to combine different forms of operationalisation in order to obtain the most complete picture of a problem. In Chapter 5 we already discussed the advantages and disadvantages of self-identification and external definition. We therefore do not favour one method over the other, but we want to emphasise that this depends on the purpose, the context and the status of the data producer.

6.3.5   Transparency: from start to finish

Transparency is essential in all work on equality data. Thus, the producers or users of data are responsible for explaining and justifying the purposes and methodology of data collection and analysis. In a survey, for example, this means that respondents must know what will happen to the data (in terms of content). In the case of data with self-identification, the need and purpose for collecting equality data must be explained so that the groups involved can actively participate in the investigation by deliberately disclosing certain identity information. There must be transparency not only about the use, but also about the other stages of working with data: what are the guarantees for the protection of privacy, how will the data be stored, etc.? Respondents must be given this information and be able to understand it.

This transparency requires good communication, which must be accompanied by pedagogical efforts towards the respondents. As stipulated by Articles 13 and 14 of the GDPR, this means that the information must be concise, accessible and easy to understand. The language used should be clear and simple, and visual elements (infographics, diagrams, graphs, etc.) should be used if necessary. Transparency and clear communication are important to ensure the participation and confidence of the respondents in the data collection; they are a guarantee of reliable and high-quality data.

Transparency is also about the accessibility of data: the procedure to be followed to have access to the data must be clearly stated and the same for everyone. It must also meet all requirements of the legislation on the protection of personal data. This also means that the information about the available data is clearly presented: which variables are there, for which years, according to which definition, which collection methodology, etc.

6.3.6   Accountability and prevention of abuse

The producers and users of equality data have a responsibility for the impact that the data or the results of the analysis can have. After all, this data is hazardous if it is used for illegitimate purposes, for example

to stigmatise and even discriminate against groups. Incidentally, this was a core concern of all our respondents: the fear of the data being misused.

After all, analytical categories are needed for data analysis and research. Attempting to capture fluid identities and heterogeneous groups in abstract concepts and categories **risks reducing people to variables that do not reflect the true complexity and heterogeneity (see Chapter 5)**. To tackle discrimination and inequality, it is important to measure these phenomenon and thus identify the discriminated groups through analytical categories. We must not lose sight of the risk of stereotyping and misuse of data for illegitimate purposes here. However, stereotyping is mainly a result of political and social discourses and interpretations of data.

Therefore, guarantees are needed against all forms of abuse. The risk of misuse must be taken into account and anticipated from the start of the reflection on data collection or analysis. First of all, misuse can be prevented by a specific and adapted methodology, for example by limiting the data collection and/or analysis only to the data necessary for the legitimate purpose in the fight against discrimination and inequality (see section 6.3.3). It is also important for researchers to devote sufficient attention and thought to the operationalisation of personal characteristics and to be aware of the possible consequences that a certain operationalisation can have on the outcome of the data collection. Operationalisation must be done in such a way that the risk of stereotyping is minimized.

Second, the responsibility also lies in the discourse that accompanies the equality data. Clear information must be prepared to avoid interpretations that could potentially harm the groups concerned. Right from the start, resources and time must be provided for clear, responsible and accessible communication of research results and data.

6.3.7 Building and maintaining trust

Trust is a *conditio sine qua non* for the development of high-quality equality data. This trust requires continuous efforts towards all stakeholders, the respondents and the groups involved who provide the personal information and public opinion in general.

The issue of trust is particularly sensitive between the governments that want to collect and produce equality data and the groups involved in this data. In Belgium there is sometimes a lack of trust in the government. For the adequate collection of equality data, it is necessary that this trust is restored, certainly for the effective use of self-identification. This trust can be restored by being transparent about data collection and processing and providing the opportunity for participation.

6.3.8 Whenever possible and appropriate: strive to use self-identification

Self-identification is a basic principle in the human rights-based approach to data: the information about personal characteristics must be provided by the people involved in this data. In Belgium, this option is used in certain polls and surveys, but not in the case of administrative data. Where possible and appropriate, efforts should be made to use self-identification, depending on the purpose and in accordance with the legal framework. However, several contextual conditions must be fulfilled to successfully use self-identification. If these conditions are met, it can be assumed that self-identification is a useful and ethical way to measure inequalities and to question personal characteristics.

Here, we list these minimum conditions for the successful use of self-identification in equality data:

- *Participation of the groups involved:* the operationalisation of each identity category (religious or philosophical belief, national or ethnic origin, gender identity, etc.) must be developed in consultation with the groups involved [45](see section 6.3.2).

- *Optional answer and multiple answers*: the collection of this type of data should remain optional. For questions about personal identity, there should always be the option not to answer or to indicate multiple identities. In addition, an open answer category is desirable in order to leave the choice up to people to define themselves as they wish. But it is a delicate balancing act, because as we explain in Chapter 5 on operationalisation, the multiplication of categories does not facilitate analysis. It is a matter of seeking a balance and this must be done in a dialogue between the groups concerned and the data producers or researchers.

- **Where necessary and appropriate, use auto-hetero perception:** this can be seen as a variant of self-identification where individuals can indicate how society would define them. It is interesting to be able to map out how a person thinks he is perceived by others and thus to what extent this person is potentially discriminated against.

6.4 More attention to intersectionality

Intersectionality is an essential perspective when we talk about equality data. More attention is therefore required in order to enable a broader and more in-depth operationalisation of intersectional analysis of data. In the full length reports in French and Dutch, we took a closer look at the concept of intersectionality, its operationalisation in data and research, its challenges, and how it appears in our mapping and in the responses to our surveys and interviews.

Here we list the main recommendations for the successful operationalisation of an intersectional perspective. We emphasise that in this report we do not provide unequivocal answers about the operationalisation of intersectionality, because each specific case requires a specific approach. There are, however, a number of general principles that can be mentioned:

- Although splitting data and crossing different personal characteristics is a good first step, an intersectional perspective should be applied **throughout the data processing process and/or the entire research cycle**: in formulating the research questions, data collection, data analysis and the communication of the results.

- It is important to consider the context in which individual cases of inequality and discrimination occur when analysing and interpreting the data. **Structural forms of inequality and discrimination** should be kept in mind and can frame the data analysis.

- Researchers or data analysts must be **reflective** about their own position in society and the implications this has on the design of the research and/or the data (analysis and collection).

- **Participation of and co-creation with** the groups involved is essential from an intersectional perspective and ties in with the minimum conditions we discussed above and with a human rights-based approach to data. In this way, one can also avoid that the frame of reference of

---

[45] OHCHR (2018), *A Human Rights-based approach to data. Leaving no one behind in the 2030 agenda for sustainable development. United Nations*, p. 3

dominant groups in society dominates the research or data analysis.

- We recommend organising **a consultative body** for reflection about the operationalisation of the concept of intersectionality in Belgian equality data. This consultative body could provide advice on specific cases and formulate more general recommendations for improving and developing intersectionality in Belgian equality data. In any case, concerned groups should be part of this consultative body, again in line with the human rights-based approach to data.

- Different organisations and institutions have only limited authority or work on certain defined themes. However, in order to apply an intersectional perspective, it is necessary to be able to **collaborate across the boundaries of these competences, target groups and themes**. A consultative body on equality data can enable this (see section 6.5).

- Proper operationalisation of intersectionality in research and data requires **resources** and is also an organisational challenge. These resources can be used, for example, to set up collaborations, to enable participation of concerned groups and a true intersectional data collection and analysis.

We would also formulate a number of more precise recommendations that apply specifically to quantitative equality data or survey-based research:

- Respect for personal data and related legislation must always be taken into account (see Chapter 3). This is all the more important because when quantitative data is broken down into different personal characteristics, the anonymity of people can be compromised. Here, therefore, intersectional working can sometimes lead to a balancing act between high-quality data on the one hand and the privacy and anonymity of people on the other.

- To split data into different characteristics, you need **sufficient data** and the data file or sample must therefore be large enough. In a sample, it is therefore important that specific target populations are sufficiently present. This may mean that the data collection or research must be carried out at an international level, if insufficient data can be collected in Belgium for a specific target group.

- To ensure that specific target populations are sufficiently present in a survey, **appropriate communication** is recommended to address them. Communicating results at the end of the research cycle is also a point for attention.

6.5 A consultative body on equality data

As a result of this project, we advocate for the development of a **coordinated and structural policy** on equality data in Belgium. This report and the associated data hub are just a first step. This policy should address the gaps that we have identified in this report and should lead to better use and development of reliable and high-quality equality data in Belgium.

In order to develop such a strategy, a **consultative body** is needed where the different actors who work with equality data can collaborate, exchange information and good practices and devise coordinated actions. At least three types of actors should participate on this consultative body: public institutions, civil society and academics. Their presence is necessary as data processers and data users. Their additional expertise would make it possible to develop an equality data approach that integrates the

limitations and sensitivities of the work in the field, of scientific analysis and management of administrative databases.

The **participation** of the groups involved must be part of the DNA of this consultative body. This principle must permeate all debates and actions of the body. The presence of civil society is a first step in this respect, but continued engagement of these actors in decision-making process is imperative. In addition, the consultative body must be accessible, organised in a clear and transparent manner and communicate about its operation and objectives in order to ensure maximum participation.

This consultative body must take into account the complex institutional architecture in Belgium, in particular in terms of languages and competences, and allow all federal entities to participate. Given the scope of the task, it cannot rely solely on the goodwill of experts. It will be necessary to provide the consultative body with its own resources, to make its existence more sustainable and to formalize its structure. This can be done, for example, by housing it in an existing institution that works with equality data. That is the only way to guarantee that it will be effective.

As already mentioned, this consultative body will have to take into account the specific sensitivities and issues that may vary between the different groups involved (and also within the groups themselves). For example, as we discussed in Chapter 5, the data collection and operationalisation of gender identity pose different challenges than those of the 'racial' criteria. To meet this need, it may therefore be necessary to organise different thematic working groups for the population groups concerned.

This consultative body will have several **missions**:

- **Raise awareness about the importance of equality data and about their use** in combatting discrimination and inequality, among civil society and public institutions. This awareness-raising must be accompanied by a pedagogical approach, because this theme can be a deterrent due to the (sometimes assumed) complexity of data or the mistrust it arouses in some.

- **Develop and promote clear information, concrete tools and good practices related to the minimum conditions** in order to develop reliable and high-quality equality data.

- **Host thematic bodies** on specific issues related to equality data that require continuous attention, in particular:

   o Thematic body on participation: this body will be in charge of the concrete implementation of the principle of the participation of the groups concerned in all steps of the processing of equality data. This can be done by establishing and promoting guidelines. This subgroup can also propose advice and concrete solutions to producers or users of data who want to apply the participation principle in their process.

   o Thematic body on **intersectionality**: this body will be a space to consider the operationalisation of the concept of 'intersectionality' in equality data. This working group can provide advice on specific cases and formulate more general recommendations for improving and developing intersectionality in Belgian equality data. In any case, concerned groups should be part of this working group.

- **Developing the data hub and keeping it up to date:** the data hub will support the activities of the consultative body and promote data sources, and will enable to objectify the progress a

clearer path to progress.

- This report can serve as a **starting point for the debates on the remaining issues related to equality data**. There is a lot undeveloped or yet to be considered in Belgium. Debates that are organised in a transparent and substantiated manner benefit the producers of data, as well as the users and the groups involved.

Colophon

Le rapport complet est disponible en français.

Het volledige rapport is beschikbaar in het Nederlands.

This publication is available on the Unia website: www.unia.be.

UNI A

unia.be