

Kai Mühleck / Christoph Meng / Martin Unger / Ute Hoffstätter

EUROGRADUATE Pilot Survey 2018

Data and Methods Report on the EUROGRADUATE Pilot
Survey conducted in 2018/2019

Data and Methods Report

July 2021



Authors:

Kai Mühleck
Christoph Meng
Martin Unger
Ute Hoffstätter

Editor:

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)
German Centre for Higher Education Research and Science Studies (DZHW)
Lange Laube 12 | 30159 Hannover | Germany | www.dzhw.eu
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960 | info@dzhw.eu

Management board:

Prof. Dr. Monika Jungbauer-Gans
Karen Schlüter

Chairman of the supervisory board:

Ministerialdirigent Peter Greisler

Registration court:

Amtsgericht Hannover | B 210251

The EUROGRADUATE Pilot Survey was conducted under a contract with the European Union and with the support of the Erasmus+ Programme of the European Union. The opinions expressed are those of the authors only and do not represent the European Union's official position. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of the information therein.

The content of this publication is licensed under the Creative Commons License "Attribution - Non-Commercial – Share Alike 3.0" (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/>



Table of Contents

List of Tables	IV
I Introduction.....	5
II Data Use Instructions	6
1 Outline of the Study	7
2 Survey Instruments.....	8
3 Target Population	8
4 Sampling, Contacting, and Return to the Survey.....	9
4.1 Data collection in Austria	10
4.2 Data collection in Czechia	11
4.3 Data collection in Germany.....	11
4.4 Data collection in Greece	12
4.5 Data collection in Croatia	14
4.6 Data collection in Lithuania.....	14
4.7 Data collection in Malta	15
4.8 Data collection in Norway	16
5 Data Preparation	17
5.1 Data Checking and Data Cleansing.....	17
5.2 Generation of Variables	17
5.3 Coding of Missing Values	17
6 Weighting	18
7 Anonymisation.....	19
8 Bibliography.....	22

List of Tables

Table 1: Statistical Anonymisation Measures for the Data of EUROGRADUATE Pilot Survey 2018.....	20
--	----

I Introduction

The EUROGRADUATE Pilot Survey aimed at laying the ground for a sustainable European-wide graduate survey. For this end, pilot surveys have been conducted among higher education graduates in eight pilot countries: Austria, Czechia, Croatia, Germany, Greece, Lithuania, Malta, and Norway. The project simultaneously surveyed the graduate cohorts of 2016/17, and 2012/13. The EUROGRADUATE Pilot Survey was funded by the European Commission.¹

The Data and Methods Report is part of the documentation of the EUROGRADUATE Pilot Survey (doi: [10.21249/DZHW:egr2018:2.0.0](https://doi.org/10.21249/DZHW:egr2018:2.0.0)). Further documentation materials for the study (e.g. codebook, questionnaires etc.) can be downloaded from the search portal of the RDC (<https://metadata.fdz.dzhw.eu/#!/en>) as well as from the webpage of the EUROGRADUATE Pilot Survey (www.eurograduate.eu). Specifically the technical assessment of the EUROGRADUATE pilot survey (Mühleck et al. 2020)² provides more details on the target population, questionnaire, and conduct of the EUROGRADUATE Pilot Survey.

Key information on the use of the data is presented in Section II. Chapter 1 introduces the content and structure of the EUROGRADUATE Pilot Survey. The remaining chapters of the report are oriented towards the stages of the research process. In Chapter 2, the applied survey instruments are described, and the survey implementation process (e.g. sampling procedure, survey operation, data preparation etc.) is detailed in Chapters 3, 4, and 5. In Chapters 6 and 7, weighting and anonymisation practices used are presented.

¹ The opinions expressed are those of the authors only and do not represent the European Union's official position. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of the information therein.

² See <https://www.eurograduate.eu/results>, technical assessments

II Data Use Instructions

[Data Use Requirements] The data of The EUROGRADUATE Pilot Survey are anonymised and made available by the RDC of the DZHW in accordance with the European General Data Protection Regulation (GDPR) and released exclusively for scientific use.³ The RDC provides a *Scientific Use File* (SUF) for scientific secondary use.

Requirements for the use of a SUF are an employment at a scientific institution and the conclusion of a data use agreement. Students or doctoral candidates without a position at a scientific institution must conclude a data use agreement together with a supervising staff member. In the course of concluding the contract, the RDC also checks whether there is any scientific interest in using the data. The data usage application form can be downloaded from the RDC website.

[Data Access] The SUF of the EUROGRADUATE Pilot Survey can be used at the local computer.

- **Download:** Data are available for download via a secure connection from the RDC website. Users can save the data on their local computer to link with data from external sources as well as perform analysis using their own software.

[Charges for Data Access] SUF are available free of charge (effective December 2019). The present fees regulation can be found on the RDC website (<https://fdz.dzhw.eu>).

[Responsibilities of Data Users] Data users are obliged to observe the following rules⁴:

- **Scientific Use:** Data must be used exclusively for scientific research purposes. Commercial use is forbidden.
- **De-anonymisation forbidden:** Any attempt of re-identification for the units of analysis (e.g. persons, households, institutions) is prohibited.
- **Duty to report security loopholes:** If data users become aware of security loopholes with respect to data protection or data security, the RDC should be informed immediately.
- **No data disclosure:** SUF may only be used by persons who have made a data use agreement. CUF may only be disclosed in the context of specified teaching activities.
- **Duty to delete:** SUF downloads must be deleted after expiry of the agreed period of use (as a rule 1.5 years) from all computers, servers and data storage devices. Likewise all backup copies, modified data sets (e.g. work-, excerpt- or help-data) as well as print-outs must be destroyed.
- **Notification/Provision of Publications:** The RDC has to be immediately notified of all types of publications that are produced using data of the RDC. An electronic version of the publication shall be provided immediately.
- **Citation rules:** The data used must be cited in publications, other work (e.g. theses) and lectures according to the RDC guidelines.⁵

³ The RDC's data protection policy is based on the portfolio approach of Lane et al. 2008, pp. 6, on upon which the Leibniz Institute for Educational Trajectories (LifBi) (cf. Koberg 2016, pp. 699) and the RDC of the Federal Employment Agency at the Institute for Employment Research (cf. Hochfellner et al. 2012, p. 9) have oriented themselves. The RDC has adapted the portfolio approach to the requirements of its own data files and uses four categories of measures in securing data protection, which are combined in various ways: legal-institutional measures, informational measures, technical measures and statistical measures.

⁴ The data use agreement regulates terms and conditions of use in detail.

⁵ See [10.21249/DZHW:egr2018:2.0.0](https://fdz.dzhw.eu/10.21249/DZHW:egr2018:2.0.0)

1 Outline of the Study

The EUROGRADUATE pilot survey is a unique survey of recent graduates in eight European countries that was launched with the intention to lay the ground for a sustainable European wide graduate research.⁶ The study was funded by the European Commission following the recommendations of the European Council on improving the tracking of graduates in Europe (European Commission 2017). The study was conducted by an international consortium of 12 organisations. Six organisations worked as international core consortium: the University of Maastricht Research Centre for Education and the Labour Market (ROA, overall coordination, the Netherlands), DZHW (Germany), the Institute for Advanced Studies (IHS, Austria), DESAN Research Solutions (the Netherlands), cApStAn linguistic quality control (Belgium), and GESIS Leibniz Institute for the Social Sciences (Germany). The surveys in the pilot countries were conducted in close collaboration of the core consortium and national teams in the countries: Austria (IHS), Czechia (The Centre for Higher Education Studies (CHES)), Croatia (Faculty of Law, University of Zagreb & Agency for Science and Higher Education (ASHE)), Germany (DZHW), Greece (The Panteion University of Social and Political Sciences (UPSPS)), Lithuania (The Research and Higher Education Monitoring and Analysis Centre (MOSTA)), Malta (The National Commission for Further and Higher Education (NCFHE)), and Norway (The Nordic Institute for Studies in Innovation, Research and Education (NIFU)).

EUROGRADUATE targeted the two graduate cohorts of 2012/13 and 2016/17 simultaneously. In four countries, samples of graduates were drawn, with different sampling procedures reflecting the availability and storage of information needed for setting up the sampling frame. In Croatia, Greece, Lithuania, and Malta all graduates with available contact information were invited to the survey. In most countries contact information was available, for the (vast) majority of graduates. It must be noted, however, that for some countries and cohorts availability of contact information was very limited (see below for more details). The field phase took place between 08.10.2018 and 17.02.2019, with some variation between countries. All contacted graduates were invited to log in to an online survey hosted centrally by DESAN.

To ensure comparability of the data, the EUROGRADUATE pilot survey strived for as much standardisation in data collection methods as possible. The survey was hosted centrally and graduates of all countries were surveyed within the same computer-assisted interviewing framework. Translations and country-specific adaptations of the English master questionnaire were done by research teams within countries to build on in-depth knowledge of the context. To guarantee highest quality and comparability of translations, they were organised and monitored centrally as well as verified by experts to ensure that they are linguistically equivalent to the source. At the same time, data collection methods needed to correspond to the different starting conditions in the countries. Countries varied considerably e.g. regarding the availability and quality of population data, the availability and quality of contact information of the graduates, or the readiness of higher education institutions to participate in the survey. Despite all efforts, differences in survey conditions between countries may have led to differences in the quality of the data between countries. Researchers using the data are advised to see if any limitations are imposed on their analyses by carefully taking note of the descriptions of the data collection in the pilot countries provided in this report as well as of

⁶ For more information see www.eurograduate.eu.

additional information provided in the technical assessment of the EUROGRADUATE pilot survey (Mühleck et al. 2020).

2 Survey Instruments

The questionnaire covered the following topics: course and characteristics of higher education, previous and further education, transition to the labour market, labour market relevance of the completed programme, labour market outcomes, skills, (international) mobility, and social outcomes.

In designing the questionnaire the team dwelled on existing surveys in order to use well-established instruments if possible. The questionnaire of the international graduate survey REFLEX was an important source (Research Centre for Education and the Labour Market, 2005). Questionnaires of national graduate surveys, such as the DZHW graduate panel (DZHW Graduate Panel, n.d.), were considered as well. Moreover, the European Social Survey (www.europeansocialsurvey.org) was used as a source for several instruments for the section on social outcomes.

Building on these sources as well as, to the extent needed, newly designed instruments, the EUROGRADUATE consortium set up English master questionnaires for both cohorts, 2012/13 and 2016/17. Before translation, the source text was submitted to three linguists from different language groups (Hungarian, Greek, and Croatian) for translatability assessment.⁷ The master questionnaires were translated by the respective country team to the official languages of the country. In this process the questionnaire was slightly adapted if necessary to reflect characteristics of the country (e.g. country-specific degrees). Linguists of cApStAn checked translations, verified, or suggested alternatives. Verifiers had the task to ensure that the translation was linguistically equivalent to the source by comparing target versus source.

For all pilot countries, two language versions of the questionnaire have been prepared: (1) one in the official language of the country and (2) one in English to account for the needs of e.g. international students.

Technically, the survey was implemented by a centrally hosted computer-assisted interview (CAI) framework. The survey method for all respondents was a computer-assisted web interview (CAWI).

For more details on the questionnaire, the translation of the master questionnaire and the CAI framework see the technical assessment of the EUROGRADUATE pilot survey (Mühleck et al. 2020).

3 Target Population

The EUROGRADUATE pilot survey consistently applied the same definition of the target population in all pilot countries. This definition set out from the common understanding to survey graduates in Autumn/Winter 2018 one and five years after graduation for covering the short-term and the mid-term development of graduates. To give a comprehensive picture all

⁷ In the beginning, Hungary was foreseen as one of the pilot countries but then cancelled its participation.

higher education graduates of the academic years 2012/13 and 2016/17 were defined as target population.

All graduates included: ISCED-2011 levels 6 (BA-level) and 7 (MA-level or long degree programmes). ISCED 5 programmes (short courses) are included, when they are regarded as a significant part of the higher education system in the respective country and are offered by institutions, which are also offering programmes at least BA-level. Among the EUROGRADUATE pilot countries, this only applied to Malta.

The EUROGRADUATE pilot survey covers persons that graduated from the above-mentioned programmes in the respective pilot country regardless of their current place of residence (in or outside of the pilot country), their location of prior education (school or first degree), their previous enrolment status (e.g. full-time, part-time, correspondence), and their country of birth or nationality.

EUROGRADUATE also involves graduates of all ages and graduates from international joint-degree programmes if part of their degree was issued from a higher education institute within the country participating in the pilot survey.

All institutions (public and private) offering programmes at ISCED levels 6 or 7 were included in the EUROGRADUATE pilot survey, except for institutions at which students are employed and that are run by an employer. This applies to 'corporate universities', military or police universities, as well as higher education institutions of the public administration. These types of institutions were excluded because the transition to the labour market, which is a key topic of the survey, is part of the admission to those institutions, and therefore very specific.

4 Sampling, Contacting, and Return to the Survey

Almost 140,000 graduates were contacted for EUROGRADUATE. More than 22,000 graduates had begun the survey yielding a gross return rate of 16.2%. Quite a number of respondents did not fill-in the questionnaire up to a defined minimum point (the section on demographic information close to the end of the survey). All in all, the pilot survey returned 16,582 usable cases in eight countries for the two target cohorts 2012/13 and 2016/17 resulting in an overall net return rate of 12.0% (measured by the invitees). This is a moderate response rate only. For a potential future full roll-out one should clearly strive for a better response. However given that it was a pilot survey carried out in partially very difficult environments, the return is seen as satisfactory under the circumstances. The field phase lasted from 08.10.2018 to 17.02.2019 with some variation between countries (see below).

Starting conditions in the eight pilot countries were quite different. This concerns the availability of population data, the need to coordinate with existing graduate surveys, the size of graduate cohorts, the willingness of higher education institutions to support the survey as well as the availability, kind, quality and place of storage of contact data (postal or e-mail addresses, in a central database or just the HEIs, or other registry databases). Thus national sampling plans were set-up individually for each pilot country. For Croatia, a random sample of graduates was drawn but finally decided to invite all graduates with available contact information in order to achieve high numbers of respondents. For three pilot countries, no sample was drawn: For Malta, no sample was drawn considering the small size of graduate cohorts in the country. In Greece, no sample was drawn, because conditions did not allow for centralised sampling and only a small number of higher education institutions agreed to

participate in the survey and sampling would have further reduced the number of (potential) respondents. For Lithuania, no sample was drawn to make full use of the scarce contact information available.

In the following the sampling, contacting, and field phase are described in brief country by country and cohort.

4.1 Data collection in Austria

Field phase: 17.10.2018-06.01.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	50,641	50,268
Sample size	7,104	6,667
Opt-in accepted	1,302	1,313
Gross response rate	18.3 %	19.7 %
Usable questionnaires	1,054	1,120
Net response rate	14.8 %	16.8 %

Sampling and stratification:

Sampling method: probability sample: stratified sample

The sample was stratified for type of institution and type of study program. In drawing the sample it was ensured that the sample matched the distribution of study fields in the population. For universities and teacher training colleges a central register, which also contains contact information, was available for central sampling. For the universities of applied sciences, a central register was available to determine the number of graduates to be sampled per participating institution and study field, but without a possibility to link the selected graduates with the contact details stored locally at the institutions. Therefore, the universities of applied sciences had to draw samples based on rules provided by EUROGRADUATE.

Contacting:

For universities and teacher training colleges it was possible to update postal addresses for graduates living in Austria. Invitations and reminders for this part of the sample were sent by postal letter. At the universities of applied sciences and four small universities, invitations and two reminders were sent by institutions themselves via e-mail, without the possibility of updating contact details.

Known limitations:

Private universities were not included in the pilot study; however they only make up for about 2% of all graduates in the cohort 2012/13. Graduates that have left the country might be underrepresented as they are not obliged to report their new address abroad.

4.2 Data collection in Czechia

Field phase: 08.10.2018-06.01.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	90,082	69,947
Sample size	5,914	6,350
Opt-in accepted	1,286	1,547
Gross response rate	21.7 %	24.4 %
Usable questionnaires	745	1,015
Net response rate	12.6 %	16.0 %

Sampling and stratification:

Sampling method: probability sample: stratified sample

The sample was stratified by ISCED level and enrolment status. In drawing the sample it was ensured that the sample matched the distribution of study fields in the population. The sample was drawn by the institutions.

Contact details and invitation mode:

Invitations and reminders were sent by e-mail for all graduates by the institutions. Quality and actuality of e-mail addresses depended on the practices of the individual institution.

Known limitations:

Non-coverage with e-mail addresses was at 28% for the cohort 2012/13 and at 19% for the cohort 2016/17.

4.3 Data collection in Germany

Field phase: 08.10.2018-06.01.2019

Sizes of population and sample, number of respondents and response rates:

	2016/17 (t+1)
Population size	417,312
Sample size	5,474
Opt-in accepted	1,083
Gross response rate	19.8 %
Usable questionnaires	914
Net response rate	16.7 %

Sampling and stratification:

Sampling method: probability sample: stratified cluster sample

A random sample of clusters of graduates was drawn centrally from a database using data of the German federal statistical office. Clusters were defined as graduates of a specific field of study and a kind of degree (Bachelors, Masters, or state examination) at a higher education institution. In case the respective institution agreed to participate, all graduates of the cluster were contacted.

Contact details and invitation mode:

Graduates were contacted by the institutions. Kind and quality of contact information differs across institutions. Mostly, graduates were contacted via postal letters informing about the survey and announcing an e-mail that contained the personalized links to access the online survey. Following the postal letter, invitations and reminders were sent by e-mail. For some institutions, graduates could only be contacted either by postal letters or by e-mails.

Known limitations:

To avoid overlap with the “German Tracer Studies Co-Operation Project” (KOAB), which would have caused double surveying, clusters of institutions participating in KOAB were replaced by similar clusters in the same or neighbouring German federal state. In replacing clusters, the type of institution and the type of degree have always been kept identical. The field of subject has usually been kept identical as well. As a consequence, no institution from North-Rhine-Westphalia, the regional focus of the KOAB project, is part of the sample. Moreover, a relatively large proportion of institutions refused to participate in EUROGRADUATE. All in all, the German sample for the cohort 2016/17 is somewhat skewed as compared to the distribution in higher education statistics, specifically regarding the representation of fields of study (e.g. the STEM fields are overrepresented). Statistical weights have been designed to account for this.

The EUROGRADUATE pilot survey in Germany does not cover the cohort of 2012/13 as this cohort was surveyed by DZHW graduate panel survey at similar time.

4.4 Data collection in Greece

Field phase: 30.10.2018-17.02.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	122,353	61,096
Sample size	Estimation: 28,907 *	
Opt-in accepted	616	1,204
Gross response rate	6.3 %	
Usable questionnaires	446	866
Net response rate	4.5 %	

* Only estimations can be given for the numbers of invitations and the response rates.

Sampling and stratification:

Sampling method: total universe/complete enumeration

A full survey, including all institutions that agreed to participate was conducted. No sample was drawn.

Contact details and invitation mode:

Invitations and reminders were sent by e-mail only. Contact details that could be used for EUROGRADUATE were stored locally at institutions. Different strategies were applied to contact graduates, reflecting that most institutions were not willing to provide contact information to the national research team and that some institutions were not able to send individualised links to graduates:

- Institutions (6 out of 36) provided the e-mail addresses to the national research team which sent invitations and reminders centrally.
- Institutions invited their graduates to participate in the survey by asking them to contact the national research team and provide their email address if they were interested (this had a very low response rate).
- Institutions that did not want to provide the national research team with contact details nor had the possibility to send individualised e-mails to graduates sent a general invitation to the EUROGRADUATE pilot survey to their graduates containing a link that led to a web platform operated by the national research team. Through this platform, individualised links for answering the EUROGRADUATE survey were distributed to the graduates who agreed to participate in the survey. For these graduates, reminders were also sent by the national research team, after they had entered their e-mail address on the platform.
- Facebook ads and contact through HEIs alumni associations were used as well to hint and link graduates to the web platform.

Known limitations:

Response rates for the EUROGRADUATE pilot survey in Greece can only be estimated, as graduates were also contacted via ads and other means. Thus the number of contacted graduates is unknown. The estimated response rate is very low. Still, comparing the distribution of the respondents with population data suggests that it is not extremely skewed.

4.5 Data collection in Croatia

Field phase: 10.10.2018-06.01.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	35,018	33,004
Sample size	10,467	22,868
Opt-in accepted	1,483	5,676
Gross response rate	14.2 %	24.8 %
Usable questionnaires	919	4,278
Net response rate	8.8 %	18.7 %

Sampling and stratification:

Sampling method: total universe/complete enumeration

Initially random samples of graduates were drawn, but for boosting the number of respondents it was decided to invite all graduates to the survey, resulting in a full-population survey.

Contact details and invitation mode:

Contact details stored at the HEIs were provided to the national research team to generate a central database. Graduates were contacted via phone and/or e-mail, depending on the availability of contact details by the national research team.

Known limitations:

Only about 60% of the higher education institutions have agreed to participate in the survey of the cohort 2012/13. For the cohort 2016/17 a more comprehensive share of 80% the institutions participated in EUROGRADUATE. The response rate for the cohort 2012/13 was considerably lower than for the cohort 2016/17 presumably to less up-to-date contact information.

4.6 Data collection in Lithuania

Field phase: 10.10.2018-04.02.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	38,074	26,821
Sample size	11,882	12,507
Opt-in accepted	878	1,542
Gross response rate	7.4 %	12.3 %
Usable questionnaires	640	1,164
Net response rate	5.4 %	9.3 %

Sampling and stratification:

Sampling method: total universe/complete enumeration

The survey was conducted with all available contact data and hence no further sampling took place.

Contact details and invitation mode:

Two central registers were used as sources for contact information: the students' register and the social fund registry (containing all taxpayers in Lithuania). Both registers have e-mail addresses as contact information. Invitations and reminders were sent by e-mail only.

Known limitations:

Contact information in the students' register is provided by higher education institutions, varies in kind (institutional e-mail addresses, private e-mail addresses or both) and is not always up-to-date. The social fund registry holds up-to-date e-mail addresses but covers only a small proportion of the graduates. Thus the coverage with up-to-date e-mail addresses was lower than expected. Graduates that have left the country are not covered by the social fund registry and may have been reached to by invitations to a lower degree. Generally, the response rates are low.

4.7 Data collection in Malta

Field phase: 24.10.2018-04.02.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	4,167	4,492
Sample size	4,167	4,492
Opt-in accepted	635	705
Gross response rate	15.2 %	15.7 %
Usable questionnaires	457	506
Net response rate	11.0 %	11.3 %

Sampling and stratification:

Sampling method: total universe/complete enumeration

In Malta a full survey at the two main higher education institutions (the University of Malta and the Malta College of Arts, Science & Technology (MCAST)) was conducted. Due to the small population size no sample was drawn. Other institutions were not included as they only make up for a very small part of the population.

Contact details and invitation mode:

Contact details were stored at the institutions which sent invitations and reminders. Graduates of the University of Malta were contacted by postal letters. Graduates of MCAST were contacted by postal letters and e-mails.

Known limitations:

The response rate is moderate only.

4.8 Data collection in Norway

Field phase: 17.10.2018-06.01.2019

Sizes of population and sample, number of respondents and response rates:

	2012/13 (t+5)	2016/17 (t+1)
Population size	40,984	48,427
Sample size	5,058	5,287
Opt-in accepted	1,322	1,538
Gross response rate	26.1 %	29.1 %
Usable questionnaires	1,124	1,160
Net response rate	22.2 %	21.9 %

Sampling and stratification:

Sampling method: probability sample: stratified sample

The sample was stratified for type of institution, kind of degree, and sex. In drawing the sample it was ensured that the sample matched the distribution of study fields in the population. Sampling was done centrally by using a register.

Contact details and invitation mode:

Contact details were available centrally. The cohort 2016/17 received invitations and two of the reminders by e-mail, and one reminder by postal letter. For the cohort 2012/13 e-mail addresses were not registered; invitations and reminders were sent by postal letters only.

Known limitations:

None.

5 Data Preparation

5.1 Data Checking and Data Cleansing

The data collected was centrally cleaned for all countries at the IHS. The different national adaptations of the questionnaire (e.g. to account for differences in the education systems) had to be merged into a coherent data file. Data was recoded to match international classifications for education (ISCED) or professions (e.g. ISCO). Open data were coded with the support of the language-competent national teams. Missing values were defined for all variables. Special attention was given to filter questions for the following quality checks. Dates (e.g. date of graduation) were converted into decimal numbers, which are easier to calculate. Data on income have been converted into purchasing power parities and per hour salaries. Various variables have been regrouped and categorized to facilitate analyses. Conversely, some variables also had to be split into separate variables for each country, as the wording differed according to national circumstances (for example grade scales).

Plausibility checks have been performed on many variables, e.g. whether previous or subsequent education courses took place in or outside the higher education system or whether amounts given are realistic.

Finally, the valid cases were identified. In order to count as a valid case certain variables needed to be filled-in completely and plausibly – mainly those used for the weighting (see below).

In some countries it was possible to add additional information from register data to the data set (for example, the exact programme completed).

5.2 Generation of Variables

Besides variables that directly contain answers of the respondents, the scientific use file of the EUROGRADUATE pilot survey holds variables that have been generated from answers of respondents for facilitating analyses. E.g. income in purchasing power parities, variables measuring the immigration background or the social background are readily available. Moreover a number of variables using international classifications were generated, e.g. ISCED 2011 level, ISCED 2013 field, ISCO, ISEI, ISIC, NACE, and NUTS. For a full documentation of all generated variables please consult the codebook of the EUROGRADUATE pilot survey which is accessible via the webpage on the dataset: <https://dx.doi.org/10.21249/DZHW:egr2018:2.0.0>.

5.3 Coding of Missing Values

The following systematic was used for coding missing values in the data set:

Code	Meaning
-66	Question not asked (filters/routing)
-77	Question not asked in country/cohort
-88	Question not asked in cohort
-99	No answer

6 Weighting

In survey data, the distribution of certain characteristics among the respondents may diverge from the distribution in the population, e.g. by design (because of an intentionally disproportionate sample) or because of non-response. To account for such differences and calculate statistics reflecting the real distribution in the population as closely as possible, statistical weighting is applied.

Statistical weights of the EUROGRADUATE pilot survey have been generated by raking. Weighting was complicated by the fact that the availability and quality of population data was quite heterogeneous across countries.

Generally, the raking focussed on variables for which all countries were able to provide the distribution in the population (i.e. the higher education graduates of the cohorts 2012/13 and 2016/17 respectively). With minor exceptions, the following variables have been considered for the statistical weights in each of the countries:

- Cohort
- Type of higher education institution (not in Lithuania for MA-level graduates)
- Kind of degree
- ISCED broad field (not in Malta and not for the non-university sector in Greece, Czechia, and Austria)
- Gender of graduate (not in Czechia for the cohort 2012/13 from the non-university sector)

In addition, specific variables have been used if they were available from the population data in the respective country:

- Austria: age groups
- Norway: age groups
- Germany: region
- Croatia: public or private institutions, region
- Czechia: age groups, full or part-time programme in non-university sector
- Lithuania: age groups
- Malta: age groups (not for graduates of the cohort 2016/17 in non-university sector)

Resulting weights have not been trimmed. Therefore, data users are advised to check if very large weights or very small weights strongly impact on their results.

For further information on the weighting see the technical assessment of the EUROGRADUATE pilot survey (Mühleck et al. 2020).

7 Anonymisation

[Data Protection Legal Framework] For personal data⁸, which is collected in voluntary surveys carried out by the DZHW, the EU General Data Protection Regulation (EU-GDPR) and the German Federal Data Protection Act (BDSG) in its most recent version dated 30 June 2017 apply.⁹ According to this provision, personal data must generally be prepared for disclosure for scientific secondary use (without a declaration of consent for secondary use of the personal data) in such a way that “the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person” (Art. 4 (5) GDPR; see also Art. 89 GDPR and recital 26 GDPR). This means that for the transfer of data from scientific research projects to third parties, the data must be made anonymous in such a way that no reference to the individual persons can be made.

[Data Access] For EUROGRADUATE, the RDC for Higher Education Research and Science Studies provides a SUF for scientific secondary use. The anonymity of the respondents is thus protected by a combination of statistical measures and technical access barriers. The SUF is made available for download after conclusion of a data usage contract.¹⁰

[Statistical Anonymization Measures] In the course of anonymisation, all information that directly allows individuals or institutions to be identified is deleted. These so-called *direct identifiers*, such as names, addresses and email addresses, were placed in a separate data set during the field phase of EUROGRADUATE and are neither contained in the CUF nor in the various SUF variants. To further prevent any re-accessing of this information, the original identification number was removed and replaced with a new randomly assigned identification number.

Additionally, *quasi-identifiers* were determined, i.e. information which, in combination with or by the allusion to external information, allows for indirect identification.¹¹ For EUROGRADUATE, the following quasi-identifiers were used: higher education institution, subject, type of degree, biographical, educational and career information, regional information (higher education institution, location where higher education entry qualifications were obtained and place of work), dates, nationality and country information. To prevent a clear association with the data of EUROGRADUATE, these key attributes were aggregated or deleted (cf. Table 1). In addition, Ebel and Meyermann recommend deleting open responses even if the respective questions are unproblematic in themselves. For there is a danger that study participants have disclosed critical information that could lead to identification when answering questions that are actually unobjectionable and openly answered (cf. Ebel &

⁸ “‘Personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” (Art. 4 GDPR, p. 1).

⁹ The GDPR applies in principle within the EU and thus also to the DZHW. The BDSG in its new version of 30 June 2017 (Act for the Adaptation of the Data Protection Act to the Regulation (EU) 2016/679 and for the implementation of the Directive (EU) 2016/680 (Data Protection Adaptation and Implementation Act EU DSAnpUG-EU)) also partially applies in addition, as DZHW GmbH is legally regarded as a public body of the Federation (§ 2 Para. 3 BDSG). The Federal Government holds the absolute majority of the shares of DZHW GmbH and the institute fulfils tasks of the Federal Government's public administration in the broadest sense.

¹⁰ <https://www.fdz.dzhhw.eu/en/data-usage> (last access: 2020-10-01)

¹¹ It is pointed out that the identification of a person is already made more difficult by the sample selection, since uncertainty arises whether a respondent has a unique combination of characteristics in the population.

Meyermann, 2015, p. 5). Most of the open responses have already been coded by the primary research project as part of the data preparation process and will be made available in this form (partly aggregated). Non-coded open responses were deleted in the SUF.

Finally, it was checked whether the data contained sensitive information, e.g. on health, sexual orientation and political attitudes. Although these are not necessarily suitable for the re-identification of individuals or institutions, the information can be useful in the case of de-anonymisation (cf. Koberg, 2016, p. 694) and are therefore particularly worthy of protection (Art. 9 GDPR, recital 51 GDPR). In EUROGRADUATE, health information and information on political opinions were collected for which no additional consent for secondary use was obtained from the respondents. These answers were therefore aggregated/deleted from the SUF.

Table 1 below summarizes the statistical anonymization measures carried out.

Table 1: Statistical Anonymisation Measures for the Data of EUROGRADUATE Pilot Survey 2018¹²

Characteristic	Access Way (Download-SUF)
Direct identifiers	Deletion
Original-ID	Deletion and allocation of a random ID
Higher education institution	Type of higher education institution
Languages	Official language(s) of the survey country (yes/no); English (yes/no) (for instruction languages only)
Grades	Grouped
Student finance	Country-specific questions combined with “others”
Countries	NEPS-Classification (German National Educational Panel Study) and world regions ¹³ (except for survey country)
Dates	Months and more detailed dates deleted; years aggregated (indicated by value labels)
School and vocational qualifications before graduation	Yes/no
Year of study programme before and after reference study programme	Deletion
Months (unemployment)	Grouped
Occupations	ISCO-88 - International Standard Classification of Occupations: Combination of 1 and 2 digit scheme
Educational degrees	ISCED 2011: levels
Field of study: Major	ISCED-F 2013: Broad fields
NACE - Nomenclature statistique des activités économiques dans la Communauté européenne	Aggregated revision 2 ¹⁴
Numbers of staff supervising	Aggregated higher numbers (indicated by value labels)
Age	Grouped (indicated by value labels)

¹² Detailed information on the anonymised variables can be found in the codebook. (last access: 2020-10-01)

¹³ see (s. code list cl-dzhw-39) (last access: 2020-10-01)

¹⁴ see https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE-REV2&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=&IntCurrentPage=1 (last access: 2020-10-01)

Characteristic	Access Way (Download-SUF)
Citizenships	First citizenship aggregated to NEPS classification and world regions; second and third citizenships: yes or no; additional variable formed: citizenship of survey country yes or no
Number of children	From three on aggregated
Age of child	Aggregated to age groups; only specified for the first three children
Information revealing health related issues, political opinions, racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, sex life or sexual orientation	Deleted or aggregated (indicated by value labels)
Others	Aggregations of certain categories (indicated by value labels)

8 Bibliography

- DZHW Graduate Panel. (no date). Accessed 2019 December 16th, https://www.dzhw.eu/en/forschung/projekt?pr_id=467.
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- European Commission (2017). Council Recommendation of 20 November 2017 on tracking graduates, [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017H1209\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017H1209(01)).
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *Datenschutz am Forschungsdatenzentrum* (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), 2–16.
- Mühleck, K., Meng, C., Oelker, S., Unger, M., Lizzi, R., Maurer, S. et al. (2020). *EUROGRADUATE Pilot Survey. Technical assessment and feasibility of a full rollout*. Luxembourg: Publications Office of the European Union. doi:10.2766/85072.
- Research Centre for Education and the Labour Market (2005). *REFLEX - REsearch into employment and professional FLExibility*. DANS. <https://doi.org/10.17026/dans-z3s-a2dh>.