



HAL
open science

Diversity Data Collection: Exploratory Mapping & Reflection

Catherine BOTO KO CLAEYSEN, Gabriele Rosenstreich, Kurt De, Tom Bekers, Aki Hagelin, Päivi Kuuppelomäki, Stella Strzemecka, Caroline Wallace

► **To cite this version:**

Catherine BOTO KO CLAEYSEN, Gabriele Rosenstreich, Kurt De, Tom Bekers, Aki Hagelin, et al.. Diversity Data Collection: Exploratory Mapping & Reflection: Una Europa Diversity Council Action Group Data Collection Final Report 2022. Una Europa. 2022. hal-03930967

HAL Id: hal-03930967

<https://paris1.hal.science/hal-03930967>

Submitted on 9 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

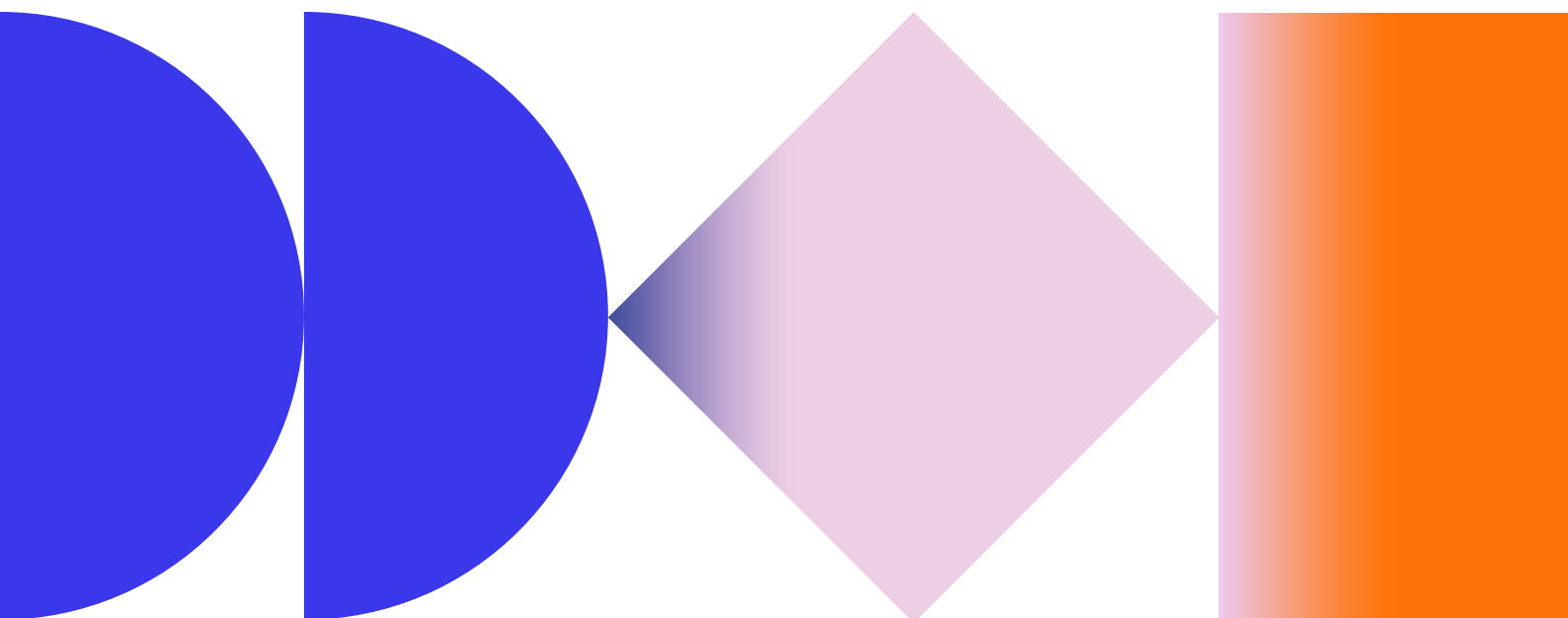
Diversity Data Collection: Exploratory Mapping & Reflection



Una Europa Diversity Council
Action Group Data Collection Final Report

2022

Gabriele Rosenstreich
Kurt De Wit
Tom Bekers
Catherine Botoko
Aki Hagelin
Päivi Kuuppelomäki
Stella Strzemecka
Caroline Wallace



Co-funded by the
Erasmus+ Programme
of the European Union





This is an Una Europa document published on September 2022
This publication is an output developed under the 1Europe project.

If you have any questions or comments relating to this publication, please contact:

Gabriele Rosenstreich
diversity@fu-berlin.de

Rosenstreich, G., De Wit, K., Bekers, T., Botoko, C.,
Hagelin, A., Kuuppelomäki, P., Strzemecka, S., Wallace, C. (2022)
Diversity Data Collection: Exploratory Mapping & Reflection. Una Europa.

<https://una-europa.imgix.net/documents/Una-Europa-Diversity-Council-Action-Group-Data-Collection-Final-Report.pdf>

Co-funded by the
Erasmus+ Programme
of the European Union



The European Commission support for the production of Una Europa content does not constitute an endorsement of the contents, which reflect the views only of the authors. The European Commission is not liable for any consequences stemming from the reuse of any and all publications of Una Europa.

Content

1. Introduction	4
1.1 Objective of this Report	4
1.2 Una Europa and the AG Data Collection	4
1.3 Key Concepts and Terms	6
1.3.1 Diversity	6
1.3.2 Diversity Indicators and Values	7
1.3.3. Good Practice	8
2. Comparative analysis of current practice in administrative data	10
2.1 Data Drivers	11
2.2 Mapping Results	13
2.2.1 Staff Data	13
2.2.2 Student Data	28
2.3 Strengths and Weaknesses in Current Practice	44
3. Issues	49
3.1 Overview of Issues Identified	49
3.2 Discussion of Issues	51
3.2.1 Needs-Driven Data Collection	51
3.2.2 Interests Underlying Data Collection	53
3.2.3 Data Protection	55
3.2.4 Definitions of Diversity Indicators (or The Problem of the Categories)	56
3.2.5 Intersectionality	58
3.2.6 Methodological Issues	59
3.2.7 Expertise & Responsibility	61
4. Gathering & analysing robust diversity data	63
4.1 Key Findings & Recommendations	63
4.2 Reflective Questions to Guide Data Collection	66
4.3 Potential Next Steps	67
5. Literature	69
6. Appendix	70
Appendix 1: Overview of Diversity Indicators per Category	70
Appendix 2: Template of Mapping Exercise	79
Appendix 3: Draft Template for Future Mapping of Diversity Data	82
Appendix 4: Triangulation	87
Appendix 5: Gender Data	88

1. Introduction

1.1 Objective of this Report

Most universities engage with diversity in some way and have their own methods of collecting data in relation to it. But it is seldom evident to those outside each university (and sometimes to those within it) what is actually being done, let alone how this practice compares with other universities and the extent to which it could be considered *good* practice.

Diversity data collection is an evolving field, and consideration of it can represent only a snapshot in time. This report presents the findings of an exploratory mapping of indicators and practices within the [Una Europa](#) alliance conducted by the Diversity Council Action Group 'Data Collection', and the reflections resulting from that mapping and from the issues that emerged during the process. The report does not claim to represent current practice or thinking beyond the scope of the Action Group and the seven participating universities, and it certainly lays no claims to providing firm conclusions about how to address the many issues identified. Nevertheless, the aspiration is that the report will provide some food for thought and discussion by others engaging with data collection in relation to diversity.

The focus of the Action Group's work, and thus of this report, is on quantitative data collection practices and in particular on the administrative data gathered by the universities relating to the composition of university students and staff across numerous diversity dimensions. The Action Group considered all dimensions of diversity deemed relevant by the member universities themselves. Both research-driven data collection and qualitative data were only peripherally considered. Importantly, the Action Group's purpose was not to collect or analyse actual data, but was directed at bringing together information on practices of data collection in each participating university, to the best of the knowledge of the Action Group's members at the time. Moreover, the purpose of the Action Group was not to define, analyse or operationalise the dimensions of diversity in themselves. Such debates are occurring fruitfully elsewhere.

Gathering data is not an end in itself, but rather a means to an end. The purpose of diversity work is to acknowledge differences and to increase equity, therefore the core purpose of diversity data collection from this perspective is to make differences visible, to show where there is exclusion and inequity, and to generate evidence that can inform policies and practices that address this. It can be the case that diversity data initially is gathered for other reasons, e.g. to fulfil reporting requirements (as will be discussed in this report), but the core objective of addressing inequity and exclusion needs to be the ultimate purpose of diversity data collection. By collating, structuring and sharing the insights that the Action Group has gained, this report aims to contribute to ongoing efforts to strengthen the evidence base supporting the achievement of diversity aims.

1.2 Una Europa and the AG Data Collection

[Una Europa](#) is an alliance of eleven universities within or alongside the European Union (UK and Switzerland are not EU members). It was established in the framework of the EU's European universities initiative 2019.

The Una Europa Diversity Council was formed 2020 with the overarching aim of contributing to diversity work within the alliance.¹ To pursue this aim the Diversity Council set up Action Groups to work on four topic areas:

- Data collection and benchmarking
- Best practices
- Education
- Research

The Diversity Council established the Action Group on Diversity Data Collection in 2021 under the leadership of a member of the Diversity Council, Gabriele Rosenstreich. Its goal was to investigate how diversity data is being addressed at the different universities and to map, reflect on and disseminate the different practices of its member universities with regard to gathering and analysing quantitative data on diversity. In particular, it aimed to

- identify issues in relation to collecting diversity data;
- provide an overview of the indicators currently used in the member universities;
- assess practices (and share some good practice examples);
- formulate recommendations that could improve practices within the Una Europa universities, or indeed in higher education institutions in general.

Terms of reference were approved by the Diversity Council and the member universities were asked to delegate staff to participate in the Action Group. It commenced its work in March 2021 and concluded September 2022 with the completion of this report.

Seven member universities of Una Europa contributed to the work of the Action Group:

- [Freie Universität Berlin](#) (FU Berlin), Germany, represented by Gabriele Rosenstreich (AG Lead)
- [KU Leuven](#), Belgium, represented by Kurt De Wit (AG Co-Lead) and Tom Bekers
- [Helsingin yliopisto / Helsingfors universitet](#) (U Helsinki), Finland, represented by Aki Hagelin and Päivi Kuuppelomäki
- [University of Edinburgh](#) (U Edinburgh), United Kingdom, represented by Caroline Wallace
- [Uniwersytet Jagielloński w Krakowie](#) (JU Krakow), Poland, represented by Stella Strzemecka
- [Université Paris 1 Panthéon-Sorbonne](#) (U Paris 1 PS), France, represented by Catherine Botoko
- [Alma Mater Studiorum Università di Bologna](#) (AMSU Bologna), Italy, represented by Azzurra Meoli²

As outlined above, the Action Group members explored their universities' practices collecting diversity data relating to the composition and representation of university students and staff across numerous diversity dimensions. Initially the Action Group had also hoped to investigate the collection of data on structures and measures to foster diversity and to reduce barriers to participation, as well as the evidence base in relation to discrimination on the basis of diversity categories. However, this proved to be beyond the scope of the group's capacity.

The Action Group never intended to collate or analyse actual diversity data from the member universities, due in particular to issues of data protection, comparability, and capacity. For example, while all universities collect and use data on their staff members and students, they all operate within a specific ecosystem with its own logic, choices, possibilities and constraints. As is discussed in this report, this results in the use of a

1 See the report produced by the Diversity Council at the end of Una Europa's 1.Europe Project (Diversity Council, 2022).

2 Unfortunately, AMSU Bologna was only able to participate in the Action Group to a limited extent, however, did provide information for the mapping exercise and thus contributed valuable input that has flowed into the findings contained in this report.

wide range of indicators and values. Therefore, the Action Group did not collect actual quantitative data, but rather collected data (i.e. information) that can shed light on the purpose and practice of data collection at the respective universities.

The Action Group met regularly online, with one physical meeting during the Una Europa week in Helsinki (June 2022). It worked within the terms of reference as agreed upon by the Diversity Council, and took an explorative approach, meaning that the actual work plan developed iteratively as work progressed.

The work of the Action Group covered two broad and interlinked areas:

- Mapping data collection practices in relation to diversity dimensions at each university using a detailed template (see Appendix 2). This encompassed the indicators and values used, the reasons why the data was gathered and issues identified. This information was then collated and analysed. The findings of the mapping exercises are presented in Section 2 and form the evidence base on which the Action Group derived its recommendations (Section 4).
- Identifying and discussing the issues faced in relation to diversity data collection. The results of the Action Group's discussions to date are presented in Section 3 and informed the recommendations (Section 4).

The Action Group members made a significant investment of capacity to carry out the agreed work alongside their regular roles (delegating tasks and consulting within their university where possible). This report thus reflects the work done by the individual Action Group members and the information and insights they have been able to gather in that process. While the goal was to be as comprehensive and rigorous as possible, this was an exploratory exercise which had various limitations and the report does not claim to reflect fully and accurately the positions of the respective universities.

The Action Group presented this report to the Diversity Council in August 2022.

1.3 Key Concepts and Terms

1.3.1 Diversity

The Action Group worked on the basis of the definition of diversity that the Diversity Council committed to as its conceptual framework and guiding vision (as developed by Gabriele Rosenstreich):

“Common to all concepts of diversity is the acknowledgement that there are multidimensional differences between people structured by social categories, such as ‘race’/ethnicity³, citizenship, migration status, religion, gender (incl. gender identity), age, class, (dis)ability and health status, sexual orientation, family/relationship status, etc. Individuals are positioned at the intersections of multiple social categories, as members

3 ‘Race’ and ‘ethnicity’, like other social categories, are social constructs. These intertwined terms refer to racialised identities (processes of attributing racial meaning to people’s identity, embedded in a hierarchical framework), to ideas of shared cultural, biological and historical belonging and/or to categorisations of skin colour and other physical attributes. There is no scientific basis for defining groups of people in terms of race or ethnicity beyond the social practice in which the groups are constructed. At the same time, these categories structure society (hierarchically) and, in terms of their impact, exist as real-world phenomena.

of social groups created by and within them. Which categories are relevant, the meaning they are accorded and their impact depends on the given society, time and context as well as their specific intersections.⁴

'Diversity' not only describes the fact that differences exist, but is also a programmatic term: It advocates for diversity as a positive aspect of society that should be acknowledged, welcomed and fostered so that all people are able to fully participate.

The vision or interests underlying diversity are socio-political. They are defined in anything from utilitarian (e.g. 'Diversity Management' approaches that aim to effectively manage difference to maximise the profitability of or benefit to organisations) to social justice or human rights terms (aiming to dismantle or at least reduce inequality and promote equity). [...]

As used here, the concept of diversity acknowledges that social categories are constructs that interlock with complex, intersecting societal and organisational power relations and inequality. This means that some social groups experience individual and structural discrimination and thus face barriers to accessing societal resources (including education, social status). People may even directly experience violence and/or the threat of violence related to their membership of a minoritised group. They are unable to participate as fully and actively in society as more privileged groups. [...]

Diversity is a fundamentally horizontal and intersectional approach that considers all social categories and forms of discrimination and addresses practices and structures of inclusion and exclusion rather than putting particular identities in the forefront. Focussing on specific groups or categories implies on the one hand that people are members of only that one group (whereas they in fact have multidimensional identities, associated with both privileges and disadvantage), and, on the other hand, risks reproducing stereotypes and power imbalances. At the same time, not naming marginalised groups can make them, their needs and actual disparity invisible, and invisibility contributes to exclusion. Targeted approaches are thus sometimes required in order to amplify marginalised voices, to identify and address specific needs and to provide spaces to support empowerment. Working in the field of diversity requires critical self-reflection of this tension between a horizontal and a targeted approach.

[...] 'diversity', as used by the Council, encompasses [a vision of] equity and inclusion. The Diversity Council [...] acknowledges that individual universities and stakeholders may use the term 'diversity' differently and may draw on other terminology to describe both the theory and practice of engaging with difference. This reflects a discourse that is dynamic, inconsistent and contested, even within single institutions and certainly across countries. As such, it is essential to critically reflect on what is respectively meant by the terms used in a given context." (Rosenstreich, 2022)

1.3.2 Diversity Indicators and Values

An indicator, generally speaking, is that which serves to indicate or give a suggestion of something. In data collection an indicator is something that stands for something else, a measurable parameter that signifies in

4 Of particular significance for European discourse on diversity were the four directives on equal treatment adopted by the Council of the European Union between 2000 and 2004, which were incorporated into the national law of each member state. They referred to discrimination on the grounds of racism/ethnic origin (Directive 2000/43/EC), religion/belief, disability, age, sexual orientation (Directive 2000/78/EC) and gender (Directive 2002/73/EC). These six categories therefore form the basis of not only most antidiscrimination legislation but also diversity policies and practices in members of the European Union. Social status/class and family status or caregiving responsibilities, are known to also play a particularly significant role in relation to access to higher education and employment, and have thus received specific attention in many universities.

a reliable way the status of a broader phenomenon, or at least some aspect of it. In other words, a statistical indicator is an observed value of a variable – it allows information about a characteristic to be represented and thus meaningfully compared and monitored.⁵

A complex concept such as diversity, with its many intersecting dimensions and the categories that relate to them, cannot be grasped or measured in a simple figure or image. Therefore, for each category it is necessary to use a range of indicators that give an approximation of the more general notion that the category is trying to encapsulate.

For example, the education levels of the parents and household income can serve as indicators of a person's socio-economic status, but neither measure 'class' as such, which includes many other aspects depending on the underlying conceptual/theoretical framework.⁶

The options for each indicator are its values, e.g. membership of a religious community may be considered the indicator for the category 'religion' and the values might be simply 'yes' or 'no', or they could be 'Muslim', 'Buddhist', 'Rastafarian', 'Atheist/Agnostic' etc. (or they might be broken down further, e.g. 'Protestant', 'Catholic' or 'Anglican', 'Baptist' etc. for Christians). The values can be thought of as the response options on a form or a questionnaire that are considered to provide information about whether the indicator is present or not and – potentially – about that indicator.

For example, sexual orientation can be considered in relation to a range of indicators, in particular sexual behaviour, identity (one's own as well as an attributed identity) and sexual attraction. These aspects of sexual orientation may or may not align with one another. Which indicator is relevant will depend on the context and this, together with the specific need driving the data collection, will determine which values are used. For example, health policy is usually most interested in sexual behaviour, and thus often uses values such as 'men who have sex with men (MSM) – exclusively', 'MSM – primarily', etc. since not all men who have sex with men identify as gay or bisexual, or are even attracted to other men. At the same time, health policy may have additional information needs. For example, identity is relevant for targeting public health campaigns, so surveys will often also include items asking about sexual orientation and using as values/response options the identity labels that many people use themselves, e.g. values such as 'gay', 'bisexual', 'heterosexual/straight' (or indeed emerging identity labels such as 'pansexual', 'demisexual' etc.). Here too though, interpretation of data gathered needs to take into consideration the impact of stigma and discrimination associated with homosexuality and bisexuality is likely to have on respondents' answers.

There is no right or wrong, but the indicators and values used determine the data collected and its meaning.

1.3.3. Good Practice⁷

The core of good practice is that it is fit for purpose. Good practices in the context of diversity are measures and actions that take diversity into account, in all of its complexity, and have positive effects in relation to achieving diversity goals such as equity (see Section 1.3.1 above).

5 See, for example, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Statistical_indicator

6 The sociologist Pierre Bourdieu, for example, considered class in relation to not only economic capital or resources but also social capital, cultural capital, symbolic capital, cultural field and habitus.

7 This section has drawn upon the work carried out by the Diversity Council Action Group on good practice. See Diversity Council (2022).

While there is, to date, no consensus on standard criteria for good diversity practice aside from fulfilling these core criteria, a range of prerequisites and quality criteria can be considered when determining whether any practice can be considered 'good' or not. For example, 'good' practice is relevant (responds to defined/ diagnosed problems, needs etc.), efficient, effective (the objectives initially set for the intervention have been achieved and the intended beneficiaries are able to benefit from the changes made) and sustainable (the measures are embedded in structures and policies, their continuation and permanence of results is ensured). There are also a range of quality criteria to consider in relation to process, including how measures need to be planned, implemented, evaluated and documented in order to fulfil the criteria noted. For example, for good practice to be effective, it should take into account the university's structures and procedures, including the organisational culture and decision-making processes.

Individual good practices work well to fulfil the identified need. To do so they are contextual but ideally they are also potentially transferable. They can target specific groups to address specific needs but, considered as a whole, diversity practice should be addressing the needs of the entire university community (not only academic and other staff, but also students), and possibly find resonance outside the university community.

In line with the vision of diversity outlined above, good practice should also strive for an intersectional approach (take the complexity of intersecting identities, forms of discrimination and power relations into account). And it should involve the participation of members of minoritised groups in a meaningful way. This is both an ethical issue in relation to the interests underlying the practice and one of quality and effectiveness. If members of the target groups are involved in every step – from defining issues, needs, goals, quality criteria and indicators as well as in the formulation of specific items/survey questions, the dissemination of instruments, through to the analysis of data, the interpretation of results and the communication of results, then there is a greater likelihood of the data collected being fit for purpose and the analysis being accurate, meaningful and useful in achieving aims of equity and inclusion.

In relation to diversity data collection specifically, good practice provides the evidence base required to identify and/or solve specific problems/issues related to diversity and does so in a way that adheres to basic principles of diversity work, such as inclusion and equity. It should enable the robust planning, evaluation and monitoring of diversity measures in a wide range of fields (ranging from recruitment, work-life balance and career development through prevention, support services and complaints procedures in the case of discrimination, teaching skills, curriculum development and student services through to IT systems and public relations). Good practice also adheres to broader ethical and methodological quality criteria for data collection more generally, e.g. protection of privacy, robust statistical analysis and representative samples (including taking diversity dimensions into account when considering representation).

2. Comparative analysis of current practice in administrative data

As outlined in Section 1, the mapping exercise collected information on *how* the seven universities collect data that relates to diversity, rather than looking at the actual data collected by the universities.

The Action Group chose to work with an open survey of universities' data collection practices in three iterations: an initial pilot survey, a follow-up survey with a slightly modified template to incorporate learning gained in the pilot (see Appendix 2), and one-on-one talks for further clarification. The collated information was used as a basis for further discussion in the Action Group.

The final dataset contains information on 13 diversity categories, which were based upon current diversity discourse and the actual practices of the universities:

- Age
- Gender
 - Gender identity - as a specific aspect of gender (including preferred names)
- Sexual orientation
- Religion
- Ethnicity/racialised identity
- Migration
 - Refugee status/experience - as a specific aspect within the broader category of migration
- Nationality
- Language
- Class/social status
 - Working student (paid work alongside studies) – as a specific aspect as several universities gather this data and it appears to be a relevant indicator in relation to socioeconomic status
- Disability or health status. While disability should not be confused with health, operationally they are often dealt with together so for the purposes of this exercise they were both considered within one category
- Family status

The mapping exercise looked firstly at whether any kind of data was being gathered for either staff or students that could be considered an indicator of the respective diversity category.

It is important to make a distinction between diversity data gathered from students and from staff members. A university is a complex organisation in which students and staff can appear to be two entirely different populations. There are often substantial differences not only in the roles, needs and composition of these populations but also in the (viable) modes and instruments of collection (e.g. student applications versus HR forms), the availability of data, the policies, the sensitivities, the legal and regulatory frameworks, and the data collection drivers that relate to each group.

The Action Group agreed on a template with which to gather information to 'map' the practice of diversity data collection within the universities (see Appendix 2). Reflecting the Action Group's exploratory approach, the template contained questions on both administrative and survey data and practices in order to gain an initial broad overview of the field, however, the analysis presented in this report focuses on the available administrative data. The systematic analysis of survey-practices would be a significant task on its own that was beyond the scope of the Action Group. Nevertheless, some examples of relevant surveys are mentioned.

The template gathered a wide range of information about the indicators, how they are defined, which values (response options) are used, why the data is collected (or not collected), what kinds of instruments and systems are utilised and, finally, what issues are associated with the collection of data in that category and any other information that might appear relevant. Follow-up interviews were conducted (by Tom Bekers) to better understand the information provided.

Note that even when data categories are recorded as collected, this usually comes with several caveats, and does not necessarily mean that data is being collected in a comprehensive or purposeful way. It only indicates that some type of data is being collected that relates to this category of diversity.

This information is available for the seven Una Europa partner universities that were members of the Action Group Data Collection. The findings cannot be generalised, and no conclusions can be drawn regarding other universities within the respective countries.

It is important to bear in mind that this dataset is the result of an initial, exploratory mapping. As the Action Group worked it became clear that finding the information was not always straightforward, and the more closely data collection practices were interrogated, the more questions arose, for example, about the interpretation of the categories and indicators as well as how they are operationalised (e.g. which population – first years, Undergraduates, all students; at which point in the career or study pathway – on application, on commencement, when leaving; and changes over time). As with any exploratory process, more questions were generated than answers found, providing a rich basis for further work.

2.1 Data Drivers

Data is not an end in itself, therefore a key question in the Action Group's work was 'why is/isn't this data collected?'. In terms of diversity goals, the core driver of data collection is to provide an evidence base on the composition of the university in relation to groups known to face structural disadvantage and on (potential) inequality and discrimination. The data should assist in the identification of needs and inform robust planning to meet those needs. This includes monitoring changes over time. Every university mentioned some examples of data collection that were driven by initiatives to support specific groups of staff and/or students with perceived specific needs (e.g. people with disabilities or refugees).

The mapping revealed that the data being collected and how it is done, depends on the context in which the universities are working. In some cases, legal and regulatory frameworks restrict the collection of data on

certain diversity categories, while in other cases universities are legally required to collect particular diversity data. For instance, in France and Germany the governments have set quotas on employees with a disability (in order to promote inclusion), and organisations that do not meet them are fined. Therefore, French and German universities are required to gather data on the numbers of employees with disabilities and to report this to government agencies. In Finland, on the other hand, the University of Helsinki interprets data protection legislation very strictly, forbidding the collection of any data that is deemed sensitive or is not 'required' so no disability data at all is collected by the university. At the same time, relevant staff data is collected by external health providers and can be accessed by the university under certain conditions for the specific purpose of organising specific adjustments/support for the individual staff member.

Some universities collect data within the framework of an external entity or an alliance of which they are a member, so that what and how information is gathered is determined by that entity rather than by the individual university. For instance, all full-time undergraduate students apply for admission to UK universities via the centralised [Universities and Colleges Admissions Service](#) (UCAS), an independent charity. Thus, the related administrative data collection is centralised and individual universities receive data from UCAS on those students that enrol at their own university. The Finnish [National Agency for Education](#) (Opiskelijavalintarekisteri) also has a centralised student application data system, which has data on applications, offers and enrolment. Other government agencies are also able to access the application data system. For instance, the Finnish social security agency ([Kela](#)) uses this data to grant study support. In the region of Flanders in Belgium, there is an agreement between universities on what and how diversity data is collected at student intake, but each university collects the data themselves and reports it to the [Flemish Interuniversity Council](#) (VLIR). In Italy, some of the data on students' careers is collected by the [AlmaLaurea](#), consortium of universities via a pre-graduation questionnaire, with the results provided to participating universities.

Such external agencies and the regulatory and legislative frameworks can be both a lever and a barrier to data collection. Individual universities can (and do) still gather additional data, beyond the external requirements, but they may need to either collect parallel data on similar indicators or to adjust the data being sent to the external agency to fulfil its requirements. They may not be able to introduce changes themselves, for example to adapt to emerging needs, and therefore may lack an incentive to come up with a clear vision on diversity data collection and future needs.

Next, there are important national differences on how the educational system is organised (e.g. entrance criteria and fees for higher education) as well as other cultural differences or educational policies in the countries in which the Una Europa universities operate. For example, the Finnish education system guarantees the right to university studies (although it might be difficult for some in practice because the system is built on assumption that university students have graduated from high school).

The core data needs for diversity policy stem from a will to monitor and reduce potential discrimination/inequity and to foster inclusion and equity. Data should be providing a robust evidence base for the development of strategies, policies and operational measures to pursue the university's diversity goals. Therefore, a clear vision and careful methodological choices are essential. The data, the policy needs, the topics and the conceptual, organisational and legal frameworks are ever-evolving, requiring a critical, reflective and iterative approach.

The data that is available on diversity in universities is often a side-product of operational needs, regulatory requirements or initiatives with another primary purpose. This means that while data might be available, it might not necessarily be suitably useful as an evidence base to assist the pursuit of diversity goals. Even when data is collected in the context of a diversity initiative, it does not always generate data that is useful for other initiatives or for policy development more broadly. In addition, although we might have useful data, we might be unable to actually use it for a certain purpose because we did not ask permission from the data subject to use

it in this way at the time it was collected, or we may not have another lawful reason to process the data. Given the operational origin of much data collection, the IT systems are not necessarily equipped to handle sensitive and/or health data in compliance with data protection regulations.

As discussed, data is often collected for a combination of reasons, but it is not always clear why certain data was *not* being collected. Possible reasons for non-collection include a lack of identified need for the data, concerns about methodological or ethical issues, and issues of data protection. It is important to stress that data protection regulations do not forbid data collection, but rather they set criteria to be met in order to protect privacy while collecting it. It may be argued that it would be too great an effort for a university to collect data in a way that conforms to these criteria, but that is a resource-based decision of the university, not due to a data protection regulation as such.

Currently, none of the seven universities explored here have an explicit policy on diversity data collection to support such strategic, operational and methodological decisions.

2.2 Mapping Results

In this section the information gathered from each university has been collated and analysed. An overview of which university gathers information on which category of diversity as part of their administrative data systems, and which indicators and values they use to do so is provided in Appendix 1. The template used to gather this information is provided in Appendix 2.

2.2.1 Staff Data

Every university collects staff data on age, sex and nationality. That does not necessarily mean there is a formulated policy need or plan. This data is mostly/originally collected for operational (identification, retirement), statistical (government, EU funding) and/or contractual reasons (wages, taxes, insurances), and its use for other purposes is often secondary. This is why it is almost always collected as part of the appointment or application process and why the data is required and verified.

In all other diversity categories, the practice of the universities varies significantly. As has been noted elsewhere in this report, in all cases, the data gathered relate to composition of the population rather than the actual experiences of the persons concerned.

Staff – Age

Staff age data is consistently gathered at all universities. It is always derived from entry of the ‘date of birth’ as per official identification documentation provided on employment.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Age:	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]

Staff – Gender

All universities gather data on the gender of their staff, primarily driven by operational needs to identify individuals and/or by internal and external efforts to monitor gender equity.⁸ This data allows monitoring of gender-based differences in various roles and levels of the organisation.

Staff gender data is collected in a binary manner, i.e. male or female, based on the person’s legal status in all universities except the FU Berlin, which – in line with the German legal framework – uses four gender markers in its data system, including intersex/non-binary and a non-response option (see Good Practice Example 1). This does not mean that the other universities do not critically reflect on their binary gender markers: there is a growing acknowledgement not only that biological sex can be defined using a wide variety of potentially divergent indicators (e.g. genitalia, hormones and chromosomes) and is certainly not binary (i.e. intersex people exist), but also that a person’s official gender can change over time and does not necessarily reflect their lived gender identity – which is the variable of core interest in relation to the pursuit of gender equity goals. Several of the universities have begun discussing provision of three or more options, but the interface with a binary legal framework poses challenges.

Also of note is the terminology universities use in relation to this dimension of diversity: The English distinction between sex (biological) and gender (social) is not found in most of the other languages of the universities and when working in English the two terms are often used interchangeably. The legal frameworks also differ considerably in relation to the ease with which people can change their legal gender marker to align with their lived gender identity. Thus the data gathered based on legal status cannot be assumed to reflect lived gender. Only U Edinburgh gathers additional data that could indicate a difference between someone’s legal gender marker and someone’s gender identity or gender expression (see below).

See Appendix 5 for a detailed discussion of the diversity category gender.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Gender:	• legal gender marker [female / male / diverse (intersex/ non-binary) / no entry]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female] <i>(third option under consideration)</i>	• legal gender marker [male / female]

⁸ These requirements include, for example, the [European Commission Horizon Europe research and innovation funding programme](#). To be eligible, it is mandatory that organisations collect and publish disaggregated data on the sex and/or gender of personnel (and students, where relevant) and carry out annual reporting based on indicators.

Good Practice Example 1: Freie Universität Berlin - Gender Markers

FU Berlin now uses four gender markers in line with 2018 changes to German civil status law (Personenstandsgesetz): female, male, intersex/non-binary and a non-response option. At the same time, under existing German law (Transsexuellengesetz 1980) it is still very difficult for people to change their official gender marker from the one assigned at birth. A change to this law was announced 2022 but until it comes into effect, it appears likely that the majority of trans and non-binary people have a legal gender status that differs from their lived gender identity.

One aim of the university's Diversity Strategy and Action Plan 2021-2023 is to strengthen gender self-determination. A working group was formed to investigate how the needs of trans, intersex and non-binary members of the university could be met within the digital data systems in relation to gender markers and name usage. The group mapped existing practices, internal and external requirements and identified needs, and on that basis it developed proposals for the university's executive board. Early 2022 modification of the data systems was completed and new procedures implemented and communicated to staff and students.

Students are now able to autonomously change their names and gender marker independently of their legal gender status as it was determined that this data was primarily collected to monitor gender inequity and there are neither legal nor operational requirements of the university to gather data on their legal gender. Therefore, it is the lived identity that is of primary interest.

Staff gender data proved more challenging as the university is required to provide data sets, including names and gender, to several external agencies. While some external agencies accept data that reflects the lived gender identity even if it differs from the legal status, others, such as health insurers and the pension scheme, require the official data for identity checks. The university therefore added additional fields to its core HR administration data system and restricted access to the official data to staff on a need-to-know basis for these specific purposes, and the data based on lived gender identity is now used for all other purposes within the university.

Aside from monitoring gender-based inequity, the key purpose for collecting gender data had been to inform how the university communicates with its members. In German, gender impacts not only how someone is addressed (e.g. Ms Schmidt, Mr Meyer) and the pronouns used (he, she, they) but also grammar (e.g. in particular what ending nouns have). Therefore, in order to pursue its aims, FU Berlin introduced regulations and guidelines on gendered language. All new templates and official communication must now use language that is gender neutral or inclusive of all genders as a default, and gradually existing templates, etc. are being updated. In terms of data, this means that gender markers are no longer required for purposes of communication and questions related to gender are being removed from forms where they only served that purpose (as opposed to monitoring representation and equity).

www.fu-berlin.de/en/sites/diversity

Staff – Gender Identity

U Edinburgh is the only university in the Action Group currently collecting data specifically on the gender identity of their staff to supplement the data collected on their official status. This data meets an internal university policy purpose by allowing staff to more accurately record their gender, recognising the limitations of the official definition of ‘sex’ in the UK, which only permits ‘male’ or ‘female’ gender markers. The data is collected electronically in-system on a voluntary basis when the staff member is appointed and held in in the staff record. Staff can update their record at any time via a self-service online portal and no verification is conducted. Gender identity is asked using the response options: ‘male’, ‘female’, ‘non-binary’, ‘other’, ‘prefer not to say’, with an additional variable on trans status (‘yes’/‘no’).

Although they only collect data on staff legal gender markers, FU Berlin, U Paris 1 PS and KU Leuven as well as U Edinburgh all offer the possibility for staff members to be addressed by their preferred name when it differs from the name on their official identification. While the policy was not specifically developed in relation to gender identity in all cases,⁹ it is particularly significant for people who have changed their name to better align to their gender identity. This preferred name is then used for all communication, except for those official documents that are required to use the legal name.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Gender Identity:			<ul style="list-style-type: none"> gender identity as per self-identification [female / male / non-binary / other / prefer not to say] trans-status [yes / no] 				
Name:	<ul style="list-style-type: none"> legal name (used for official documents and data migration to external agencies where required) preferred name [first name &/ or surname of choice] (used for all documents, communication, reporting that does not explicitly require the legal name) 	<ul style="list-style-type: none"> legal name 	<ul style="list-style-type: none"> preferred name [first name of choice] legal name 	<ul style="list-style-type: none"> legal name 	<ul style="list-style-type: none"> legal name 	<ul style="list-style-type: none"> legal name (used for official documents and data migration to external agencies where required) preferred name [first name of choice] (used for communication) 	<ul style="list-style-type: none"> legal name preferred name [first name of choice]

⁹ At KU Leuven, for example, the option of recording a preferred name was initially introduced some time ago not in relation to gender identity but rather to address abbreviations (such as ‘Tine’ instead of ‘Martine’ or ‘Frans’ instead of ‘Franciscus’) but the policy was then also expanded to meet the specific needs of trans people. Similarly at U Edinburgh, there is no specific policy but staff can use any name they wish to. At FU Berlin the driver was the strategic goal of being more inclusive of trans and nonbinary people so it is particularly targeted to them, however, other staff can also change their names if they wish.

Staff – Sexual Orientation, Religion, Ethnicity / Racialised Identity

U Edinburgh is the only university in the Action Group currently collecting data on the diversity categories sexual orientation, religion and ethnicity/racialised identity. There is a legislative requirement for data on these categories to be gathered (UK Equality Act) and a requirement to report to the UK Higher Education Statistics Agency on these dimensions.

This data is collected electronically in-system on a voluntarily basis when the staff member is appointed and is held in the staff record. Staff can update their record at any time via a self-service online portal and no verification is conducted. The question on sexual orientation provides several possible responses based on how the person identifies themselves and the questions on religion and ethnicity allow for a very detailed registration with a wide range of possible answers being offered (multiple responses), including a 'prefer not to say' and an 'other' option, again, based entirely on self-identification.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Sexual Orientation:			• sexual orientation as per self-identification [bisexual / gay man / gay woman / lesbian / heterosexual / other / prefer not to say]				
Religion:			• religion or belief as per self-identification [list ^A , including: 'prefer not to say']				
Ethnicity/ Racialised Identity:			• ethnicity as per self-identification [list ^B , including 'prefer not to say']				

List ^A (U Edinburgh - religion or belief as per self-identification): No religion / Buddhist / Christian / Christian - Church of Scotland / Christian - Roman Catholic / Christian - Presbyterian Church in Ireland / Christian - Church of Ireland / Christian - Methodist Church in Ireland / Christian - Other denomination / Hindu / Jewish / Muslim / Sikh / Spiritual / Any other religion or belief / Prefer not to say.

List ^B (U Edinburgh - ethnicity as per self-identification): White / White – Scottish / Irish Traveller / Gypsy or Traveller / Other White background / Black or Black British – Caribbean / Black or Black British – African / Other Black background / Asian or Asian British – Indian / Asian or Asian British – Pakistani / Asian or Asian British – Bangladeshi / Chinese / Other Asian background / Mixed - White and Black Caribbean / Mixed - White and Black African / Mixed - White and Asian / Other mixed background / Arab / Other ethnic background / Prefer not to say.

Staff – Migration

KU Leuven is the only one of the seven universities that systematically gathers data on the migrant backgrounds of its staff (in a regular, anonymous staff survey, see Good Practice Example 2). Most other universities mention that there is no identified justified need (often referring to data protection regulations or their interpretation by the university's Data Protection Officer) and no legal requirement for collecting data related to migration. That does not necessarily imply that the university has no relevant policy in relation to staff with migrant backgrounds, just that it is not evidence-based / data-driven, or at least not by their own data. KU Leuven found that there is a big overlap between considering oneself as having an 'international background' and a 'migrant background'. In some cases, terms such as 'migrant background' are defined by the national statistics agency, and the university draws on this – often problematic - definition to ensure comparability with reference data gathered in national surveys or a census.¹⁰ This is one of many examples on the complexity of our data and the difficulty in comparing data sets.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Migration:						• migrant background as per self-identification [yes / no]	

¹⁰ For example, in Germany the national statistics agency defines someone to have a 'migration background' if they or at least one of their parents did not have German citizenship when they were born. Given the barriers in place to gaining German citizenship, a significant proportion of the population falls in this category. It is frequently used as a proxy variable to indicate a 'non-German' ethnicity or racialised identity (a practice that is frequently criticised but for which there are currently no widely accepted alternatives). In Berlin, a law change in 2021 (Partizipations- und Migrationsgesetz) introduced the requirement for (state) universities to gather data on the 'migration backgrounds' of their staff in order to monitor representation and thus structural barriers to employment, to inform policies and practices to reduce them and to promote accountability. This has not yet been implemented at the FU Berlin.

Good Practice Example 2: KU Leuven - Employee Satisfaction Monitor

KU Leuven hires an independent, external firm to collect anonymous survey data on its staff every three years via the 'employee satisfaction monitor'. This data provides a valuable supplement to the administrative data gathered. Among other things, it includes questions on the self-identification of a staff member as an 'international employee', as a person with a 'migrant background', and as a person with a 'disability or a chronic/long term illness'. Having a 'migrant background' is defined as having lived in Belgium for a long time, but have (had) a nationality other than Belgian yourself or having parents and/or grandparents who had a nationality other than Belgian. 'International employees' are defined as having come to Belgium from another country specifically to take up employment at KU Leuven and having a nationality other than Belgian.

This data provides a sound basis for policy initiatives, as the survey allows the findings in relation to membership of social groups or indicators such as 'migrant background' to be cross-referenced with the results of questions on discrimination, harassment, inclusion and other dimensions of wellbeing. The survey results are used as input for university-wide policy discussions (e.g. through information sessions or workshops) and for decision-making at the level of the university's Executive Board and Academic Council, but are also the starting point for discussions at the 'local' level of departments, faculties and campuses. A feedback report with the results is issued at that local level, to provide insight in how employees experience their local working context. The aim is to encourage dialogue on priority actions at the local level

(for example, regarding job content, career development, leadership styles, or wellbeing). On a more personal level, employees with a migrant background and/or a disability who wish to be contacted to discuss policy needs on these topics are invited to provide their e-mail address to be shared with the university for contact in relation to policy initiatives (e.g. invitations to participate in focus groups).

Self-identification as having a disability is particularly important for the university because the only other information available on staff with disabilities is based on a very narrowly defined status, determined by the Flemish Agency for Persons with Disabilities (VAPH), and – for reasons of privacy – this information is not easily accessible within the university.

www.kuleuven.be/english/about-kuleuven/respect/support

Staff – Refugee Status / Experience

None of the universities have identified any policy needs that would justify associated risks to privacy and potential stigmatisation (data protection) related to refugee status or experience and therefore collect no data on this category from their staff members.¹¹

It is important to note that there's also a difference between refugee status and experience. A person could no longer officially be a 'refugee', i.e. could have citizenship, but could still be impacted by those experiences of having to flee from their home country, seek asylum and live with a refugee status, often for a long period of time.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Refugee Status / Experience:							

Staff – Nationality

All seven universities collect staff data on nationality. Both citizenship and residency status are used as indicators of this diversity category (citizenship at all seven universities, residency status in at least two).

There is not necessarily a formulated policy need in relation to this category, but in some cases there is, such as a policy on recruiting international staff. It is, however, not always clear how 'international' is defined. Usually the policy relates to recruiting 'the best' researchers or academic staff, who are often based in other countries. The focus is thus not nationality as such, since in most European countries there is a significant domestic population who have a citizenship other than that of their country of residence.

This data is mostly collected for operational (identity checks), regulatory (government reporting requirements) and/or contractual reasons (checking eligibility to work in the country), and use for other purposes is at most secondary. This is why collection is almost always mandatory as part of the appointment or application process and why the data is verified against identity documentation in all universities.

The universities have different approaches to the registration of citizenship. U Helsinki and KU Leuven allow for more than one nationality in each staff record, whereas the other universities allow for only one (even if you are a dual or multiple-national). FU Berlin, for example, focuses on the German nationality (or not), only recording and reporting another nationality if the person is not a German citizen. At U Paris 1 PS, a person chooses which nationality is entered in their record, but they must provide an official document of proof (e.g. identity card or passport). Therefore, the group of dual- or multiple-nationals is invisible in terms of data reporting.¹²

¹¹ It should, however, be noted that although they do not gather data on refugee status, several universities deliver programmes targeting academics in exile. For example U Paris 1 PS is a member of the '[Migrants in Higher Education Network](#)' (MEEnS) which connects 44 higher education institutions and other organisations, all committed to the welcome and integration of students and researchers in exile in France. Similarly, FU Berlin is a member of the German network '[Academics in Solidarity](#)' and provides a wide range of financial and other support for students and researchers in exile, and KU Leuven has opened a number of temporary positions for Ukrainian refugees as well as joining the European Commission's initiative '[Science4Refugees](#)', a database of job offers.

¹² While significant barriers to getting German citizenship remain in place in comparison with other countries, it is relatively easy for nationals of other EU countries who reside in Germany to get dual citizenship. The number of dual-/multiple nationals can be presumed to be rising as EU mobility increases. Given the strategic significance of international policies and mobility for the universities, it appears worthwhile to investigate more closely what nationality data would be of value, in particular in terms of identifying needs/barriers and monitoring success.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Nationality:	<ul style="list-style-type: none"> current citizenship [German / If not German, country of citizenship] (verified against identity document) 	<ul style="list-style-type: none"> current citizenship [country of citizenship] (verified against identity document) residency status [resident in Italy / resident abroad] 	<ul style="list-style-type: none"> current citizenship [country of citizenship] (verified against identity document) 	<ul style="list-style-type: none"> citizenship [country or countries of citizenship] (Finnish overrides the others; verified against identity document) residency status [one or multiple countries in which have a valid residency permit] (Finnish overrides the others) 	<ul style="list-style-type: none"> current citizenship [country of citizenship] (verified against identity document) 	<ul style="list-style-type: none"> current citizenship [country or countries of citizenship] (verified against identity document) international employee as per self-identification [yes / no] 	<ul style="list-style-type: none"> current citizenship [country of citizenship] (verified against identity document)

Staff – Language

Only KU Leuven and U Helsinki systematically collect staff data on language. This should not, perhaps, come as a surprise since Belgium is officially trilingual (Dutch, French and German) and Finland is officially bilingual (Finnish and Swedish), while the other universities operate in an environment with only one official language. KU Leuven and U Helsinki both collect the data for operational reasons, primarily to define in which language the staff member is to be addressed. U Helsinki collects data on both first language and preferred business language (see Good Practice Example 3). At KU Leuven the situation is somewhat different as a result of the political sensitivity regarding the use of languages in different parts of Belgium. There are very strict regulations on which language can be used and this means that the official language in the region of Flanders, and hence at KU Leuven, is Dutch. Therefore KU Leuven offers its staff the choice of receiving university communications in either Dutch (as the only official language permitted) or English (as an ‘international language’).

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Language:				<ul style="list-style-type: none"> • first language [Finnish / Swedish] • preferred business language [Finnish / Swedish / English] 		<ul style="list-style-type: none"> • language of communication [Dutch / English] 	

Good Practice Example 3: Helsingin yliopisto/ Helsingfors universitet - Multilingualism

Finland is officially bilingual (Finnish and Swedish) and U Helsinki places great value on multilingualism, with English also used extensively. The university formulated a [language policy](#) in 2014, in which it noted its particular responsibility for safeguarding the national languages (Finnish and Swedish) as languages of science, as well as its responsibility for internationalisation. “It supports students and staff in their efforts to improve their language skills because it sees language skills as a requirement for participating in international research, a path to understanding other cultures, and a tool for promoting one’s own culture. Multilingualism and knowledge of one’s own and other cultures are understood to promote creative thinking and a community spirit”. The policy notes linkages with other diversity-related internal and external regulations and policies and aligns with the government’s [Strategy for the National Languages of Finland](#) (2012).

Implementation of the policy is part of the university’s operations management process. In order to effectively operate multilingually, the university registers the preferred language of business (Finnish, Swedish, or English) in the data system. This data is used, for example, to determine in which language communications with students and staff are sent.

The Academic Affairs Council (ONE), the Research Council (TINE) and the committee for Swedish-language affairs (Svenska verksamhetsnämnden) are responsible for monitoring the implementation of the language policy annually.

Language data is gathered regularly to support implementation of the policy. For example, the university’s language policy includes a wide range of requirements, such as that Swedish-language courses and services are provided for Swedish-speaking students and that units responsible have a sufficient number of staff members who are competent in Swedish. Course, service and staff planning requires robust, accessible information on numbers of staff and students within the respective language groups and language skill levels.

The university also gathers data on the number of credits completed in different languages, the languages of publications through the TUHAT research database, and staff participation in language training. Faculties and departments are required to report on their implementation of the language policy as part of their annual reporting. At the rector’s discretion surveys can also be conducted on particular topics or target groups.

Staff – Class / Social Status

JU Krakow is the only university that collects data related to their staff members' class or social status: the amount of gross monthly income per family member. This data is collected only if the staff member applies for a social benefit and is recorded only in the individual file, i.e. this data is not collated or linked to other data.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Class / Social Status:					<ul style="list-style-type: none"> • eligibility for social benefit [amount of gross monthly income per family member] <i>(staff file only, not collated)</i> 		

Staff – Disability and Health

None of the universities collect specific health data for staff, except for some inclusion of long-term illness as a value within the disability indicator or monitoring the amount of sick leave taken. However, all universities, with the exception of U Helsinki¹³, collect some data on disability, although the definition of what is meant by disability generally relies on a medical model – which can be at odds with the social model of disability often used in diversity work. The definitions used vary widely. For example, at the FU Berlin, the only indicator used is 'severely disabled' as defined by law: a person who has been officially assessed by an accredited physician as having physical or mental (intellectual or psychological) limitations such that they can function at a level of less than 50% of what would be expected from a typical person of their age.

Disability data is mostly locally collected, which means that data is not collated or linked to other data sets. KU Leuven supplements the administrative data (based on a narrow legal definition) with data gathered in a staff survey (based on self-identification) (see Good Practice Example 2).

Most universities produce some sort of report with generalised statistics, but this is often only provided to the regulatory body to fulfil reporting requirements.

In contrast to age, sex and nationality, data on disability is only collected for those staff members who decide to make themselves known to the university, mostly because they are seeking assistance/adjustment, access to financial support or preferential consideration in recruitment (in those countries with quota systems).¹⁴ This means that these numbers are probably under-reported.

¹³ As noted above, the law in Finland prohibits the collection of data that is not explicitly required and to date the U Helsinki has not defined a policy need to collecting disability data.

¹⁴ In Germany, France and Italy, employers must collect disability data to comply with government policy that a percentage of staff members (5%, 6% and 7% respectively) must meet a certain threshold in relation to disability status. In all three countries, universities who do not meet this minimum percentage are fined (See Good Practice Example 4). Some organisations decide to pay the (small) fine rather than to address the reasons behind this lack of representation.

Additionally, in Germany, job applicants with an officially recognised status as 'severely disabled' must be interviewed if they fulfil formal requirement. However, no data is systematically collected on the numbers of applicants with this status or similar.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Disability & Health Status:	<ul style="list-style-type: none"> • severe disability as per legal definition (<50% 'normal' capacity) [yes / no entry] 	<ul style="list-style-type: none"> • disability as per government definition [no / yes, if so: type of disability, temporary / long-term] 	<ul style="list-style-type: none"> • disability or long-term health condition as per self-identification [list^c, including 'prefer not to say' / no entry] 		<ul style="list-style-type: none"> • degree of disability [light / moderate / significant / no entry] 	<ul style="list-style-type: none"> • disability as per governmental definition [yes / no entry] • disability, long-term disorder or illness as per self-identification [yes / no] 	<ul style="list-style-type: none"> • disability status [category of disability as per list / no entry]

List ^c (U Edinburgh – disability or long-term health condition as per self-identification): No known disability / Two or more impairments and/or disabling medical conditions / A specific learning difficulty such as dyslexia, dyspraxia or AD(H)D / General learning disability (such as Down's syndrome) / A social/communication impairment such as Asperger's syndrome/other autistic spectrum disorder / A long standing illness or health condition such as cancer, HIV, diabetes, chronic heart disease, or epilepsy / A mental health condition, such as depression, schizophrenia or anxiety disorder / A physical impairment or mobility issues, such as difficulty using arms or using a wheelchair or crutches / Deaf or serious hearing impairment / Blind or a serious visual impairment uncorrected by glasses / A disability, impairment or medical condition that is not listed above / Prefer not to say.

Good Practice Example 4: Université Paris 1 Panthéon-Sorbonne - Disability Strategy and Action Plan

The Board of Directors of U Paris 1 PS adopted a Disability Strategy and Action Plan 2019, relating to both students and staff. It is based on a proposal by the Academic Council and aims to improve the welcome/reception, support, accessibility of premises, digital accessibility and the provision of adjustments to facilitate access to employment and studies (e.g. note-taking assistance, online courses). It builds on the vision formulated in a disability charter 2012 and strives to take into account the entire career path, in the case of students from pre-study information to post-graduate professional orientation (including teaching, examinations, university life, internships and international mobility, etc.). The Plan represents a shift from a logic of individualised assistance to one of global accessibility of the institution.

The President of the university is required to submit an annual report on the implementation of the Disability Strategy and Action Plan to the Board of Directors, including data on monitoring indicators. The Disability Strategy and Action Plan fulfils the requirement to establish a comprehensive multi-year plan for disability policy under the 2013 Law on Higher Education and Research (Law 2013-660). It also relates to requirements under the 2005 Law for Equal Rights and Opportunities, Participation and Citizenship of Disabled People, which reaffirms in particular the obligation for public and private employers to employ at least 6% disabled staff.

Staff and job applicants provide information regarding their disability status on a voluntary basis if they wish to have adjustments made to accommodate their disability. This is most straight forward if they have the 'recognised status of disabled worker' (RQTH). RQTH is granted by a government agency if the person is deemed to have reduced possibilities of obtaining or retaining a job due to the deterioration of at least one physical, mental or psychological function. See www.service-public.fr/particuliers/vosdroits/F1650. The university itself collects data on the RQTH status and which category of disability it relates to (verified by the government documentation), but not on the precise disability the staff member has.

Since 2015, universities have had to submit a report on the number of disabled staff employed to the government. The report must include the total number of employees by gender, age and category of disability. The university must also clearly indicate how many people with disabilities have been recruited through competitive assessment processes (examinations), contracts, etc. and is required to open competitive examinations specifically for people with disabilities (1 to 2 per year for administrative staff). The university is also required to provide specific promotions for teaching staff with a disability and provide a contract to a doctoral candidate with a disability from its own budget (the government funds about 30 per year itself).

If the university did not reach the 6% rate, it is required to pay a fine. As the university is part of the civil service, this fine goes into a fund for the integration of disabled people in the civil service (Fonds d'insertion pour les personnes handicapées dans la fonction publique).

www.education.gouv.fr/handicap-tous-concernes-99935

Staff – Family Status

Like age, nationality and gender, all universities gather some data on family status, albeit often indirectly. They do so mainly for operational reasons to do with remuneration or similar (wages, taxes, insurances, leave provisions) or, in the case of the U Edinburgh, because it is a legislative requirement as part of government efforts to monitor equity (Equality Act). Two types of indicators are used: relationship status (e.g. civil union, married, single, divorced, cohabiting) and children (e.g. number, ages). KU Leuven collects this data to inform the planning of childcare places. U Helsinki and KU Leuven also collate some information provision of maternity/parental leave and leave to care for a sick child in their staff data systems.

Although some universities have provisions for staff to be accommodated to fulfil caring responsibilities for family members other than the staff members' children (e.g. disabled partners, unwell parents), none of the universities appear to collect relevant data.

Staff: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Family Status	<ul style="list-style-type: none"> official relationship status [married / divorced / widowed / single] 	<ul style="list-style-type: none"> relationship status [living with partner / not living with partner] children [number] 	<ul style="list-style-type: none"> relationship status [living together / married/ civil partner / separated/ divorced / single / widowed / prefer not to say] 		<ul style="list-style-type: none"> married [yes / no] children [date of birth per child] 	<ul style="list-style-type: none"> relationship status [married / legally cohabiting / non registered cohabiting / unmarried] children [date of birth per child] formal relationship to children [child / step child or foster child / child of the partner] 	<ul style="list-style-type: none"> relationship status [single / married / divorced / separated / cohabiting / PACS (civil partnership) / widowed] children [date of birth per child] dependants [number]

2.2.2 Student Data¹⁵

Each of the seven universities collects student data on age, nationality, gender, and preferred name. This mostly stems from operational (core business, administration, identification) or statistical interests, but in some cases this data is collected to fulfil a legislative reporting requirement.

In all other diversity categories, the practice of the universities varied significantly, but at least one university collected at least some data for each of the categories considered.

As has been noted elsewhere in this report, in all cases, the data gathered relate to composition of the population rather than the actual experiences of the persons concerned.

Students – Age

All universities gather age data, usually as part of their own student enrolment/matriculation procedure (at U Edinburgh the data collection is conducted externally for undergraduates, see Good Practice Example 6). JU Krakow, FU Berlin and U Helsinki are required to report student age data to a central statistical office. U Helsinki is required to report this data to both Statistics Finland and the Ministry of Education.

Age is always derived from date of birth (and verified against official identification documents); although data analysis at U Edinburgh focuses on age at enrolment.

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Age:	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]

¹⁵ Please note that many of the universities also conduct a range of student surveys that are not conducted by the central university administration and are not collated within the central data system. They do, however, often provide valuable additional diversity data. This mapping exercise was not able to take such survey data into account.

Students – Gender

All universities collect data on the gender composition of their students, primarily for the purpose of monitoring gender equity, e.g. across disciplines or faculties. JU Krakow, FU Berlin and U Helsinki are required to report student gender data to a central statistical office. U Helsinki is required to report this data to both Statistics Finland and the Ministry of Education.

Although a critical reflection of practice in relation to gender markers is starting to take place at many universities, student gender is defined in a binary manner in almost all universities, i.e. male or female. There are two exceptions: as with staff data, FU Berlin allows four options (based on the German legal framework: ‘female’, ‘male’, ‘diverse’ i.e. intersex/non-binary and ‘no entry’, see above); In contrast to its staff data, U Edinburgh allows a third gender option for students (‘other’) to allow for citizens of countries that allow more than male and female gender markers in their legal identification.

FU Berlin introduced a policy in 2022 that allows students to change their gender in the data system to better reflect their lived identity. There are no legal requirements for the student gender data reported to external agencies to reflect the official gender marker and, in fact, the gender markers entered on matriculation were not previously verified against official identification documents, even prior to the new policy. This effectively means that the gender data reflects gender identity rather than sex in either a biological sense or as per legal status (like many other languages, German does not distinguish between sex and gender). This flexibility was not possible for staff because data is linked to external agencies that rely on official gender markers for identification purposes (see Good Practice Example 1 and Appendix 5).

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Gender:	• gender identity [female / male / diverse (intersex/non-binary) / no entry]	• legal gender marker [male / female]	• legal gender marker (“ <i>legal sex</i> ”) [male / female / other]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female] (<i>third option under consideration</i>)	• legal gender marker [male / female]

Students – Gender Identity

As with staff data, only the U Edinburgh has specific data on students’ gender identity. This is gathered on enrolment via the UCAS form (see Good Practice Example 6). Students can update/change this information upon arrival. In the case of students, the data collected relates to whether the person changed their gender.

As noted above, since 2022 FU Berlin’s student sex data effectively measures gender identity since there is no verification against official sex and individuals are able to change their entry to better reflect their lived identity, however, it does not specifically measure which students identify as trans or non-binary (see Good Practice Example 1).

Each of the universities offers the possibility to students to enter their preferred name into the university’s data system, enabling them to be addressed by that name, at least in internal university documents and identification cards (see Good Practice Example 5). These policies address the needs of trans and non-binary students but are sometimes open to other students too (e.g. at U Edinburgh and FU Berlin). Without an official

change of name, however, the legislative frameworks appear to currently make it impossible for any of the universities except the FU Berlin to use this preferred name on legal documents, such as degree certificates.¹⁶

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Gender Identity:			<ul style="list-style-type: none"> gender identity in relation to gender assigned at birth, as per self-identification [same as gender assigned at birth / different to gender assigned at birth / refused to answer / not available] 				
Name:	<ul style="list-style-type: none"> legal name preferred name [firstname &/ or surname of choice] <p><i>(student can choose whether the preferred name is used for all purposes or only for online platforms)</i></p>	<ul style="list-style-type: none"> legal name preferred name [first name of choice] 	<ul style="list-style-type: none"> legal name preferred name [first name of choice] 	<ul style="list-style-type: none"> legal name 	<ul style="list-style-type: none"> legal name preferred name [firstname &/ or surname ending of choice] 	<ul style="list-style-type: none"> legal name <i>(used for official documents and data migration to external agencies where required)</i> preferred name [first name of choice] 	<ul style="list-style-type: none"> legal name preferred name [first name of choice]

¹⁶ In June 2022 the FU Berlin received confirmation from the State Government that it could issue degrees using the graduate's preferred name. This extends the provisions already enshrined in the state higher education law that require universities to use students' preferred names and genders in all internal university communication and documentation. While some debate continues in Germany as to whether there is potential for legal or other problems for either the issuing university or the graduate, several other universities in Germany already have policies to issue degrees in the preferred name. This is in line with a recent shift in public policy in Germany, including proposed legislation to allow an official change of name and gender based solely on a self-declaration.

Good Practice Example 5: Uniwersytet Jagielloński w Krakowie – Preferred names

As the Covid pandemic began in 2020, like many universities, JU Krakow transferred most classes to distance learning. The university identified and addressed the needs of transgender and non-binary people in relation to the digital services offered by the university. This group potentially faces significant stigma, discrimination and structural barriers in relation to their gender identity. Online learning platforms can increase involuntary outing of transgender and non-binary users by displaying official data that no longer reflects their day-to-day lived identity – in particular their former name.

The university looked at ICT systems in particular, and developed solutions enabling the use of the student's preferred name in their studies (including online and face-to-face learning environments) and internal communication. The preferred name is uploaded to the university's student management data system by the IT service staff. Both first names and the endings of surnames can be changed (many Slavic surnames have gendered endings). The chosen names are used in all internal university documents, including course

participant lists. However, the legal/official names must still be used for official documents such as degree certificates.

JU Krakow is a pioneer in Poland with regard to this solution, and other Polish universities are looking to adapt their system in line with JU Krakow's policy.

Transgender and non-binary students who wish to add a preferred name (and/or would like receive support more generally) can email or call a special unit – Department of Security, Safety and Equal Treatment – Safe JU. After presenting a certificate from the specialist supporting their transition or after a brief consultation (on- or offline), the applicant completes a form and the changes are entered into the data system. They are not required to provide any further documentation or other evidence of their gender identity.

A communication plan was developed to inform both staff and students of this solution, which not only enabled access for students and ensured staff were using the preferred names appropriately, but also raised awareness of the needs of trans and non-binary people in the university. Top-down support by the university leadership enabled smooth implementation.

At the end of each academic year, the JU Krakow evaluates its support services by sending an anonymous survey to all those who added a preferred name to the ICT system. This allows the university to improve this solution and further improve its support to transgender and non-binary students at the university more generally. The survey thus assists the university to identify needs of transgender and non-binary students, evaluate implemented practices to address those needs, and increase participation of the target group in the development of the academic community.

www.safe-student.uj.edu.pl/en_GB/support-for-transgender

Students – Sexual Orientation, Religion, Ethnicity / Racialised Identity

As with staff data, only the U Edinburgh has data on the sexual orientation, religion and ethnicity/racialised identity of their students. As mentioned before, they do not collect it themselves, but receive these data from the UCAS form (see Good Practice Example 6). Students can update/change this information upon arrival.

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Sexual Orientation:			<ul style="list-style-type: none"> sexual orientation as per self-identification [bisexual / gay man / gay woman / lesbian / heterosexual / other / prefer not to say] 				
Religion:			<ul style="list-style-type: none"> religion or belief as per self-identification [list^D, including: 'prefer not to say'] 				
Ethnicity / Racialised Identity:			<ul style="list-style-type: none"> ethnicity as per self-identification [list^E, including 'prefer not to say'] 				

List^D (U Edinburgh - religion or belief as per self-identification): No religion / Buddhist / Christian / Christian - Church of Scotland / Christian - Roman Catholic / Christian - Presbyterian Church in Ireland / Christian - Church of Ireland / Christian - Methodist Church in Ireland / Christian - Other denomination / Hindu / Jewish / Muslim / Sikh / Spiritual / Any other religion or belief / Prefer not to say.

List^E (U Edinburgh - ethnicity as per self-identification): White / White – Scottish / Irish Traveller / Gypsy or Traveller / Other White background / Black or Black British – Caribbean / Black or Black British – African / Other Black background / Asian or Asian British – Indian / Asian or Asian British – Pakistani / Asian or Asian British – Bangladeshi / Chinese / Other Asian background / Mixed - White and Black Caribbean / Mixed - White and Black African / Mixed - White and Asian / Other mixed background / Arab / Other ethnic background / Prefer not to say.

Good Practice Example 6: University of Edinburgh – UCAS Enrolment data

U Edinburgh is located in the UK, where registration is organised by UCAS (Universities and Colleges Admissions Service). When applying for university or college places, undergraduate students must fill in the UCAS form. This means that data is collected and analysed on a central level, but once a student is enrolled at a specific university, this university receives the data and can also use it for their own policy and analysis. They are required to provide the Higher Education Statistics Agency (HESA) with a yearly update of these data. HESA provides national benchmarking data.

Data on various diversity categories is gathered, including sexual orientation and gender identity (LGBT identity), ethnic group, religion, care leaver status (having been in the care system, relevant to the category family status) and living/schooling area background (relative disadvantage is used as an indicator of social status/class). An intersectional approach is taken, considering members of minoritised groups in relation to other categories of diversity too.

UCAS has devised the 'multiple equality measure' (MEM) as its principal measure of equality. It brings together several categories of diversity, for which large differences in the progression into higher education are known to exist. These include gender, ethnic group, where people live, secondary education school type and income background (as measured by whether a person was in receipt of free school meals, a means-tested benefit while at school).

Supplementing the reports and data sets provided to the individual universities, UCAS also publishes collated data, including breakdowns by diversity data. For example, since the 2018 application cycle, UCAS has published regular report on offer rates and placed applicants by sex, area background and ethnic group at 132 larger universities.

In addition to the application data, UCAS surveys particular groups of applicants who had identified as members of minoritised groups, gathering more in-depth data on their experiences and publishing findings and recommendations derived from them as 'Additional Insights' These have included data sets and reports on the experiences of disabled students in education and influences on the choices international students make (2021), on the experiences of LGBT+ students in education and student mental health (2020), and mature (21 years old and over) students (2018).

In some cases these reports are produced in collaboration with civil society organisations, for example the report 'Next steps. What is the experience of LGBT+ students in education' was published in partnership with Stonewall, a leading UK NGO representing and advocating on behalf of LGBTQ+ people.

www.ucas.com

Students – Migration

Only three of the universities collect any student data that relates to migration.

FU Berlin is required to report (and verify) data on the *country of university entrance qualification* to an external government agency. When linked to nationality, this data shows the number of non-mobile foreign students, i.e. those who have a foreign nationality but are domestic students. KU Leuven derives information about students' *migrant backgrounds* from data gathered in the student registration process: *country of university entrance qualification*, the *current country of residence* and the *birth nationality of the student, the parents and the grandparents*. The external AlmaLaurea consortium graduate survey provides AMSU Bologna with several indicators that relate to students' migrant backgrounds by collecting data on the *country of university entrance qualification*, the *current country of residence* and the *birth nationality of the student and their parents*.

U Edinburgh sees neither a legislative requirement nor a policy need and thus does not collect any data on migration. Neither does JU Krakow, where it is argued that this data is too sensitive in regard to data protection requirements and thus that there is no justified need. At U Paris 1 PS the Data Protection Officer shares JU Krakow's position, referring to article 9 of GDPR as justification for not collecting migration data.

Some universities offer pre-study programmes for foreign students who do not (yet) fulfil regular matriculation requirements, usually without feeding the data from these programmes into the main student data base. For example, U Helsinki offers the Academic Readiness Educational Programme (AKVA) for immigrants with a foreign higher or secondary education degree, however, immigration status itself is not registered in the university's data systems. The programme coaches participants applying to a higher education institution and neither leads to a university degree nor guarantees the right to study in a degree programme.

In addition, some universities have indirect information at their disposal because they provide support services that relate to migration. For example, AMSU Bologna offers *legal support for students in relation to immigration procedures and health care (on request)*. The data collected by these support services is, however, not collated and thus not available for further policy development.

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Migration:	<ul style="list-style-type: none"> • non-mobile foreign student, derived from the indicators: • Country of university qualification [value: university entrance qualification in Germany] • nationality [value: not German] 	<ul style="list-style-type: none"> • non-mobile foreign student, derived from the indicators: • Country of university qualification [value: university entrance qualification in Italy] • birth nationality [value: not Italian] • birth nationality of parents [value: not Italian] • current country of residence [value: Italy] 				<ul style="list-style-type: none"> • migrant background [EU1 country (List^f) / non-EU1 country / no migrant background / unknown / other] derived from the indicators: • nationality of the student • birth nationality of the parents • birth nationality of the grandparents • country of residence • country of university entrance qualification 	

List ^f EU1 countries (KU Leuven – migrant background): Denmark, Germany, Finland, France, Ireland, Iceland, Liechtenstein, Luxemburg, the Netherlands, Norway, Austria, United Kingdom, Sweden, and Switzerland.

Students – Refugee Status / Experience

U Helsinki and JU Krakow state that they do not collect data on refugee status/experience because they see no justified need to do so that would outweigh data protection/privacy concerns. KU Leuven collects information on residency status, which includes categories on refugee status (recognised refugee and asylum seeker). This is collected as part of the enrolment procedure in order to fulfil external reporting requirements, but is also used to invite refugee students to make use of student services and to follow up their study progression (as a group, not individually). U Edinburgh has unverified data on refugee/seeking asylum in the UK stemming from the external UCAS form.

U Paris 1 PS is the only university with a policy relating to students with refugee status, in line with French government policy. They have verified data, collected at enrolment, on students who enjoy special protection and have a status recognised by OFPRA (the French office for protection of refugees and stateless persons), and have a [degree programme](#) specifically for students with refugee/asylum seeker status or under subsidiary protection. The university also has a specific policy of welcoming refugees from Syria and now from Ukraine (in agreement with government authorities). The Ministry of Foreign Affairs has asked the university to provide information on the number of students from Syria who have been accepted, as it offers subsidies for new students of Syrian nationality.

Some universities, e.g. FU Berlin, have pre-study programmes specifically for recently arrived refugees, however, as noted above in relation to foreign students participating in pre-study programmes, the participants are not formally matriculated in the university so the admission data is not integrated into the universities' student data system.

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Refugee Status / Experience:			• refugee status [refugee / limited leave to remain in the UK / seeking asylum in the UK / none]			• residency status [List ⁶ , including recognised refugee / asylum seeker / other]	• official status as refugee / stateless [yes / no]

List ⁶ (KU Leuven - residency status): European Economic Area (EEA) nationality / unlimited residency permit in Belgium / non-EEA student visa / permanent family reunification / temporary family reunification / employment visa non-EEA / recognised refugee / asylum seeker / other / diplomat status.

Students – Nationality

All seven universities gather data on student nationalities at enrolment based on citizenship (verified against official documentation). FU Berlin and U Helsinki are required to report this data to a central statistics office, with U Helsinki also reporting it to the Ministry of Education. AMSU Bologna and KU Leuven also gather data on students' residency status.

FU Berlin also notes a specific policy need in relation to its profile as an international university and draws on the data to monitor representation of international students. However, the data is only informative to a limited extent since it is assumed that a significant number of domestic students do not have German nationality. The university does also gather data (as per reporting requirements) on the country in which the student gained their university entrance qualification, which in principle would enable most domestic foreign nationals to be identified, but to date no multivariate analysis appears to have been conducted.

As with staff nationality data, some universities allow for more than one nationality in each student record, others record only one.

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Nationality:	<ul style="list-style-type: none"> • current citizenship [country or countries of citizenship] <i>(multiple possible; German overrides the others in reporting)</i> 	<ul style="list-style-type: none"> • current citizenship [country of citizenship] • current residency status [resident in Italy / resident abroad] • citizenship at birth • citizenship of parents at birth 	<ul style="list-style-type: none"> • current citizenship [country or countries of citizenship] <i>(dual possible)</i> 	<ul style="list-style-type: none"> • current citizenship [country or countries of citizenship] <i>(multiple possible)</i> 	<ul style="list-style-type: none"> • citizenship [country of citizenship] 	<ul style="list-style-type: none"> • citizenship [country of citizenship] • international student [yes / no] • residency status [List ⁶] 	<ul style="list-style-type: none"> • citizenship [country of citizenship]

List ⁶ (KU Leuven - residency status): European Economic Area (EEA) nationality / unlimited residency permit in Belgium / non-EEA student visa / permanent family reunification / temporary family reunification / employment visa non-EEA / recognised refugee / asylum seeker / other / diplomat status.

Good Practice Example 7: KU Leuven – common framework for mapping student diversity data

As universities began to develop diversity policies in Flanders, starting around 2008, it became apparent that there was little data available to inform the development of these policies. Although every university had 'some' data available, often stemming from administrative or service needs, every university used different indicators, most of which also did not align with those used in secondary education and labour market organisations.

An Action Group in the Flemish Education Council (VLOR) had begun work in 2008 to map diversity themes and formulate suggestions for defining these themes; Based on an update of this work in 2015, in 2018 all universities in Flanders agreed to map diversity in their student influx through the use of similar instruments and definitions.

They also agreed to add this data to the Higher Education Databank (DHO), which contains all data on students' study pathways in Flanders. Some data was already recorded in the DHO (gender, nationality, age), however, in the case of additional indicators (e.g. regarding disability, migrant background, educational background of the parents), so far no agreement has been found with the government regarding the ownership of the data. Therefore the data is not (yet) part of the DHO. Despite this, the Flemish council of universities (VLIR) decided to collect these data themselves and to provide an annual infographic on student intake at Flemish universities and all universities commenced gathering the data 2018-2019.

Almost all of the agreed diversity categories and definitions in Flanders were already being used and collected at KU Leuven to provide an evidence-base of diversity measures prior to 2018. The introduction of a common framework not only affirmed KU Leuven's good practice, but also enabled comparability and benchmarking over time and across universities.

Students – Language

KU Leuven and U Helsinki collect data on the language of students as part of the enrolment procedure. At U Helsinki this is the *first language*, the *language of business* and the *language of high school education*. At KU Leuven, next to the required choice of a *communication language* (Dutch or English) for operational purposes, students are asked about their *home language*: At home, does the student only speak Dutch, another language next to Dutch or only another language? This question is surveyed as agreed upon with all Flemish higher education institutions (see Good Practice Example 7) and answers a policy need.

AMSU Bologna also gathers data on language, but only at graduation through the (external) AlmaLaurea consortium surveys: the *mother tongue* (one language) and (*level of*) *knowledge of other languages*. U Paris 1 PS provides the opportunity for students to tick a box on/at their examination to clarify French is not their *mother tongue*. This data is further not collected or collated, but is part of the university's policy. JU Krakow, U Edinburgh and FU Berlin do not have data on language at their disposal.

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Language:		<ul style="list-style-type: none"> • mother tongue [one language] • knowledge of other languages [language(s)] • level of knowledge of other language 		<ul style="list-style-type: none"> • first language [Finnish / Swedish] • language of business [Finnish / Swedish / English] • language of high school education [Finnish / Swedish / other] 		<ul style="list-style-type: none"> • language of communication [Dutch / English] • language spoken at home [only Dutch / Dutch + other / only other / no answer] 	

Students – Class / Social Status

All of the universities other than U Helsinki collect data related to class/social status (or socioeconomic status), usually derived from two or more indicators. The data is often collected as part of an application for financial assistance but also as part of a commitment to widening participation in higher education. Parental income or occupation and parental education, in particular whether they graduated from university, are the most commonly used indicators in relation to the diversity category class/social status. For example, U Paris 1 PS collects data on the occupational category of the parents at matriculation, in line with internal university policy and government reporting requirements.

Items on both indicators are included in the UCAS form used to collect data for U Edinburgh (see Good Practice Example 6): Students are required to provide information on parental education and parental employment to identify whether they are first generation students and to identify the socio-economic group of the student or their parents. It is also mandatory for students to provide their addresses, from which the Scottish Index of Multiple Deprivation (SIMD) is derived.

FU Berlin records data on the *university entrance qualification*, which they are legally required to do as part of the enrolment process. This is sometimes used as a proxy for social status in Germany, based on an

assumption that a vocational qualification is more likely to be an indicator of a working class background, whereas the more academic university entrance examination, 'Abitur', is more likely to indicate a middle class background. Further data, on subjective social status and on 'first generation' students, is gathered in regular student surveys.

Socio-economic status is considered in JU Krakow's policies (including in relation to the JU Scholarship Fund) and derived from data gathered on five parameters. However, information is only gathered from students who apply for financial assistance and the data is not collated.

KU Leuven uses the three types of data relating to class/social status, including the student's perception of their parents' financial background. This data is neither mandatory nor verified.

AMSU Bologna differentiates between the social class of the family, which is derived from a comparison between the socio-economic position of the father and that of the mother (a function of their last profession and qualification) and the economic situation of the family (this data is available only for students who have applied for financial benefits).

Students: Indicator(s) [& values]	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Class / Social Status:	<ul style="list-style-type: none"> • university entrance qualification (UE) [General UE ("Abitur") / General UE (foreign) / subject-related entrance qualification / preparatory college / technical college certificate / vocational qualification] 	family social class, derived from <ul style="list-style-type: none"> • parents' last profession [unemployed / retired / ...] • parents' qualification (<i>socioeconomic status of the parents is compared</i>) • family economic situation [ISEE "Equivalent Economic Situation Indicator" certification (<i>submitted by students applying for financial benefits</i>)] 	Socioeconomic status of family, derived from <ul style="list-style-type: none"> • parental occupation [list ^H, occupation categories] • first generation student, derived from • parents' higher education qualification [yes / no / don't know / information refused / no response given] • Scottish Index of Multiple Deprivation [home postcode] 		socio-economic status, derived from: <ul style="list-style-type: none"> • gross monthly income per family member [amount] • family composition [number and type of family members] • main occupations of adults in family [occupation categories] • social assistance [type of benefit] • alimony [no / yes, if so, amount] • farm [no / yes, if so size] • scholarship [no / yes, if so, amount] 	<ul style="list-style-type: none"> • first generation student [level of schooling parents completed] • student's perception of difficulty faced by parents paying for higher education [very difficult, difficult / quite difficult / quite easy / easy / very easy] • study grant from Flemish government [yes / no / discount in study fee] (<i>criteria include income of family members, number of children studying</i>) 	<ul style="list-style-type: none"> • official occupational category of the parents • eligibility for social benefits [financial situation of parents (copy of 'avis d'imposition')]

List H (U Edinburgh - parental occupation): Higher managerial & professional occupations; Lower managerial & professional occupations; Intermediate occupations; Small employers & own account workers; Lower supervisory & technical occupations; Semi-routine occupations; Routine occupations; Never worked & long-term unemployed; Not classified.

Students – Working Students

Both KU Leuven and U Paris 1 PS collect data on working students as a normal part of the enrolment procedure. While employment status is not a dimension of diversity as such, students often need to work due to their socio-economic status. Given that working students are likely to be disadvantaged in their studies, collecting data on student employment can assist in the development of policies to address related needs.

At KU Leuven there are three types of data available:

- the number of students who enrol in programmes that specifically target working students.¹⁷
- the number of students who apply to KU Leuven for status as a working student (in addition to governmental programmes, KU Leuven offers its own specific programmes and facilities for students who are registered as working students), and
- the number of students who indicate in the survey at registration that they combine studying with being an employee or entrepreneur

U Paris 1 PS has (voluntary) data on working students, which is verified to allow access to provisions in the Student Employees' Charter (see Good Practice Example 8). JU Krakow also offers provisions for working students, on application, but this data is only available in the individual student file and is not collated since no justified need has been identified. Finnish law sets strict limits on the collection of personal data and U Helsinki has identified no justified need to collect data on working students, who are treated the same as all other students.

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Working Students						<ul style="list-style-type: none"> • employment status [fulltime employment / jobseeker / self-employed / student assistant / none of the above] • enrolment in targeted study programs for working students [yes / no] • KU Leuven working student status [yes / no] 	<ul style="list-style-type: none"> • having a paid job during the academic year [yes, a regular job / yes, a casual job / no]

¹⁷ The regional government of Flanders encourages universities to provide programmes targeting working students by giving those students extra weighting in the governmental funding model. Entry requirements are that students do not yet have a masters degree and are currently employed or registered as jobseekers. The programmes must fulfil various requirements, for example, they must be timed to fit alongside employment (in the evenings, during weekends, grouped together on one day of the week or over a couple of weeks during the academic year) and use appropriate teaching formats (e.g. distance learning or digital education alongside or instead of face-to-face teaching).

Good Practice Example 8: Université Paris 1 Panthéon-Sorbonne - Data on working students

Each year, ORIVE (Observatoire des Résultats, de l'Insertion professionnelle et de la Vie Etudiante) conducts a survey on behalf of U Paris 1 PS. Respondents are first-year students (during their first week at the university), third year students and fourth year students. The survey includes questions on the respondents' living conditions and, more specifically, on whether they are in paid employment, on how they finance their studies and on their housing conditions.

In relation to their paid work, they are asked two multiple choice questions:

1) Do they have a paid job during the academic year? Response options: yes, a regular job; yes, a casual job; no. If it is a regular job, they are asked how many hours a week they work (responses: less than 10 hours, between 10 and 20 hours, more than 20 hours).

2) Do they have a paid employment during the university holidays?

The focus lies on those students needing to work during the academic year, as this is deemed to have a greater impact on their studies.

Analysis of the results of these surveys identified a policy need in relation to working students. It led to the establishment of a 'Charter for Students in Professional Activity', which aims to promote success and equal opportunities for students obliged to work in order to finance their studies (see www.pantheonsorbonne.fr/sites/default/files/202009/2019_regimes_speciaux_etudiants.pdf). Affected students are offered the possibility of pedagogical adjustments to facilitate their studies (e.g. early distribution of teaching schedules and knowledge control tests, priority choice of enrolment in tutorial groups, possibility of exemption from attendance at tutorials, online availability of course materials, etc.).

<https://data.pantheonsorbonne.fr/data>

Students – Disability and Health

Except for U Helsinki and FU Berlin, all universities collect data on disabilities in their student body. This data is only collected from applicants to the university, often after a verification process. It is not necessarily available for policy analysis since data is either not collated at all (JU Krakow), only available as generalised statistics (U Edinburgh) or not available in full detail (U Paris 1 PS). AMSU Bologna, U Paris 1 PS and U Edinburgh mention a legislative requirement: they refer to the [UN Convention on the Rights of Persons with Disabilities](#) and national laws, as well as HESA reporting requirements in the UK. U Helsinki outsources social services to student organisations (and national social security) so sees no policy need that would justify collection of this data.

None of the universities collect health data for students as part of their administrative data collection other than long-term health conditions sometimes collected within the disability category, although some do in research-driven surveys.

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Disabilities & Health Status		<ul style="list-style-type: none"> • type of disability [physical, mental or sensory impairments / intellectual disabilities / blindness / deaf and dumb / temporary illness / no entry] • Satisfaction with university support (<i>targeted survey</i>) 	<ul style="list-style-type: none"> • disability or long-term health condition [list¹] 		<ul style="list-style-type: none"> • nature of disability or health problem [visual / hearing / motor skills / mental health / autism spectrum / chronic illness / reading and writing difficulties / other / no entry] • degree of disability [light / moderate / significant/ no entry] 	<ul style="list-style-type: none"> • type of disability [list¹, including multiple] 	<ul style="list-style-type: none"> • eligibility for adjustments [yes / no]

List¹ (U Edinburgh – disability or long-term health condition): No known disability / Two or more impairments and/or disabling medical conditions / A specific learning difficulty such as dyslexia, dyspraxia or AD(H)D / General learning disability (such as Down's syndrome) / A social/communication impairment such as Asperger's syndrome/other autistic spectrum disorder / A long standing illness or health condition such as cancer, HIV, diabetes, chronic heart disease, or epilepsy / A mental health condition, such as depression, schizophrenia or anxiety disorder / A physical impairment or mobility issues, such as difficulty using arms or using a wheelchair or crutches / Deaf or serious hearing impairment / Blind or a serious visual impairment uncorrected by glasses / A disability, impairment or medical condition that is not listed above / Prefer not to say.

List² (KU Leuven – type of disability): physical disability / hearing impairment / visual impairment / chronic disease / developmental disorder (learning disability / autism spectrum disorder / AD(H)D / tic disorder / developmental coordination disorder / developmental dysphasia / developmental stuttering / combination developmental disorders) / psychiatric disorder / other / multiple / no entry.

Good Practice Example 9: Alma Mater Studiorum Università di Bologna – Disability Data Collection

At the AMSU Bologna, data is gathered by the Service for Students with Disabilities and Specific Learning Disorders (SLD) within the Sector of the Right of Education. This office carries out support activities for students with disabilities and SLD, from entry to the University and throughout their studies. The data is collected for internal university reporting and the national reporting required by law. It is used to assess and establish the service and support for the student with disabilities.

Furthermore, every year, students with disabilities are asked to fill out a questionnaire of satisfaction concerning university support. The questionnaire is an instrument to evaluate and improve the services offered to students with disabilities.

www.unibo.it/it/ateneo/amministrazione-trasparente/servizi-erogati/carta-dei-servizi-e-standard-di-qualita

Students – Family Status

Social services for students at the U Helsinki are provided primarily by student organisations rather than as part of the core operations of the university, therefore the university has identified no policy needs around collecting data on student family status. FU Berlin and AMSU Bologna also gather no data on students' family status (but do provide support services). JU Krakow collects data on the composition of students' family members (whether they have parents, siblings, children and spouses). KU Leuven, U Edinburgh and U Paris 1 PS gather data on relationship status, but all three also use further indicators: KU Leuven asks whether students live with their parents, their mother, their father, independently or otherwise. U Edinburgh provides the option for students to make themselves known as a carer of dependent young people/children or other relatives/friends and asks about their care leaver status (care experienced at children's home, foster care, etc. for any time between ages 0-18).¹⁸ U Paris 1 PS also gathers data on students with dependent children.

In relation to data collection on family members (in this category as well as others), some universities mentioned discussions taking place regarding the replacement of the gendered terms 'father' and 'mother' by 'parent 1' and 'parent 2' in their data systems, acknowledging that not all families are headed by a heterosexual couple.

¹⁸ Collection of data on students' Care Leaver Status is a legislative obligation in the UK. The data is gathered via HESA and the UCAS-form (see Good Practice Example 6).

2.3 Strengths and Weaknesses in Current Practice

Each university has its individual areas of strength, some of which have been illustrated in the good practice examples included in Section 2.2 above. These examples are by no means exhaustive and there are many further examples where individual universities have effectively gathered and used diversity data to progress the achievement of their diversity-related goals.

When general tendencies across the seven universities are considered, several areas of strength can be identified. Most universities consistently monitor at least some diversity data, in particular indicators of sex, age and nationality as part of their regular administrative data collection for both students and staff. This data often stems originally from administrative operational needs (e.g. identity checks) and/or external reporting requirements and was later drawn upon by the university for its own diversity monitoring and in some cases expanded upon in ways that reflect the university's own priorities. Some universities also gather data in relation to students' social status or class, often in surveys. Given that sex/gender and social status in particular are acknowledged to be major barriers to accessing higher education, this consistent monitoring is a strength that can be further built on. The data, when analysed, allows evidence-based institutional policies to be developed and effective measures to be implemented to reduce the barriers faced by some groups of people in higher education – both (potential) students and staff.

Most universities have considerable data on students at their disposal, in part due to legal and regulatory requirements, or other (semi) external reporting commitments, such as membership in university consortiums, etc. In some cases, sensitive data is collected by trustworthy external institutions, like UCAS in the UK or AlmaLinea in Italy, boosting trust that data privacy will be adhered to and potentially increasing response rates and the validity of the data.

Surveys, specifically designed by university departments or individual researchers, often add significant depth to administrative data or fill in gaps. Student surveys are relatively common, and provide important information that can be drawn on to identify and address institutional needs.

The focus of the Action Group's work was, however, to investigate issues that currently hinder the collection and use of robust diversity data. The following interrelated areas appear to be weaknesses in current practice. They were not systematically analysed but do appear to be relevant in some way to most of the seven universities (noting that in each case there are exceptions) and were taken into consideration when drafting a possible template for future use in diversity data mapping (see Appendix 3).

- Weak Reasons for (Not) Collecting Data

Perceived legal constraints regarding data protection is probably the most frequent reason put forward for why data is not collected on a wide range of diversity dimensions. The entry into force of the European General Data Protection Regulation (GDPR) in 2018 has been interpreted by some of the universities as a hard barrier to the collection and processing of personal data. There appears to be a common misunderstanding that it is unlawful to collect personal data that is not specifically required by external/ government agencies.

In fact, the GDPR merely stipulates there must be a legitimate interest in the processing of personal data, and that where such an interest exists, measures are taken to protect the privacy of the individual in relation to the

collection storing, analysis and reporting of the data. There must therefore always be reasons to justify the collection of personal data and the collection can only start once this question has been answered.

It appears that in at least some cases limiting data collection to that which is required by external agencies is considered a more comfortable default position than undertaking a differentiated consideration of risk mitigation and benefits and developing data collection approaches that comply with both the letter and the spirit of data protection requirements.

While the often gaping gaps in diversity data collection are most often explained with reference to data being 'sensitive' i.e. constraints due to data protection, there are other reasons too. In particular, most of the data collected is driven by the needs of the government (and thus reporting requirements) rather than the university's own needs.

This is no doubt associated with the fact that most universities have only begun developing strategic diversity goals recently, if at all. Without such a framework, it is difficult if not impossible to determine robustly what types of data is needed to support the development and implementation of measures to achieve diversity goals.

Even when it is known that there are good reasons to gather data, most universities gather no data on some dimensions of diversity. For example, no data collection is done in all but one of the member universities on racialised identities despite the significant policy need to monitor and understand racism within the university – a form of discrimination that is known to have a significant impact at both an individual and a structural level in higher education. While the reasons given often relate back to questions of sensitivity and the protection of privacy, there are likely to be other reasons at play too, such as the inherent difficulty of developing robust and meaningful indicators of highly complex and politically fraught social constructs such as 'race', let alone defining questions and response options (values) that would be acceptable and comprehensible to a highly diverse target group. Any attempts to gather data that relates to identities and deeply rooted social inequity will be imperfect and on some level problematic, and many stakeholders will be wary of engaging with it. At best, weak proxy indicators of 'race', such as nationality or, in some surveys, 'migrant background' are collected in the universities. Limited data on related indicators is sometimes collected for situational reasons (e.g. on refugee status to support implementation of measures to support students or academic staff fleeing war-torn countries in specific programmes) but not in a structural manner, i.e. to assist the development and implementation of measures to reduce inequality or foster the participation of all.

Where data is collected it usually relates to the composition of the staff or the student body, i.e. measuring the representation of members of certain social groups in the university (not measuring those who have been excluded from the university). The data most closely related to experiences, and in particular to discrimination is often not collected both for data protection reasons and because it is methodologically more challenging, even though it is precisely the data that we most need to identify, monitor and try to resolve (potential) discrimination.

- **Weaknesses Related to (Lack of) Utilisation of the Data Collected**

Data collected to fulfil external reporting requirements is often not really 'used' once that requirement has been complied with. Again, this relates in part to weak connections to actual policy and operational needs (if data is not fit for purpose it is hardly surprising that it is not utilised).

Diversity is a cross-cutting issue in universities that needs to be addressed by all staff within their own areas of responsibility. Yet diversity data collected (for any reason) is often not disseminated widely within the university. Too often, data is not actually available for policy analysis, monitoring, reporting and benchmarking, or it is

available in principle but hard to find even if the person knows to look for it. This often relates to a scatter gun or incoherent approach to data collection whereby a wide variety of stakeholders collect data and there is no central, pooled information about it (see below). But it can also have to do with other factors, such as privacy concerns resulting in data only being made accessible to a limited number of persons under specific conditions. This is not necessarily a weakness – in terms of data protection it is entirely appropriate to limit access to a needs-to-know basis, however, often the potential to utilise even sensitive data in a way that protects privacy while also serving the broader interests and needs of the university is not fully exhausted.

Diversity data is collected in a range of contexts by different stakeholders in each of the universities, who are sometimes unaware of each other's work (see below). However, the respective data sets produced must be stored safely and are seldom accessible to other stakeholders, even after being cleaned and anonymised (see Section 3.2.3.). Therefore, while linkage between the data sets does occur in some cases, it is rare and, in many cases, technically and methodologically impossible in the universities, despite there being obvious benefits in terms of adding depth to the evidence base, especially in relation to intersectionality. Few universities appear to have procedures in place to enable data sharing while ensuring that quality criteria are adhered to (e.g. an application and approval process). Similarly, while triangulation of methodologies was not explicitly sought in the mapping exercise and does likely exist, especially within research projects, the impression gained was that it is not a common approach in the field of diversity data collection (see Appendix 4).

Sometimes data is available to staff on internal databases, however, comparatively few stakeholders have either the technical skills to access the data (even when efforts are made to create user-friendly tools as an interface) or to analyse and interpret the findings (even in those cases where information and training are available). The mapping exercise indicated that few universities have interactive tools that allow comparatively unskilled users to generate reports on data of interest to them (such tools do exist, but must overcome methodological challenges, such as meeting pseudonymisation standards that ensure individuals cannot become identifiable in small populations). Even when the data is made more accessible, for example, in reports, such documents frequently present collated results without analysing them, let alone linking data and relating findings to the work and objectives of the university or even communicating exactly what the data means and what its limitations are.

Weak Diversity Indicators

Definitions of the dimensions of diversity are often non-existent or weak in the universities, and the phenomenon of interest to the university in relation to them are often not clearly defined. The indicators used in data collection are frequently not defined at all, are not clearly defined, are not communicated to the end users and/or are inadequate for providing information about the complex phenomenon of interest. It appears that in some cases the indicators are considered an end in themselves rather as a means to gain insight into the complexity of lived social reality. Like other issues, this also relates to the weaknesses in formulating clear policy and thus information needs. If we do not know exactly what phenomenon we are looking for, we can hardly be expected to define robust indicators for them.

This area of weaknesses relates in part to operational issues - in each of the universities a range of stakeholders were engaged in the collection of data, often based on quite different drivers and areas of expertise. The Action Group also identified various constraints on data collection, such as respondent burden, accessibility/comprehensibility for the target group, and data protection, which posed practical challenges and limited the scope of instruments to include differentiated items or a range of items that could then be linked. This affects both administrative data collection and research-driven survey data.

The lack of robust and agreed-upon indicators for most categories of diversity was also evident in the lack of comparability across the universities given significant differences in their legal, social and academic contexts.

More importantly however, this area of weakness also reflects a lack of scientific consensus on robust indicators for the various categories of diversity that extends beyond these seven universities. It relates to a fundamental conceptual challenge, for the dimensions of diversity are dynamic, interlinked social phenomena and any indicator is inherently inflexible, providing inadequate proxies for the complex reality on the ground or the experiences of members of the university..

Even seemingly straightforward indicators such as 'gender' (see Appendix 5), 'international student' or 'international staff member' quickly become porous when looked at carefully: Many people do not have the nationality of the country they are permanent residents of, so using nationality as an indicator will include both locals and people who have moved to the respective country for the specific purpose of studying or working at the university. There may even be different definitions of the term within the same university. The weak definitions (and the non-harmonisation of definitions between the different universities) means that results are often not robust and any analysis requires numerous caveats.

- **Lack of a Coherent Approach to Data Collection**

The ownership, collection and utilisation of data generally and diversity data specifically was often scattered across a university (and beyond). This appears likely to contribute to a lack of transparency and consistency (and thus comparability) and to pose a challenge to efforts to improve data quality.

A wide variety of practice within one organisation is not necessarily a weakness, in fact, in some cases it is a strength, and demonstrates, for example, that the practice is highly targeted and thus relevant to the specific context and need. However, differences of practice within one university can be a weakness when it reflects the lack of a coherent approach and in turn contributes to it. The AG identified significant heterogeneity in relation to the following factors, in addition to those already discussed above:

- **Location and ownership of data collection.** The responsibility for the formulation of the need underlying the collection is often located in a different part of the organisation than the implementation. The users and targets of the data are also often not involved in the design or evaluation of the methodology. Data is collected at different levels, e.g. university-wide, within a faculty, or within a specific programme or service as part of its own documentation, within a specific research project, etc. Responsibility for collection can lie outside the university, within a specific administrative or academic unit, with an individual researcher (e.g. as part of their dissertation), etc. There are also significant differences in the subject matter expertise, and the skills and knowledge of the people responsible for the different aspects of data collection so the typical communicative and translational challenges faced in any type of interdisciplinary work are evident in diversity data collection too, or at least they are when the different areas actually work together (this is frequently not the case).

There is seldom any central overview of who is collecting what, and why, let alone a framework of quality control. This may contribute to poor comparability, inefficiency and under-utilisation of data and, where data on similar indicators is being gathered in parallel processes, it adds to respondent burden or the need for individuals to have to 'retell their story' numerous times, something that can be very stressful.

- **Instruments with which data is collected and stored.** Data is collected in a wide range of ways, which each have some impact on the results and their quality. For example, it can be gathered in survey questionnaires, in application forms, in HR staff forms or derived from letters of request or service notes. It can be documented in excel files or even word documents (e.g. individual staff or

student files) or entered into sophisticated integrated data management systems, such as SAP. The instruments used can be on paper, a digital file or an online platform (e.g. student self-service portals, surveys and polls). Often the choices made about instruments are pragmatic rather than reflecting robust consideration of what is fit for purpose.

- **Timing.** Data is collected in different phases, repeatedly or one-off. This is often not transparent and not taken into account in interpreting the data. Some types of data are continuously updated to reflect changes (e.g. if it has an impact on income tax, staff in many organisations will inform their HR department of changes in their family status) while other types of data are only collected once, reflecting not only a specific moment in time but also a specific position in the person's career or study pathway. Data collected in different phases is seldom differentiated. Surveys sometimes target students at a specific point in their studies, e.g. at intake, in 2nd year or at graduation, but often do not, meaning that the responses of first-year students might be equated with those at the end of their studies. This might mean, for example, that questions about whether students know of support services in case of discrimination (a useful item to gain insights into service access and thus needs) are interpreted incorrectly if the sample is skewed to those who have already been in the university for several years.
- **The relationship to external data.** Data can be linked to, migrated to or verified with data in other internal or external (often government) databases, or it can be kept separate, either deliberately and carefully or simply because no one has considered linkages. Aside from implicit reference to approximately half the population being female, universities seldom report or interpret their data or set benchmarking goals with reference to whole of population data. This may be entirely appropriate as the indicators and values used may not in fact be comparable, however, it appears likely that this is not always a decision based on methodological or strategic considerations. In some universities a lot of the student data was gathered collectively across most or all universities in the region or country. In these cases, reference data is available at that level, which can be of great value. Comparable data on university staff in relation to sex/gender are often available, and accessibility will likely increase as the new EU Horizon funding programme criteria (including the requirement to publish sex/gender data) have an impact, however, there does not appear to be any comparable data on other diversity indicators among staff available.¹⁹ All of the universities participate in at least some international and national rankings,²⁰ but do not consistently drill down in detail and utilise the results for diversity-related strategic or policy development, e.g. diversity-related benchmarking.

19 The FU Berlin is a member of a national working group within the German Network '[Diversity in Universities](#)' seeking to develop and implement a national instrument for diversity climate surveys of university staff, however, as of mid-2022 the group has been unable to acquire the funding required to develop the concept further and design the instrument.

20 All seven universities participate in the [Times Higher Education \(THE\) World University Rankings](#), for example (and, incidentally, are either the highest ranked overall in their respective countries or at least in the top seven). The THE Rankings include various items with relevance to diversity. For example, the THE Impact Rankings assess universities against the United Nations' Sustainable Development Goals (SDG), including SDG 3 good health and well-being (indicators include the health of students and staff), SDG 5 gender equality (indicators include the recruitment and promotion of women) and SDG 10 reduced inequalities (indicators include the recruitment of staff and students from under-represented groups).

3. Issues

As the Action Group conducted its exploratory mapping, members continuously identified and discussed issues that appeared to go beyond the specifics of practice within these seven universities.

3.1 Overview of Issues Identified

Conceptual Issues

- The meaning & significance of diversity dimensions/categories change according to context and over time
- Associated with this, different meanings are ascribed to different terms in different disciplines, languages²¹ and societal contexts
- Indicators only inadequately reflect the actual complexity of the respective categories and are often assumed to equate to the category as a whole rather than one aspect of it
- The concept of intersectionality contrasts not only with existing structures and legal frameworks, which are generally based on an additive model, but also with the necessity to separate phenomenon into clearly delineated categories in quantitative research
- ...

Ethical and Strategic Issues

- Data is often treated as an end in itself rather than a means to an end
- In whose interests are data gathered? When is it perhaps in the interests of minoritised groups not to collect data?
- If numbers are small then the needs of the group are often not considered a policy issue. Potentially data could be used to hinder rather than promote diversity initiatives.
- High levels of sensitivity/distrust around the collection of ethnicity data given the history of racist misuse of such data by governments and other agencies
- Tensions between data protection and the need to provide granular data to promote action
- What happens to the data? Do those from whom it was collected have access to it?

²¹ Quite apart from differences between a lay person's understanding and use of a term such as 'class', within the academic community understanding will differ depending on whether it is located in economics, political science or social psychology, and even within those disciplines meaning is contested. When we consider linguistic differences the issue becomes even more complex. For example, in English 'sex' and 'gender' carry interrelated but different meanings (even if they are often used interchangeably), however, not all languages make this distinction. Similarly, the term 'race' is commonly used in the UK and USA and it could be argued that in mainstream discourses in those contexts is at least a basic understanding that this is a social construct. In many European countries the discourse around race and racism, in contrast, led to a rejection of the term, due mainly to its implication of biological differences and the consequences this had in the holocaust. As a result the language used to describe this social phenomenon is contested and diverse. Such differences in terminology are intertwined with conceptual debates within both academia and society at large and not only influence the discourse within a particular language but also pose challenges for international engagement, which usually occur in English.

- How to involve the students and staff from whom we are defining, collecting, analysing and using data? Especially members of minoritised groups. Participation should be equitable (e.g. who can participate in paid time? who gets what out of it?), safe (who risks what?) and, potentially, empowering.
- How long data should data be kept?
- Without top-down commitment it is hard to prioritise diversity related activities
- Data is often not really 'used', reports seldom analyse findings and relate them to the university's work and goals (often they just present them)
- To what extent do universities prioritise a robust data for evidence-based policy development?
- How does diversity data collection align with the institution's strategic goals?
- ...

External Legal and Regulatory Environments

- Legal requirements to report specific data to various agencies
- Diverse legislative and regulatory frameworks impact on comparability
- Antidiscrimination legislation provides a framework for the categories of interest and justification for data gathering, but also some constraints, e.g. in relation to the categories perceived as relevant and concerns that the collection or use of data may itself be discriminatory
- Data protection law at EU and national level is often perceived as a barrier to the collection of diversity data
- The wording of collection categories and indicators required by external agencies may differ from that accepted and/or understood by respondents and data users
- Universities might have data that they are not legally permitted to use because they had not asked permission to use it for that particular purpose (permission is a requirement of data protection regulations)
- ...

Methodological Issues

- There is a lack of (robust) reference data for the society as a whole to aide data cleaning, interpretation, benchmarking, etc.
- Demographic data is easier to gather than data about experiences of inclusion or exclusion
- Qualitative data vs quantitative data - how can it be effectively interlinked?
- Data from administrative units, research-driven surveys, and international projects is seldom linked and often not comparable. Among other things, the depth of data varies significantly
- Usability: are the methods generating data fit for (operational) purpose and is the data reported in a way that is comprehensible for and accessible to all stakeholders (e.g. visualisation)
- (In)consistent and often weak definitions, items and indicators across the university and between institutions.
- Loss of data quality vs loss of comparability when definitions, items and indicators are inconsistent with external (e.g. government) definitions (hinders benchmarking and monitoring) (*see also external environment*)
- A policy for definitions or a data dictionary can ensure comparability but also reduce self-critical reflection and can be inflexible to contextual issues and changes over time, including lessons learnt. When does improving data quality outweigh comparability?
- Intersectionality is hard to operationalise/measure. Multivariate analysis offers potential to provide insight into intersectional phenomena but is seldom conducted.

- Self-identification of respondents in relation to membership of social groups, the terms used, etc. vs external criteria and terminology. This is a particular issue where external (but problematic) items and indicators provide the only available reference data
- Administrative statistics (whole of population) vs surveys (samples)
- Some data is collected on a voluntary basis, which means the data might be incomplete or inaccurate (especially in the case of sensitive data).
- High rates of non-disclosure of some equality data
- The balance between gathering differentiated information on complex topics and limiting respondent burden
- The instrument and context of data collection as well as the perception of the organisation collecting the data impact on the responses
- ...

Practical Issues

- Where does the responsibility for diversity data collection sit within the institution? Is it firmly anchored and prioritised?
- Stakeholders' skills to interpret and work with quantitative data are often weak
- Dedicated resources are required to collect, interpret, communicate and use data, including staff capacity to do this core work and financial costs (both gathering and buying data)
- Integration of diversity indicators into existing instruments (e.g. course evaluation forms) is efficient but often faces barriers
- ...

3.2 Discussion of Issues

This section shares some thoughts in relation to selected issues that the Action Group was able to delve into in more detail, based on but moving beyond its reflection on the findings of the practice-mapping exercise. It lays no claims to being exhaustive or reflecting current debate or research on these topics or, indeed, even all discussions and perspectives of members of the Action Group.

3.2.1 Needs-Driven Data Collection

An increasing emphasis is being placed on acquiring and analysing data within university administrations. However, as noted above, data is not an end in itself but rather a means. Universities collect data for various reasons, e.g. (1) to gain a general overview of the organisation, including key performance indicators, (2) to inform organisational reform and development, by identifying or clarifying needs, gathering specific information that assists the planning of measures to meet known needs, or monitoring the impact of measures (3) to fulfil external reporting requirements, which are often national legal obligations (4) to participate in external programmes e.g. university rankings or funding programmes, (5) because it has 'always' been done, and (6) because it 'seems interesting', i.e. for no specific purpose.

The last two reasons, data collected just because it has been done that way in the past or for its own sake, should be avoided. Of course, sometimes specific needs are only identified as a result of the data collection rather than being clear at the outset, but generally even in this case the data has been collected purposefully,

either as exploratory research to identify needs in order to inform organisational reform and development (2 above) or for another purpose and the findings have indicated another area of need not previously considered, as an unexpected bonus. Unsystematic data collection that is not associated with any specific purpose not only potentially wastes organisational resources and represents an unnecessary respondent burden, but also yields data sets that often end up posing more problems than providing answers. For example, inadequately considered needs often result in poorly considered indicators and instruments and, in some cases, the organisation retains these in the interests of consistency and thus comparability over time, despite the data generated having little or no value or potentially even being misleading.

The question of who defines needs is significant here. Marginalised groups need to be involved in the definition and assessment of needs if they are to be identified accurately and in the interests of the groups themselves (in line with the vision of diversity outlined in Section 1.3.1). In order to ensure that those needs remain at the forefront – and to ensure that the data collection is fit for purpose - this participation should extend to all stages in the process, including decision making, survey design and implementation through to interpretation of findings and evaluation.

Diversity data is a relatively new area of interest and growing field of data collection. Just as with other types of data, it is useful to first collect and analyse data to gain a general overview of the possible diversity-related patterns that exist. This relies of course on a clear conceptual understanding of what is meant by diversity, i.e. what the object of investigation is. Patterns can be, for example, positive or negative trends of something that is desired (e.g. representation of minoritised groups within the university) or something that is unwanted (e.g. discrimination, barriers to participation). If a pattern has been found, further data may be required to confirm the existence of that pattern and/or to gain some more information about it.

Diversity data can be both a driver of institutional reforms or interventions (i.e. data indicates a policy issue that needs to be addressed) and an empirical basis with which to inform institutional developments (where the drivers may, for example, be political or strategic rather than empirical). It is good practice to at least try to evaluate the effects of interventions or other reforms. On this basis, decisions can be made regarding future planning. Incidentally, the same applies to the data collection itself – its quality should also be evaluated, in particular in relation to whether it was fit for purpose, i.e. met the needs it was intended to.

Most diversity data collection currently measures indicators of representation, i.e. how many people with a certain characteristic are present in the university (e.g. women, people with disabilities, first-generation students). This enables the university to monitor who has access to the institution at all or, for example, who has access to (or is excluded from) leadership roles, etc. This logic underlies many of the government reporting requirements in relation to diversity. Knowing who is there is helpful if we seek to identify barriers to access or factors that lead to leaving or career stagnation and to assess the impact of measures to reduce them. However, this is only one aspect of the information needs associated with diversity. Other needs relate in particular to understanding the actual experiences that people have within the university. Just because people of colour have made it into an institution, for example, does not mean that they do not experience significant racism that harms their wellbeing and makes it much harder for them to do well in their role or studies. To date, little diversity data is gathered that meets the need to understand and reduce individual and structural discrimination beyond questions of access.²²

22 Akel (2019) is a comparatively recent example of data collection that does drill more deeply into issues, however, while supported and published by the university, like much available data on experiences, it is a one-off research project rather than being embedded in regular administrative data collection. Based on data available on representation and attainment levels of Black and minority ethnic (BME) students at Goldsmiths, University of London, Akel's study used several methods to gather data on BME students' experiences in relation to four areas: decolonality and representation, racism and microaggressions, academic attainment, hate crimes and mental health support (and impact on confidence in the institution).

In practice, diversity data is in fact too often an end in itself, in that is not really 'used'. For example, when data that has been collected is never reported anywhere or analysed, let alone related clearly to the work and goals of the university and the needs associated with that (such as understanding what barriers block people from accessing higher education).²³ For example, open-ended answers to questions about experiences are sometimes collected in a survey but not analysed or reported, perhaps because text analysis requires significantly more resources than quantitative statistical analysis. Sometimes findings are reported but no action is taken even if analyses highlight diversity-related problems. This can easily occur if the purpose of data collection is to fulfil external reporting requirements, but the university has not identified any internal need for the data.

Given the significance of evidence-based decision making, quality assurance and strategic development at universities, a systematic and purposeful engagement with diversity data is essential. The knowledge derived from the data can contribute to the improvement of the functioning of universities, including the improvement of the quality of education, the reduction of barriers to equitable participation and strengthening the wellbeing of both staff and students.

Often, universities gather significantly more data on their students than on their staff. These groups, presumably, have very different compositions in relation to dimensions of diversity as well as distinct roles and structures within the university. They thus have their own characteristics and needs, and the same degree of focus should be placed on collecting, analysing and using data on both staff and students. Not only because this is in line with the vision of the university as an environment that fosters diversity and combats discrimination but also because the success of the university in terms of high-quality education and research is dependent on reaching and engaging effectively with a diverse student body, and this in turn depends on the staff that deliver its activities.

We need to ensure that the data is purposefully collected and used, based on the needs of the whole of the university community.

3.2.2 Interests Underlying Data Collection

As noted above, external reporting requirements drive much diversity data collection. Much government monitoring of diversity data serves the aims of reducing inequity, and planning resource allocation. This usually aligns fundamentally with universities' own goals, however, a range of interests can underlie data collection and they are not always clear cut, transparent or identical.

Data is often collected in a unilateral way, to meet an organisational need sometimes far removed from the needs of the people providing the data. Whereas companies can provide discounts, membership or loyalty bonuses to those giving them data, universities and government agencies often request or even require data from people with no obvious or tangible return. Ideally institutions use (or try to use) the data in a way that benefits people, but aside from the provision of data being a requirement of certain procedures (e.g. applications to study or work in the organisation or to obtain a benefit), there is often no direct or observable gain for the people providing their data. This has been criticised as 'data-colonialism'. The Action Group emphasises the moral duty to use personal data for the tangible benefit of the people themselves, as well as the moral and legal duty to communicate the purpose underlying the data collection when collecting it, for example, monitoring inequity within the organisation. Encouraging members of privileged or majority groups to

²³ Of course there are times when respondent numbers are so small that information on them cannot be published due to privacy issues, at least at a high level of granularity (detail).

voluntarily disclose personal data may well pose specific challenges because it has fewer tangible benefits for them than for members of minoritised groups.

Not only is it crucial to be transparent about the purposes and benefits of data collection, it is equally important to make sure that data collected reflects people's lived experience as accurately as possible and provides an evidence base that actually assists their needs to be addressed (within the overarching framework of diversity goals). Social reality is complex, and the same diversity issue may mean different things to different universities and/or people. Diversity is a multidimensional concept (see Section 1.3.1) and the data collected and the indicators used necessarily simplify reality, providing limited insight into the actual issue. This is sometimes exacerbated by reporting requirements that oversimplify reality even more. The indicators and values used seldom do justice to people's day-to-day experiences (e.g. the distinction male/female does not reflect all gender identities), which can make it harder for people to see the relevance of the data collection to their own interests.

Related to this is the fundamental question touched on above as to whether both the collection process itself and the use of the data actually serve the interests of minoritised groups and empower them, or whether they perhaps themselves contribute to exclusion, minority stress²⁴ and individual and structural discrimination. For example, when stigmatising terms are used in forms, when response options invalidate or render invisible respondents' experiences, when people are asked to explain themselves and their circumstances repeatedly or to provide information related to negative experiences, such as discrimination, violence or health issues, this has a negative impact. Often surveys that include items on potentially triggering topics include information about support services to assist in such cases. Even if such mitigating measures are taken, the core question of cost-benefit remains.

Universities commonly distinguish the populations of interest by status group, i.e. students and staff, or by students, academic staff/faculty and non-academic staff, etc. The Action Group found that the data collection on staff members and students differs both in terms of the number of issues that are considered relevant and in relation to the reason why data collection occurs. In a number of cases, there are more legal requirements for universities regarding collecting data on students, including diversity indicators, than there are regarding staff. This could be interpreted as prioritising student diversity over staff diversity, but it is crucial to stress that inclusion and equity are relevant policy issues with regard to staff as well. Quite apart from broader diversity goals such as equity and social justice, a diverse staff body is more likely to foster an inclusive and equitable environment for diverse students that enables them to succeed. Thus, gaining the knowledge required to acknowledge and foster diversity (including reduction of inequity) is in the interests of a university seeking to fulfil its core aim of effectively providing higher education.

It is widely acknowledged that there is great need for diversity policies and most universities appear to have good intentions in this respect, but currently these (emerging) policies too often tend to be based on vague definitions of diversity, to lack clear (measurable) objectives, and to be grounded in assumptions rather than facts, while at the same time failing to be of day-to-day relevance to people, in particular people from minoritised social groups. The Action Group therefore feels there is a need to move away from the question "What do we need to report?" and to take steps towards taking a more relevant question as point of departure: "What data can we collect to create value for our communities (and make data collection an empowering experience)?"

24 Minority stress is a conceptual model that describes stressors related to stigma, exclusion and discrimination faced by members of minoritised groups as causes of both physical and mental health problems and impacting negatively on wellbeing in general. See for example Meyer/Frost 2012.

3.2.3 Data Protection

The [European General Data Protection Regulation](#) (GDPR) came into force in 2018 and has been implemented in national legislation throughout the European Union. The GDPR introduced a framework of compliance and liability mechanisms to protect the privacy of individuals in relation to how their data is gathered, used and stored. It appears to place constraints on the data we can collect in relation to diversity, and therefore on its processing. As a result, many universities explain their lack of diversity data by stating that collection is not possible for reasons of data protection.

However, the collection, analysis and use of personal data²⁵ is possible if a number of factors are adequately addressed, in particular:

- Consent: before collecting personal data, the consent of the data subjects must be obtained
- Transparency: the data subjects must be informed in clear and understandable terms about the use of their personal data
- Lawfulness: a legal obligation or a legitimate interest (including public interest) as per GPDR Article 6.

Article 9 GDPR deals with the processing of particularly sensitive data (such as racial or ethnic origin, political opinions, religious beliefs, genetic or biometric data, etc.). Such data must be accorded additional protection: the consent given must be explicit and the interest must be substantial, well founded in law, with processing proportionate to the purpose.

When the collection of personal data appears to present a high risk to the rights and freedoms of individuals, the Data Protection Officer is required to conduct a privacy impact assessment. This includes examining potential risk scenarios (e.g. breach of confidentiality, availability or integrity of data), and their potential impact on the rights and freedoms of individuals, the threats that might result in such an event occurring, and an estimation of the likelihood and the severity for the individuals concerned.

The main requirement is to be transparent and explicit about why we need to collect personal data (this is a matter of accountability and ethics) and to take data privacy issues into account. Indeed, respect for people's private lives, their homes and communication, their freedom of conscience, religion, freedom of expression, information, enterprise etc. are rights enshrined in EU primary law. It is therefore important that the protection of individuals with regard to the processing of personal data is regulated.

We must have a reason to collect the data (the processing of the data must therefore have a purpose, a determined, explicit and legitimate aim), collect only what is necessary for the aim pursued (only adequate, relevant data can be collected), keep the data only for as long as necessary (the maximum retention period must be defined), and store it securely (and visible only to authorised persons). We should allow access to personal data only to those who really need it, for example, professors may need to know what adjustments are required for a student but they do not need to know what disability they have. There should be systems in place around access, for example to make it clear to all stakeholders who does (and does not) have access to what information and signed confidentiality declarations.

²⁵ Personal data is any information relating to a person that enables them to be identified, directly or indirectly. For example: a name, a photo, a fingerprint, a postal address, an e-mail address, a telephone number, a social security number, an internal personnel number, an IP address, a computer connection identifier, a voice recording, etc. It does not matter whether this information is confidential or public. All personal data must be made anonymous in such a way as to make it impossible for unauthorised people to identify the person concerned, e.g. ensuring data items are not linked, masking names, blurring faces, etc. If it is possible to identify a person by cross-referencing several pieces of information (age, sex, city, qualification, etc.) or by using other means, the data is still considered personal. See <https://gdpr.eu> for more detailed information on DPDR compliance.

While some diversity data needs to be linked to personal information in data systems, some does not, for example when the underlying interest is monitoring structural inequity. In the latter case, numerous measures can be taken to either fully anonymise data in the system or in reports (resulting in it no longer being personal data) or to at least enhance the privacy of data contained in data sets by partial anonymisation or pseudonymisation.²⁶

In addition, as noted in the definition of diversity, a person's position in relation to social categories can change over time. A person might not belong to a vulnerable group this year but may belong to one next year. Individuals should be able to make/inform changes to the data stored.

The GDPR can be seen as a constraint and, when it came into force, the initial inclination of states and institutions was to believe that the processing of personal data was prohibited. However, this is not the case. In fact, the GDPR only regulates the collection and especially the processing of data and it is possible to find a balance between respect for privacy and the collection and processing of data. Data protection guidelines create a transparent – and useful – framework of trust and allow for ethical data collection by clearly formulating not only criteria to protect privacy but also people's right of access, rectification and erasure of their personal data.

It often appears as though the focus is on what is – seemingly – prohibited, as though the constraints defined in data protection regulations are a hindrance to good quality data collection. Whereas in fact, if the protection of privacy is understood as a criterion for good practice, these regulations provide guidance that can assist in the collection of good quality data.

It is not only entirely possible but in fact desirable for universities to collect personal data in order to provide a robust evidence base for policies to protect and empower vulnerable or minoritised groups within the institution – so long as this is done in line with good practice.

3.2.4 Definitions of Diversity Indicators (or The Problem of the Categories)

Evaluating progress towards equality, diversity and inclusion in our institutions requires clear and robust measures. A common approach to evaluating progress is to measure the demographic composition of staff and student communities in broad diversity categories or indicators and identify where certain groups are underrepresented. The collection of data about staff and student diversity characteristics – via monitoring forms, surveys or other methods – is a continually evolving area of work that can provide important insights into the barriers and inequalities faced by particular groups. However, it is important to recognise the limitations and challenges of this type of approach.

The design of diversity monitoring questions is often contested, arising in part from a tension between the need to align with various reporting requirements while also ensuring that staff and students can, as far as possible, describe themselves in ways that reflect how they identify. Grouping people into simplistic categories can never

26 Anonymisation of data is a process whereby any personally identifiable data or information that may allow an individual to be connected to stored data is removed or other steps taken to ensure the subject's privacy, e.g. encrypting it. Once data has been fully anonymised it is no longer personal data, however, data can often not be fully anonymised, meaning that data protection regulations still apply. Partial anonymisation and pseudonymisation (separating the data from direct identifiers) may be sufficient to enhance privacy. See U Edinburgh: www.ed.ac.uk/data-protection/data-protection-guidance/specialised-guidance/anonymisation-personal-data. Anonymisation is usually carried out as part of data cleaning or cleansing, a systematic process of identifying and then correcting, validating or removing incorrect, duplicate or incomplete raw data within a dataset.

fully reflect the complexity of individuals, their lived experience, and the intersecting barriers to inclusion and participation that they face (see Section 3.2.5).

The dimensions of diversity and the categories formed are not objective and static but are rather social constructs that reflect the current socio political context. Comparisons across the seven universities reveal that we are not always defining and/or collecting data on either categories of diversity or their indicators in the same way. Cultural and political drivers or sensitivities, legislation, regulation, and changes to the meaning and significance of diversity categories across context and over time work in combination to shape what we understand as relevant differences, what we collect and how we collect it.

The different linguistic contexts also play a role here. Language is not an objective descriptor, but rather both reflects and contributes to social discourses and how people understand their worlds. So even if the terms used are literal translations, e.g. 'race' in English and French, 'rasse' in German, 'razza' in Italian, 'rotuun' in Finnish, their use and the meaning ascribed them vary enormously even within Europe. The increasing use of English as a de facto lingua franca in higher education has also had an impact on not only on accuracy and shifts in meaning resulting from translations back and forth but also on discourses within other linguistic environments.

The Action Group noted stark differences in the meaning ascribed to some diversity indicators. For example, in Germany the legal system recognises three types of gender and allows for people to have no gender marker at all, whereas most other countries rely on a male/female binary, while in the UK sex and gender identity are distinct concepts defined and collected separately (see Appendix 5). In another example, the definition of 'language' varies across our institutions with meanings spread across 'mother language', 'business language' and 'British Sign Language User'. We found that U Edinburgh is the only university in the group that collects data on ethnicity/racialised identities despite clear policy needs across all institutions to identify and address individual and structural racism. At best weak proxies like 'nationality' or 'migrant background' are being used. In these cases and in others it is questionable whether the indicator being used can be considered as a valid measure of that category of diversity, such as the use of parental educational background or the average socioeconomic level of the area people live in as an indicator of class.

Diversity is a matrix of intersecting social constructs without clearly delineated boundaries. Who defines the categories and their boundaries has a significant impact. For example, instruments used to collect diversity data often include being Jewish only in relation to the diversity category 'religion', whereas many Jewish people would consider it both an ethnic and a religious identity (e.g. most secular Jews still consider themselves Jews, and antisemitism generally discriminates on the basis of racial conceptions of Jewishness).

The legislative and regulatory context in which institutions are operating can mean that some diversity categories and indicators are tightly defined in law and/or by national statistics agencies and these definitions are linked to external criteria such as government funding and the fulfilling of legislative quotas. The indicators and values used, and the language used to communicate them, may differ significantly from the understanding and language used by members of minoritised groups, potentially resulting in both misunderstandings (and thus inaccuracy) and, more importantly, offence to and alienation of the respondents. It also impacts on the quality of the findings. For example, in Germany institutions over a certain size are required to meet employment quotas under a narrow definition of 'severe disability' based on a medical model that focuses on deviations from a postulated norm and associated limitations. However, this is not only at odds with the social model of disability often used in diversity work but also has little bearing on how people experience disability and define it for themselves, or indeed on the specific needs of individuals, and therefore additional indicators may be required if useful information is to be gleaned.

This is linked to substantial differences in expectations around the validity of data that impact on what data is collected and how indicators are defined. Some institutions allow self-identification of some or most diversity characteristics while other institutions collect mainly or only data that can be verified against official documentation.

The dynamic nature of diversity also poses challenges for the definitions of indicators. Not only do the categories change over time and context but individual's identities and memberships of social groups change too. When does someone stop being a refugee? What does it depend on? Residency status, the impact of lived experience (both in the past and how the person is being treated now), or how the person self-identifies? Structural disadvantage is often evident in domestic students who are the grandchildren and great grandchildren of migrants and refugees.

The indicators are a means not an end, so again, we return to the question of the purpose behind the data collection as the deciding factor determining the definition of indicators. The issues we seek to build an evidence base to address are complex and entangled and the indicators often inadequate proxies that can easily be misinterpreted or used in ways that do not reflect the concept or goals of diversity as outlined here. The definitions of categories of diversity are contextual and quickly touch on stigmatising attributions. What assumptions underly efforts to identify how many staff or students have a migrant background? On the one hand, this is known to be a group that experience structural discrimination but in fact it is often implicitly interpreted as if deficits lie within the group. And in fact, in many countries it is used as an indicator of a racialised identity, which is not accurate, especially given increasing mobility within the EU.

The Action Group's mapping exercise has revealed that, even within institutions, the categories of diversity data and the definitions of the indicators can be inconsistent. While in some cases this may reflect fragmentation and lack of harmonisation of data collection, in other cases it may reflect justified differences in needs, contexts (such as staff or students), and freedom from legislative/regulatory constraints in some types of data collection.

Comparisons of diversity data between and within institutions therefore present significant challenges. While consensus on definitions would be helpful to support comparability, it may be that no single definition of each indicator is possible or perhaps even desirable, given that it allows a greater appreciation of context and of the complexity of individuals. It is essential then to provide clarity and transparency on how indicators are defined to enable those analysing, interpreting and reporting the data to have a full understanding of the limitations of any comparison. All data reporting requires extensive caveats and annotations on underlying assumptions and definitions.

3.2.5 Intersectionality

As outlined in section 1.3.1 above, the concept of diversity acknowledges that the differences between people are structured by social categories that intersect in multidimensional ways. Each individual is positioned in relation to multiple social categories and thus at the intersection of complex societal and organisational power relations. The US legal scholar Kimberlé Crenshaw built on Black feminist theory and activism when formulating the concept 'intersectionality' to describe the experiences of Black women experiencing discrimination in their workplace.²⁷ Their experiences differed significantly from both the sexist discrimination faced by white women and the racist discrimination experienced by black men in the same organisation. The legal framework, however, only allowed for single, stand-alone forms of discrimination, which could at most be

²⁷ Crenshaw (1989). See also Collins/Bilge (2016).

added together rather than interrelating. As a result, not only were the specific experiences of Black women – positioned at the intersections of gender and race – rendered invisible, but they were not protected by law. This challenge also faces data collection, which by its very nature sorts information into stand-alone categories.

Actual and potential discrimination (and privilege) cannot be adequately addressed if each social category or dimension of diversity is regarded in isolation. The methodological implications of intersectionality for data collection appear to relate in particular to the data gathered and the analysis conducted:²⁸

- As outlined in this report, there are significant gaps in the data collection currently undertaken by the member universities. If we do not gather data on, for example, disability or sexual orientation, then we cannot identify whether these groups experience higher rates of sexualised violence than other people – a phenomenon research in other contexts has identified.²⁹ Thus potentially specific vulnerabilities and needs of these groups are invisible and not being addressed.
- Multivariate analysis that links selected characteristics or indicators is rarely conducted, even where relevant data has been gathered.

It is, however, not as straightforward as simply asking more questions and conducting multivariate analysis. Each individual is positioned not at the intersection of two or three social categories, but at the intersection of a multitude of categories, which themselves not only shift over time but whose relevance and association with power is contextual. Conceptually therefore, identity is enormously complex, and the experiences associated with it even more so. How differentiated do we need to be in order to provide accurate information that accurately reflects this complexity, while at the same time reducing complexity enough in order to provide findings that can be translated into coherent policies and actions?

A further challenge is posed by existing diversity policies and practice frequently being structured vertically, i.e. in relation to individual dimensions of diversity in isolation to one another. This frames the defined data needs and becomes a catch-22.

3.2.6 Methodological Issues

Standard ways of collecting diversity data include administrative data obtained in particular in student enrolment forms and new staff's personnel forms and survey data. Occasionally data is also obtained from official government registers. Administrative data has several advantages. Firstly, it usually provides information on the population as a whole, i.e. all members of that group, with low rates of missing data (often due to instruments being mandatory forms). It is gathered primarily (but not exclusively) at the point of entry into the university and stored in comprehensive and well-maintained university data systems, so that it is quite reliable and accessible.

Administrative data, however, also has its limitations. For example, information provided by new staff members within the recruitment and onboarding process is often not gathered anonymously and they are unlikely to

28 While these considerations most obviously relate to quantitative data, it can equally be argued that they also relate to qualitative data collection and analysis.

29 A German study found that women with disabilities experienced two to three times more sexualised violence than women on average, both in childhood and as adults ([Bundesministerium für Familie, Senioren, Frauen und Jugend, 2014, Lebenssituation und Belastungen von Frauen mit Beeinträchtigungen und Behinderungen in Deutschland](#)) The 2021 Australian National Student Safety Survey found that 14% of students with disabilities reported having experienced sexual harassment at university in the preceding 12 months as compared with 7% of students without disabilities. The differences were even more marked for sexual orientation: pansexual 21%, bisexual 18%, gay or lesbian students 12%, heterosexual students 6% (Heywood et al. 2022).

feel comfortable providing sensitive information, especially when it is not obvious to them how it relates to their primary reason for filling in the forms (e.g. applying for a job or study place, being onboarded). Given the dependency and associated power relationship between the organisation asking for information and the person providing it, combined with the potential for stigmatisation and discrimination that can result in information being known about a person's identity (e.g. in relation to disability, gender identity), it is possible that the information provided is not always accurate. While, as noted, staff and students are required to provide a significant amount of the data gathered by the university administration, some is gathered on a voluntary basis or only gathered when an individual proactively applies for a specific programme or benefit. It cannot be known whether other members of the university might be equally eligible but either choose not to apply or are unaware of the possibility. Thus, the potential for underreporting (and inaccurate responses) must be taken into account in administrative data sets too.

In addition, the indicators of some diversity categories, like ethnicity, migration or gender, are sometimes pre-defined by external regulatory frameworks that the university has to report against, but which do not necessarily reflect current discourse and/or the understanding of members of the groups themselves, potentially resulting in both inaccurate data and alienation of the respondent (see Section 3.2.4). Nonetheless the potential of administrative data lies in its ability to provide a quantitative evidence base for the population as a whole, i.e. for all staff and/or students of the university.

Surveys on the other hand are almost always voluntary so they frequently suffer under low response rates and self-selection of respondents, which can impact on how representative the results are for the organisation as a whole. This is especially problematic in the case of diversity data, when some minoritised groups may be completely missing from the data. There is, for example, some evidence of higher levels of distrust on the part of minoritised groups towards the institution that is likely to impact on their willingness to respond to a survey. While data can be cleaned to strengthen how representative they are, this is usually done on the basis of existing reference data, e.g. adjusting response to reflect the known number of women or proportion of students in different study programmes. However, given the existing gaps in knowledge about minoritised groups – especially in the administrative data system – it is currently near impossible to adjust data to ensure they are accurately represented in the findings. And on top of that, surveys bring their own challenges, such as usually requiring significant resources to conduct.

On the other hand, surveys are highly adjustable to reflect current needs, and they can be used to collect qualitative data that supplements the quantitative data and can assist in its interpretation, enabling a more explorative approach to be taken.³⁰ Diversity is often related to complex experiences of inclusion or discrimination, which qualitative data can assist us to understand more fully. Surveys can also be used to collect quantitative data in cases when, for example, legal obstacles hinder using administrative data to do that, although one must bear in mind that usually quantitative survey data is not as reliable as similar quantitative administrative data.

There are additional numerous methodological issues associated with diversity data collection quite apart from the instruments used. Some have been listed in section 3.1 above.

One major problem with the analysis of diversity data is that the methodology varies a lot from one country to another, and even within a country or university. A seemingly obvious way to overcome this problem and enable comparability would be to try to harmonise diversity definitions and diversity data collection practices.

30 The Australian National Student Safety Survey took this approach, including both quantitative and qualitative items in the instruments and reporting on them both in depth. See Heywood et al. (2022) and Nisbet et al. (2022). Akel (2019) triangulated methodologies, conducting a survey, six individual interviews and three focus groups in a study on the experiences of Black and minority ethnic students.

This is, however, often not possible. There might, for example, be legal obstacles in different countries and, possibly more importantly, the concept of diversity, the discourses, the socio-political contexts and the norms of what is considered socially acceptable differ so significantly even within Europe comparability is often an illusion. For example, while in the UK talking about ethnicity is fairly normalised, in Germany there is often discomfort around this, especially among the white population. This has methodological implications, for example in relation to how indicators are defined, the impact of how items are formulated (and thus the results), which response options are provided (some universities for instance provide a 'no response', an 'I do not know' and/or a 'rather not tell' option on some questions) and how they are interpreted by respondents, etc.

The current reality of diversity data is that we have different definitions of core concepts in different countries, different ways of collecting it and some fundamental methodological challenges that need to be self-critically addressed if we are to improve the quality of diversity data.

We need a way to make a good use of existing data and to operate within current limitations, all the while pushing at their boundaries and being as transparent as possible about the methodologies and caveats in relation to the data produced. One way to assist operating in the current reality of diversity data is to triangulate data where possible: By combining different kinds of related diversity data, and different types of methods of analysis, one can aim for a more coherent analysis of diversity data (see Appendix 4).

3.2.7 Expertise & Responsibility

In the mapping exercise, it became apparent that gathering diversity data and making sense of it hinges on two distinct capabilities: First, having the necessary expertise to understand the data and the methodological and ethical factors associated with all aspects of its collection, starting with the definitions of indicators through item design and data cleansing, anonymisation and analysis, to issues associated with where and how the data is stored and who has access to it. Second, having subject matter expertise that enables you to identify the needs that the data aims to meet, define categories of diversity and indicators that might give meaningful insights into them, understand the meanings accorded to terms by the target groups and assess what values will provide robust information, to be able to interpret and use the data, to assess how the data can be made accessible and useful for a wide variety of users. It is still rare that individuals have expertise in both data and diversity.

The Action Group's own mapping exercise reflected the tension between the two areas of expertise and sought to make best use the various skills of the members by taking an interactive and dialogue-based approach. If limited to merely an enumeration of data collected in the registration systems of the universities, the mapping exercise would have run the risk of stating similarities or differences that in actual fact, when looking at the content of the data, would be incomparable. Similar labels might hide very different definitions, and different labels might actually refer to the same reality. If limited to a technical perspective of data analysis, the exercise would have lacked adequate understanding of what is actually being registered in the systems (and what not).

This complexity is not only relevant with regard to the mapping exercise that the Action Group carried out, but also more broadly. Diversity data collection cannot be only about the nuts and bolts of how data is recorded, nor only about being aware of, and knowledgeable about diversity issues and themes. The two must go hand in hand. Often however, the responsibility for the two areas sits in quite different parts of the university, if it is designated at all.

The Action Group's conclusion, in short, is that both methodological data expertise and subject matter expertise in relation to diversity are crucial when developing a diversity data collection strategy and designing,

implementing and maintaining a diversity data collection system. If the responsibility for these areas lies in separate units within the university, there should be systematic collaboration between them. And if there are gaps in expertise in-house, then that expertise should either be developed or bought in. In that way, it can be ensured that the purpose of diversity data collection is actually fulfilled by the way in which data is collected, analysed and made available for use.

The expertise of *data users* is also an issue the Action Group identified. As noted, the data that is collected often appears to be underutilised. At least some potential users appear to lack not only an awareness of the need for and uses of robust data, but also the skills required to understand, interpret and apply the data to their work area. Many universities have online platforms through which at least some staff can access data from administrative data systems but these often still require considerable technical understanding and the operational relevance of the data is seldom elaborated. Few universities appear to make data easily accessible to a wider range of stakeholders or specifically engage with the data subjects in a meaningful way, or indeed to see this as their responsibility.

Providing a document or webpage on which information about data use is given would satisfy data protection requirements but would not engage the data subjects in a meaningful way. Similarly, simply publishing reports on findings does not necessarily make that data useful and usable for stakeholders. An innovative example can be found at KU Leuven, where some of the student data collected is given back to the students in the form of learning dashboards.³¹ The learning dashboards provide visual representations of the students' study skills (data from standardised questionnaires) and study pathways (data from administrative data records) that require little prior knowledge to understand. They provide students with insight into their skills and pathways in an accessible way, but at the same time enable them to see in a direct and relevant way how the data collected from them is analysed and used within the university for the evaluation of study programmes and development of policies on study progression.

31 The learning dashboards project at KU Leuven is explained in this short video: www.kuleuven.be/onderwijs/learninglab/projecten/scale-up-projecten/learning-dashboards

4. Gathering & analysing robust diversity data

4.1 Key Findings & Recommendations

Key Findings	Recommendations
<p>The focus in most universities is on fulfilling external reporting requirements rather than monitoring and promoting equity, and it is often not clear how data collected is actually used to support policy development and practice within the university.</p> <p>There are gaps in data collection in some areas known to have significant impact in higher education, e.g. racism, impact of class, gender identity.</p> <p>Often the data that is gathered is at best collated, with little analysis in relation to the diversity goals of the organisation.</p>	Use 'why' rather than 'what' as a starting point. Ensure that purpose drives data collection and is clearly defined at the outset. When making decisions, critically reflect on the reasons behind (not) gathering data together with other stakeholders and ensure that the data gathered is fit for purpose.
	Determine the way in which data will be used and made accessible as part of the planning phase.
	Place the purpose of reducing discrimination and promoting equity at the forefront of data collection (for example, gaining information to identify needs and to inform the response to these needs).
	Where possible collect data in a 'design for all'/ mainstreaming logic, rather than focussing on categories of people as target groups.
	Supplement the data required by external regulatory frameworks with needs-driven data collection. In some cases, an interface can translate internal data into the form required to meet external reporting requirements.
<p>Each university operates in a very specific context (e.g. legislation, public discourse, education system, data management, organisational culture) and conceptualises diversity issues differently. This limits comparability of data and the opportunities to adopt good practices from other universities.</p>	Treat any comparison of data with the utmost caution.
	When reporting on diversity data collection, provide extensive and explicit notes on definitions, contextual factors and methodological decisions. This enables assessment of comparability and robust consideration of potential transferability of practice. ³²

32 Eurostat compares available European data despite differences between countries, but relies on a complex system with many footnotes in order to do so.

Key Findings	Recommendations
<p>There is a lack of knowledge on what privacy and data protection entail and how they can be operationalised. This often leads to data simply not being collected due to 'data protection' concerns.</p>	<p>Use the guidance offered by the legal data protection framework to develop clear, accessible internal guidelines and a strict policy that defines quality criteria and requirements to ensure that all personal data is collected, processed, stored and reported in a way that protects the privacy of individuals. Unless there is a justified need to link the data to individuals, gather data anonymously or, where that is not possible, anonymise it.</p>
	<p>Provide clear information from the outset about the privacy policy, the purpose of the data collection and how the data will be used.</p>
	<p>Limit access to personal data on a transparent needs-to know basis (with a signed confidentiality declaration).</p>
	<p>Consistently and proactively inform individuals about how to access all personal data collected on them, how long data will be stored (as per data protection regulations), and how they can make changes as required.</p>
<p>The universities vary enormously in their approach to collecting, analysing, and using diversity data. There is seldom a specific policy on diversity data, demographic data or dealing with sensitive data more generally (beyond general data protection policies).</p>	<p>Develop a strategic policy on data for the university as a whole that provides direction based on accepted good practices and the university's own strategic goals, including those related to diversity. Include guidance on what data should be prioritised and what it will be used for (e.g. which diversity categories and the areas of focus), principles and standards (e.g. definition of purpose, data protection/privacy), the instruments/methods used to gather data (e.g. consistent definitions across the university), how often data should be gathered and who it should be accessible to and how.</p>
	<p>Acknowledge and proactively communicate the limitations and potential for error, and take a pragmatic approach in combination with a continuous learning approach (reflecting on and improving practice).</p>
<p>Both administrative and survey data almost always focus exclusively on singular identity indicators (measuring representation of members of defined social groups within the university), with little data gathered on experiences (e.g. of inclusion or discrimination) or on multidimensional/intersectional identities.</p>	<p>Monitor the experiences of university members and relate them to their membership of minoritised groups. For example, conduct regular surveys on how respondents experience the organisational environment.³³</p>
	<p>Formulate survey items/questions on the basis of currently accepted good practices, including acceptance by the target group. Where possible, use them consistently across time to maximise comparability, unless they are misaligned with current best practice, in which case revision should be considered.</p>
	<p>Conduct multivariate analysis where possible to provide insights into intersectional identities and experiences.</p>

33 Recommended are discrimination and/or climate surveys that collect data on the composition of the members of the university as well as their experiences of discrimination and other factors associated with a positive, inclusive and equitable environment and linking the data in multivariate analysis. Climate surveys that gather data on perception in isolation cannot be adequately interpreted to provide meaningful findings. For example, if findings indicate that respondents perceive the university to be a supportive and inclusive environment but all or most respondents are members of the dominant/privileged groups, this finding is likely to reflect their experiences and biases, whereas members of minoritised groups may have very different experiences.

Key Findings	Recommendations
Indicators are a mix of self-identification (i.e. the person determining their own identity) and other external criteria such as legal status or official assessments. The two do not always align. For example, a person's own understanding of whether they have a disability may differ from whether they meet defined legal thresholds.	Where possible, enable respondents to determine for themselves how they relate to diversity categories (self-identify). Take a flexible approach that puts the purpose of the data collection at the forefront, recognises that there are sometimes discrepancies between documents and lived experience, and makes it transparent to respondents exactly how indicators are being defined. ³⁴
The definitions used for indicators and values are not always clear, increasing both potential variance in respondents' interpretation and barriers to the robust interpretation (and utilisation) of findings.	<p data-bbox="802 439 1447 533">Formulate items so that they are easily comprehensible to the target group (both questions and response options), including descriptions to explain key terms.</p> <p data-bbox="802 533 1447 633">Include comprehensive notes in all reporting, with clear definitions of the indicators and values used and any caveats relating to the findings.</p>
There are significantly different practices in relation to the publication of data, and when it is made available, it is seldom communicated in an accessible way.	<p data-bbox="802 633 1447 790">Make the results of data collection should be made available to those from whom it has been collected and to those who can derive practice implications from it. If that is not possible, at least be communicate transparently why the data is not available.</p> <p data-bbox="802 790 1447 860">Communicate data in a way that is accessible and that the subjects can understand.</p>
Data collection is generally carried out by practitioners with either expertise on diversity issues or expertise in data analysis, seldom both.	<p data-bbox="802 860 1447 1016">Build dedicated and qualified staff capacity for diversity data collection, analysis and reporting, combining expertise in both data analysis and diversity policy. Address gaps in expertise, e.g. by staff development or buying in external expertise.</p> <p data-bbox="802 1016 1447 1086">If responsibility lies in separate units, foster systematic collaboration between them.</p>
Similar data is often collected in different contexts and using different instruments.	Adhere to the principle of data minimisation. For example, record personal data once and make it accessible where it is required so that people do not need to tell their stories repeatedly.
There is little linkage between administrative and survey data.	Triangulate different instruments and methods, in particular in relation to sensitive data (see Appendix 4).
There is often significant variance between data collected for the different populations within the university. Often different indicators are used for staff and student data.	<p data-bbox="802 1285 1447 1359">Consider the needs of the university community as a whole (staff and students).</p> <p data-bbox="802 1359 1447 1505">While there might be different data needed for the different groups, if data is being gathered on the same category of diversity, use consistent indicators where possible, based on accepted best practice at the time (multiple indicators where necessary).</p>

34 For example, collect sex/gender data based on lived identity to measure gender-based differences rather than official gender markers (given the barriers in place to changing gender markers in most countries).

4.2 Reflective Questions to Guide Data Collection

Derived from these recommendations, consideration of the following questions can assist in the development and implementation of robust and useful diversity data collection:

1. Why do we intend to collect this data? (purpose) What exactly are the policy needs we seek to address and how would data help us to do that? In whose interests is the data being collected? How will it contribute to the goals of the university in relation to diversity?
2. Do we have the right to collect this data? Legally? Ethically?
3. Do we have an ethical responsibility to collect this data?
4. How will members of minoritised groups/target groups participate in the definition of purpose, the development of instruments, the implementation and the analysis, interpretation and reporting of findings?
5. Is the data fit-for-purpose? (see question 1)
6. What will be done with the data? Who will have access to it and why? How exactly will it be used? How will findings be made available to stakeholders, including the respondents?
7. How is data privacy being addressed? How is this communicated to (potential) respondents?
8. How does this data link to other data already being collected? (Within our own organisation, external/official reference data). Are the definitions and questions comparable? If not, why not?
9. How are the indicators being defined? Does this reflect current understandings of good practice? How does it relate to understandings within minoritised groups and to the conceptual framework of the university in relation to diversity?
10. Are the questions being asked/the forms formulated in a way that is clearly understandable for all respondents and unambiguous? Does the terminology reflect the language used by the respondents themselves? Is the language used in the instrument as a whole non-discriminatory?
11. Is the instrument being used accessible to all members of the target group? Which groups may face barriers (e.g. language, technology, distrust) and how will those barriers be addressed?
12. How will the data collection process be evaluated? What understanding of quality underlies that evaluation? How will the target groups be involved? How will learning derived be incorporated into future data collection planning processes?

4.3 Potential Next Steps

This report has raised a wide range of issues that need further attention if we wish to improve the quality and thus the usefulness of diversity data to pursuing the vision of diversity outlined above. In addition, the Action Group suggests the following specific actions as beneficial next steps to prioritise.

At Una Europa Level

The concept of mobility needs to specifically address refugee experiences. This has taken on a particular urgency in the light of the ongoing movement of people fleeing conflict or seeking to do so. For example, since 2015 many people have sought refuge from the war in Syria, including university students and staff, and since February 2022 a large number of people have managed to flee the war in Ukraine to the countries of Una Europa members, including both Ukrainian citizens and international students and academics, mostly without EU citizenship. To date universities have been responding to this situation in a well-intentioned but ad hoc manner, with little underpinning with data at either a country or a university level. When the conceptual framework and associated aims have been formulated, there needs to be consideration of what data is required to provide a robust evidence base for decision-making and the development of effective policies and measures at both the Una Europa level and within individual universities to achieve those aims.

Within Individual Universities

The EU Horizon Funding Programme has new requirements for Gender Equality Plans and reporting on gender data. Universities should use this opportunity to strengthen the quality, visibility, availability and use of gender disaggregated data for their universities, based on current international best practice (some of which has been noted in this report). For example, including non-binary gender identities and putting lived experience rather than legal gender marker at the forefront.

This also provides an opportunity to look more broadly at the diversity data gathered in addition to gender. On the one hand, this is necessary if universities are to take the intersectional approach noted in the EU Horizon guidelines and advocated by the Una Europa Diversity Council, and on the other hand, a differentiated consideration of all key dimensions of diversity for the university provides the evidence base that is needed to effectively deliver on organisational goals.

At the Level of a Working Group

The following activities would be useful next steps to build on the exploratory work conducted by the Diversity Council Action Group.

- A workshop with other stakeholders to present and critically reflect on the findings in this report and to develop a strategic approach to progressing them further.
- A systematic literature review on data collection in relation to each of the main diversity dimensions, including both demographic data/representation of social groups but also in relation to experiences of discrimination and inclusion (related in turn to membership of social groups as per demographic data).

- Extension of the systematic mapping to surveys, including consideration of qualitative data.
- Detailed analysis of data collection practice in relation to each diversity dimension.
- Mapping data collection in relation to measures to foster diversity and reduce barriers to participation (as opposed to data on group members and peoples' experiences).
- In-depth meta-analysis of the instruments used for data collection (at a methodological level), including definitions of indicators, the items (questions respondents are asked), how representation is addressed, etc.
- A glossary of key terms to assist communication and cooperation between stakeholders in different countries and with different areas of expertise.
- Mapping and evaluation of practice and organisational structures in relation to diversity monitoring and benchmarking.
- Development of detailed guidelines for good practice in diversity data collection.

Data collection is not an end in itself, but rather a means to an end. The Action Group calls on universities and researchers to take a deliberate and reflective approach when engaging with diversity data to ensure that it is fit for purpose and all people benefit from policies, activities and research, whether they are developed under the label of 'diversity' or not. Our findings and recommendations are neither final nor exhaustive. By grappling with the issues outlined in this report as we formulate data needs or develop research questions, design instruments and methodologies, analyse data, and report and use the results, we can continuously improve quality, deliver ever more rigorous findings and contribute to a robust evidence base that supports us to achieve our vision of diversity.

5. Literature

Akel, S. (2019) Insider-Outsider. The Role of Race in Shaping the Experiences of Black and Minority Ethnic Students. Goldsmiths University of London. www.gold.ac.uk/media/docs/reports/Insider-Outsider-Report-191008.pdf Last accessed 14th August 2022

Blackless, M., Charuvastra, A., Derryck, A., Fausto-Sterling, A., Lauzanne, K., Lee, E. (2000) How sexually dimorphic are we? Review and synthesis, in *American Journal of Human Biology* 04/2000; 12(2):151-166. DOI:[10.1002/\(SICI\)1520-6300\(200003/04\)12:2<151::AID-AJHB1>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1520-6300(200003/04)12:2<151::AID-AJHB1>3.0.CO;2-F)

Collins, P.H., Bilge, S. (2016) *Intersectionality*. Polity Press, Cambridge

Crenshaw, K. (1989) Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Policies. *University of Chicago Legal Forum* 1: 139-167.

Engaging with Diversity in European Universities (ed. M. Karjalainen). *Una Europa*.
<https://una-europa.imgix.net/documents/Engaging-with-Diversity-in-European-Universities.pdf>

Heywood, W., Myers, P., Powell, A., Meikle, G., Nguyen, D. (2022) National Student Safety Survey: Report on the prevalence of sexual harassment and sexual assault among university students in 2021. The Social Research Centre, Melbourne.
www.nsss.edu.au Last accessed 23rd April 2022

Jones, T. (2017) Intersex and Families: Supporting Family Members with Intersex Variations. *Journal of Family Strengths*. 2017 Sep 7;17(2). <http://digitalcommons.library.tmc.edu/jfs/vol17/iss2/8> Last accessed 2nd August 2022

Meyer, I., Frost, D.M. (2012) Minority Stress and the Health of Sexual Minorities, in Patterson, C.J., D'Augelli, A.R. (eds) *Handbook of Psychology and Sexual Orientation*. Oxford University Press; DOI: [10.1093/acprof:oso/9780199765218.003.0018](https://doi.org/10.1093/acprof:oso/9780199765218.003.0018)

Nisbet, L., Halse, G., van Esbroek, E., Heywood, W., Powell, A., Myers, P. (2022) National Student Safety Survey: Qualitative research on experiences of sexual harassment and sexual assault among university students in 2021. The Social Research Centre, Melbourne. www.nsss.edu.au Last accessed 23rd April 2022

Rosenstreich, G. (2022) Defining diversity, in Karjalainen, M. (ed.) *Engaging with Diversity in European Universities*. *Una Europa*.
<https://una-europa.imgix.net/documents/Engaging-with-Diversity-in-European-Universities.pdf>

6. Appendix

Appendix 1: Overview of Diversity Indicators per Category

Tom Bekers, KU Leuven, and Gabriele Rosenstreich, Freie Universität Berlin

The following tables provide overviews of the indicators used in the universities' administrative data systems per category of diversity for staff and students respectively. The values for each indicator are also provided where known.

Please note that these overviews do not include data collected and stored outside the universities' administrative data systems.

Table 1: Staff

Staff: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Age	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]
Gender	• legal gender marker [female / male / diverse (intersex/ non-binary) / no entry]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female] <i>(third option under consideration)</i>	• legal gender marker [male / female]
Gender Identity			• gender identity as per self-identification [female / male / non- binary / other / prefer not to say] • trans-status [yes / no]				
Name	• legal name <i>(used for official documents and data migration to external agencies where required)</i> • preferred name [first name &/or surname of choice] <i>(used for all documents, communication, reporting that does not explicitly require the legal name)</i>	• legal name	• preferred name [first name of choice] • legal name	• legal name	• legal name	• legal name <i>(used for official documents and data migration to external agencies where required)</i> • preferred name [first name of choice] <i>(used for communication)</i>	• legal name • preferred name [first name of choice]
Sexual Orientation			• sexual orientation as per self- identification [bisexual / gay man / gay woman / lesbian / heterosexual / other / prefer not to say]				

Staff: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Religion			<ul style="list-style-type: none"> religion or belief as per self-identification [list^a, including: 'prefer not to say'] 				
Ethnicity/ Racialised Identity			<ul style="list-style-type: none"> ethnicity as per self-identification [list^b, including 'prefer not to say'] 				
Migration						<ul style="list-style-type: none"> migrant background as per self-identification [yes / no] 	
Refugee Status / Experience							
Nationality	<ul style="list-style-type: none"> current citizenship [German / If not German, country of citizenship] <i>(verified against identity document)</i> 	<ul style="list-style-type: none"> current citizenship [country of citizenship] <i>(verified against identity document)</i> residency status [resident in Italy / resident abroad] 	<ul style="list-style-type: none"> current citizenship [country of citizenship] <i>(verified against identity document)</i> 	<ul style="list-style-type: none"> citizenship [country or countries of citizenship] <i>(Finnish overrides the others; verified against identity document)</i> residency status [one or multiple countries in which have a valid residency permit] <i>(Finnish overrides the others)</i> 	<ul style="list-style-type: none"> current citizenship [country of citizenship] <i>(verified against identity document)</i> 	<ul style="list-style-type: none"> current citizenship [country or countries of citizenship] <i>(verified against identity document)</i> international employee as per self-identification [yes / no] 	<ul style="list-style-type: none"> current citizenship [country of citizenship] <i>(verified against identity document)</i>
Language				<ul style="list-style-type: none"> first language [Finnish / Swedish] preferred business language [Finnish / Swedish / English] 		<ul style="list-style-type: none"> language of communication [Dutch / English] 	
Class / Social Status					<ul style="list-style-type: none"> eligibility for social benefit [amount of gross monthly income per family member] <i>(staff file only, not collated)</i> 		

Staff: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Disability & Health Status	<ul style="list-style-type: none"> • severe disability as per legal definition (<50% 'normal' capacity) [yes / no entry] 	<ul style="list-style-type: none"> • disability as per government definition [no / yes, if so: type of disability, temporary / long-term] 	<ul style="list-style-type: none"> • disability or long-term health condition as per self-identification [list^c, including 'prefer not to say' / no entry] 		<ul style="list-style-type: none"> • degree of disability [light / moderate / significant / no entry] 	<ul style="list-style-type: none"> • disability as per governmental definition [yes / no entry] • disability, long-term disorder or illness as per self-identification [yes / no] 	<ul style="list-style-type: none"> • disability status [category of disability as per list / no entry]
Family Status	<ul style="list-style-type: none"> • official relationship status [married / divorced / widowed / single] 	<ul style="list-style-type: none"> • relationship status [living with partner / not living with partner] • children [number] 	<ul style="list-style-type: none"> • relationship status [living together / married/ civil partner / separated/ divorced / single / widowed / prefer not to say] 		<ul style="list-style-type: none"> • married [yes / no] • children [date of birth per child] 	<ul style="list-style-type: none"> • relationship status [married / legally cohabiting / non registered cohabiting / unmarried] • children [date of birth per child] • formal relationship to children [child / step child or foster child / child of the partner] 	<ul style="list-style-type: none"> • relationship status [single / married / divorced / separated / cohabiting / PACS (civil partnership) / widowed] • children [date of birth per child] • dependants [number]

List^a (U Edinburgh - religion or belief as per self-identification): No religion / Buddhist / Christian / Christian - Church of Scotland / Christian - Roman Catholic / Christian - Presbyterian Church in Ireland / Christian - Church of Ireland / Christian - Methodist Church in Ireland / Christian - Other denomination / Hindu / Jewish / Muslim / Sikh / Spiritual / Any other religion or belief / Prefer not to say.

List^b (U Edinburgh - ethnicity as per self-identification): White / White – Scottish / Irish Traveller / Gypsy or Traveller / Other White background / Black or Black British – Caribbean / Black or Black British – African / Other Black background / Asian or Asian British – Indian / Asian or Asian British – Pakistani / Asian or Asian British – Bangladeshi / Chinese / Other Asian background / Mixed - White and Black Caribbean / Mixed - White and Black African / Mixed - White and Asian / Other mixed background / Arab / Other ethnic background / Prefer not to say.

List^c (U Edinburgh – disability or long-term health condition as per self-identification): No known disability / Two or more impairments and/or disabling medical conditions / A specific learning difficulty such as dyslexia, dyspraxia or AD(H)D / General learning disability (such as Down's syndrome) / A social/communication impairment such as Asperger's syndrome/other autistic spectrum disorder / A long standing illness or health condition such as cancer, HIV, diabetes, chronic heart disease, or epilepsy / A mental health condition, such as depression, schizophrenia or anxiety disorder / A physical impairment or mobility issues, such as difficulty using arms or using a wheelchair or crutches / Deaf or serious hearing impairment / Blind or a serious visual impairment uncorrected by glasses / A disability, impairment or medical condition that is not listed above / Prefer not to say.

Table 2: Students

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Age	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]	• age in years [date of birth]
Gender	• gender identity [female / male / diverse (intersex/ non-binary) / no entry]	• legal gender marker [male / female]	• legal gender marker (" <i>legal sex</i> ") [male / female / other]	• legal gender marker [male / female]	• legal gender marker [male / female]	• legal gender marker [male / female] <i>(third option under consideration)</i>	• legal gender marker [male / female]
Gender Identity			• gender identity in relation to gender assigned at birth, as per self-identification [same as gender assigned at birth / different to gender assigned at birth / refused to answer / not available]				
Name	• legal name • preferred name [firstname &/or surname of choice] <i>(student can choose whether the preferred name is used for all purposes or only for online platforms)</i>	• legal name • preferred name [first name of choice]	• legal name • preferred name [first name of choice]	• legal name	• legal name • preferred name [firstname &/or surname ending of choice]	• legal name <i>(used for official documents and data migration to external agencies where required)</i> • preferred name [first name of choice]	• legal name • preferred name [first name of choice]
Sexual Orientation			• sexual orientation as per self- identification [bisexual / gay man / gay woman / lesbian / heterosexual / other / prefer not to say]				
Religion			• religion or belief as per self-identification [list ^p , including: 'prefer not to say']				

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Ethnicity / Racialised Identity			<ul style="list-style-type: none"> • ethnicity as per self-identification [list^E, including 'prefer not to say'] 				
Migration	<ul style="list-style-type: none"> • non-mobile foreign student, derived from the indicators: • Country of university qualification [value: university entrance qualification in Germany] • nationality [value: not German] 	<ul style="list-style-type: none"> • non-mobile foreign student, derived from the indicators: • Country of university qualification [value: university entrance qualification in Italy] • birth nationality [value: not Italian] • birth nationality of parents [value: not Italian] • current country of residence [value: Italy] 				<ul style="list-style-type: none"> • migrant background [EU1 country (List^F) / non-EU1 country / no migrant background / unknown / other] derived from the indicators: • nationality of the student • birth nationality of the parents • birth nationality of the grandparents • country of residence • country of university entrance qualification 	
Refugee Status / Experience			<ul style="list-style-type: none"> • refugee status [refugee / limited leave to remain in the UK / seeking asylum in the UK / none] 			<ul style="list-style-type: none"> • residency status [List^G, including recognised refugee / asylum seeker / other] 	<ul style="list-style-type: none"> • official status as refugee / stateless [yes / no]
Nationality	<ul style="list-style-type: none"> • current citizenship [country or countries of citizenship] <p><i>(multiple possible; German overrides the others in reporting)</i></p>	<ul style="list-style-type: none"> • current citizenship [country of citizenship] • current residency status [resident in Italy / resident abroad] • citizenship at birth • citizenship of parents at birth 	<ul style="list-style-type: none"> • current citizenship [country or countries of citizenship] <p><i>(dual possible)</i></p>	<ul style="list-style-type: none"> • current citizenship [country or countries of citizenship] <p><i>(multiple possible)</i></p>	<ul style="list-style-type: none"> • citizenship [country of citizenship] 	<ul style="list-style-type: none"> • citizenship [country of citizenship] • international student [yes / no] • residency status [List^G] 	<ul style="list-style-type: none"> • citizenship [country of citizenship]

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Language		<ul style="list-style-type: none"> • mother tongue [one language] • knowledge of other languages [language(s)] • level of knowledge of other language 		<ul style="list-style-type: none"> • first language [Finnish / Swedish] • language of business [Finnish / Swedish / English] • language of high school education [Finnish / Swedish / other] 		<ul style="list-style-type: none"> • language of communication [Dutch / English] • language spoken at home [only Dutch / Dutch + other / only other / no answer] 	
Class / Social Status	<ul style="list-style-type: none"> • university entrance qualification (UE) [General UE ("Abitur") / General UE (foreign) / subject-related entrance qualification / preparatory college / technical college certificate / vocational qualification] 	<ul style="list-style-type: none"> family social class, derived from • parents' last profession [unemployed / retired / ...] • parents' qualification (<i>socioeconomic status of the parents is compared</i>) • family economic situation [ISEE "Equivalent Economic Situation Indicator" certification (<i>submitted by students applying for financial benefits</i>)] 	<ul style="list-style-type: none"> Socioeconomic status of family, derived from • parental occupation [list ^H, occupation categories] first generation student, derived from • parents' higher education qualification [yes / no / don't know / information refused / no response given] • Scottish Index of Multiple Deprivation [home postcode] 		<ul style="list-style-type: none"> socio-economic status, derived from: • monthly income per family member [amount] • family composition [number and type of family members] • occupation of adults in family [occupation categories] • social assistance [type of benefit] • alimony [no / yes, if so, amount] • farm [no / yes, if so size] • scholarship [no / yes, if so, amount] 	<ul style="list-style-type: none"> • first generation student [level of schooling parents completed] • student's perception of difficulty faced by parents paying for higher education [very difficult, difficult / quite difficult / quite easy / easy / very easy] • study grant [yes / no / discount in study fee] (<i>criteria include income of family members, number of children studying</i>) 	<ul style="list-style-type: none"> • official occupational category of the parents • eligibility for social benefits [financial situation of parents (copy of 'avis d'imposition')]

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Working Students						<ul style="list-style-type: none"> • employment status [fulltime employment / jobseeker / self-employed / student assistant / none of the above] • enrolment in targeted study programs for working students [yes / no] • KU Leuven working student status [yes / no] 	<ul style="list-style-type: none"> • having a paid job during the academic year [yes, a regular job / yes, a casual job / no]
Disabilities & Health Status		<ul style="list-style-type: none"> • type of disability [physical, mental or sensory impairments / intellectual disabilities / blindness / deaf and dumb / temporary illness / no entry] • Satisfaction with university support (<i>targeted survey</i>) 	<ul style="list-style-type: none"> • disability or long-term health condition [list '] 		<ul style="list-style-type: none"> • nature of disability or health problem [visual / hearing / motor skills / mental health / autism spectrum / chronic illness / reading and writing difficulties / other / no entry] • degree of disability [light / moderate / significant/ no entry] 	<ul style="list-style-type: none"> • type of disability [list', including multiple] 	<ul style="list-style-type: none"> • eligibility for adjustments [yes / no]

Students: Indicator(s) [& values] per diversity category	FU Berlin	AMSU Bologna	U Edinburgh	U Helsinki	JU Krakow	KU Leuven	U Paris 1 PS
Family Status			<ul style="list-style-type: none"> relationship status [single / married / separated / divorced / widowed / co-habiting / information refused / not known] caring responsibilities [young people/ children / other relatives/friends / no dependents / both young people/ children & other relatives/friends] <i>(voluntary information)</i> care leaver status [yes / no / prefer not to say] 		<ul style="list-style-type: none"> family composition [parents / siblings / children / spouse] 	<ul style="list-style-type: none"> living situation [main residence with both parents / mother / father / independent / other] relationship status [single / legally married / officially cohabiting / factually cohabiting / widow(er) / legally divorced / separated] 	<ul style="list-style-type: none"> relationship status [single / couple / married] dependent children [number]

List^d (U Edinburgh – religion or belief as per self-identification): No religion / Buddhist / Christian / Christian - Church of Scotland / Christian - Roman Catholic / Christian - Presbyterian Church in Ireland / Christian - Church of Ireland / Christian - Methodist Church in Ireland / Christian - Other denomination / Hindu / Jewish / Muslim / Sikh / Spiritual / Any other religion or belief / Prefer not to say.

List^e (U Edinburgh – ethnicity as per self-identification): White / White – Scottish / Irish Traveller / Gypsy or Traveller / Other White background / Black or Black British – Caribbean / Black or Black British – African / Other Black background / Asian or Asian British – Indian / Asian or Asian British – Pakistani / Asian or Asian British – Bangladeshi / Chinese / Other Asian background / Mixed - White and Black Caribbean / Mixed - White and Black African / Mixed - White and Asian / Other mixed background / Arab / Other ethnic background / Prefer not to say.

List^f EU1 countries (KU Leuven – migrant background): Denmark, Germany, Finland, France, Ireland, Iceland, Liechtenstein, Luxemburg, the Netherlands, Norway, Austria, United Kingdom, Sweden, and Switzerland.

List^g (KU Leuven – residency status): European Economic Area (EEA) nationality / unlimited residency permit in Belgium / non-EEA student visa / permanent family reunification / temporary family reunification / employment visa non-EEA / recognised refugee / asylum seeker / other / diplomat status.

List^h (U Edinburgh – parental occupation): Higher managerial & professional occupations; Lower managerial & professional occupations; Intermediate occupations; Small employers & own account workers; Lower supervisory & technical occupations; Semi-routine occupations; Routine occupations; Never worked & long-term unemployed; Not classified.

Listⁱ (U Edinburgh – disability or long-term health condition): No known disability / Two or more impairments and/or disabling medical conditions / A specific learning difficulty such as dyslexia, dyspraxia or AD(H)D / General learning disability (such as Down's syndrome) / A social/communication impairment such as Asperger's syndrome/other autistic spectrum disorder / A long standing illness or health condition such as cancer, HIV, diabetes, chronic heart disease, or epilepsy / A mental health condition, such as depression, schizophrenia or anxiety disorder / A physical impairment or mobility issues, such as difficulty using arms or using a wheelchair or crutches / Deaf or serious hearing impairment / Blind or a serious visual impairment uncorrected by glasses / A disability, impairment or medical condition that is not listed above / Prefer not to say.

List^j (KU Leuven – type of disability): physical disability / hearing impairment / visual impairment / chronic disease / developmental disorder (learning disability / autism spectrum disorder / AD(H)D / tic disorder / developmental coordination disorder / developmental dysphasia / developmental stuttering / combination developmental disorders) / psychiatric disorder / other / multiple / no entry]

Appendix 2: Template of Mapping Exercise

Gabriele Rosenstreich, Freie Universität Berlin

The following template was used by the Action Group members to map diversity data collection practice within their own university (in an Excel table).

Name of University:							
Contact person							
Q1 Administrative Data (whole of population): What quantitative data do you collect on the composition of your staff and/or students in relation to diversity in your university's administrative data/statistics? (enter into table)							
Diversity Category	Is data collected	How do you define this? Indicator(s)	Values/ response options (per indicator)	Status group	Why is this data (not) collected?	Notes (e.g. issues)	Instruments & systems
Age							
Gender (e.g. official or self-reported gender)	Drop-down menu: • Yes • No • Do not know			Drop-down menu: • Staff • Students • Staff & Students • Do not know			
Gender identity (as a specific dimension of gender that may supplement the gender status, e.g. trans status, cis-gender, intersex status, endo, nonbinary)							
Sexual orientation (e.g. heterosexuality, homosexuality, bisexuality, asexuality)							
Religion							
Ethnicity/racialised identity							
Migration							
Refugee status/experience							
Nationality (e.g. citizenship, residency status)							
Language							
Class/social status (e.g. first generation to attend university, socioeconomic status)							
Working student (paid work alongside studies)							
Disability & health status (e.g. physical or learning disability, mental health, chronic illness)							
Family status (e.g. relationship status, children, caring responsibilities/carer status)							
Other:							

Q2 Survey Data (samples): What surveys does your university carry out that may be relevant for diversity data collection? (enter into table)										
Survey	Target group	Frequency	url (if applicable)	Diversity categories included (demographic data) include item/question asked and notes (e.g. consistency with other surveys; regarding issues/challenges)	Why is this data collected?	Instruments & methodological notes	Ownership of survey (organisational unit)	Survey items on experiences in relation to diversity/discrimination include specific questions asked, type of data (quantitative or qualitative) and notes (e.g. consistency with other surveys; if/how qualitative data relates to quantitative data; regarding issues/challenges)	Data use (how) is the data accessible? (e.g. is there a system in place to share data?) How is it reported and used? Please note any observations/issues.	Notes (e.g. primary interest, issues)
		Drop-down menu: <ul style="list-style-type: none"> • One-off • Monthly/quarterly • Irregular (every few years) • Regular (annual) • Regular (every 2 years) • Regular (every 3-5 years) • Regular (every 5-10 years) • other 								

Q3: Does the university have a data collection policy, e.g. about using consistent definitions for items/indicators across surveys?			
		Details/Comments:	

Q4: Does the university have a policy about unofficial names within the data system in relation to gender identity that does not align with the official gender marker?			
		Details/Comments:	

Q5: What issues have you identified in relation to the collection of diversity data?					
Methodological Issues	Legal Issues	Practical Issues	Conceptual Issues	Ethical Issues	Other Issues
Q6: Other thoughts/comments/concerns?					

Appendix 3: Draft Template for Future Mapping of Diversity Data

Tom Bekers, KU Leuven

The following template incorporates some of the factors that emerged as pertinent during Action Group discussions. It may provide a useful starting point for future mapping. As with the Action Group's initial mapping exercise, this template focuses on demographic data in relation to diversity categories in administrative data systems. Mapping practice in relation to data collection on experiences of discrimination, organisational culture, etc. would require further consideration.

Q1: Please specify the name of your university, your name and your function at the university.

University:	
Name:	
Function:	

Q2.1: Please indicate for each of the following diversity categories whether there is any data systematically being collected about your university's student/staff population relating to this diversity category.

	Student population	Staff population
1 age	Yes/no	Yes/no
2 gender	Yes/no	Yes/no
3 gender identity	Yes/no	Yes/no
4 sexual orientation	Yes/no	Yes/no
5 religion	Yes/no	Yes/no
6 ethnicity / racialised identity	Yes/no	Yes/no
7 migration	Yes/no	Yes/no
8 refugee status/experience	Yes/no	Yes/no
9 nationality	Yes/no	Yes/no
10 language	Yes/no	Yes/no
11 class / social status	Yes/no	Yes/no
12 working student	Yes/no	Yes/no
13 disability & health status	Yes/no	Yes/no
14 family status	Yes/no	Yes/no

Q2.2: Is there any other relevant diversity category that does not belong to or strongly relate to the categories already listed and that you would like to add to this list? Please note that you can specify the population later on. There will also be an open question for every diversity characteristic where you can specify nuances, details or other additional information.

(add diversity category)
(add diversity category)

[Note that the population for added diversity categories can be specified in Q4.]

[Go to Q3 for the first 'no' that was being selected in Q2.1. Offer Q3, Q19 and Q20 for the first 'no'-diversity category, before offering Q3, Q19 and Q20 for the second 'no'-diversity category. Continue until all 'no'-diversity categories per population are handled.]

Q3: Why not?

1. no legislative and/or regulatory requirement
2. no justified need
3. data protection officer declined registration
4. costs of ensuring data protection outweigh the benefits of having the data
5. explicitly forbidden by law/external regulations
6. other: ...

(please specify 'other')

[After offering Q3, Q19 and Q20 for each 'no'-diversity category per population: go to Q4 for each 'yes' that was being selected in Q2.1 and for each diversity characteristic that was added in Q2.2. Offer Q4 to Q20 for every 'yes'-diversity category separately, giving note to further technical instructions.]

Q4: For which population is this diversity category being collected? [Multiple answers possible]

1. students – whole of population
2. students – sample
3. students – for those who request something related to this diversity category
4. staff – whole of population
5. staff – sample
6. staff – for those who request something related to this diversity category
7. other: ...

(please specify 'other')

[Offer Q5 and all the following questions for every population selected in Q4. Make sure to specify the selected population for every run]

Q5: What indicator(s) is/are being collected at your university concerning this diversity category and this population?

(please specify indicator)

(add indicator)

[Offer Q6 and all the following questions for every indicator that is being added separately]

Q6: Which values does the indicator currently have?

(please specify values)

Q7: What purpose is the indicator being collected for? [multiple responses possible]

1. to provide information to determine payments (e.g. wages/taxes/leave provisions)
2. to verify identity
3. to comply to government reporting requirements related to equity monitoring (e.g. legislation, state statistics agencies)
4. to comply with reporting requirements of non-governmental agencies related to identification (e.g. health insurance providers)
5. to comply with voluntary external agreements (e.g. among universities)
6. to monitor representation of social groups within the university
7. to identify needs in relation to structural discrimination
8. to provide an evidence base for the planning of specific services and measures

9. to verify entitlement to adjustments, social benefits or financial benefits
10. for academic research
11. other: ...

(please specify 'other')

Q8: How is the data currently being used within the university? [multiple responses possible]

1. for administrative purposes
2. for descriptive policy purposes (e.g. reports on the number of students/staff members that applied/were helped/...)
3. for analytical policy purposes (e.g. equity-analysis of job applications, selection procedures, etc.)
4. other: ...

(please specify 'other')

Q9: What instrument is the data being collected with? [multiple responses possible]

1. application form
2. survey/questionnaire
3. personnel form on appointment
4. self-service online portal
5. other: ...

(please specify 'other')

Q10: When is the data being collected? [multiple responses possible]

1. students: on application
2. students: on enrolment
3. students: on graduation
4. students: at a defined phase of studies
5. students: any time, in the student-file
6. students: on demand, when applying for a specific status, benefit or specific program
7. students: on an exam form
8. staff: on application
9. staff: on appointment
10. staff: on demand, when applying for a specific status, benefit or specific program
11. staff: any time, in the staff-file
12. other: ...

(please specify 'other')

Q11: Is the data updated/revised after initial registration? [multiple responses possible]

1. yes, anytime, through the student-/staff-file
2. yes, annually/systematically/periodically
3. yes, for every new study program the student takes up
4. yes, for every new position the staff member applies for
5. yes, for every new position the staff member is appointed to
6. yes, on every new round of the questionnaire
7. yes, in interaction with governmental database
8. no, not necessary

9. no (reason unspecified)
10. other: ...

(please specify 'other')

Q12: Is the data being collected on a voluntary or mandatory basis?

1. mandatory
2. mandatory, with a 'prefer not to say' response option
3. voluntary but prerequisite for provision of services or benefits
4. voluntary

Q13: Is there any kind of verification of the collected data?

1. yes, verification of data per legal document
2. yes, self-identification
3. yes, verification per governmental database
4. no, no verification

Q14: Who is the data being collected by? [multiple responses possible]

1. a unit within the central university administration with specific responsibility for this diversity category (possibly among others)
2. a unit within the central university administration, other
3. a faculty/faculties at the university
4. an official student organisation affiliated with the university
5. an external partner/organisation/consortium (please specify in the 'further explanation'-column at the end of this row)
6. a government agency (please specify in the 'further explanation'-section later on)
7. other: ... (please name here, and specify in the 'further explanation'-section later on)

(please name 'other' and specify at the 'further explanation'-section later on)

Q15: Does your university have a specific policy about who has or can have access to this specific data?

1. Yes
2. No

Q16: Who has access to the collected data? [multiple responses possible]

1. specific functions within the central university administration
2. all members/large groups of the central university administration
3. specific member(s) of the collecting faculty/faculties of the university
4. specific member(s) of the official student organisation
5. specific member(s) of the external partner/organisation/consortium: ...
6. specific member(s) of the governmental agency: ...
7. (diversity) policy officer(s) at the university
8. other: ...

(please specify 'other')

Q17: Under which circumstances do(es) the (diversity) policy officer(s) have or could gain access to this data?
[multiple responses possible]

1. on simple request (no special circumstances)
2. having reasonable grounds (e.g. diversity policy questions)
3. having authorisation from leadership (e.g. rector/vicerector/dean)
4. signing a confidentiality agreement
5. having signed a code of conduct
6. agreeing to the unit's/faculty's conditions concerning access
7. explicit permission of the students/staff members to use their data for policy analysis
8. none, not available

Q18: What would access to this data for a (diversity) policy officer(s) most likely look like? Note that the answer might depend on the actual need or policy question. Please specify the most detailed level in which data would be made available.

1. raw, detailed data, linkable to other datasets
2. pseudonymised raw data, unlinkable to other datasets
3. descriptive reports, tables and/or numbers
4. reports with detailed analysis
5. access to physical files (e.g. cupboards with paper files or boxes with written exams)
6. other: ...

(please specify 'other')

Q19: Further explanation of this indicator and the context in which the data is being collected and used (or not).

(please specify 'other')

Q20: What issues/problems have been identified in relation to this indicator and, where applicable, what solutions are under consideration?

(please specify 'other')

*[see instructions following Q2, Q3, Q4 and Q5]
[when instructions completed, offer Q21 as a general final question]*

Q21: Is there anything else you would like to add or specify concerning this mapping exercise?

(please specify 'other')

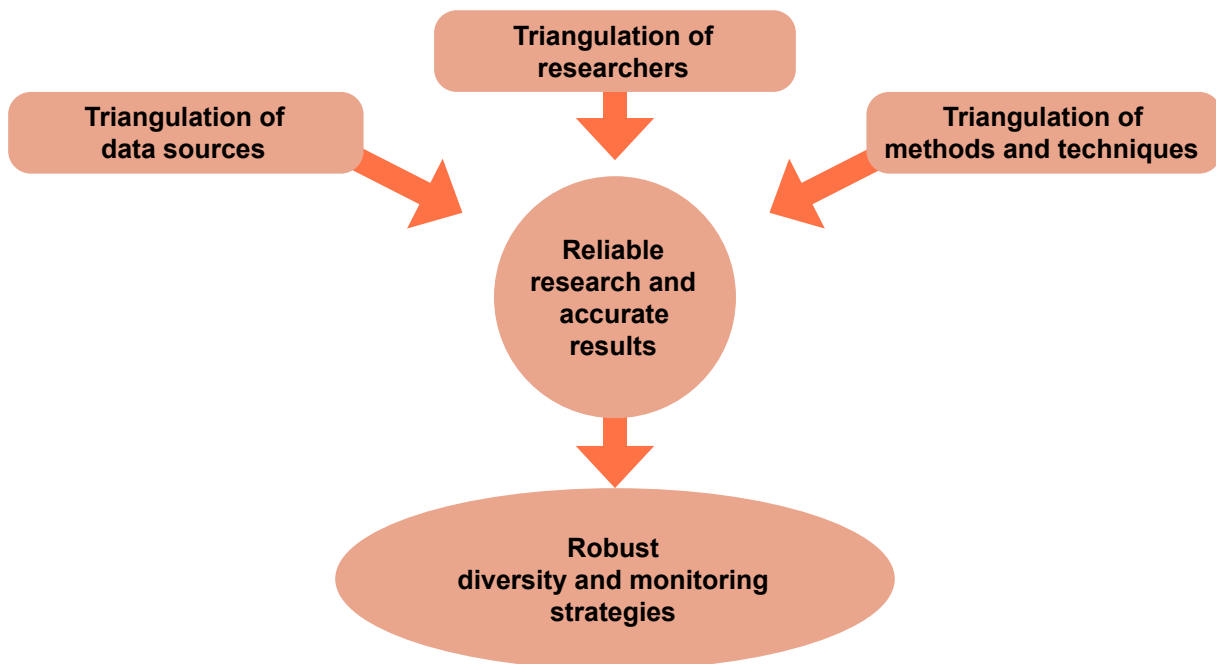
[end of survey]

Appendix 4: Triangulation

Stella Strzemecka, Jagiellonian University in Kraków

Social research uses the term triangulation to describe using two or more methods, data sources and/or researchers in conjunction to investigate a research question in order to increase the reliability and validity of the research results. It enables both researchers and university authorities to look at the problem under analysis from multiple perspectives.

Three levels of triangulation in diversity research in a university setting



Source: own work based on K. Konecki (2000). *Studia z metodologii badań jakościowych. Teoria ugruntowana*. Warszawa: PWN.

The interdisciplinarity of researchers, integration of multiple research methods and data sources fosters a comprehensive analysis of phenomena and processes related to diversity at universities, bridging the limitations of individual research methods. The combination of qualitative and quantitative methods offers particular potential. It is possible to reach respondents, for example, both in in-depth interviews and by means of a survey questionnaire. Linking a variety of complementary data sources enables connections to be made, data to be verified and gaps to be filled.

Researchers specialising in different research methods and disciplines can also be involved in the research process, contributing their respective perspectives.

Through the triangulation approach, the conclusions and recommendations for the future can be more rigorous and valuable.

At the same time, triangulation requires a greater commitment of resources than use of one method. The strategy of linking quantitative and qualitative diversity data may require the design of a university research schedule, along with the distribution of roles and responsibilities among those units involved in data collection and analysis.

Appendix 5: Gender Data

Gabriele Rosenstreich, Freie Universität Berlin

Sex is assigned to babies at birth, enshrined in legal documents and used to structure anything from legal rights through to toilet access. Sex/gender data has been part of administrative data collection in numerous (but not all) universities in the EU for many years for the following reasons, and possibly more:

- Individual proof of identity
- Monitoring sex/gender-based representation and thus structural inequality
- Planning targeted measures or services based on sex/gender
- Facilitating gender-specific language in communication.

In addition, of course, research often gathers data to identify disparities between the sexes/genders³⁵ and/or correlations between other factors and sex/gender.³⁶

At first glance gender may appear to be one of the most straight forward categories of diversity (alongside age), however, when considered in more depth, a number of the interrelated issues discussed in this report become evident. The reflection of some of those issues can be considered a paradigmatic example of the complexity to be taken into account when gathering or using diversity data.

Some Issues

Firstly, sex and gender are usually conceptualised within a binary model, i.e. data is collected using the values 'female' and 'male'. However, peoples' lived realities sometimes differ from this binary, and social and biological sciences increasingly acknowledge the complexity of sex and gender. For example, a significant number of people are intersex, i.e. born with physical sex characteristics that differ from common understandings of male or female.³⁷

The terms themselves demonstrate another layer of complexity. In English the terms 'sex' and 'gender' are often used interchangeably but are in fact two concepts used to differentiate between a physical/biological state (sex) and non-material or social dimensions (gender).³⁸ Sex can be defined using a range of indicators, including chromosomes, reproductive organs and hormones, which may or may not align (e.g. a woman whose uterus is removed remains female and a person may have a level of testosterone or chromosomes assigned the value 'male' while the other indicators are assigned the value 'female').

35 In the past much research was conducted without differentiating data by sex (sometimes using only male subjects), resulting in inaccurate conclusions in relation to other sexes, sometimes with life threatening consequences. As a result in 2016 the US National Institutes of Health introduced a [policy on sex as a biological variable](#) that should be factored into research designs, analyses, and reporting in clinical medical research.

36 Each individual is positioned in relation to multiple social categories in addition to gender and thus at the intersection of complex societal and organisational power relations. See section 3.2.5 above and for more indepth discussion of the concept intersectionality, Collins/Bilge (2016).

37 The exact numbers of intersex people cannot be reliably estimated as there are significant limitations to the data available. Based on a systematic review, Blackless et al. (2000) estimated that around 1.7% of live births are intersex. See [Intersex Human Rights Australia](#) for a discussion of intersex population figures.

38 See for example, the differentiation and interpretation of the terms 'sex' and 'gender' that underlies the work of the UK Office for National Statistics (www.ons.gov.uk/economy/environmentalaccounts/articles/whatisdifferencebetweensexandgender/2019-02-21)

Gender, the socially constructed aspect, can also be differentiated according to, for example, the subjective sense of self (gender identity),³⁹ how gender is expressed through clothing, behaviour, etc. (gender expression), what societal norms and expectations are associated with gender (gender roles). In everyday life (and in most cases at birth), sex is usually ascribed on the basis of external anatomy (genitals), and, after puberty, so-called secondary sex characteristics (e.g. breasts, muscle mass, hair distribution) in combination with gender expression and other external factors such as gendered names. Again, the ascribed sex or gender may or may not align with the actual sex and/or gender of the person based on other indicators.

Whether the label 'sex data' or 'gender data' is used, is not necessarily reflective of which of the two concepts is being drawn upon. Many languages do not make any distinction between sex and gender (e.g. German = 'Geschlecht') so translations into English vary (and are often not based on a differentiated understanding of the terms in English) and, indeed, scientific developments have begun to question the distinction too as the socially constructed nature and instability of the concept of sex becomes more evident. Increasingly, the term 'gender' is being used as the general default to encompass all aspects of this dimension of diversity. For that reason, and because data collection actually seldom relates directly to physical sex characteristics, this report uses 'gender' as a general default term unless explicitly referring to a specific indicator.

Gender data collection has traditionally been grounded on two core assumptions: That sex and (social) gender are congruent and that gender is static, i.e. does not change over time. This is why gender has long been used as a key indicator for proof of identity in a wide range of contexts. Of course, neither assumption is accurate for all people.

While physical sex and social gender are often congruent, this is not always the case and they interact in various ways, e.g. trans people have a gender identity that differs from the sex that they were assigned at birth and they may or may not make changes to their hormones and/or sex organs to better align the two; most intersex people do not have a non-binary gender identity but rather identify as male or female⁴⁰; and people with a non-binary gender identity may or may not have a gender expression or fulfil gender roles that are commonly associated with their physical sex, and this may change day to day. But even people whose gender identity aligns with the sex assigned at birth (cisgender) often have a gender expression or gender roles that do not conform to social norms (which are of course embedded in a specific historical and social context and thus change).

Gender data is usually explicitly or implicitly based on the person's gender marker, i.e. the official entry in identity documents.⁴¹ In most cases this equates to the sex assigned at birth and recorded in the person's birth certificate (which is used as a basis for issuing other key official documents such as government identity cards, passports and drivers licences in many countries). It could, therefore, be assumed that most gender data relates to the indicator physical sex. This is, however, not necessarily the case.

39 There are a wide range of gender identities. The terms used to describe them vary over time, societal context and of course language. In the UK at this time they include, for example, female, male, non-binary, cis, queer, gender-queer, gender-fluid, trans, trans feminine, trans man and agender/nongender people, who have no gender identity. And the meanings accorded each term vary significantly too. Often trans is used as an umbrella term to describe all people who are not cis, i.e. whose gender is not the same as the sex assigned at birth. See, for example, the TGEU glossary: <https://tgeu.org/glossary/>

40 For example, a 2015 Australian study of intersex people found that 52% identified as female, 23% as male and only 7% as 'X'. A further 6% were unsure and 12% identified in some other way. Multiple responses were possible (Jones 2017).

41 Medical and biological research are the major exception, since the core interest is usually in physical characteristics. In all other cases, it is likely that most respondents will assume that the official gender marker is required unless otherwise specified, especially in the case of formal documents, such as application forms.

Firstly, while numerous countries have introduced gender markers for people who are neither male nor female (based on various criteria),⁴² most countries still allow for only two gender markers. Infants are therefore categorised as male or female at birth, generally based on their visible genitalia. Therefore, even those people who are identified as intersex at birth are almost always assigned male or female (and in many cases medical interventions are carried out to better align them with ideas of male or female bodies).⁴³ Gender markers therefore do not necessarily accurately reflect physical sex.

Secondly, trans, non-binary and intersex people sometimes change their official gender marker. The laws around gender recognition vary significantly across countries, even within the EU.⁴⁴ Many countries either do not permit changes or have very high barriers in place, often requiring significant physical/medical interventions, while other countries allow changes to the gender marker based on a simple declaration.

In short, no consistency can be assumed between physical characteristics, gender identity and the official gender marker and anyone of these indicators can change over time.⁴⁵

This means that gender data is not actually suitable for identification purposes. It also, however, poses challenges for the other purposes underlying gender data collection.

Diversity data is, above all, used to monitor the representation of minoritised groups within an organisation - and thus structural inequality - and as an evidence base for planning targeted measures or services (in particular those that address the needs of minoritised groups). The collection of gender data for these reasons has become common practice in most European countries over recent decades and is mandatory for most universities (both in terms of government reporting requirements and as a prerequisite for third party funding programmes, such as the EU Horizon research and innovation programme⁴⁶).

The focus of such efforts has been to measure the presence of women in order to combat sexism (that continues to result in significant structural inequality). It is, however, not always, clear what is meant by 'women'. The term can, as noted, refer to physical characteristics, to gender identity, to legal status, etc. It is, for example, often not clear whether trans and intersex women and/or trans men are included in such measurements and if so on what basis (e.g. self-identification or official gender marker). Usually, it is lived

-
- 42 For example, Argentina, Austria, Australia, Bangladesh, Canada, Colombia, Denmark, Germany, Iceland, Malta, the Netherlands, New Zealand, Pakistan, India, Nepal, and the USA all allow for a separate gender marker for intersex or non-binary people or those of unspecified gender, at least under some circumstances (conditions vary). For example, German civil status law (Personenstandsgesetz 2018) defines 3 gender markers: 'male', 'female', 'diverse' (German: 'divers') and enshrines the right to have no gender marker. The gender marker 'diverse' does not equate to any identity labels and is officially only for intersex people, however, this is contested and increasingly the gender marker is also used to encompass non-binary people in a range of contexts. In other countries, some jurisdictions or organisations also allow for three or more gender markers even if it is not officially mandated at a national, federal or state level. The markers themselves vary but in many cases 'M', 'F' and 'X' are used on official identity documentation in line with regulations for machine readable passports introduced by the International Civil Aviation Organization.
- 43 Intersex activists advocate against the practice of medical interventions on infants and children when these are not necessary for health reasons (few intersex conditions actually require intervention). Gradually medical practice is changing, however, interventions on infants are still common in most countries (sometimes even without the consent of parents).
- 44 Almost all European countries allow some form of legal gender recognition for trans people (in the EU Bulgaria and Hungary are the exceptions). See the ILGA Europe Rainbow Map for detailed information on the specific conditions of each: www.rainbow-europe.org/#o/8701/o
- 45 In some cases people can even have more than one official gender marker. In the USA, for example, it is possible to self-select the gender marker that will appear on the passport. The options are currently male (M), female (F), or unspecified or another gender identity (X). It is not required to match the gender marker on supporting documentation, such as a birth certificate or state identification documents. See <https://travel.state.gov/content/travel/en/passports/need-passport/selecting-your-gender-marker.html>
- 46 The Horizon programme notes that an increasing number of countries and organisations are collecting "gender statistics" according to gender identity, with usually at least three categories considered - female, male, and non-binary -, and provides refers to recommendations on sex/gender disaggregated data in [Gendered Innovations 2: How inclusive analysis contributes to research and innovation](#) (European Commission, 2020)

experience as a woman, as a non-binary person, etc., that is key to understanding inequality, rather than an official gender marker, yet, as noted, much data will reflect the later rather than the former. In addition, there is an increasing awareness of the discrimination faced by intersex people, trans people, and non-binary people, groups that are clearly also minoritised on the basis of gender. Their experiences often overlap with those of cis women (e.g. high rates of sexualised violence), but there are also significant differences (e.g. stigma). Although the discourse around gender-based inequality is shifting to include all groups minoritised on the basis of gender, currently trans, intersex and non-binary people are rarely included in actual measures, services and programmes designed to address gender-based inequality.

In some countries and within some organisations, questions regarding the inclusion and exclusion of trans and non-binary people, and to a lesser extent intersex people (who often remain particularly invisible), are hotly contested political debates. This discourse is often at odds with current scientific understandings of gender as a complex concept and itself stigmatises and impacts negatively on the wellbeing of people who do not conform to traditional binary gender norms. It should not, however, be forgotten that these heated debates are not universal and in many contexts the aim of gender-inclusion is uncontroversial. In either case, the question of which genders are included and how they are defined, can pose challenges for policy development and operational planning, as well as for the collection of data that can usefully and accurately assist it.

Another challenge relates to 'numbers'. As noted, most diversity related data collection measures the existence of members of the minoritised groups within the organisation. In the case of intersex, trans and non-binary people the numbers are likely to be very low, leading to particular caution regarding privacy. The more granular, or detailed data is and the greater the level of linkage, the easier it might be to identify an individual. For example, if a university reports that there is one non-binary bachelor degree student in a small faculty, it may lead to people guessing who that could be. If that data is further broken down, and for example, shows that of the bachelor's students in that faculty only one of the 'first generation' students is non-binary, the guessing becomes more accurate and the stigma potentially greater. On the one hand, intersectional data is highly valuable as a differentiated reflection of complex relationships between categories of diversity, but on the other hand it can quickly result in very small numbers, with associated data protection needs that may lead to data not being able to be reported and thus used effectively.

In addition, because we have no reliable reference data on population numbers (regardless of the indicators) we cannot estimate how low they are likely to be and assessing how representative data is becomes impossible. Also – and the two aspects are linked - because there is such a high level of stigma (and associated risk of discrimination and violence), a high level of trust and data protection is generally required before intersex, trans and non-binary people will disclose their gender. Poorly defined values that do not align with the terminology people use for self-identification contribute to distrust. Many organisations consider data on gender identity to be so sensitive that it is not collected at all (and in other cases the organisation does not acknowledge these groups as existent or as having policy needs and does not collect data for that reason). Even when it is collected, underreporting can almost always be assumed.

Low numbers also pose an ethical and political challenge. Often policy development draws on 'numbers affected' to define the significance of an issue, to justify and plan targeted measures and services and to set priorities, i.e. the more people who are affected, the more important it is. Even in the unlikely event that there was only a single trans person in a university, would that justify that person not being addressed correctly and their needs (e.g. being treated with respect, access to toilets and other services structured by gender) not being met? And is not the lack of trans people in itself an indication of barriers in place to them entering the organisation, feeling able to disclose and/or having a clear articulation of their own identity? How many is 'enough' to justify action being taken? If low numbers are made public, can that not be instrumentalised for transphobic arguments? Is it, in fact, in the interests of trans and non-binary people to be counted if the

numbers are low? These questions touch on the heart of the definition and vision of diversity underlying an organisation's work.

Sometimes gender indicators are collected in digital data systems to facilitate gender-specific language in communication. In English this relates primarily to forms of address (Ms/Mr, etc) and pronouns (she/he/they, etc), however other languages also have a grammatical gender. German, for example, declines all nouns according to male, female, neutral (der, die, das), i.e. the ending of the word changes according to the gender of the person or people being referred to. This impacts on, for example, job titles (e.g. cleaner, lecturer, technician), other titles (e.g. Dr, Dean, Chair) and other nouns referring to people (e.g. student, complainant, applicant) so that data systems require accurate gender information if they are to use gender-specific words correctly, e.g. in official documents. If the official gender marker is used, this may result in people being misgendered, i.e. contribute to an organisational culture that renders some genders invisible and in doing so fosters discrimination.

Information about gender is often signalled via names, i.e. names are a form of nonnumerical data. Trans and non-binary people whose official documents contain their previous or 'dead' name rather than their preferred name, which reflects their gender identity, face involuntary disclosure every time this data is used. This applies in particular to first names, which are frequently (but not always) associated with one gender in particular, but in some languages, surnames also vary depending on gender (in Europe, Polish and Czech for example).

As noted, most gender data collected relates to the gender of the respondents/members of the organisation. Data is much more rarely collected on the actual experiences they have in relation to that gender or on the impact of measures and programmes aiming to reduce gendered inequality beyond monitoring changes in the numbers of women in the organisation or in certain roles. Organisations thus have a very limited evidence base on how barriers to the full and equitable participation of minoritised genders work and what determines whether measures to combat them are successful.

Some implications for data collection

The complexity of gender has several implications for data collection. At the most fundamental level, it requires us to consider the purpose of the data collection very carefully, e.g. given that the gender marker can change, other data is more likely to provide a robust proof of identity. We should be collecting and using data in the light of that need. This means that we need to carefully consider which indicators are actually of interest and which values can provide meaningful and robust insight into them. In particular, if we are measuring gender-based inequity, it relates more to the lived experience than necessarily to the official gender marker on identity documents.

Universities can respond to the issues in a range of ways, depending primarily on the identified needs. For example, when considering which indicator(s) to use in relation to sex/gender there are at least four options:

a. Self-defined gender identity.⁴⁷

Pro: strengthens self-determination; can be a more accurate indicator of lived reality than other indicators; protects the privacy of trans individuals

Con: Potential problems at the interface with external systems if they require the legal gender identity

47 This approach was taken at Freie Universität Berlin in relation to students. See Good Practice Example 1.

- b. Gather two (or more) gender indicators, in particular self-defined gender identity and legal gender marker.⁴⁸
 Pro: enables interface with external systems and internal self-determination; most accurate indicator of lived reality; data linkage provides insights into the relationship between the different gender indicators; affirming for people with minoritised genders
 Con: technical challenges in some digital data systems; issues remain in relation to when the legal gender marker is used and when the actual gender identity is used; potential for involuntary disclosure/outing of trans people;
- c. Gather only legal gender marker.
 Pro: clear basis.
 Con: does not reflect lived reality and therefore does not meet some data needs; discriminates against people whose gender marker differs from their gender identity (potential for forced disclosure/outing); renders people whose gender is not reflected in those markers invisible and signals to them that they are unimportant
- d. Gather no gender data at all.
 Pro: avoids the problem – we often gather gender data without a defined need
 Con: no longer able to identify and monitor gender-based inequity; makes it harder to target services/measures to actual needs; non-adherence to existing reporting requirements; Depending on the language(s) used, formulating all communication gender neutrally may pose challenges.

In all four options, instruments and reports need to explicitly communicate which indicator is being used. For example, many (but not all) applicants will assume that the official gender marker is required unless otherwise stated, potentially leading to inaccurate results. We also require a range of values for indicators of gender (i.e. more than just male and female) in order to more accurately reflect lived reality. At minimum data systems should allow values for

- gender markers aligned with those used on official identity documentation in that country (e.g. in Poland it would be ‘female’ and ‘male’ while in Germany it would be ‘female’, ‘male’ and ‘diverse’)
- ‘other’ for those genders not included in the official gender marker options and for other gender markers on identity documents from other countries⁴⁹
- ‘no entry’, ‘unknown’ and/or ‘prefer not to say’.

This approach ensures as well as comparability with external data sets, however, it has the disadvantage of othering those gender identities that do not fall within the existing legal confines. It also does not provide information on the numbers of people who are trans, since the definitions of male, female and non-binary would include trans people, who of course identify as a gender, just one that differs from the gender assigned at birth. If data specifically on trans people or changes in gender identity or markers are needed, then this would have to be explicitly gathered.⁵⁰ Similarly, most intersex people identify as male or female and these

48 Collecting actual ‘sex’ data, that relates specifically and exclusively to physical sex rarely useful in terms of the needs underlying data collection (except of course in the case of data for medical purposes or biology research). It therefore appears more useful to distinguish between the legal gender marker and (lived) gender identity, rather than between sex and gender in the context of university data collection.

49 For example, even if the national legal system only allows for a male/female binary (as in Poland and the UK), there may be staff or students who are citizens of countries that do allow for further categories (as in Germany) and thus have other gender markers. U Edinburgh added the value ‘other’ to gender data specifically to accommodate foreign students.

50 The University of Edinburgh, for example, collects data on changes in gender identity in addition to data on the legal gender marker (see above). The 2021 Australian National Student Safety Survey, for example, asked both about the sex assigned at birth (providing the options “female”, “male”, “Another term (please specify)” and “prefer not to say” as well as about the lived gender identity (item: “How do you describe your gender? Gender refers to current gender, which may be different to sex recorded at birth and may be different to what is indicated on legal documents” Response options: “Woman or female”, “Man or male”, “Non-binary”, “I use a different term (please specify)” and “Prefer not to say”). The data collected was linked and when reporting the findings and distinctions were made between cis and trans respondents. See Heywood et al. (2022).

values will provide little or no insight into how many respondents have an intersex variation so a separate question would be required for that.⁵¹

A further problem with this approach is that it merges indicators (physical sex with gender identity) in a way that does not necessarily reflect people's own understandings or lived realities. Most intersex and non-binary people would see little overlap given that the former relates specifically to physicality and the latter to gender identity, however, these groups are frequently conflated in so-called 'third gender' options.

It would appear most accurate to provide a range of gender options (e.g. female, male, intersex, trans, non-binary, cisgender) and to allow multiple responses (i.e. a person could enter cisgender and male, trans and non-binary, trans and female, or intersex and female)

The terms used in instruments determine the data collected.⁵² For example, in Germany the term 'diverse' was introduced in law 2018 to designate a gender marker for intersex people but has never been a term of self-identification, and ideas about who it includes as well as who uses it varies, so that when used in a form it is likely 'ticked' by both intersex and non-binary people as well as not ticked by numerous intersex people who identify as and are legally designated men or women. The responses would potentially be quite different if the terms 'intersex' and/or 'non-binary' were used.

The terms also influence the respondents' perception of and relationship with the organisation collecting the data. When language is used that respondents find alienating or even offensive, it will not only likely lead to inaccurate responses (increasing response bias) and higher dropout rates, but it will also erode trust in the organisation. For example, use of the term 'transsexual', while common in the past, is now widely considered inappropriate, by younger trans people in particular. A further example: In Germany, organisations are now legally required to take all three gender markers (and no marker) into account, and there have been examples of forms which provide options of how the person would like to be addressed that include 'Mr, Ms, Diverse'. Correspondence generated from this form is then addressed to "Diverse Smith". This is a clear signal to trans, non-binary and intersex people in particular that the organisation has not seriously reflected on the topic of gender and most certainly has not consulted with community members.

In turn, the respondents' perceptions of and trust in the organisation influence their willingness to disclose potentially stigmatising or otherwise sensitive data, such as a minoritised gender. For this reason, some organisations outsource some or all diversity data collection to external agencies.

Ideally, regardless of the values used, the instruments would also include a field for comments so that respondents could elaborate on their entry if they wish. This could assist in the interpretation and cleaning of the data.

To maximise self-determination, gender data could be entered via a text field so that people's own terms are used. These could be coded and categorised as part of the data analysis. Such an approach, however, has the disadvantage of preventing comparability with external data sets, requiring a high burden of analysis, having significant scope for variance in interpretation and thus less accuracy of the findings. It is most robust and useful when included as an option in addition to predefined values (see for example Heywood et al. 2022).

51 See the recommendations by Intersex Human Rights Australia: <https://ihra.org.au/forms/>

52 Even seemingly self-evident terms such as 'woman' require explicit descriptions to ensure that the respondent understands which underlying indicator (e.g. official gender marker, gender identity) the question refers to, i.e. whether trans women are meant too. And certainly not all respondents will identify with or even understand quite common terms such as trans or intersex, requiring items in forms or surveys to be descriptive and in plain language. For example, Intersex Human Rights Australia's recommends "Intersex is a term for people born with atypical physical sex characteristics. There are many different intersex traits or variations. Do you have an intersex variation? Yes/No" (<https://ihra.org.au/research/>)

Consulting with people who have minoritised genders about the terms used when defining values assists in ensuring that they are robust. Testing instruments with diverse respondents enables us to evaluate acceptance and understand how people are interpreting the items and response options. Further, we should be consulting with people who have minoritised genders at all stages of the data collection process as part of quality assurance, commencing with the formulation of data needs, through to the interpretation of the results.

Intersections with external agencies pose a significant challenge to efforts to gather gender data in a more robust and inclusive manner, beyond the issue of comparability already noted. Universities intersect with external systems in a range of ways, for example, funding programmes that target women, official data collection agencies, government identification documentation, health and social services. There may be discrepancies in the respective indicators and values⁵³ used and this can not only lead to problems in terms of accuracy and interpretation, but also to technical issues. For example, if an employer is required to migrate some staff data to pension funds, health insurers, etc. (as is the case in Germany for example) and a staff member is listed as female in the HR data system while being recorded as male in the external agency and/or is recorded under their preferred name rather than their official name, this may result not only in a data error being recorded but may potentially lead to the involuntary outing of the person concerned and/or them not receiving benefits or services due to them. Some organisations address this issue by having parallel fields in their digital data systems, one for the official gender marker and one for the gender identity, one for the official name and one for the preferred name (see Good Practice Example 5). They draw on the field with official data when reporting data to external agencies who stipulate this indicator and use the other field for all other purposes. Because there are usually a large number of interfaces between data systems, any changes to fields such as name and gender (e.g. the addition of new fields), requires careful checking and adjustment to ensure smooth data transfer and avoid unintentional consequences.⁵⁴

One of the ways that gender data can be used, is, as noted above, to enable gender-specific communication. Given that this purpose relates to communicating with people respectfully, gender identity is the most logical indicator to use. Alternatively, communication templates linked to central data systems can be gender neutral or inclusive of all genders, so that they do not rely on accurate gender data.⁵⁵

In addition to information about the gender of respondent/members of the organisation, data is also required on experiences that people have in relation to their gender and the impact of measures and programmes aiming to reduce gendered inequality. A range of instruments exist that can be used to collect such data and examine the relationship to existing gender diversity and/or structures in the organisation. For example,

53 This should, however, always be checked. Sometimes it is assumed that the external agency requires data on official gender markers and/or names, whereas in reality that is often not the case.

54 Organisations also face other types of practical challenges when using external software and platforms. They may, for example, provide default options or use gendered terms that do not align with the approach taken internally. For example, Freie Universität Berlin currently uses an online learning platform in which the terms for ‘guest’, ‘participant’, ‘facilitator’ etc. are in a masculine form (the so-called ‘generic masculine’, which, supposedly applies to everyone). This contradicts the policy of gender-inclusive language the university strives to promote. Sometimes the software or platform can be adjusted to individual needs, however, this may require significant resources; sometimes it can be more effective to lobby the provider to adjust its product to be more inclusive.

55 Freie Universität Berlin introduced a formal policy on ‘gender-inclusive’ language 2022, expanding on and formalising practice already being used for communication with students: All official communication on behalf of the university is now required to use language that is gender-neutral or encompasses all genders. For example, templates now use the first name and surname rather than Ms/Mr and surname as the standard form of address, inclusive phrases such as ‘Dear colleagues’ replace ‘Dear ladies and gentlemen’ (German: ‘sehr geehrte Damen und Herren’), and inclusive or gender neutral terms are used (e.g. for students ‘Student*in’ or ‘Studierende’). Staff are also encouraged to include their preferred pronouns in their email signatures. Communication can, however, be gender specific if only one gender is meant or an individual’s preferred pronouns are known. See www.fu-berlin.de/en/sites/diversity/antidiskriminierung/sprache. In terms of data collection, this means that while the field ‘gender’ remains in the data system for other reasons, it is not linked to templates to determine language use.

gender impact assessments,⁵⁶ satisfaction and organisational climate surveys,⁵⁷ and, of course, evaluation methodologies. Evaluations, in so far as they are conducted, are often limited to the collection of data on the numbers of participants (by gender) and their level of satisfaction or success on completion of the measure or programme. This provides little insight into how barriers to the full and equitable participation of minoritised genders work and what determines whether measures to combat them are successful. Participative evaluation models offer potential in relation to diversity. They include members of the minoritised groups - in this case not only women but also intersex, trans and nonbinary people – in the definition of criteria for success and quality and in the assessment of whether they have been fulfilled. Participation of minoritised people in the generation of evaluation data, must however, occur in a safe way that addresses existing power imbalances and inequity (including in compensation for work done), dependence and mitigates for risks.

This discussion of issues related to gender and their implications is by no means exhaustive. And, as noted, most of the issues discussed here can also be related to other categories of diversity. In all cases, no matter how straight forward the category may appear, a deliberate and reflective approach is required to critically reflect on the assumptions and challenges underlying data collection and their implications.

-
- 56 Gender impact assessment is a tool within gendermainstreaming approaches. The European Commission defines it as the process of comparing and assessing, according to gender relevant criteria, the current situation and trend with the expected development resulting from the introduction of the proposed policy. It is the estimation of the different effects (positive, negative or neutral) of any policy or activity implemented to specific items in terms of gender equality. See <https://eige.europa.eu/gender-mainstreaming/toolkits/gender-impact-assessment/what-gender-impact-assessment>
- 57 Satisfaction surveys gather anonymous data on how respondents (often staff) perceive and judge the organisation (see Good Practice Example 2) while organisational diversity climate surveys collect data from members of the organisation (in the case of universities: students, faculty and other staff) about their attitudes towards and experiences relating to diversity within the organisation, including organisational culture, discrimination, inclusion, and equity. Campus climate surveys are beginning to be used in some European universities but appear to be more common in North America. See for example, the University of California, Santa Barbara <https://diversity.ucsb.edu/climatesurvey>.



Una Europa vzw

KU Leuven Campus Brussels
Warmoesberg 26
1000 Brussels – Belgium
© Una Europa vzw, 2022

info@una-europa.eu ↗

www.una-europa.eu ↗

**Keep up to date with 1Europe and
Una Europa online**

[Twitter](#) ↗

[LinkedIn](#) ↗

[Instagram](#) ↗

[Facebook](#) ↗