European Tertiary Education Register (ETER) Handbook for Data Collection

Contract No. EAC-2021-0170

Brussels, 2023

Education and Training

European Tertiary Education Register (ETER)

Handbook for data collection

Contract No. EAC-2021-0170

Disclaimer

This document has been prepared for the European Commission; however, it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Document summary

This document provides an in-depth description of the data collection methodology for the ETER database, i.e. the reference database on European Higher Education Institutions. It provides basic definitions, the perimeter for data collection, classification schemes for data, the definition of variables and guidelines for data collection and quality control.

This version is updated as of summer 2023 and will be adopted for the data collection for the year 2021 (as well as for updating the former data collected). This current version was updated with information on the changed handling of special codes.

Acknowledgements

This handbook is an outcome of the collective work by the ETER team since 2014, which has been coordinated by Benedetto Lepori. We acknowledge specifically contributions and comments by Andrea Bonaccorsi, Renato Bruni, Giuseppe Catalano, Alessandro Daraio, Cinzia Daraio, Hebe Gunnes, Elisabeth Hovdhaugen, Agata Lambrechts, Giorgio Matteucci, Lisa Schön, Michael Ploder, Monica Scannapieco and Daniel Wagner-Schuster.

The authors also acknowledge support and input from the European Commission, Directorate-General Education, Culture, Youth and Sports, Joint Research Centre and from EUROSTAT, as well as representatives of National Statistical Authorities and Higher Education Ministries involved in the development of the ETER database.

Table of Contents

Table 1 List of abbreviations

Abbreviation	Full Name
DG EAC	Directorate-General Education, Youth, Sports and Culture
DG RTD	Directorate-General for Research and Innovation
EC	European Commission
EEA	European Economic Area
EFTA	European Free Trade Agreement
ERA	European Research Area
ETER	European Tertiary Education Register
EU	European Union
EUROSTAT	the Statistical Office of the European Union
FOE	Fields of Education
FTE	Full Time Equivalents
HC	Headcount
HEI	Higher Education Institution
ISCED	International Standard Classification of Educational Degrees
NE	National Experts
NIFU	Nordic Institute for Studies in Innovation, Research and Education
NSA	National Statistical Authority
OECD	Organisation for Economic Cooperation and Development in Europe
UAS	Universities of applied sciences
UOE	UNESCO-UIS/OECD/EUROSTAT data collection on formal education manual
USI	Università della Svizzera italiana
AY	Academic year

1 Introduction

The European Tertiary Education Register (ETER) is a European-level database providing a reference list of Higher Education Institutions (HEIs) in Europe and data at the institutional level on HEIs' activities and outputs, such as students, graduates, personnel and finances.

As of spring 2023, ETER included 40 countries in Europe and provided data from 2011 to 2020 for a total number of 28,158 observations. Data for the year 2021 will be collected in fall 2023 and published in spring 2024. The countries covered so far are the EU-27 countries, the UK, EEA/EFTA countries (Iceland, Liechtenstein, Switzerland and Norway), as well as candidate and potential candidate countries to the European Union (Albania, Kosovo, Montenegro, Serbia, Turkey, The Republic of North Macedonia), as well as Andorra and the Holy See. ETER data can be accessed and downloaded from the ETER web interface at www.eter-project.eu.

ETER has been developed through a series of contracts by the European Commission from 2014 onwards. The current contract by the Directorate-General for Education, Youth, Sport and Culture of the European Commission (contract no. EAC-2021-0170) runs for three years, from May 2021 to April 2024. It is a joint undertaking of five partners - USI – Università della Svizzera italiana, Lugano, Center for Organisational Research, JOANNEUM RESEARCH, Graz, NIFU – Nordic Institute for Studies in Innovation, Research and Education, Oslo, University of Rome La Sapienza, Department of Computer, Control and Management Engineering Antonio Ruberti, Rome, and the Austrian Institute of Technology (AIT), Vienna – together with a network of experts in the concerned countries.

ETER is implemented together with a network of data providers in the National Statistical Authorities and National Education Ministries in the participating countries. While not part of the European Statistical System, ETER is largely compliant with statistical regulations and manuals, specifically with the UNESCO-UIS/OECD/EUROSTAT (UOE) data collection on formal education manual (UOE manual)[1] and the OECD Frascati Manual on research and experimental development (R&D) statistics[2]. It is also

[1] http://uis.unesco.org/en/files/uoe-data-collection-manual-2020-en-pdf

[2] https://www.oecd.org/sti/inno/frascati-manual.htm

embedded within the OECD Analytical Database of Higher Education Providers (ADHEP) project.

ETER is also closely cooperating with different tools providing data and analyses on European Higher Education, such as the Research Infrastructure for Science and Innovation Studies (RISIS https://www.risis2.eu), the Database of European Quality Assurance Results (DEQAR https://www.eqar.eu), the EURYDICE network on educational policies (https://eacea.ec.europa.eu/national-policies/eurydice/), the U-Multirank tool (UMR https://www.umultirank.org/) and the World Higher Education Database (WHED https://www.whed.net).

This handbook provides a detailed description of the methodology and process of the data collection in ETER, covering the following main items:

- The basic conceptual principles for ETER data collection (chapter 2).
- The ETER data sources and data collection process (chapter 3).
- The definition of the perimeter of Higher Education Institutions to be included (chapter 4).
- The list of classifications to be used for different types of data, including students, graduates and personnel (chapter 5).
- The definitions of the variables to be collected (chapter 6).
- The definition and procedures for calculating indicators in ETER (chapter 7).
- The procedures for data validation, quality control and production of metadata (chapter8).
- The description of the ETER technical infrastructure (chapter 9).

Besides a number of minor changes, the 2023 edition of the ETER handbook includes a major simplification of special codes in data collection (section 5.5).

2 Core elements of ETER

ETER is built around a number of core elements, which define the statistical units, the perimeter of data collection, the main dimensions for which data are collected, the methodology and the data sources mobilised. These core elements also a better understanding of the relationships and complementarity of ETER with statistical data on (higher) education and Research and Development (R&D) collected by EUROSTAT.

These are in summary:

- A focus on educational institutions (HEIs) as the primary statistical unit.
- A definition of the perimeter of analysis, smaller than tertiary education as defined in educational statistics, broadly corresponds to (national) definitions of higher education,
- An understanding of HEIs as multifunctional organisations displaying a diversity of profiles according to different dimensions.
- A methodological approach that builds on existing statistical frameworks (specifically the UOE data collection on education) but adapts definitions to the specific needs of ETER, introducing new variables whenever necessary.
- A data collection process based on secondary data, largely building on the existing data collections for educational and R&D statistics and complementarily on existing data sources at the European and international levels.

2.1 Institutions as the statistical unit

The basic design of choice of ETER is to focus on whole educational institutions as the statistical unit for which data are collected and published. This marks the main distinction and complementarity with educational statistics based on the so-called UNESCO-UIS/OECD/EUROSTAT (UOE) manual[3], where the primary statistical unit is an educational program with data aggregated at the country or regional level[4].

This design choice reflects extensive changes in the organisation and management of higher education in Europe in the last three to four decades. On the one hand, many educational institutions, and particularly universities, have been increasingly granted autonomy from the public administration and have become 'managed' organisations with stronger central structures and the ability to develop an institutional strategy (de Boer et al., 2007; Fumasoli & Lepori, 2011; Bonaccorsi & Daraio, 2007a). Universities

[3] http://uis.unesco.org/en/document/uoe-data-collection-manual-2020.

[4] https://ec.europa.eu/eurostat/web/education-and-training/data/database.

have also increasingly started competing at the European and international levels (Hazelkorn, 2015), as evidenced by emerging policies such as the European Universities Initiative[5].

On the other hand, the higher education system continuously expanded in the last decades with the integration of new types of institutions (Kyvik, 2004), while traditional distinctions between the university and non-university sectors have become blurred (Huisman et al., 2007). This implied that policymakers need a better understanding of the internal differentiation between profiles and types of institutions in order to design tailored policies (Daraio et al., 2011) since conventional distinctions between types or sectors of higher education do not any more provide a faithful representation of reality (Lepori, 2021).

This principle has far-reaching consequences for the ETER infrastructure. First, the first step in the data collection process is establishing an agreed list of institutions to be included or excluded; as discussed below in chapter four, identifying the units is not always straightforward and includes some elements of conventionality. Second, data and indicators are provided at the whole higher education institutions' level rather than its constitutive units, such as faculties and departments or campuses. ETER has been designed for analyses for which an institutional level is a meaningful unit of analysis (Lepori et al., 2019); it is not suited for studies of differentiation within institutions. Third, the principle has implications for defining variables, particularly for what concerns financial variables. Even if basic definitions might be the same, the different unit of analysis generates different measures and counting methods compared with official statistics.

2.2 A specific focus on Higher Education

The focus on educational institutions also implies a second important distinction, i.e. a definition of the perimeter of interest smaller than 'tertiary education' as defined in educational statistics, which can be broadly related to national definitions of higher education. In most countries, higher education comprises a specific set of institutions that are officially recognised as part of a national system and, frequently, subject to accreditation and quality assurance at the institutional level. While there are many commonalities between countries – for example, public universities are always included – there are also significant differences in the inclusion of other institutions, such as

[5] https://ec.europa.eu/education/education-in-the-eu/european-education-area/european-universities-initiative_en.

institutions of professional education and private HEIs. This implies that the ETER perimeter is, to some extent, conventional and based on national political choices on how to govern the educational system and is primarily based on decisions by national authorities.

Therefore, the ETER perimeter is conceptually different from tertiary education, defined as all educational programs delivering degrees at least at level 5 of the International Standard Classification of Educational Degrees (ISCED5). Indeed, some tertiary education degrees, particularly the short ones (ISCED5 level), are delivered by institutions whose primary mission is to provide secondary education. Others are awarded through exams without an institution delivering formal courses (particularly in vocational education (Lepori, 2020)). However, there is a large overlap, as most tertiary education degrees are delivered by educational institutions included in the ETER perimeter.

2.3 HEIs as multi-functional institutions and profiling

ETER comes with an understanding of higher education institutions as multi-input multi-output organisations, which use sets of inputs – financial resources, human resources, and infrastructure – to produce multiple sets of outputs, including research output, educational outputs, transfer activities towards society and the economy (Bonaccorsi & Daraio, 2007b). We consider that the production processes for these outputs cannot be easily separated, so we made a choice to characterise the whole set of inputs and outputs at the level of the whole institution instead of looking at individual activities. Thus, the ETER framework overcomes the traditional distinction between educational and R&D statistics and does not attempt to separate these activities but instead provides distinct indicators concerning educational and research production. ETER builds in that respect on the literature on the differentiation of HEI profiles (van Vught, 2007) to identify a set of core dimensions for the characterisation of individual HEIs (Huisman et al., 2015), i.e. the educational profile (subject domains offered, importance of different degree levels), student's profile (the composition of the student's body in terms of gender, nationality, mobility), research activities, involvement in knowledge exchange activities (technology transfer), international orientation of HEI activities, and engagement of HEIs in regional development. While many of these dimensions are already included in ETER, this framework also provides guidance to further expand ETER by addressing core information gaps, notably concerning knowledge exchange and regional engagement. Additionally, ETER provides core information on the level and

composition of resources available for HEI activities, notably the composition of academic personnel and institutional revenues.

An important application of the profiling approach is the identification and characterisation of the main groups (or 'types') of HEIs in Europe, where legal types of HEIs, such as universities and universities of applied sciences, are becoming less distinctive and comparable across countries. Unlike in the US with the Carnegie classification (McCormick and Zhao, 2005), until recently, there was no consistent classification of European HEIs due to the important national differences (Van Vught et al., 2008; Borden and McCormick, 2020). ETER might allow to fill in this gap (see Lepori, 2021).

2.4 Building on existing methodological frameworks

A core design choice of ETER is to rely on, as far as possible, existing methodologies, specifically from official statistics, for the definition of variables and indicators. This allows for the reusing of data collected in the framework of educational and R&D statistics for ETER and guarantees the possibility of comparison between ETER and international statistics.

More specifically, the definition of ETER variables largely builds on the UNESCO-UIS/OECD/EUROSTAT (UOE) manual[6] for the data collection on education. ETER definitions concerning students and degrees are based on the UOE with few adaptations related to different statistical units. ETER definitions concerning HEI personnel also largely comply with UOE definitions with some adaptation to the specific context of higher education (such as the explicit inclusion of educational personnel). As for financial data, ETER has its own definitions since the unit of analysis is different (national budgets vs institutions).

As for the definition of Research & Development (R&D) expenditures, ETER follows definitions and rules in the Frascati Manual (OECD, 2015), also adopted by the EUROSTAT for R&D statistics.

Finally, ETER uses its own definitions and methodologies for variables of non-statistical nature (most institutional descriptors and geographical information) and variables from other sources (credit mobility, quality assurance).

[6] http://uis.unesco.org/en/document/uoe-data-collection-manual-2020.

Noticeably, even if definitions might be the same as adopted in UOE, different statistical units might imply different counting methods. For example, in UOE, students enrolled at multiple institutions (for example, in the framework of joint programs) are counted only once, while in ETER, they will be counted in each individual institution. Accordingly, the sum of students enrolled in ETER HEIs might exceed national totals in the UOE data collection. The focus on the institutional level might generate other discrepancies, such as institutions enrolling students but not awarding degrees.

2.5 ETER as a secondary data collection

A final design choice is that ETER currently does not undertake primary data collection but relies on secondary data sources provided by other entities. This choice is to reduce the burden of collecting data but requires intensive efforts to harmonise data collected on different grounds. Data in ETER undergoes extensive quality checks (see chapter eight), but the ETER project team assumes no responsibility for errors in the original data.

In terms of the sources, ETER data can be divided into three main groups:

- Data derived from the UOE data collection on educational statistics. Primary data are the same as those in the country and regional education and training statistics[7] published by EUROSTAT but disaggregated at the institutional level. These data are provided by National Statistical Authorities (NSAs) or Higher Education Ministries; for a few countries, they are collected by the ETER team from official sources, such as the NSAs' websites.
- Data collected by the ETER project team from public sources, such as institutional websites and Wikipedia. These include data such as foundation years, demographic events, and institutional addresses.
- Data derived from existing databases at the European and international level, such as credit mobility supported by the Erasmus program, data on researcher's mobility funded by European Framework Programs from the EUPRO database[8], and data on quality assurance from the DEQAR database[9].

[7] https://ec.europa.eu/eurostat/web/education-and-training/data/database.

[8] https://rcf.risis2.eu/dataset/4/metadata.

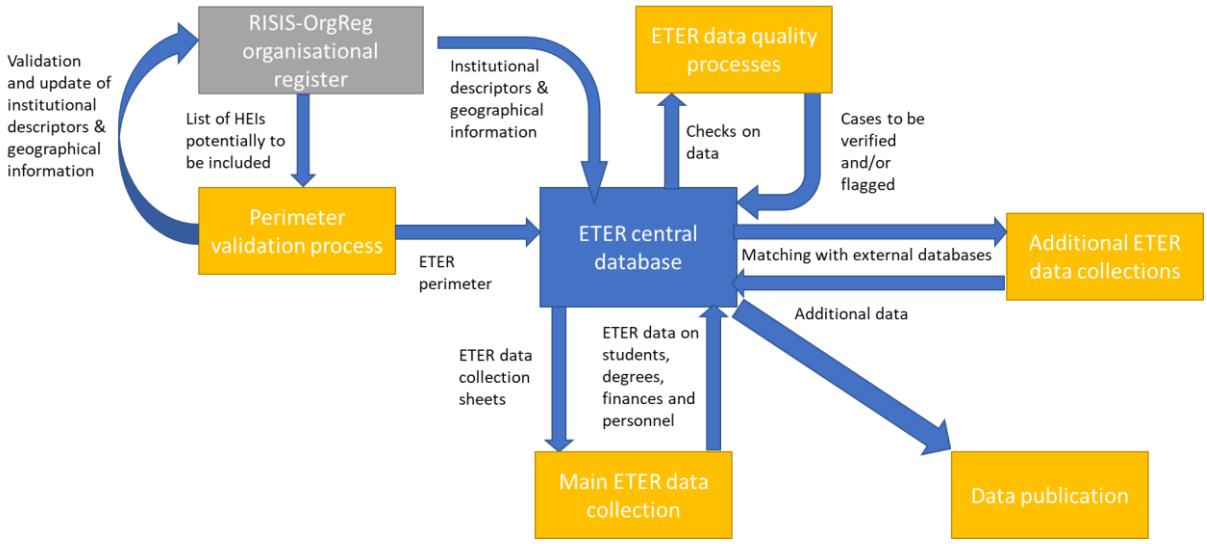
[9] https://www.eqar.eu/qa-results/search/by-institution/.

3 Data sources and the data collection process

ETER has a complex data infrastructure that is constructed from various sources and through different processes. The main data collection involves compiling statistical data on HEIs, including information on students, graduates, finances and personnel from NSAs or national higher education ministries. This is complemented by additional data collected from various European and international databases. This chapter provides detailed information on these processes and the integration of data into the ETER data infrastructure.

The ETER data collection and management process is presented in Figure 1 below. It is organised around the ETER central database and data infrastructure, described in chapter nine. The ETER central database stores all ETER variables and indicators based on ETER ID per year as a primary key (e.g. AT001.2015), i.e. all data and indicators are stored for each HEI and year.

Figure 1. The ETER data collection and data management process



The main processes can be shortly described as follows:

- The goal of the perimeter validation process is to establish - for each year and country participating in ETER - the list of HEIs to be included in the database and any changes over time. It is based on the criteria described in chapter three and builds on the RISIS-OrgReg register of public research organisations

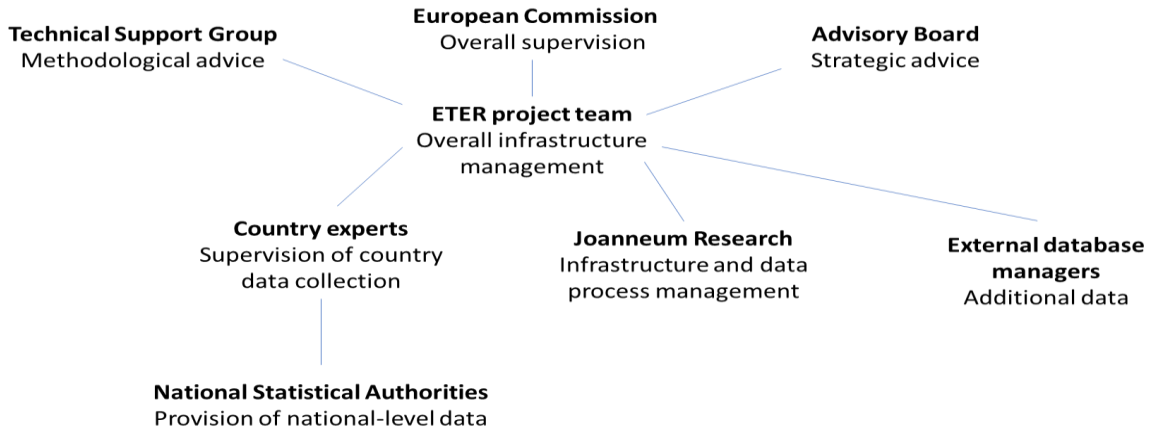
(https://register.orgreg.joanneum.at[10]). Alongside the perimeter validation, institutional descriptors and geographical information (see section 3.2) are updated in OrgReg and copied into ETER.

- The main ETER data collection process involves the collection and validation of data relating to students, degrees, finances, personnel and research activities, i.e. the core of the ETER dataset, from NSAs and national higher education ministries. It is performed on a country-by-country basis in close cooperation between the ETER project team and country correspondents.
- Supplementary data collection involves integrating additional variables derived from existing databases at the European and international levels, such as data concerning Erasmus mobility or quality assurance variables. It is performed directly by the ETER team in partnership with the source databases.
- When all data have been integrated into the ETER database and validated, data are made available through the public database interface (see chapter 9).

3.1 Actors and their roles

Figure 2 provides an overview of the main actors involved in the development and management of the ETER infrastructure.

Figure 2. Actors and their roles in ETER



The overall project supervision is by the *Directorate-General for Education, Youth, Sport and Culture of the European Commission* (DG EAC) in close cooperation with the

[10] Access to OrgReg requires previous registration at the RISIS Central facility https://rcf.risis2.eu/.

European Statistical Office EUROSTAT. It is supported by an Advisory Board (AB) and a Technical Support Group (TSG).

The *Advisory Board* is composed of European Commission services, international organisations, representatives from national higher education authorities and other relevant stakeholders in the field of higher education data use and management. The Advisory Board provides leadership and steering on strategic directions for further development and facilitates partnerships to foster ETER's contribution to the European Education Area and develop international cooperation on higher education data.

The *Technical Support Group* is composed of representatives from National Statistical Authorities, international organisations such as the OECD and other higher education data experts. This group provides advice on statistical methodologies, availability of data, data sources, data analysis and feasibility of inclusion of new data and promotes communication and dissemination of ETER reports through national systems.

The *ETER project team* (members of personnel from the involved consortium organisations), composed of people of the involved organisations, manages the overall infrastructure.

The main functions of the project team are as follows:

- Coordinating the entire process and developing its methodology, as stipulated in this Handbook.
- Preparing perimeter and data collection templates and sending them to NSAs.
- Performing data validation and quality analysis on collected data and reporting any problems to NSAs.
- Integrating the data in the ETER database and publishing them.

Within the project team, the USI is responsible for the overall coordination, while JOANNEUM RESEARCH is responsible for managing the data collection and the ETER infrastructure.

For each country, the ETER consortium has indicated a person responsible for a country with the following tasks:

- Maintaining contact between the ETER consortium and the NSA.

- Assisting the NSA in the delivery of data. In particular, the country responsible will provide information on guidelines, help the NSA in the collection of non-statistical data (descriptors, geographical information) and assist the NSA in inserting the data in the ETER datasheets.
- Perform a preliminary check of the data and clarify any relevant issues with NSAs.

In a few countries, national experts have been hired for the project (for example, because of linguistic reasons) to assist in communications with the national contacts and to support the project team's understanding of the specificities of national systems and the available data.

National statistical authorities or *national higher education ministries* are the main providers of data in ETER, except for descriptors and geographical information. They provide data to ETER based largely on existing data collections, specifically the educational statistics following the UOE manual.

3.2 Perimeter validation process

The goal of the perimeter validation process is to define the list of HEIs included in ETER for a specific year. This list is the basis for all data collections, implying that data will be collected only for the HEIs included in the ETER perimeter in that specific year. As already discussed in chapter two, ETER perimeter is partly conventional, for example, in deciding whether an entity should be included or the specific years of inclusion (for example, in the occurrence of demographic events). The process builds on the register of public research organisations OrgReg (https://orgreg.joanneum.at), managed by the RISIS infrastructure (RISIS-OrgReg).

The process is organised as follows (see
Figure 3 overleaf):

1) First, the list of HEIs for each country participating in ETER in the reference year is extracted from OrgReg. The list also includes institutional descriptors and geographical information for that year.

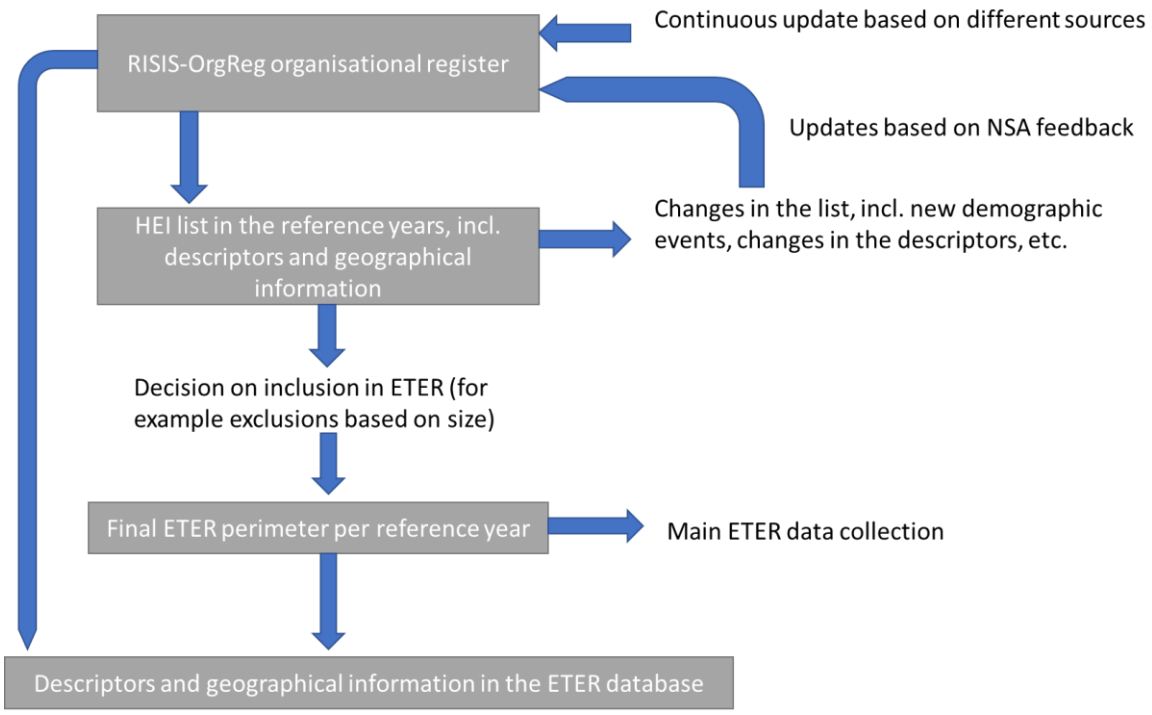
2) The list is sent from the country responsible persons to NSAs for verification of the following points:

- Whether HEIs in OrgReg should be included in ETER, as the OrgReg list also includes some HEIs which are not part of the national system (for example, private non-accredited HEIs).
- Any changes in the list, for example, demographic events not included in OrgReg, or missing HEIs to be added.
- Any corrections to the information provided (for example, changes in institutional names, location, etc.).

3) Changes are then reported by the ETER team in OrgReg. Then, the final ETER list for the reference year is generated as a basis for the other data collections. The new entries (ETER ID Year) are inserted in the ETER database.

4) Descriptors and geographical information are then directly copied into the ETER database so that there is no need for collecting these in the main data collection.

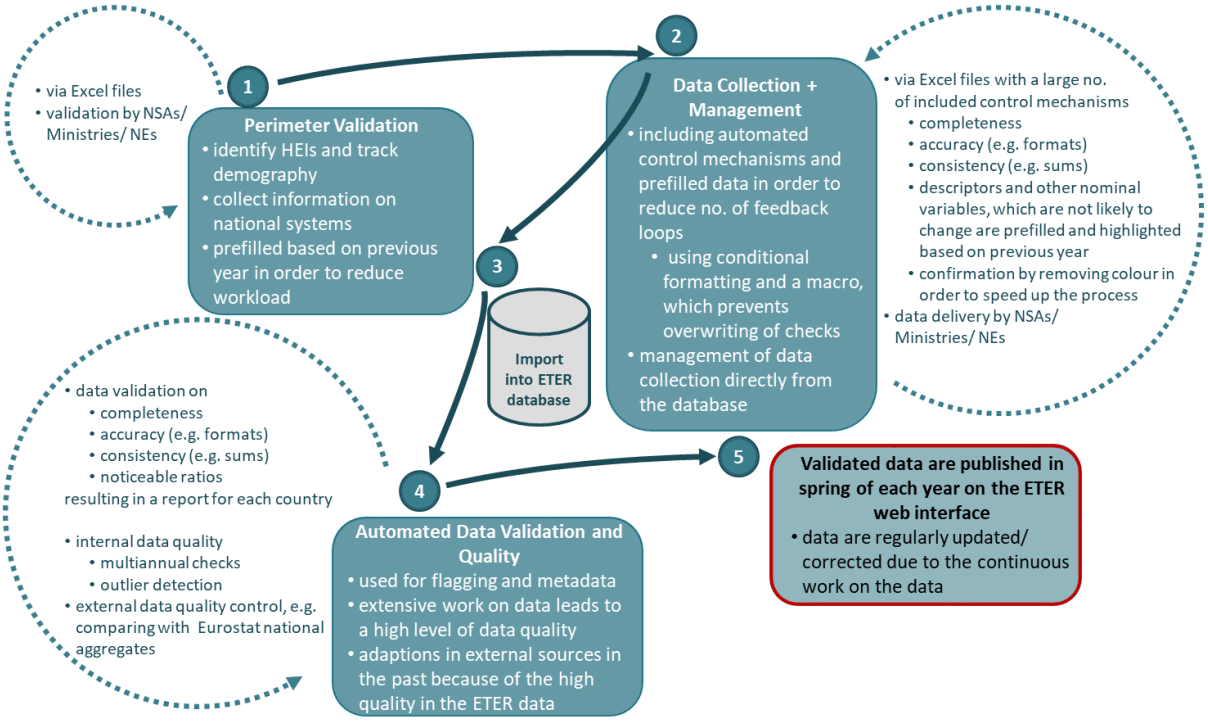
Figure 3. ETER perimeter validation process



3.3 Main ETER data collection

The main data collection collects the core of ETER data for every reference year, i.e. the variables on students, degrees, finances, personnel and part of the variables concerning research. It also includes an extensive process of data checks and quality checks to ensure a high quality of ETER data (see figure below and chapter 8).

Figure 4. ETER data collection process



The data collection process (2) builds on the perimeter validation process (1), and is organised in the following way (see Figure 4 above):

1) First, data collection files for every country participating in the data collection and for each individual year are generated in Excel. A high level of data quality should be ensured by implementing the first phase of quick checks of the data collection files to call the attention of the NSAs/NEs, enabling them to correct the data immediately.

The implemented checks alert the deliverer of data about:

- incomplete data and blank cells,
- format accuracy problems, i.e. wrong data formats or irregularities, and
- mistakes in sums and inconsistencies between variables.

For detailed information, see section 8.2.5 on the checks performed on the data in chapter eight.

Data collection files include prefilled cells with values from the previous year to support NSAs/NEs and reduce the burden of data collection. These include information that is not expected to change systematically from year to year. Prefilled data include:

- Basic and geographical information,

- nominal variables which are not likely to change and the resulting 0 values (e.g. where no ISCED 5 level programmes are available in a country, 0 is prefilled in the respective categories),
- flags and notes (partly), and
- metadata at the country level.

Prefilled variables ensure that the persons completing the data get some help in the process.

2) Data collection files are shared with NSAs, together with detailed methodological guidelines on how to fill in the data. This is done using a file-sharing platform named Nextcloud hosted by JOANNEUM RESEARCH. Once the data collection files have been filled, they are either saved on Nextcloud or sent to a country's contact person.

3) The incoming data collection files are screened for any mistakes detected by the implemented automatic checks in the Excel files to provide reasonably clean data for the first upload into the database. Problematic cases are analysed by the project team member responsible for the country. They need to decide whether:

- to correct the data (when the source of problems is clear),
- to flag the data, when the reason for the problem is already known or explained in the metadata. Flags are introduced in the data collection phase in order to simplify their management (this will also avoid detecting problems that have been flagged in the previous year again) or
- to report back to NEs/NSAs when the explanation of the issues is not straightforward.

4) The resulting datasets are uploaded to the database to enable the tracking of all changes. The imported data are in a further step separately checked country by country using a script, which produces a report about the consistency and format accuracy of the data. At the end of the data collection, a report about the consistency and format accuracy of the whole dataset will be produced. The script will detect:

- missing values,
- inconsistencies regarding format accuracy (data formats, irregularities), and
- mistakes in sums.

The ETER project team corrects the obvious mistakes identified in the report. All other issues are clarified in cooperation with the NSAs/NEs. Additionally, the report about the

consistency and accuracy of the data forms the basis for the first flagging of data. On that basis, values which do not meet the requirements of quick checks could be approved based on a manual check. After data corrections or new data delivery, the concerned data will be uploaded to the database again and automatically checked. This step should ensure that the detected inconsistencies are corrected and the data include no new inconsistencies. This phase of checks aims to spot problems that were not detected by the previous phase and thus systematise checks and data validation to ensure the best possible data quality.

5) Once data have been integrated into the ETER database, a set of more advanced data checks is performed. The checks are performed once a sufficient number of countries have been collected and validated as they exploit observed statistical regularities in the data to detect potentially problematic cases.

Data quality checks are of two types (see chapter eight for more details):

- Noticeable ratios, such as students per personnel, which are outside some lower and higher bounds determined from the data.
- Multiannual checks of variation in variables across years for the same HEI, as compared with the observed variation in the whole dataset.

Data quality produces a report on deviant cases to be checked, including the data, the observed anomaly and any information which might help to explain the deviant case, such as demographic events and data flags.

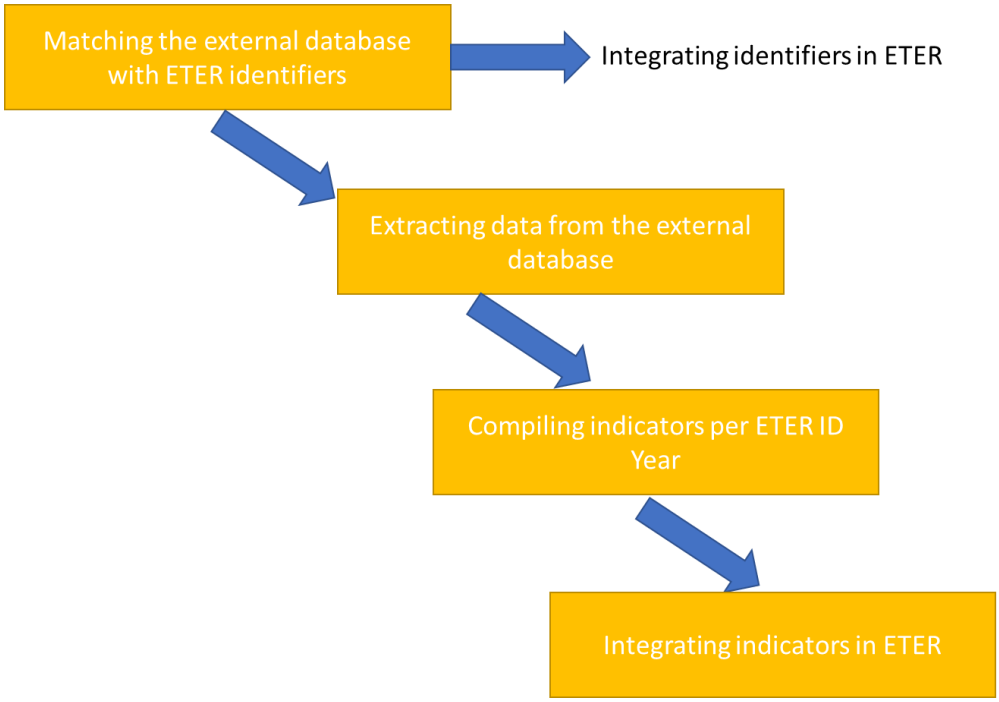
The list of deviant cases is checked by country responsible persons together with NE/NSAs. This can lead to different outcomes:

- data are corrected, or
- data are flagged, and an explanation of the deviant case is provided.

3.4 Additional ETER data collections

Additionally to the data collected from NSAs, ETER is enriched with data extracted from other European and international databases.

Figure 5. ETER additional data collection



For these databases, the process is organised as follows (see Figure 5):

1) First, entities in the external database are matched with ETER based on country, institutional name, location and website.
2) Second, the external database identifiers are integrated into ETER; this allows a 1:1 correspondence between the databases for extracting indicators.
3) Third, data are extracted from the external database and, when needed, summarised to provide indicators by ETER ID*Year.
4) Fourth, the indicators are integrated into the ETER database.

The process is repeated every time the ETER database undergoes a major extension, such as a new year of data collection and by major changes in the external database.

Table 2Table 1***Fehler! Verweisquelle konnte nicht gefunden werden.*** below lists the additional indicators available in ETER.

Table 2 List of indicators from external databases

Indicators	External database	Remarks
Quality assurance indicators	DEQAR database	
Credit mobility indicators	European Commission, Erasmus executive agency	
Researchers' mobility indicators	AIT EUPRO database	
Scientific publications (Web of Science)	CWTS Publication database	Indicators published on OrgReg
Patents from PATSTAT	IFRIS Patstat database	Indicators published on OrgReg
Participations to EU Framework programmes	AIT EUPRO database	

4 Definition of the perimeter

A central issue for ETER is establishing the list of institutions (HEIs) to be included in the data collection, the so-called *perimeter*. The definition of the perimeter is complex as there is no widely accepted definition of Higher Education Institutions, while the concept of Tertiary Education, as defined in educational statistics, refers to the level of educational programs but not to institutions directly. As this chapter will discuss, there is some unavoidable element of conventionality in the perimeter definition, which requires *ad hoc* agreements with National Statistical Authorities; however, a common framework is needed in order to achieve some comparability (both across countries and with EUROSTAT educational statistics).

Issues of feasibility and workload for data collection are also relevant in the perimeter definition since, in most countries, most of the tertiary education activities are concentrated in a relatively small number of entities, but there are also large quantities of very small institutions.

In this respect, two functions of the definition of a perimeter should be distinguished:

- The first one is the *register function*, i.e. providing a list of higher education institutions in a country, including some essential information like history, demography, and linkages between institutions. This function is fulfilled by the organisational register OrgReg, developed with the RISIS infrastructure project and described in section 4.2 below. The basic unit in OrgReg is, therefore, the HEI with its unique identifier (e.g. AT0001).
- The second one is the *statistical database* function, i.e. providing statistical data for an HEI concerning students, graduates, finances, personnel, research, etc. This function is fulfilled by the ETER database, where the basic unit is the HEI*year (e.g. AT001.2011) since data are collected yearly. While the register should be in principle as comprehensive as possible, the statistical perimeter of ETER might be smaller because of issues of data availability and/or workload, for example, excluding HEIs outside the national statistical system or smaller HEIs.

The perimeter definition process is, therefore, two-step (see section 3.2): first, a list of institutions in the Organisational Register (OrgReg) is established; then, the NSA decides whether HEIs in OrgReg should be included in the ETER data collection.

4.1 Principles for the construction of the perimeter

In general, ETER collects data on *higher education institutions*, defined as entities:

- which are recognisable as distinct organisations,
- which are nationally recognised as HEIs, and
- whose major activity is providing education at the tertiary level (ISCED 2011 level 5, 6, 7 and/or 8). R&D activities might be present but are not a necessary condition for inclusion in the perimeter.

A HEI is *nationally recognised* if it is officially accredited by a legitimate organisation in a country. *Recognisable as a distinct organisation* means that it is possible to identify the perimeter of these institutions rather unambiguously; they have an internal organisational structure and, at least in principle, their own budget.

4.1.1 Delimitation criteria

To operationalise this definition, the following three main criteria should be used:

- *Major activity*. Institutions for which education at ISCED 2011 level 5, 6, 7 and/or 8) is a major activity and constitutive part of their mission. These institutions can have a large share of research activity, but it should not be their only purpose.
- *Graduation at ISCED 2011 level 5, 6, 7 and/or 8.* In principle, institutions delivering courses in curricula where other institutions attribute the degree are excluded.
- *National recognition as a higher education institution*, for example, in national laws, is a relevant criterion of inclusion. This might include the specific mention of the institution in the law ruling higher education, accreditation by a specific body, or a mention in the list of officially recognized HEIs.

Further, for unclear cases, the following two additional criteria might be used:

- *Size and visibility* are further criteria, also for practical purposes. Institutional size should be measured through personnel data, and the number of students enrolled as a reference. As a general rule, institutions with less than 30 FTEs of academic personnel and less than 200 students should only be included in exceptional cases, specifically if institutions are primarily awarding degrees at ISCED level 8.
- *Continuity*, hence the intention of being a stable organisation, is another delimitation criterion. Thus, structures that are, in principle, transitory, for

example, by offering just one training cycle, should not be included in the perimeter.

4.1.2 Examples of higher education institutions

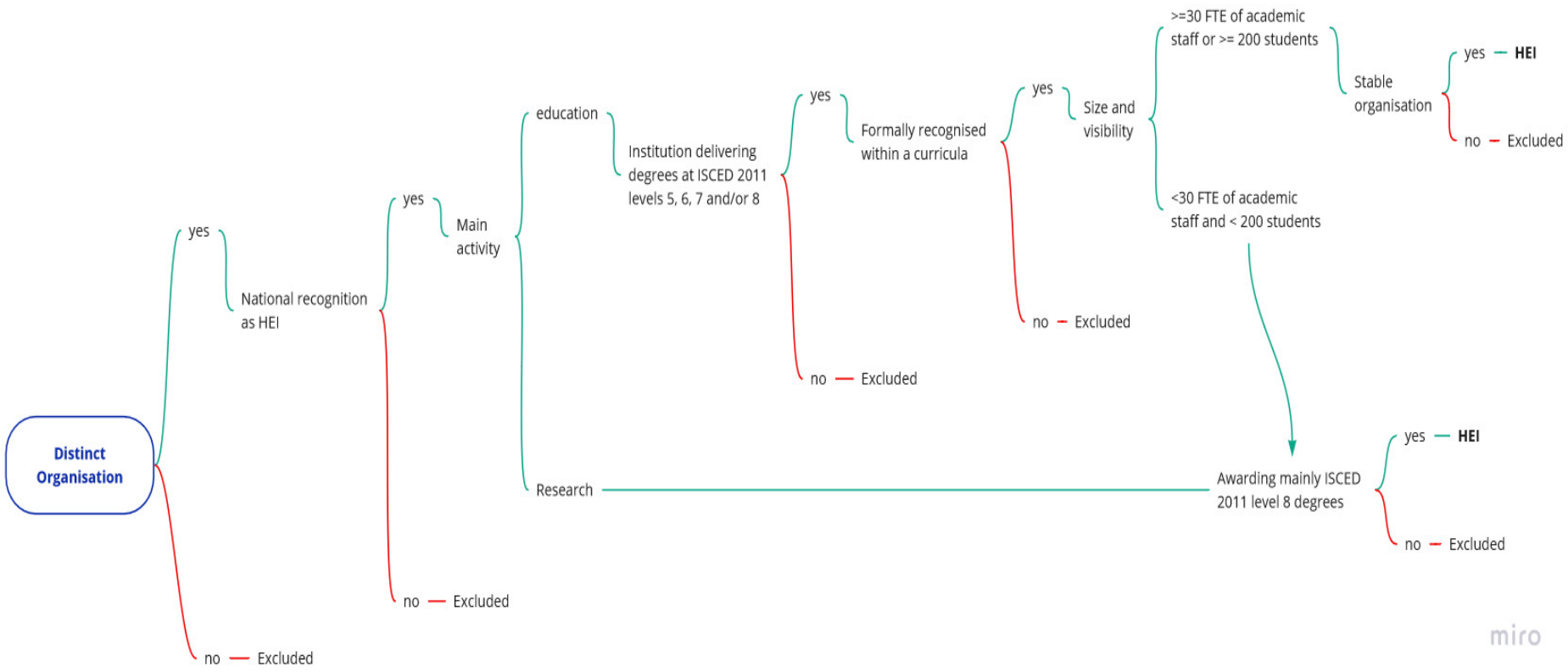
Examples of higher education institutions to be included are universities (PhD awarding) and universities of applied sciences (Fachhochschulen, Polytechnics). Other examples are colleges (e.g. of arts and music), theological schools, schools of pedagogy, and distance education universities.

Military academies should be included as separate institutions unless other institutions in the same country already account for their expenditures.

4.1.3 Exclusion cases

- Institutions offering only services for education (but no curricula) are not included in the ETER data collection despite their inclusion in the UOE data collection (non-instructional institutions; see UOE handbook).
- Institutions whose primary mandate is to provide secondary education but which also provide some tertiary education (such as preparatory classes in France) are generally excluded, following the criterion that tertiary education should be the main activity.
- Research institutions, whose principal mandate is performing R&D, like public research institutes and academies of sciences, are excluded (based on the main activity criterion) even if they deliver some educational activities - unless they deliver a large number of doctoral degrees.
- The same applies to organisations providing other professional services, which offer training courses as a side activity, specifically in professional education in countries such as Germany or Switzerland.
- Similarly, degrees obtained without a formal education, such as ISCED5 or ISCED6 professional exams, are generally excluded from ETER since, in most cases, the educational institution cannot be identified.

Figure 6. Decision tree for identifying institutions as HEIs



4.2 RISIS-OrgReg

The register of public research organisations (RISIS-OrgReg) has been developed within the RISIS European Union infrastructure project. It is closely linked to ETER, as it provides the basic list of institutions to be considered for inclusion in ETER. For full information about OrgReg structure and methodology, please refer to the RISIS Website (https://www.risis2.eu/registers-orgreg/).

OrgReg covers all 27 European Union member states, EEA-EFTA countries (Iceland, Liechtenstein, Norway and Switzerland), as well as candidate countries (North Macedonia, Kosovo, Montenegro, Serbia and Turkey), the United Kingdom, Andorra and the Holy See. It includes organisations undertaking research and/or delivering higher education, which are not exclusively market-oriented, corresponding to the government, higher education and private non-profit organisations in the Frascati Manual. More precisely, OrgReg includes Higher Education Institutions (HEI), Public Research Organisations (PRO), and Research Hospitals (RH). Public Administration Research Units (PA) and Private Non-Profit Research Organisations (PNP).

For each entity, OrgReg provides a stable identifier, basic descriptive information, such as foundation year and history, and geographical information. This information is automatically copied into ETER.

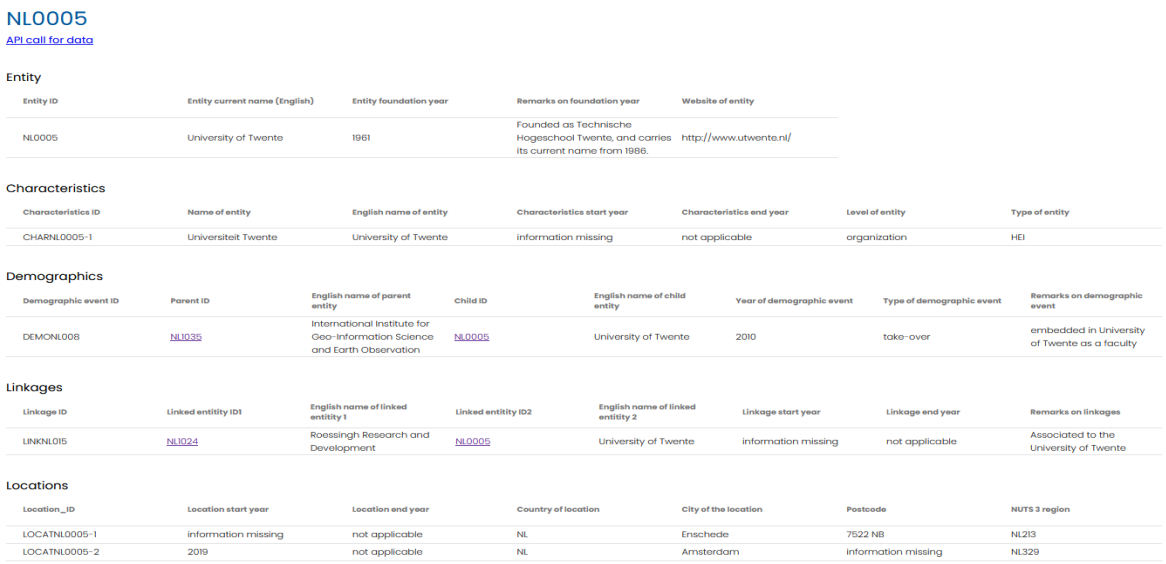
Additionally, a significant feature of OrgReg is to cover demographic events, such as mergers (see section 4.4) and to provide information on linked or associated centres to HEIs, such as affiliated institutes and university hospitals. This information can be accessed directly from ETER through the OrgReg institution's page (Figure 7 overleaf).

The relationships between OrgReg and ETER can be described as follows:

- ETER uses the same identifiers as OrgReg.
- The ETER perimeter is by construction smaller than the OrgReg one; HEIs newly introduced in ETER are also inserted in OrgReg, but there are entities classified as HEIs in OrgReg that are not included in ETER. Examples include private schools hosted but not accredited in the national system or affiliated schools included in the parent HEI in ETER. OrgReg might also include groups of HEIs for which no statistical data are available.

- While the OrgReg/ETER team decides on inclusions in OrgReg, NSAs have complete control over which institutions to include in ETER.
- OrgReg provides basic descriptive information, such as the institutional name and foundation year and geographical information to ETER – implying that changes remarked in the ETER data collection are copied into ETER. Similarly, OrgReg provides information on associated university hospitals.
- Finally, OrgReg provides information on demographic events affecting HEIs; that information is stored in a relational format in OrgReg and then copied in a suitable format for multiannual data collection in ETER (see below section 4.4).

Figure 7. OrgReg Institution's page



NL0005

API call for data

Entity

Entity ID	Entity current name (English)	Entity foundation year	Remarks on foundation year	Website of entity
NL0005	University of Twente	1961	Founded as Technische Hogeschool Twente, and carries its current name from 1986.	http://www.utwente.nl/

Characteristics

Characteristics ID	Name of entity	English name of entity	Characteristics start year	Characteristics end year	Level of entity	Type of entity
CHARNL0005-1	Universiteit Twente	University of Twente	information missing	not applicable	organization	HEI

Demographics

Demographic event ID	Parent ID	English name of parent entity	Child ID	English name of child entity	Year of demographic event	Type of demographic event	Remarks on demographic event
DEMONL008	NL1035	International Institute for Geo-Information Science and Earth Observation	NL0005	University of Twente	2010	take-over	embedded in University of Twente as a faculty

Linkages

Linkage ID	Linked entitity ID1	English name of linked entitity 1	Linked entitity ID2	English name of linked entitity 2	Linkage start year	Linkage end year	Remarks on linkages
LINKNL015	NL1024	Roessingh Research and Development	NL0005	University of Twente	information missing	not applicable	Associated to the University of Twente

Locations

Location_ID	Location start year	Location end year	Country of location	City of the location	Postcode	NUTS 3 region
LOCATNL0005-1	information missing	not applicable	NL	Enschede	7522 NB	NL213
LOCATNL0005-2	2019	not applicable	NL	Amsterdam	information missing	NL329

4.3 Relationships with the definition of tertiary education

As explained above, the ETER perimeter is related to but not identical to the tertiary education defined in international educational statistics (which includes all programs at the ISCED levels 5, 6, 7 and 8).

The major difference is generated by the program-oriented approach of educational statistics vs the institution-oriented approach of ETER. While the two broadly coincide in most cases, there are cases (especially in professional education) of institutions whose main missions and activities are outside the educational sector – like professional associations – but they also deliver tertiary education diplomas to a limited extent. These educational providers are not included in ETER. Further, in professional education,

degrees might be awarded without graduates following regular courses, such as in the case of on-job training and, accordingly, no educational institutions could be identified.

A second difference is related to feasibility issues and the rationale to include small-scale educational units or parts of the public administration when their size is relatively small. This is relevant since, in ETER, data have to be provided and published at the level of individual HEIs. In contrast, for international educational statistics, data are first aggregated by NSAs (and, in some cases, regional authorities or private providers) and then transmitted in aggregated form to EUROSTAT. Thus, in practical terms, including too small units would significantly increase the burden for data collection, validation, and management without substantially adding information to the data collection. Descriptive information on the characteristics of national systems shows that this is particularly the case for diplomas at the ISCED5 level.

If HEIs exist and fulfil the perimeter requirements, but data are missing or cannot be delivered because of confidentiality issues, they should nonetheless be included in the perimeter. Information about the missing data will be included in the metadata.

In terms of the number of students, the coverage of ETER is expected to be slightly smaller than the whole tertiary education for undergraduate students (ISCED 5-7), whereas it should be almost identical for doctoral students given their concentration in large universities. Concerning personnel, the perimeter of ETER might be larger than the tertiary education perimeter, owing to the inclusion in university personnel data of research centres.

As a general principle, coverage in ETER of at least 90% of tertiary education at the country level and 95% for the whole data set should be envisaged. Experience from data collection shows that, with the chosen definition, coverage of HEIs graduating students at least at level ISCED 6 (bachelor) is reasonably complete, whereas this is less the case for HEIs graduating students only at level ISCED 5 (professional diplomas of less than three years). This is largely explained by the different structure of professional education and the fact that it is usually provided by a very large number of small institutions.

4.3.1 Perimeter description

Since ETER does not necessarily cover the whole of tertiary education, an accurate description of the coverage is essential to match data with EUROSTAT national

aggregates and ascertain whether individual cases of HEIs are excluded. This description provides a list of the main institutional sectors of tertiary education in the concerned country, and for each sector list the following information:

- A general description of the perimeter.
- Sector description.
- Curricula delivered by the sector (with reference to the ISCED mapping http://uis.unesco.org/en/isced-mappings) and their coverage in ETER.
- Important excluded cases of HEIs in ETER.
- A reference, for example, a link to the official list of HEIs in the country.

Information is partially provided by NSAs, partially derived from the UOE ISCED mappings and from other sources like Eurydice.

An example of mapping is provided on the following page.

4.4 Demographic events affecting the perimeter

Demographic events are occurrences which involve a change in the basic organisational units included in the register: examples of demographic events are the birth of an organisation, its closure, the merger between two organisations, which gives rise to a new organisation, etc.

Demographic events do not include transformations of an organisation that do not alter its identity, like the change of the name, relocation or the opening or closure of activities. They also do not include transformation in the legal status of existing organisations, for example, as an outcome of legal reforms when organisations keep their continuity.

In some cases, it might be challenging to decide whether an event represents just a change of an existing organisation or the closure of one organisation and the foundation of another one.

In general, the main criteria to distinguish an organisational change from a demographic event should be the following:

- Continuity in the management and organisational structure.
- Continuity in the main activities and production factors (for example, personnel).
- Continuity in location.
- Continuity in the name and official recognition.

Table 3 Perimeter description for Switzerland (example)

Perimeter description	The perimeter includes all institutions which are officially recognized as part of the Higher Education System by law, including universities, universities of applied sciences and teacher-trainign institutions; are excluded a few very small-scale HEIs. The perimeter does not include professional schools outside the higher education sector, which enrolled in 2011 about 20% of total students at the tertiary level (however with a strongly decreasing share of tertiary education in the last ten years). About half of these students are enrolled in public higher professional schools (largely full time), the rest in part-time professional courses. There are about 150 higher professional schools with an average number of students below 200; no disaggregated data are available at national level.						
Switzerland	**ISCED level**	**N. of Institutions (estimated)**		**ETER coverage by students' numbers**	**Sector description**	**Programme (ISCED mapping) included in ETER**	**Programme (ISCED mapping) not included in ETER**	**Most relevant exclusion cases in ETER (institutions)**
		All	Included in ETER					
University sector	6,7,8	12	12	Nearly 100%	This sector includes the 10 Cantonal universities, two federal institutes of technology as well as a small number of HEIs recognized by the law.	7.1 University diploma / 6.3 University/UAS bachelor / 7.02 University master /8.01 University doctorate	7.03 university post-graduate	Facoltà di Teologia di Lugano, Frankin College Switzerland, Theologische Hochschule Chur, Institut Universitaire Kurt Bösch excluded because of size threshold
Universities of applied sciences	6,7,8	9	8	Nearly 100%	This sector includes the seven public and the two private universities of applied sciences currently recognised by the Confederation.	6.2 university of applied science diploma / 6.04 Fachhochschule, post-graduate		UAS Les-Roches-Gruyere excluded since it is below the size threshold
Teacher training universities	6,7,8	15	15	Nearly 100%	This sector includes the 15 teacher training universities not integrated in the UAS.	6.1 Pedagogical university diploma (bachelor included in figures for the UAS sector)		Hochschule für Logopädie Rorschach excluded since it is below the size threshold
Federal PET Diploma examination / higher vocational education, stage I	5,6	some hundreds (estimate)	0	Not covered	Exams regulated at the federal level, no mandatory curricula, but preparatory courses offered by a large number of mostly private providers.		6.05 Federal PET Diploma examination / higher vocational education, stage I	No reliable data on students as there are not mandatory and structured curricula.
higher vocational education, stage I (no regulation on the federal level)	5,6	some hundreds (estimate)	0	Not covered	Exams regulated at the federal level, no mandatory curricula, but preparatory courses offered by a large number of mostly private providers.		5.15 Higher vocational education, stage I (no regulation on the federal level)	No reliable data on students as there are not mandatory and structured curricula.
PET College /technical school	6	150	0	Not covered	Cantonal schools (Höhere Fachschulen) delivering professional tertiary education degrees (duration 2-3 years). Older ISCED-1997 5B, classified at ISCED-2011 level 6.		6.06 PET College /technical school / 6.07 Postgraduate course PET college	This sector is not included in ETER since it is too fragmented and only aggregated data at regional level are available.
Advanced Federal PET diploma examination / higher vocational education, stage II	7	some hundreds (estimate)	0	Not covered	Exams regulated at the federal level, no mandatory curricula, but preparatory courses offered by a large number of mostly private providers.		7.04 Advanced Federal PET diploma examination / 5.16 higher vocational education, stage II	No reliable data on students as there are not mandatory and structured curricula.
Source								
Official list of recognised institutions by the swiss law from the Swiss rectors conference (last consulted 24.06.2016)								
https://www.swissuniversities.ch/en/higher-education-area/recognised-swiss-higher-education-institutions/								

When most of these characteristics change, an entity should be subject to a demographic event (for example, closure and birth of a new entity). The establishment of a new legal entity might be an indication of demographic events, but there are cases where it should be considered as an organisational change – for example, the transformation of a PRO in a private legal entity while keeping its mission, funding source and activities, does not constitute a demographic event.

Following are examples of cases that should *not* be considered as demographic events:

- Changes in the name when the entity keeps most of its activities and employees.
- Extension of activities and/or opening of new locations, like new educational offerings, the opening of new research centres and new campuses.
- Changes in legal status when keeping activities and structure, such as the recognition of colleges as universities (even if accompanied by a change in name).

Demographic events can be divided into the following groups:

- Events which imply the creation of entities (foundation) or the closure of existing entities (closure). Since they are unique to each entity, they are handled directly by attributing the corresponding foundation and closure years.
- Events which imply the grouping of entities (mergers) or their division (split). They are recorded as separate events, while the involved entities receive the corresponding foundation and closure years.
- Events which imply the change of relationships between entities; in most cases, an organisation becomes a component of another organisation (a take-over) or a component becomes independent (a spin-out). They are handled as demographic events, while the involved organisation disappears from the register in case of take-over (closure year at the take-over), or it is included newly from the spin-out (foundation year at the spin-out). This is consistent with the rule that, in general, components of organisations are not included in the register.

Table 4 List of demographic events

Demographic Event		**Change of identifier (ID)**
Birth or foundation. The creation of a new institution. (Defining the name, place). The change of status is not handled as a birth or a foundation.	*Death.* The complete closure of activities of an existing institution.	New independent identifier (ID) is introduced in the case of births. Identifiers of disappearing HEIs are reserved and not reused.
A *Merger* between two institutions into a new institution (death of the merged institutions, new name, legal status, accreditation etc.). The (date of the) merger corresponds to the (year of) birth (foundation) or entry.	*Split* of an existing institution into two or more independent institutions (two or more new institutions, new name, accreditation, maybe legal status).	New independent identifier (ID) introduced analogously to births. Identifiers of antecedent HEIs are reserved and not reused. Metadata of the register allow the connection of interlinked IDs.
Take-over of one institution by another one (death of the institution which was taken over, the taking over institution continues to exist). The new institution retains the identity of the institution taking over if name and location are not changed. In order to show that data are not comparable across years after a take-over, data for the institution taking-over have to be flagged with "de".	*Spin-out (Spin-off).* Splitting of a section of an institution to become a separate institute (old institution exists unchanged, new institution has new name, legal status, accreditation, governance, etc.). The remaining unit keeps name, legal status and location of the previous institution. As required for a take-over the flag "de" has to be added for the parent institution (the HEI losing a part) in order to show that data are not comparable across years after a spin-out.	In the case of take-overs, the identifier of the dominant HEI taking over another will be held (even if this HEI is smaller than the HEI taken over). The identity of the HEI taken over is reserved and not reused. In the case of spin outs, the ID of the original HEI will continue. Analogous to a birth, a new ID is created for the new HEI spinning out. Metadata of the register allow the connection of interlinked IDs.

Demographic events are recorded in a specific table in OrgReg, which includes information on parent and child entities (with the respective identifiers), year, type of event, as well as some descriptive notes (see Table 5).

Table 5 Examples of demographic events

Demographic event ID	Parent ID	Child ID	Year of demographic event	Type of demograph	Remarks on demographic event
Demo_ID	Demo_parent_ID	Demo_child_ID	Demo_year	Demo_type	Demo_remarks
Text+integer ("Demo"+country code+3 digits, start with 001 for each entity)	Entity ID(s), code ZZ0001 if organization is not included in the register	Entity ID(s), code ZZ0001 if organization is not included in the register	Year	nominal (merger=5, split=6, take-over=7, spin-out=8)	Text
DEMOCH001	CH0025	CH0017	2009	7	Take-over
DEMOCH002	CH0027	CH0036	2011	6	Split into three cantonal schools
DEMOCH002	CH0027	CH0037	2011	6	Split into three cantonal schools
DEMOCH002	CH0027	CH0038	2011	6	Split into three cantonal schools
DEMOCH003	CH1011	CH0011	2008	7	Since 2008 ISREC was integrated in
DEMOCH004	CH0801	CH0013	2003	7	Integration

These events are automatically transferred to ETER and inserted in two consecutive years, i.e. the year before the event (for the parent institution) and the year after the event (for the child institution). This procedure ensures that complete information is available within the yearly strata of ETER.

Additionally, ETER includes information on the entry of an existing institution in the data collection perimeter, for example, because of a change of status or national accreditation, as well as the exit from the perimeter, for example, when the HEI, while still existing, has become too small to be still included in the data collection.

4.4.1 Exemplary cases of demographic changes

- *Merger*. Aalto University in Helsinki (Finland) was founded in 2010 as a merger of three pre-existing institutions, Helsinki University of Technology, Helsinki School of Economics and University of Arts and Design Helsinki. In EUMIDA (the reference year 2009), the three schools are included in the perimeter with their respective IDs. These three IDs are not active anymore in the 2011 perimeter (but they will be reserved), whereas Aalto University received a new ID.
- *Take-over*. The teacher-training University of Southern Switzerland was merged with the University of Applied Sciences in the same region (SUPSI) in 2010. It was therefore included in the EUMIDA 2009 perimeter with an individual ID. In the 2011 perimeter, the ID was not used anymore. SUPSI keeps its 2009 ID.
- *Spin-out*. The teacher training university of the Swiss Canton Zug was part of the teacher training university of central Switzerland until 01.01.2013. It is not recorded as an individual institution in the 2009 and 2011 perimeters, but it was given a new individual ID in the 2013 perimeter. The parent HEI keeps its ID.

4.5 Multi-site institution

Even if most higher education institutions are one-sited, with most of their personnel and activities concentrated in a single location. There are, in fact, a few cases of truly multi-site institutions. The ETER-perimeter has to face two types of multi-site:

- *National campuses* (usually different establishments in the same region) – in this case, they are treated as a unique HEI, and no disaggregated data are collected, but ETER includes a dummy for multi-site and geographical information for campuses located in other NUTS3 regions than the main campus. This information is derived from OrgReg.
- *Foreign campuses* – consistently with UOE, these are treated as self-standing HEIs in the country where they are established. This should be restricted to larger units meeting the delimitation criteria explained above. Foreign campuses should be treated as self-standing HEIs in the country where they are established and included if they comply with the definitions for the perimeter of the ETER data collection. Decisions in this respect should be taken by the national statistical authority of the country where the delimitation criteria are identical to those described above. Conforming to UOE practice, data for foreign campuses are collected by the national statistical authority of the country where the campus is located. The linkage between the foreign campus and their legal institution in the parent country is recorded on the institutional page in OrgReg.

5 Classifications

This chapter introduces the basic conventions and classifications that are used for the ETER variables and indicators. It covers the following items:

- Classification by educational fields (Fields of Education and Training).
- Classification by educational level (ISCED classification).
- Geographical classification (NUTS regions).
- Conventions concerning the reference period for variables.
- Conventions concerning the use of monetary units.
- Special codes and data flags.

While the chapter has been written so that these classifications can be used directly for data collection, for a detailed description of classification schemes, references are made to the official descriptions provided by UNESCO-UIS/OECD/EUROSTAT.

As a general rule, in the ETER data collection, when data are divided by subcategories, totals should also be provided to account for unclassified cases and check for consistency. The 'unclassified' category is included in all classification schemes.

5.1 ISCED Fields of Education and Training classification

For the classification of students, degrees and academic personnel, the Fields of Education and Training 2013 classification (ISCED-F) should be used at the first level (broad fields).

In ETER, first-level classifications of ISCED-F 2013 are used. The subfields and ISCED-1997 FOE are only provided for mapping purposes (see Table 6 overleaf).

Please note: when data are available based on the previous Fields of Education classification 1997, which does not break down fields 04 and 06 of ISCED-F, data for field 04 (Business, administration and law) should be set to "included in other columns" (xc) and the following remark should be added: "included in field 03". Similarly, data for field 06 (ICT) should be set to "included in other columns" (xc) and the following remark to be added: "included in field 05".

Reference:

UNESCO, International Standard Classification of Education: Fields of Education and Training 2013, *http://uis.unesco.org/en/topic/international-standard-classification-education-isced*

Table 6 Fields of Education and training classification

Code	Name	Subfields	ISCED 1997 FOE
00	General programmes and qualifications	001 Basic programmes and qualifications 002 Literacy and numeracy 003 Personal skills	01 Basic programmes 08 Literacy and numeracy 09 Personal development
01	Education	011 Education	14 Teacher training and education science
02	Humanities and Arts	021 Arts 022 Humanities 023 Languages	21 Arts 22 Humanities
03	Social sciences	031 Social and behavioural science 032 Journalism and information	31 Social and behavioural science 32 Journalism and information
04	Business and law	041 Business and administration 042 Law	34 Business and administration 38 Law
05	Natural sciences, mathematics and statistics	051 Biological and related sciences 052 Environment 053 Physical sciences 054 Mathematics and statistics	42 Life sciences Part of 62 (natural parks and wildlife) 44 Physical sciences 46 Mathematics and statistics
06	Information and communication technologies	061 Information & Communication Technologies	48 Computing
07	Engineering, manufacturing and construction	071 Engineering and engineering trades 072 Manufacturing and processing 073 Architecture and construction	52 Engineering and engineering trades (plus most of 85 environmental protection) 54 Manufacturing and processing 58 Architecture and building
08	Agriculture, forestry, fisheries and veterinary	081 Agriculture 082 Forestry 083 Fisheries 084 Veterinary	62 Agriculture, forestry and fishery (minus natural parks and wildlife) 64 Veterinary
09	Health and welfare	091 Health 092 Welfare	72 Health 76 Social services
10	Services	101Personal services 102 Safety services 103 Security services 104 Transport services	81 Personal services Part of 85 environmental protection (community sanitation and labour protection and security) 86 Security services 84 Transport services

5.2 Levels of education (ISCED)

Data on students and degrees will be divided by the level of education of the program to which they are enrolled, using the International Standard Classification of Education (ISCED) in its 2011 version. This version is particularly well suited for ETER purposes, as it includes the distinction between "Bologna" levels of education (Bachelor, Master and Doctorate). Relevant ISCED levels for ETER are ISCED 5 (short-cycle tertiary), ISCED 6 (bachelor), ISCED 7 (Master) and ISCED 8 (Doctoral level).

The table below provides the main categories and descriptions for the relevant levels for ETER data collection. For full details, please refer to the ISCED-2011 handbook.

Table 7 Levels of education

ISCED-2011 level	Definition	Criteria
ISCED 5 short-cycle tertiary education	Programmes at ISCED level 5, or short-cycle tertiary education, are often designed to provide participants with professional knowledge, skills and competencies. Typically, they are practically based and occupationally specific and prepare students to enter the labour market. However, these programmes may also provide a pathway to other tertiary education programmes. Academic tertiary education programmes below the level of a Bachelor's programme or equivalent are also classified as ISCED level 5.	Duration: 2-3 years Entry requirements: ISCED 3 or 4
ISCED 6 Bachelor's or equivalent levels	Programmes at ISCED level 6, or Bachelor's or equivalent level, are often designed to provide participants with intermediate academic and/or professional knowledge, skills and competencies, leading to a first degree or equivalent qualification. Programmes at this level are typically theoretically-based but may include practical components and are informed by state-of-the-art research and/or best professional practice. They are traditionally offered by universities and equivalent tertiary educational institutions.	Duration: 2-3 years Entry requirements: ISCED 3 or 4 Usually: first degree at tertiary level
ISCED 7 Master of equivalent level	Programmes at ISCED level 7, or Master's or equivalent level, are often designed to provide participants with advanced academic and/or professional knowledge, skills and competencies, leading to a second degree or equivalent qualification. Programmes at this level may have a substantial research component but do not yet lead	Duration: 2-3 years Entry requirements: ISCED 6

	to the award of a doctoral qualification. Typically, programmes at this level are theoretically based but may include practical components and are informed by state-of-the-art research and/or best professional practice. They are traditionally offered by universities and other tertiary educational institutions.	Usually: second degree at the tertiary level Direct access to ISCED 8 level
ISCED 7X6 Master or equivalent level long degrees	ISCED 7X6 programmes are long first-degree programmes at a Master's or equivalent level with a cumulative theoretical duration (at the tertiary level) of at least five years (that does not require prior tertiary education). Data for these programs is presented separately in ETER (not included in ISCED 7) wherever possible, given their different characteristics and impact on the number of diplomas.	Duration: at least 5 years Entry requirements: ISCED 3 or ISCED 4 Usually: first degree at the tertiary level Direct access to ISCED 8 level
ISCED 8 Doctoral or Equivalent level	Programmes at ISCED level 8, or doctoral or equivalent level, are designed primarily to lead to an advanced research qualification. Programmes at this ISCED level are devoted to advanced study and original research and are typically offered only by research-oriented tertiary educational institutions such as universities. Doctoral programmes exist in both academic and professional fields.	Duration: at least 3 years Entry requirements: ISCED 7 Research-based programs (not only courses).

Reference:

UNESCO, International Standard Classification of Education (ISCED) 2011, http://uis.unesco.org/en/topic/international-standard-classification-education-isced

5.3 Geographical classification

5.3.1 NUTS regions

For the geographical localization of HEIs, the NUTS classification (Nomenclature of Territorial Units for Statistics) is adopted in its most recent 2021 classification at levels NUTS2 and NUTS3. In ETER, NUTS regions are derived directly from institutional addresses through the RISIS geocoding service (https://www.cortext.net/) based on geographical coordinates. The EUROSTAT website provides a correspondence table with national structures as well as with postcodes. This correspondence table can be used to crosscheck the consistency of the NUTS classification with postcode information in ETER.

Currently, there are no NUTS regions agreed for Bosnia-Herzegovina and Kosovo among the ETER countries.

Reference:
https://ec.europa.eu/eurostat/web/nuts/background

5.3.2 Postcodes

National standards for postcodes should be used for the ETER database. In most countries, the postcodes are fully numeric. The following table provides information on the format of postcodes in ETER countries.

Table 8 Format of postal codes in ETER countries

Countries	Format
Iceland	3-digits numerical (XXX)
Austria, Belgium, Bulgaria, Denmark, Hungary, Latvia, Liechtenstein, Luxembourg, North Macedonia, Norway, Portugal, Slovenia, Switzerland	4-digits numerical (XXXX)
Bosnia, Croatia, Czech Republic, Estonia, Finland, France, Germany, Greece, Italy, Kosovo, Lithuania, Montenegro, Poland, Serbia, Slovakia, Spain, Sweden, Turkey	5-digits numerical (XXXXX)
Romania	6-digits numerical (XXXXXX)
Andorra, Ireland, Malta, Netherlands, UK	Alfa-numeric postcodes. Andorra; 5 characters, 2 letters and 3 digits Ireland: 7 characters, mixes letters and numbers. Netherlands: 6 characters, 4 digits and 2 letters Malta: 7 characters, 3 letters and 4 digits. UK: different formats by area.
Holy See	3 letters

5.4 Reference period for data collection

In principle, ETER provides a yearly data collection - all data should be collected every year, including those data that generally are not expected to change in almost all cases (name of the HEI, foundation year, etc.).

Depending on the nature of the data and the practices of data collection, individual data refer to slightly different periods, as detailed in Table 9 below. Specific national departures from these practices should be detailed in the metadata.

When data are not collected annually, like in the case of R&D expenditures for some countries, data for missing years should be coded as missing ('m'). A remark should be inserted that data are missing because data are not collected every year and are available for other years (for example, 2011 instead of 2012).

Table 9 Units of measure and reference periods

Variable	Reference period/date
Descriptors and geographical information	Last day of the calendar year (31st of December)
Expenditures	Calendar year (1st January – 31st of December)
Revenues	Calendar year (1st January – 31st of December)
Personnel	Calendar year based on person-years for FTEs. Last day of the calendar year for HC.
Students	End of the first month of begin of the academic year.
Degrees (including PhD degrees)	Academic year or calendar year (to be specified)
Age of students and graduates	First day of the year following the reference year for data collection (e.g. 01.01.2017 for 2016 data).

It is important to note that these practices slightly differ from those adopted by EUROSTAT: in EUROSTAT year '2012' refers to the academic year 2011/2012 for students and graduates, whereas in ETER, it refers to the academic year 2012/2013. In both cases, the year '2012' refers to the calendar year for finances, personnel and graduates.

5.5 Special codes

The ETER project included a set of special codes for missing variables in order to add more information on why data are missing. Although useful, the system required many efforts from data collectors. Therefore, the system of special codes was changed in summer 2023. It now includes the following special codes:

- Code 'a' is used for basic descriptors and geographical variables, when data are not applicable. It is not used anymore instead of numerical variables (for example, the number of PhD students for a HEI, which does not have the right to award doctorates). Zero (0) is used instead of 'a' for numerical values from summer 2023 onwards.
- Code 'm' refers to the fact that the data in question is missing.

- Code ‘x’ was applied until summer 2023, when a specific breakdown was not available, but the data were included in the total. Instead of ‘x’, ‘m’ is used from summer 2023 onwards.
- Code ‘xc’ was used when a value was included in another subcategory (e.g. private funding, which are included in third party funding but cannot be separated). Instead of ‘xc’, ‘m’ is used from summer 2023 onwards.
- Code ‘xr’ was used for data that was included in other rows, which can occur when an institution is part of another institution. Instead of ‘xr’, ‘m’ is used from summer 2023 onwards.
- Code ‘nc’ was used for data that has not been collected in the reference year (e.g. the gender breakdown of full professors was not collected for the academic year 2011/2012 but was introduced in the following year’s data collection. Instead of ‘nc’, ‘m’ is used from summer 2023 onwards.
- Code ‘c’ is used in the public dataset for confidential data (see section 9.4).
- Code ‘s’ is used in the public dataset cell values below 3 to keep the anonymity of individuals (see section 9.4).

In general, no blank cells are allowed in the data collection, except for ‘Notes’-fields.

An ‘unclassified’ category is provided for all breakdowns (students, personnel, revenues, expenditures). This should be handled as follows:

- When all cases are classified in a subcategory, this column should be set to a ‘0’.
- When some cases cannot be classified, this column should provide the exact number of unclassified cases.
- When all data for a breakdown are missing, this column should be ‘m’.

5.6 Data flags and remarks

Flags signal problems or specificities of data relating to format accuracy, consistency, completeness and comparability.

Flags can be attributed to:

- individual cells,
- one dimension or group of variables in a country (e.g. all variables concerning revenues), and
- all variables for one or more HEIs in a specific country (e.g. all private HEIs).

Data flags 'accompany' an existing figure and act as a warning or an explanation. These are provided for each record in a column next to the corresponding variable, where multiple flags should be separated by commas (','). Where relevant, short explanations are included in the corresponding "Notes" columns of the dataset in order to quickly identify the reasons for the flags. Detailed explanations of the flags are included in the metadata sheets (particularly for the country-level flags).

The ETER flag system builds on a simplified and reduced version of the one adopted by EUROSTAT. However, it introduces a few additional codes for cases relevant at the level of individual HEIs (Table 10).

Table 10 List of flags

Code	Description	Definition
b	a break in time series	When changes in definitions or data collection procedures imply that the data are not comparable across years. The change is explained in the remark section.
de	a break in time series due to a demographic event	When changes in the perimeter due to demographic events (the same ETER ID, but institution changed, i.e. spin-outs and take-overs) imply that data are not comparable across years.
d	definition differs	Differences in definitions adopted for data collection imply that figures significantly differ from those complying with the ETER methodology and are not comparable across countries.
i	see metadata	There are specific conditions that imply that the value of a cell should be interpreted differently or not directly compared with others.
ic	inconsistent	Either when the sum of the breakdown differs from the total or if another semantic rule is violated.
rd	rounded	When data have been rounded by the data provider and thus are included in this format in the database.
c	confidential	When data are available but restricted to public access (this flag is relevant only for users with unrestricted access).
ms	missing subcategory	This flag is applied to totals to warn users that the total does not include one relevant subcategory (for example, where total expenditures do not include capital expenditures).
p	provisional	Data quality checks highlight some anomalies, like abnormal ratios or large fluctuations between years. Either these anomalies are not explained or a generated by data issues that could not be resolved. The corresponding data may be revised in the future.
r	remark	While the data are methodologically correct, some special event generates data anomalies, like an unusually large number of graduations in a single year. The remark field explains the source of the anomaly.

5.7 Monetary units

All monetary data in the data collection should be expressed in the national currency (Euro for countries in the Eurozone) in use for the reference year at the current prices of the reference year. This also implies that if countries change their currency from year to year, data should be recorded with different currencies. In ETER, this applies specifically for Latvia and Lithuania, where Euro was adopted on 01.01.2014 and 01.01.2015, respectively.

ISO_4217 codes for the currency are included in a separate field in the data collection (see Table 11 overleaf) and reported in the country-level metadata in ETER.

For purposes of comparative analysis, all financial data are converted into:

- *Euros* at nominal exchange rates from national currency to euro for non-euros countries based on national averages as published by EUROSTAT (https://ec.europa.eu/eurostat/web/exchange-and-interest-rates/data).
- *Purchasing power parities* in euro (average of EU-27) for each country and by year. This allows for the conversion of monetary data while taking into account the average price level in each country, as published by EUROSTAT (https://ec.europa.eu/eurostat/web/purchasing-power-parities/data/database).

Please note: for Euro countries, PPPs have been computed with reference to the Euro also in the past and hence have to be multiplied by the exchange rate as well. In the ETER database, this concerns Latvia and Lithuania.

5.7.1 ISO country and currency codes in ETER

For the United Kingdom, the code UK is adopted.

See Table 11 overleaf for a full list of ISO country and currency codes.

Reference:
http://www.iso.org/iso/country_codes

Table 11 ETER country and currency codes

ISO country code	Country	Currency	ISO currency code
AD	Andorra	Euro	EUR
AL	Albania	Albanian Lek	ALL
AT	Austria	Euro	EUR
BE	Belgium	Euro	EUR
BG	Bulgaria	Lev (pl. leva)	BGN
CH	Switzerland	Swiss Franc	CHF
CY	Cyprus	Euro	EUR
CZ	Czech Republic	Czech Koruna (pl. koruny)	CZK
DE	Germany	Euro	EUR
DK	Denmark	Danish Krone (pl. kroner)	DKK
EE	Estonia	Euro (1)	EUR
GR	Greece	Euro	EUR
ES	Spain	Euro	EUR
FI	Finland	Euro	EUR
FR	France	Euro	EUR
HR	Croatia	Kuna (inv.)	HRK
HU	Hungary	Forint (inv.)	HUF
IE	Ireland	Euro	EUR
IS	Iceland	Króna (pl. krónur)	ISK
IT	Italy	Euro	EUR
LI	Liechtenstein	Swiss Franc	CHF
LT	Lithuania	Litas (pl. litai) (3)	LTL
LU	Luxembourg	Euro	EUR
LV	Latvia	Lats (pl. lati) (3)	LVL
ME	Montenegro	Euro	EUR
MK	The Republic of North Macedonia	Denar (pl. denars)	MKD
MT	Malta	Euro	EUR
NL	Netherlands	Euro	EUR
NO	Norway	Norwegian Krone	NOK
PL	Poland	Zloty (pl. zlotys)	PLN
PT	Portugal	Euro	EUR
RO	Romania	Romanian Leu (pl. lei)	RON
RS	Serbia	Serbian Dinar	RSD
SE	Sweden	Swedish Krona (pl. kronor)	SEK
SI	Slovenia	Euro	EUR
SK	Slovakia	Euro(2)	EUR
TR	Turkey	Turkish Lira (inv.)	TRY
UK	United Kingdom	Pound Sterling (pl. pounds)	GBP
VA	Vatican city state (the Holy See)	Euro	EUR

6 Variables

This chapter provides detailed information on the variables included in the ETER database, while the following chapter seven identifies and defines indicators constructed by combining different ETER variables.

All variables refer to a single year, even if, in some cases, variables are not expected to change across years, as in the case of the HEI foundation year. Please refer to chapter 5 for a detailed description of the classification schemes used in collecting these data.

6.1 Overview: dimensions of ETER variables

Table 12 lists the dimensions in which ETER variables are organised and provides synthetic descriptions and information on the main data source. The following sections provide a full overview of the used variables and their definitions.

Table 12 Dimensions of ETER variables

Dimension	Description	Main data source
Metadata at the country level	Country-level metadata describe characteristics of the ETER data.	ETER
Identifiers and demographic information	These variables identify the HEI in ETER, matching information with other datasets and information on demographic events.	RISIS-OrgReg, ETER
Institutional descriptors	Descriptive information on the HEI, such as legal status, foundation year, history, national label.	RISIS-OrgReg, ETER
Geographic information	Address and geographical coordinates of the main and satellite campus.	RISIS-OrgReg
Enrolled students and degrees	Lowest and highest degree delivered and the number of enrolled students and graduates by gender, citizenship, mobility, fields of education, age group and part-time/full-time. Includes also a variable on distance education.	ETER
Research activities	Includes information on research activity, the number of ISCED 8 (PhD) students and graduates, R&D expenditures and researcher mobility.	ETER, EUPRO
Expenditures	This dimension includes different types of HEI expenditures (personnel expenditure, non-personnel expenditure, total current expenditure, capital expenditure) and a variable of the HEIs accounting system.	ETER
Revenues	Includes detailed breakdowns on HEI revenues (core budget, third-party funding, student fees funding, current and non-recurring revenues).	ETER

Personnel	This dimension included detailed information on academic and non-academic personnel, including breakdowns on gender, citizenship and fields of education and the number of full professors.	ETER
Student and personnel credit mobility	Information on incoming and outgoing Erasmus students and personnel by ISCED class.	Erasmus executive agency
Quality assurance	Variables providing information on whether an HEI was subject to a quality assurance process following European guidelines and reported in the DEQAR database.	DEQAR
ETER indicators	Secondary indicators are constructed from the ETER variables; see chapter seven for a list and definitions.	ETER

6.2 Metadata

According to the ISO standard, *metadata* are defined as data that defines and describes other data and processes. The ETER data collection includes the possibility to indicate differences in data, which can occur on a country or institutional level. Country-level metadata are collected in order to characterise the quality of ETER data. Within country-level metadata, we distinguish between general and in-depth information. The table below summarises collected metadata and their coverage.

In addition to general metadata at the variable level, specificities at the level of individual institutions can be reported in the 'Notes'-fields of the different data collection sheets. This possibility is provided for groups of data so that information about deviations from variable definitions can be linked with the correspondent HEI data.

Table 13 List of country-level metadata

Variable	Data source
Content and departure from ETER definition	ETER
Reference period/date	ETER
Source release date	ETER
Reference to the data source	ETER
Type of providing a data source	ETER
Date of update	ETER
Change to the previous year	ETER
In-depth information	ETER

6.2.1 General Metadata

	Type of variable	Data source
Content and departure from ETER definition	text	NSA/NE

This variable requires the exact national definition of the data filled-in and a remark of possible differences with ETER variable definitions (if they exist). This variable requires the exact national definition of the data filled-in and a remark of possible differences with ETER variable definitions (if they exist). It is useful for the creation of a common data dictionary and to address the issue of geographical comparability that is very relevant in ETER. Therefore the most detailed description is required, as the information will be standardised during the data quality and validation phase. The variable definition is required only when readily available from national statistical sources and only for quantitative variables. For descriptors, it is sufficient to report possible departures from the definitions and additional comments that highlight problems of comparability with other countries.

	Type of variable	Data source
Reference period/date	text	NSA/NE

This metadata is required with reference to the international standard. Please be careful to update this information for the current year of data collection:

- Financial data (expenditure, revenues, and R&D expenditures) are reported with reference to a period of time, i.e. a whole calendar year. If data are collected according to the international standard, it is enough to report the year (i.e. 2017); otherwise, the reference period should be specified (i.e. academic year 01/10/2017-30/09/2018).
- Personnel data in full-time equivalents (FTE) are reported with reference to the calendar year. In this case, it is enough to insert the year (e.g. 2017); otherwise, the reference period should be specified.
- Personnel data in headcounts (HC) are reported with reference to an exact date (i.e. last day of the calendar year), which should be specified.
- Enrolled students are reported with reference to the end of the first month of the academic year (e.g. 31/10/2017). The exact date should be reported.
- According to international standards, graduates are reported with reference to a period of time corresponding to the calendar year. In this case, it is enough to insert the year; otherwise, the reference period should be specified.
- Descriptors and other status variables (e.g. 'research active', 'highest degree awarded', 'inclusion of PhD students') should be reported with reference to the last day of the calendar year.

	Type of variable	Data source
Source release date	date	NSA/NE

This is the date when the value was officially made available by the providing source and is necessarily previous to the date of ETER data collection compilation. For data that are published at a national level, the source release refers to the date of publication. For unpublished data, reference can be made to the date when the data collection is considered completed. When it is not possible to identify an exact date, a proxy to month/year is sufficient.

For example, in country X, data on academic personnel in headcount are collected taking into account the number of persons employed at the 31/12/2017 (reference date), and the results are published by NSA on 20/09/2017 (source release date).

For descriptors and other data collected directly from the institution's website or other internet sources by national experts, the source release date cannot be defined. In these cases, the date of data collection should be inserted instead. Since descriptors refer to structural features of HEIs that are usually stable over time, it is not necessary to update them every year. The column 'Source release date' contains the date when the descriptor was collected the first time and in most cases, has already been filled in during the past data collection.

6.2.1.1 Regulatory compliance

	Type of variable	Data source
Reference to data sources	text	NSA/NE
Type of providing data source	Nominal (0=survey, 1=administrative, 2=other)	NSA/NE

The variable 'Reference to data source' includes the exact name of the data source used and a link to the relevant website when publicly available online. Three types of data sources can be distinguished for the type of providing data source: survey, administrative, and other.

	Type of variable	Data source
Date of update	date	NSA/NE

When the collection of descriptors or other status variables is updated with reference to, e.g. 31/12/2017, it is required to report the new release date (date of update). Therefore, these columns should be filled in only when the corresponding descriptor contents have been modified compared to the previous year.

	Type of variable	Data source
Changes compared to the previous year	text	NSA/NE

Please describe any changes compared to the metadata from the last data collection, including but not limited to descriptor contents updates.

6.2.2 In-depth information

The ETER project showed that there are a few variables that are particularly problematic in terms of comparability among countries. For these variables, a more specific collection of information on national practices/situations is envisaged as an input to analysis, report and recommendations. Thus, this section contains the list of concerned variables and the issues to be clarified for each of them.

6.2.2.1 Expenditure

Expenditure perimeter: total expenditure measures the overall level of expenditures of a higher education institution included in its balance sheet. Different institutional settlements or national practices may lead to different perimeters of items in the HEIs balance sheet. Please describe here the coverage of expenditure figures reported in ETER.

Coverage of ancillary services and related expenditure: ancillary services (for example, meals or transportation) are a typical example of activities that may be included/excluded by the HEI perimeter according to different institutional settlements. A clear distinction of services provided by the HEI and accounted for in its balance sheet by services provided by other institutions (e.g., government and local authorities) should be provided here.

Coverage of expenditures in clinical medicine: this is another typical case of fading perimeter borders. Please specify the criteria and the contents of financial figures related to clinical medical activities when HEIs work with university hospitals.

Coverage and account system for capital expenditures: expenditures related to capital assets may be reported in ETER in different ways according to different accounting systems applied at a national or HEI level. Please specify the coverage of the figures provided here.

Classification of expenditures: for ETER, only a very broad breakdown of expenditures between current (personnel and non-personnel) and capital expenditures are required. To better understand expenditures breakdown and comparison among countries, it is necessary to specify the corresponding budgetary categories at the national level, which are included at the most fine-grained detail available. Please also report national categories not matched with any ETER breakdown and not included at all, if existing, together with any other relevant remark and additional information.

6.2.2.2 Revenues

Distinctions where the institution has partial fees: rules on access to higher education and relative costs for students and households widely differ among European countries, within each country among different HEI categories (i.e. state vs private institution) or even within each institution among cycles or curricula. A description of the fees regime is required here when institutions have partial fees, for example, limited to post-graduate curricula or a specific segment of the student population (e.g. mature students).

Revenue perimeter: please describe here the coverage of revenue figures reported in ETER.

Classification of revenues: similar to expenditures, please specify the corresponding budgetary categories at the national level at the most fine-grained detail available. Please also report national categories not matched with any ETER breakdown and not included at all, if existing, together with any other relevant remark and additional information.

6.2.2.3 Personnel

Personnel perimeter and coverage: please describe the coverage of personnel figures reported in ETER. Please make specific reference to the inclusion/exclusion of external personnel and temporary positions with a contract of autonomous work. In some cases, there could be autonomous entities connected with the HEI (e.g. spin-off, foundations), employing personnel who collaborate with academic personnel and contribute to the production of HEI output but are not accounted for. Specific information on this phenomenon should be provided, if relevant.

Counting method of personnel (FTE): please specify how FTE figures are calculated.
Counting method of personnel (HC): please specify how FTE figures are calculated.

Inclusion of PhD students: please describe the inclusion of PhD students among academic personnel, which depends on the contractual position of PhD students in the respective country.

Inclusion of support and professional personnel: please describe how personnel involved with ancillary services are covered in ETER.

Coverage of medical personnel: please describe how personnel involved in clinical medicine is covered in ETER.

National classification system for academic personnel by field: here, it is required to describe the national classification system of personnel by scientific or academic domain applying in each country. In addition, to improve transparency and comparability across countries, it is required to list the national categories and labels matched with each of the UOE fields of education and training (FoE) as described in the ETER handbook.

Number of professors: please specify the national categories of personnel corresponding to the definition of full professor in ETER.

6.2.2.4 Education

Typologies of degree awarded and levels of education: Detailed breakdown of the number of students by level of education is required for the data collection. It is required to describe the actual levels of education provided in each country and eventually distinguish among sub-groups of HEIs and typologies of degrees awarded. When special rules on the length of curricula apply only to specific subjects (e.g., medicine, engineering), they should also be specified here if they can impact international comparability on the number of graduates.

6.2.2.5 R&D Expenditure

Method for calculating R&D expenditure: R&D expenditures provided for the data collection should be computed according to international practices in the R&D statistics based on the Frascati Manual. It is recognised that only a minority of the countries has been able to devise a statistical system following the guidelines of the Frascati Manual. Here it is required to describe the concrete method for calculating R&D expenditures in each country, eventually describing the reasons for the non-availability of data at a HEI level.

6.2.2.6 General Remarks

This space is reserved for additional general remarks on the data collection, which do not refer specifically to one of the categories listed above.

6.3 Identifiers and demographic information

These variables provide unambiguous identification of institutions, information on demographic events, entries and exits into the ETER perimeter, and matching information with other datasets.

Table 14 Variables included in identifiers and demographic information

Variable	Data source
ETER ID	ETER
Reference year	ETER
ETER ID - Year	ETER
National identifier (optional)	NSA
Institution name (in own language)	RISIS-OrgReg
English institution name (if available)	RISIS-OrgReg
Acronym	RISIS-OrgReg
Demographic events	Demographic event between the past year and reference year
	Affected institutions (past)
	Demographic events between reference year and the following year
	Affected institutions (following year)
	Short description of the event
	OrgReg demographic event ID
ROR/GRID identifier	RISIS-OrgReg
WHED identifier	RISIS-OrgReg

For a description of which institutions are to be included in ETER, see chapter four.

6.3.1 ETER identifier

	Type of variable	Data source
ETER ID	ISO code (two-digit country code and four digits numeric)	assigned by ETER team in cooperation with RISIS-OrgReg

Country code (ISO 3166) + integer (four digits code, e.g. AT0001). Identifiers are assigned in the ETER project in cooperation with RISIS-OrgReg and are unique through the whole data set and multi-annual data collection, i.e. identifiers used in the past for HEIs that do not exist anymore are not reused.

6.3.2 Reference year

	Type of variable	Data source
Reference year	four digits numeric	assigned by ETER

MMMM (e.g. 2011). The year to which the data in ETER refer. For rules concerning the reference period, see chapter 5.4.

6.3.3 ETER ID – Year

	Type of variable	Data source
ETER ID - Year	string	assigned by ETER

ETERID.reference year (e.g. AT0001.2012). This variable is the primary (unique) key in the ETER database, which identifies the HEI and the year to which all other variables refer.

6.3.4 National ID

	Type of variable	Data source
National ID	string	NSA

This (optional) variable includes the ID used in official national registers to simplify matching between the ETER database and national data sources. When there is no such official register, it should be 'not applicable'.

6.3.5 Legal name of the institution

	Type of variable	Data source
Legal name of the institution	text, using ISO/IEC 8859 encoding	RISIS-OrgReg, NSA might correct these data if relevant

The legal name of the institution as stipulated in the legal registry or founding act or, if it is not known, the name used by the institution itself. Importantly, the name of the institution should be in the reference year for data collection; changes of names across years without a change of IDs is possible when they simply imply a relabelling of the same institution (keeping its location, most of the activities, personnel, etc.). See chapter four for the distinction between demographic events and name changes.

This also includes name changes related to changes in status (Hogeschool > University).

6.3.6 English translation

	Type of variable	Data source
Official English translation	text, using ISO/IEC 8859 encoding	RISIS-OrgReg, NSA might correct these data if relevant

Official English translation of the name of the institution when available or commonly used translation. This variable is 'not applicable' when no commonly accepted translation is available.

6.3.7 Institution acronym

	Type of variable	Data source
Institutional acronym	text, using ISO/IEC 8859 encoding	RISIS-OrgReg, NSA might correct these data if relevant

The acronym of the institution as is frequently used, for example, in official communication or on the website. This notion of an acronym is broader than the official ones; it can also refer to widely used short names, which can be found on institutional websites.

When no commonly accepted acronym can be found, this variable should be 'missing'.

6.4 Demographic events

In ETER, demographic events are directly integrated into the data for each year; this approach allows users to immediately recognise demographic events, which might affect comparisons across years. They are derived from RISIS-OrgReg but are coded differently.

While in OrgReg demographic events are recorded with a specific date, in ETER, they are recoded across years, usually the years across which the event has an impact on data collection. So, for example, if a demographic event takes place on 01.06.2018 but impacts data from 2019 onwards, the event will be recorded across the years 2018 and 2019.
To this aim, ETER includes the following variables.

6.4.1 Demographic event (past)

	Type of variable	Data source
Demographic event (past)	nominal	RISIS-OrgReg

This variable identifies whether the considered HEI in a specific year was affected by a demographic event between the previous year's data collection and the current year. The following list of codes is applied:

Table 15 Codes for demographic events

Demographic events	
0	no demographic event
1	entry
2	exit
3	birth
4	death
5	merger
6	split
7	take-over
8	spin-out

6.4.2 Affected HEIs (past)

	Type of variable	Data source
Affected HEIs (past)	ISO code (two-digit country code and four digits numeric)	assigned by ETER team in cooperation with RISIS-OrgReg

The ETER ID of the affected HEIs in the previous year. When no demographic event happened or when the event is 'entry' or 'birth', the code 'a' should be applied. Multiple HEIs should be separated by a semicolon (;).

6.4.3 Demographic event (future)

	Type of variable	Data source
Demographic event (future)	nominal	RISIS-OrgReg

This variable identifies whether the considered HEI in a specific year is affected by a demographic event between the current year of data collection and the next year. The same list of codes as for past events is adopted.

6.4.4 Affected HEIs (future)

	Type of variable	Data source
Affected HEIs (future)	ISO code (two-digit country code and four digits numeric)	assigned by ETER team in cooperation with RISIS-OrgReg

The ETER ID of the affected HEIs in the following year. When no demographic event has happened, or the event is 'exit' or 'merger', the code 'a' should be applied. Multiple HEIs should be separated by a semicolon (;).

Example: In Austria, the two HEIs AT001 and AT002 were merged to AT0033 between data collection in 2011 and 2012. At the same time, AT0003 gave rise to a spin-out HEI AT0034. This led to the following demographic notations in the ETER dataset.

6.4.5 Demographic event ID

	Type of variable	Data source
Demographic event ID	ISO code (two-digit country code and three digits numeric)	RISIS-OrgReg

The OrgReg identifier of the demographic event in the form DEMO+[country code]+XXX (for example, DEMOAT001).

Table 16 Demographic Events - Examples

ETER ID	Year	Demographic event (past)	Affected HEIs (past)	Remarks (past)	Demographic event (future)	Affected HEIs (future)	Remarks (future)
AT0001	2011	0	a		5	AT0033	
AT0002	2011	0	a		5	AT0033	
AT0033	2012	5	AT001; AT002		0	a	
AT0003	2011	0	a		8	AT0003; AT00034	
AT0003	2012	8	AT0003		0	a	
AT0034	2012	8	AT0003		0	a	

Remarks and specific cases:

Demographic events are repeated for the past and future years, but codes are managed differently in the two cases:

1) Future year demographic events are added only retrospectively, i.e. those for the reference year 2012 will be added during data collection 2013. In the meantime, they will be coded as missing ('m').
2) Past demographic codes for the ETER first year (2011) refer to events since the EUMIDA data collection for 2008.

6.4.6 ROR identifier

	Type of variable	Data source

ROR identifier	string	RISIS-OrgReg

https://ror.org/ (+alphanumeric string) (e.g. https://ror.org/01h0zpd94). These identifiers are attributed to research organisations by the Research Organisation Registry (ROR) community project (https://ror.org/), a large initiative aimed at establishing interoperability between databases of research activities and outputs. ROR functions as a gateway to other widely used identifiers in the bibliometric community, such as GRID (https://www.grid.ac), CrossRef (https://search.crossref.org) and Dimensions (https://www.dimensions.ai/). The availability in ETER of ROR IDs is therefore essential for the study of the research activity of HEIs.

6.4.7 WHED identifier

	Type of variable	Data source
WHED identifier	string	RISIS-OrgReg

IAU-(+six digits) (e.g. IAU-004210). These identifiers are derived from the World Higher Education Database (WHED; https://www.whed.net/), the largest and most comprehensive global database of HEIs, which provides information on 19,800 higher education institutions (HEIs) in 196 countries and territories. The database is managed by the International University Association in collaboration with UNESCO. The availability of WHED identifiers is, therefore, key for the international outreach of ETER.

6.5 Basic institutional descriptors

Table 17 Variables included in identifiers and demographic information

Variable	Data source
Country of establishment	ETER
Legal status	NSA
Institution category (national classification)	NSA
Institution category (standardised)	NSA
Foreign campus	NSA
Foundation year	RISIS-OrgReg
Legal status year	RISIS-OrgReg
Ancestor year	RISIS-OrgReg
University hospital	RISIS-OrgReg
Institutional website	RISIS-OrgReg
RISIS-OrgReg institutional page	RISIS-OrgReg

6.5.1 Country of establishment

	Type of variable	Data source
Country of establishment	ISO code (two digits)	ETER

The country where the institution is established using official ISO 3166 country codes (http://www.iso.org/iso/english_country_names_and_code_elements). The country of establishment is where the institution develops most of its activities, for example, where the largest part of the personnel is located, even if this is not the legal seat of the institution.

Foreign campuses (branch campuses) should be included in the country where they are located; distance institutions should be included in the country (and region) where their headquarters are located (independent from the location of the students).

6.5.2 Legal status

	Type of variable	Data source
Legal status	nominal (public=0, private=1, private government-dependent=2)	NSA

The classification between *public* and *private* is made according to whether a public agency or a private entity has *ultimate control* over the institution. *Ultimate control* is decided with reference to who has the power to determine the institution's general policies and activities and appoint the officers managing the school. Ultimate control will usually also extend to the decision to open or close the institution. As many institutions are under the operational control of a governing body, the constitution of that body will also have a bearing on the classification.

Private institutions should be further divided between *government-dependent* – either receiving more than 50% of their core funding from government agencies or whose teaching personnel is paid by a government agency – and *independent private*.

Please refer to the UOE manual for details on how to classify institutions and for a discussion of special cases.

6.5.3 Institution category (national classification)

	Type of variable	Data source
Institution category (national classification)	text	NSA (preferred)/National Experts (optional)

Specification of the type of institution, using the national types of institutions. It should be provided in the national language and English if a translation is available.

References might be, for example, the categories used in the national higher education act or commonly used national categories.

6.5.4 Institution category (standardised)

	Type of variable	Data source
Institution category (standardised)	nominal (0=other, 1=UNI, 2=UAS)	NSA (preferred)/National Experts (optional)

This variable specifies a European-level standardised classification of higher education institutions built on top of the national categories. It is relevant in order to allow for a comparative analysis of higher education systems and to analyse subgroups.

The following categories are used:

- *UNI (university).* These HEIs display a largely academic orientation (without excluding some focus on applied research). They have the right to award a doctorate and can bear the full name of 'university' (including variants like technological university). In general, awarding doctorates should be the main criterion to classify HEIs in this category, even if a few doctoral-awarding HEIs might be included in the two following categories.
- *UAS (university of applied sciences/college).* These institutions are officially recognised as a part of higher education, though not as universities (see definition above). Commonly these institutions focus on professional education. In most cases, they do not have the right to award a doctorate (exceptions are possible). National names are, for example, Fachhochschule (Austria, Germany), Hogescholen (Netherlands), colleges (Norway) and Polytechnics (Finland). In most cases, such institutions constitute the second sector of higher education. However, exceptions are possible (e.g. where the college sector is now merged into a unitary

HE system, but colleges still have a distinct mission and constitute a distinct sector, even when ruled by the same law, like in the case of Sweden).

- *Other.* All institutions that do not fit the university/university of applied science description will be categorised as 'other'. This may apply to institutions like art academies or military schools. In addition, technological and professional schools in countries without a binary system (like the UK or France) should be classified in this way.

Please note: the other category will be analysed more in-depth by the consortium. The distinction between universities and universities of applied sciences is usually stipulated in legal documents like the national higher education acts.

6.5.5 Foreign campus

	Type of variable	Data source
Foreign campus	binary (0=no, 1=yes)	NSA/pre-assigned in ETER
Name and country of the mother institution	text	NSA/pre-assigned in ETER

Variable to be checked, probably of no use / we might refer to linkages in OrgReg and insert in OrgReg the few data we have in ETER.

If the institution is a foreign campus of another institution, for instance, the Maltese Campus of Middlesex University, please fill in the following variables.

If foreign campus=1 then please fill in a text field with the name and country of the mother institution.

6.5.6 Year of establishment

The foundation year is highly relevant in order to understand the context of the considered institutions (for example, it might be assumed that older institutions are more prestigious or well-known). However, since many HEIs underwent significant changes during their history, identifying the foundation year might prove difficult in many cases. Hence, ETER will collect data for the main variable:

	Type of variable	Data source
Foundation year	integer (four digits)	RISIS-OrgReg, NSA might correct these data if relevant

Defined as the year when the institution was established (corresponding to a birth in demography).

Additionally, two optional variables will be collected if needed:

	Type of variable	Data source
Legal status year	integer (four digits)	RISIS-OrgReg, NSA might correct these data if relevant

The year when the institution received its legal status. Examples of legal status events include accreditation as an HEI, and accreditation as a university.

	Type of variable	Data source
Ancestor year	integer (four digits)	RISIS-OrgReg, NSA might correct these data if relevant

The foundation year of the oldest ancestor which can be identified. This variable is relevant when a HEI claims to descend from earlier institutions, but the current HEI is quite different from its ancestor. If there is no ancestor prior to the foundation year of the current HEI, this variable should be set to 'not applicable'.

Please note: In some cases, it might be difficult to decide whether some change gave rise to the birth of a new institution or was it a gradual change. Criteria for decisions include a completely different name, a different scope of the subject covered or a completely new internal organisation.

Examples:

- University of Lucerne (CH) – legal status and foundation year 2000, when the institution was officially founded as a comprehensive university. First ancestor: Faculty of Theology, the year 1600.
- University of Stavanger (NO) – legal status year 2005, when the institution was accredited with university status. Foundation year: 1994, when the State University College of Stavanger was established. Ancestor year: 1968 with the establishment of the District University College of Stavanger.
- University of Aalto (FI) – legal status and foundation year 2010, when the merger was undertaken – the ancestor year 1849 when the oldest of the three merged schools (Helsinki University of Technology) was founded.

The exact calendar year should be used. A short description of the change of status should be included in the metadata worksheet. All three variables should refer to the legal event

(for example, the accreditation decision) and not to the effective 'beginning of activities' (even if, for older institutions, it might be difficult to trace back to this distinction).
If the legal status year is the same as the foundation year, the same date should be inserted in the data collection sheet. If they are missing because this information is not available, the code 'm' should be used.

6.5.7 University hospital

Variables included in this section include information on the presence of a university hospital, defined as a hospital that has a close link with the HEI, based on the following criteria:

- Shared people (i.e. medical doctors also appointed as HEI personnel/professors).
- Shared facilities and research activities.
- Involvement of the hospital in medical education.
- University hospitals are not necessarily legally part of the HEI, but might be independent organisations closely linked with the HEI.

The Following variables are included:

	Type of variable	Data source
University hospital	binary (0=no hospital, 1=yes)	RISIS-OrgReg, NSA might correct these data if relevant

	Type of variable	Data source
University hospital RISIS-OrgReg ID	text, country code + four digits (for example, AT2001), code 'a' should be used if the dummy variable is 0	RISIS-OrgReg, NSA might correct these data if relevant

The identifier in the RISIS organisational register of the hospital (https://orgreg.joanneum.at). This identifier allows retrieving further information on the hospital, like name, foundation year and location information.

	Type of variable	Data source
University hospital name	text, code 'a' should be used if the dummy variable is 0	RISIS-OrgReg, NSA might correct these data if relevant

In English when available, otherwise national language.

	Type of variable	Data source
University hospital remarks	text	RISIS-OrgReg, NSA might correct these data if relevant

Including details on the nature of linkage.

6.5.8 European University Alliance

Variables included in this section include information on the membersip in an alliance within the framework of the European Universities initiative (https://education.ec.europa.eu/education-levels/higher-education/european-universities-initiative), based on the following criteria:

- HEIs which are full members of the alliances;
- Associate members (HEIs only) are included where they are de facto full members (in particular, Swiss institutions participation of which is funded by the Swiss government, and British institutions, which self-fund participation in the alliances).

The Following variables are included:

	Type of variable	Data source
Member of European University alliance	binary (0=not a member, 1=yes)	RISIS-OrgReg

	Type of variable	Data source
European University Alliance RISIS-OrgReg ID	text, INT + four digits (for example, int6001), code 'a' should be used if the dummy variable is 0	RISIS-OrgReg

The identifier in the RISIS organisational register for the alliance (https://orgreg.joanneum.at). This identifier allows retrieving further information on the alliance, like full list of member HEIs, foundation year and location information.

	Type of variable	Data source
European University Alliance name	text, code 'a' should be used if the dummy variable is 0	RISIS-OrgReg

In English.

	Type of variable	Data source
European University Alliance acronym	text, code 'a' should be used if the dummy variable is 0	RISIS-OrgReg

6.5.9 Institutional website

	Type of variable	Data source
Institutional website	text	RISIS-OrgReg, NSA might correct these data if relevant

Registration of the official website. *Format www.website_name.*

6.5.10 RISIS-OrgReg institutional page

	Type of variable	Data source
RISIS-OrgReg institutional page	hyperlink	RISIS-OrgReg

https://register.orgreg.joanneum.at/#/entity-details/(ETERID). This variable provides direct access to the RISIS-OrgReg institutional page, which provides detailed information concerning history, demographic events and linked entities of the HEI under consideration. Note that access to RISIS-OrgReg requires registering at the RISIS Central Facility at https://auth-risis.cortext.net/user/register.

6.6 Geographic information

Table 18 Variables included in geographical information

Variable	Data source
Region of establishment, NUTS2 code	RISIS-OrgReg
Region of establishment, NUTS3 code	RISIS-OrgReg
Name of the city	RISIS-OrgReg
Postcode	RISIS-OrgReg
Geographical coordinates	RISIS-OrgReg
Multi-site institution	RISIS-OrgReg
Cities of satellite campuses	RISIS-OrgReg
Geographical coordinates of satellite campuses	RISIS-OrgReg
NUTS3 codes of satellite campuses	RISIS-OrgReg
Postcodes of satellite campuses	RISIS-OrgReg

6.6.1 Region of the main seat - NUTS code

	Type of variable	Data source
Region of the main seat - NUTS code	text (4 characters for NUTS 2 region, 5 characters for NUTS 3 region)	RISIS-OrgReg

This variable requires the NUTS2 and NUTS3 regions where the institution's main seat is established (see perimeter chapter four). For the NUTS2 and NUTS3 codes, see classification chapter 5.3.1.

When HEI activities are distributed in more than one NUTS3 region, meaning no main seat can be identified, this variable refers to the legal seat.

Distance education institutions have to be classified according to the location of their main seat, independent of where the students live (and the existence of decentralized student support services).

6.6.2 Postal code

	Type of variable	Data source
Postal code	text	RISIS-OrgReg

The postal code of the official address of the HEI. Different postal code systems throughout Europe are described in chapter 5.3.2.

This variable is meant to identify the geographical location of the HEI. A remark should be added when the legal address is located far away from the place of the main activities (for example, in a different city).

6.6.3 Name of city

	Type of variable	Data source
Name of city	text	RISIS-OrgReg

The city/town where the main seat and most of the activities are located. This variable should be 'not applicable' when no central seat can be found, as the HEI is too dispersed. The English name should be added if possible.

6.6.4 Geographic coordinates

	Type of variable	Data source
Geographic coordinates	numeric	RISIS-OrgReg

Geographic coordinates, i.e. longitudes and latitudes, are based on the postcode of the official address.

This variable is not meant to allow for micro-localisation within a single city but rather for a broader localisation at the national and European levels. When the HEI activities are divided between different sites in different locations and no central place can be identified, the localisation of the legal seat should be used, but the variable should be flagged, and a remark should be added.

6.7 Multi-site institutions

Multi-sited institutions are defined as institutions with local establishments in the NUTS3 region(s) that are different from the main seat. ETER provides the following variables on multi-campus HEIs.

	Type of variable	Data source
Multicampus	binary (0=not multi-sited, 1=multi-sited)	RISIS-OrgReg

	Type of variable	Data source
Cities of the campuses	text, code 'a' should be used if the dummy is 0	RISIS-OrgReg

A list separated by semicolons (;) of the cities where the campuses are located.

	Type of variable	Data source
Geographical coordinates of the campuses	numeric, code 'a' should be used if the dummy is 0	RISIS-OrgReg

A list separated by semicolons (;) of the longitude, respectively latitude of the campuses locations (two variables).

	Type of variable	Data source
NUTS codes of the campuses	text, code 'a' should be used if the dummy is 0	RISIS-OrgReg

A list separated by semicolons (;) of the NUTS codes of the campuses location.

	Type of variable	Data source
Postcodes of the campuses	text, code 'a' should be used if the dummy is 0	RISIS-OrgReg

A list separated by semicolons (;) of the postcodes of the campuses location.

6.8 Educational activities

Table 19 Variables included in educational activities

Variable	Data source
Lowest/highest degree delivered	NSA/NE
Students enrolled at ISCED level 5 by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Students enrolled at ISCED level 6 by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Students enrolled at ISCED level 7 by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Students enrolled at ISCED level 7 long degree by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Graduates at ISCED level 5 by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Graduates at ISCED level 6 by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Graduates at ISCED level 7 by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Graduates at ISCED level 7 long degree by gender, citizenship, mobility, fields of education, age group, full-time/part-time	NSA
Distance education institutions	NSA/NE

The variables relating to educational activates (table 19) provide a rather simple but complete view of the main dimensions of educational activities without a particular burden for data collection. Description of the ISCED-2011 classification of educational degrees can be found in chapter 5.

Please note: when a HEI does not have the right to award degrees at some ISCED level, the corresponding fields for students and degrees should be set to 0.

6.8.1 Lowest and highest degree delivered

	Type of variable	Data source
Lowest and highest degree delivered	nominal	NSA (preferred)/National Experts (optional)

The lowest and highest degrees delivered by institutions should be defined using the following classification:

- ISCED 5 diplomas with duration of less than three years,
- ISCED 6 bachelor (3 or 4 years),
- ISCED 7 master or equivalent diploma in the pre-Bologna system (for example 4/5 years), and
- ISCED 8 qualification equals doctorate.

6.8.2 Students

The number of students enrolled per institution at ISCED level 5, 6, 7 levels (with separate figures for long ISCED 7 students, which should not be included in ISCED 7) should comply with the UOE manual for definitions and data collection procedures. Data should reflect the number of students enrolled at the beginning of the academic year (end of the first month of the academic year) and be based on the count of students. Exceptions to this rule should conform to UOE data collection practices.

In general, ETER data should reflect the number of students enrolled at a specific HEI and, therefore, double counting of students enrolled in joint degrees is possible. The preferred method is apportioning of students (so that the sum of proportions gives 1), but simple double counting or pro-rating (and thus fractional numbers) are allowed. However, the exact method should be explained in the metadata.
Breakdowns by gender, fields of education, citizenship, mobility, part-time/full time and age groups will be provided separately by the level of education but not combined in order to reduce the burden of data collection.

In addition, variables are included in the dataset for the total number of students at levels 5-7 by fields of education, gender, citizenship, mobility, part-time/full time and age groups.
When long ISCED 7 students and degrees cannot be separated, they should be included with the other ISCED 7 students and degrees (and the code 'm' used for these cells).

6.8.2.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.2.2 Citizenship

	Type of variable	Data source

Citizenship	numeric	NSA

A breakdown between national and foreign students. Students are considered foreign if they do not have the citizenship of the country for which data is collected (see UOE manual).
Foreign students are defined as non-citizens of the country in which they study. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.2.3 Internationalisation

	Type of variable	**Data source**
Internationalisation	numeric	NSA

International/Mobile students are defined as foreign students who have physically crossed a national border and moved to another country with the objective of studying (see UOE manual). In other words, the student has moved from what we in this context call the country of origin to the reporting country of study (also called country of destination). The country of origin is defined as the country of prior education, i.e. country where the upper secondary diploma was obtained. The status of a mobile student is maintained throughout the whole education at the tertiary level (i.e. students who entered at the bachelor level are still considered mobile at the PhD level).

Special case: distance education students enrolled in a HEI located in a country different from the one where the learner is based are *not* considered as mobile. These students are entered into the unclassified category of mobility. In this case, an explanatory remark should be added ('unclassified are distance education students based in other countries'). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.2.4 Fields of Education

	Type of variable	**Data source**
Fields of Education	numeric	NSA

Students should be divided by fields of education. The categories used in this data set are listed in chapter 5. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.2.5 Part-time and full-time students

	Type of variable	Data source
Part-time and full-time students	numeric	NSA

Data on students should be broken down between part-time students and full-time students. Following the UOE manual, a full-time student is one who is enrolled in an education programme and who is intended study load amounts to at least 75% of the typical full-time annual study load. A part-time student is one who is enrolled in an education programme whose intended study load is less than 75% of the typical full-time annual study lead. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.2.6 Students by age groups

	Type of variable	Data source
Students by age groups	numeric	NSA

Students should be broken down by age groups. Following the UOE manual, students and graduates are classified by their age as of 1st January of the year in which the school or academic year ends. Following ETER conventions, this amounts to the year following the reference year for data collection (i.e. for ETER data collection 2016, corresponding to the academic year 2016/2017, and 01.01.2017). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

To reduce the amount of data, students are grouped by the following age groups:

- Below 20
- 20-21
- 22-24
- 25-29
- Above 29

6.8.2.7 Total number of students ISCED 5-7

	Type of variable	Data source
Total number of students ISCED 5-7	numeric	NSA

The sum of students at levels ISCED 5-7.

6.8.3 Graduates

Numbers of graduates per institution at ISCED level 5, 6, 7 levels by fields of education (with separate figures for long ISCED 7 degrees, which should not be included in ISCED 7), gender, mobility, citizenship and age groups. Breakdowns by gender, fields of education and citizenship/mobility and age groups, are provided separately by the level of education but not combined in order to reduce the burden for data collection.

In addition, variables for the total number of graduates at levels 5-7 by fields of education, gender, citizenship and mobility and age groups are included in the dataset.

When long ISCED 7 graduates cannot be separated, they should be included with the other ISCED 7 graduates (and the code 'm' used for these cells).

6.8.3.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.3.2 Citizenship

	Type of variable	Data source
Citizenship	numeric	NSA

A breakdown between national and foreign graduates. Graduates (that is, students who have successfully completed a course of study) are considered to be foreign if they are not citizens of the country for which data is collected (see UOE handbook). Foreign graduates are defined as non-citizens of the country in which they studied. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.3.3 Internationalisation

	Type of variable	Data source
Internationalisation	numeric	NSA

International/Mobile graduates are defined as foreign graduates who have crossed a national border and moved to another country with the objective of studying (see UOE

handbook). In other words, the student has moved from what we in this context call the country of origin to the reporting country of study (also called country of destination). To the extent possible, definitions based on the country of prior education should be adopted. Information on the exact criterion adopted should be provided in the metadata.

Special case: distance education graduates who were enrolled in a HEI located in a country different from the one where the learner was based are *not* considered as mobile. These graduates are entered into the unclassified category of mobility. In this case, an explanatory remark should be added ('unclassified are distance education students based in other countries'). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.3.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Graduates should be divided by fields of education, using the ISCED-F 2013 classification. The categories used are listed in chapter 5. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.8.3.5 Graduates by age groups

	Type of variable	Data source
Graduates by age groups	numeric	NSA

Graduates should be broken down by age groups. Following the UOE manual, students and graduates are classified by their age as of 1st January of the year in which the school or academic year ends. Following ETER conventions, this amounts to the year following the reference year for data collection (i.e. for ETER data collection 2016, corresponding to the academic year 2016/2017, and 01.01.2017). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

To reduce the amount of data, students are grouped by the following age groups:

- Below 20
- 20-21
- 22-24
- 25-29
- Above 29

6.8.3.6 Total number of graduates

	Type of variable	Data source
Total number of graduates	numeric	NSA

The sum of graduates at levels ISCED 5-7.

6.8.4 Distance education institution

	Type of variable	Data source
Distance education institution	Nominal (0,1,2 – see below for details)	NSA (preferred)/National Experts (optional)

This variable is introduced to identify those institutions in which distance education represents a substantial component of the educational offer. The rationale is that distance education has a different delivery structure and economic model than on-campus education, with a visible impact on indicators like the student to personnel ratios.

Distance education covers educational offers where students are not regularly coming to the HEI campus, except for workshops or short seminars. Blended learning with substantial on-campus activities supported by IT is not included in the definition.

- 0 = no or limited amount of distance education (<20% of the students in distance programs).
- 1 = a substantial share of distance education. An example is Anadolu University in Turkey.
- 2 = mostly distance education (>90% of the student in distance programs). Typical examples are Fernuniversität Hagen or the UK Open University.

6.9 Research activities

Table 20 Variables included in research activities

Variable	Data source
Research active institutions	NSA/NE
Students enrolled at ISCED level 8 by gender, citizenship, mobility field of education, age group, full-time/part-time	NSA
Graduates at ISCED level 8 by gender, citizenship, mobility, field of education, age group	NSA

Inclusion in official R&D statistics	NSA/NE

6.9.1 Research active institutions

	Type of variable	Data source
Research active institutions	Binary (0=non-research active; 1=research active)	NSA (preferred)/National Experts (optional)

Research-active institutions are those that have institutionalised research activities. Criteria for inclusion are then the following:

- The existence of an official research mandate.
- The existence of research units is institutionally recognised (for example, on the institutional website).
- The inclusion in R&D statistics (availability of R&D expenditure data) as a sign of institutionalised research activity.
- Awarding doctorates or ISCED 8 degrees.
- Consideration of research in institutions strategic objectives and plans.
- Regular funding for research projects either from public agencies or from private companies.

Institutions fulfilling at least three of these criteria should be included. It is generally expected that non-research active HEIs have no or very low numbers of ISCED 8 students and graduates.

6.9.2 ISCED 8 students

The number of students enrolled per institution at ISCED 8 level should comply with the UOE manual definitions and data collection procedures. Data should reflect the number of students enrolled at the beginning of the academic year (last day of the first month of the winter semester academic year) and can be provided in full-time equivalents. Exceptions to this rule should conform to UOE data collection practices. Note that in ISCED-2011, research masters are included in the level ISCED 7.

Note that ISCED 8 students are attributed to the HEI where they are enrolled, which does not necessarily correspond to the one where they are under an employment contract. If a different criterion than enrolment is adopted, this should be noted in the metadata, and the data should be flagged. Breakdowns by gender, fields of education, citizenship and mobility will not be combined in order to reduce the burden for data collection.

6.9.2.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.9.2.2 Citizenship

	Type of variable	Data source
Citizenship	numeric	NSA

A breakdown between national and foreign students. Students are considered foreign if they are not citizens of the country for which data is collected (see UOE handbook). Foreign students are defined as non-citizens of the country in which they study. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.9.2.3 Internationalisation

	Type of variable	Data source
Internationalisation	numeric	NSA

International/Mobile students are defined as foreign students who have crossed a national border and moved to another country with the objective of studying. In other words, the student has moved from what we in this context call the country of origin to the reporting country of study (also called the country of destination). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.9.2.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Students and awarded degrees should be divided by fields of education. The categories used in this data set are listed in chapter 5.

6.9.2.5 Part-time and full-time students

	Type of variable	Data source
Part-time and full-time students	numeric	NSA

Students should be broken down between part-time students and full-time students. Following the UOE manual, a full-time student is one who is enrolled in an education programme and who is intended study load amounts to at least 75% of the normal full-time annual study load. A part-time student is one who is enrolled in an education programme whose intended study load is less than 75% of the normal full-time annual study lead. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.9.2.6 Students by age groups

	Type of variable	Data source
Students by age groups	numeric	NSA

Students should be broken down by age groups. Following the UOE manual, students and graduates are classified by their age as of 1st January of the year in which the school or academic year ends. Following ETER conventions, this amounts to the year following the reference year for data collection (i.e. for ETER data collection 2016, corresponding to the academic year 2016/2017 and 01.01.2017). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

To reduce the amount of data, students are grouped by the following age groups:

- Below 20
- 20-21
- 22-24
- 25-29
- Above 29

6.9.2.7 Total number of students ISCED 8

	Type of variable	Data source
Total number of students ISCED 8	numeric	NSA

The sum of students at level ISCED 8.

6.9.3 ISCED 8 degrees

This variable asks for the number of degrees delivered at ISCED 8 qualification. It excludes intermediate-stage ISCED 8 programs like the DEA in France and ISCED 8 post-doctorate qualifications (like the German *Habilitation*). Definitions and data collection should comply with the UOE manual. Following the UOE manual, graduations should be based on the

academic year. Breakdowns by gender, fields of education, citizenship and mobility will not be combined in order to reduce the burden for data collection.

6.9.3.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.9.3.2 Citizenship

	Type of variable	Data source
Citizenship	numeric	NSA

A breakdown between national and foreign graduates. Graduates are considered foreign if they do not have the citizenship of the country for which data is collected. Foreign graduates are defined as non-citizens of the country in which they studied. If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

6.9.3.3 Internationalisation

	Type of variable	Data source
Internationalisation	numeric	NSA

International/Mobile graduates are defined as foreign graduates who have crossed a national border and moved to another country with the objective of studying. In other words, the graduates have moved from what we in this context call the country of origin to the reporting country of study (also called country of destination).

6.9.3.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Students and awarded degrees should be divided by fields of education. The categories used in this data set are listed in chapter 5.5.

6.9.3.5 Graduates by age groups

	Type of variable	Data source
Graduates by age groups	numeric	NSA

Graduates should be broken down by age groups. Following the UOE manual, students and graduates are classified by their age as of 1st January of the year in which the school or academic year ends. Following ETER conventions, this amounts to the year following the reference year for data collection (i.e. for ETER data collection 2016, corresponding to the academic year 2016/2017, and 01.01.2017). If this breakdown is not possible, please use the appropriate special code (see chapter 5.5).

To reduce the amount of data, students are grouped by the following age groups:

- Below 20
- 20-21
- 22-24
- 25-29
- Above 29

6.9.3.6 Total number of graduates

	Type of variable	Data source
Total number of graduates	numeric	NSA

The sum of graduates at level ISCED 8.

6.9.4 Inclusion in R&D statistics

	Type of variable	Data source
Inclusion in R&D statistics	binary (0=not included in R&D statistics; 1=included in R&D statistics	NSA (preferred)/National Experts (optional)

This variable identifies those HEIs included in the national perimeter for R&D statistics to compute R&D expenditures in the higher education sector (HERD). This does not per se imply that R&D expenditures at the institutional level are available; however, if this variable is false, the corresponding R&D expenditures variable should be either 'a' or 'm'.

6.10 Participation in European Framework Programs

This section includes a set of variables derived from the participation of HEIs in European Framework Programs (FP); these are relevant to characterise HEIs' research activities, as

well as their collaboration with regional actors and industry. These data are derived from the EUPRO database[11] managed by the Austrian Institute of Technology and are, therefore, not included in the NSA data collection.

EUPRO is a unique dataset providing systematic and standardized information on R&D projects, participants and resulting networks of the EU FP, starting from FP1 and recently integrating H2020 (until 2016) and other European funding instruments, such as EUREKA, COST and selected Joint Technology Initiatives (JTIs). EUPRO covers cleaned and standardised information on R&D projects (such as project objectives and achievements, project costs, total funding, start and end date, contract type, and information on the call) and their participants (standardised name of the participating organisation, organisation type, and geographical location).

Table 21 Variables included in EU-FP participations

Variable	Data source
Project participation and coordination by Fields of Education and Training and by subprogramme	EUPRO
Industrial collaboration in EU Framework Programmes	EUPRO
Regional collaboration in EU Framework Programmes	EUPRO
Researchers' mobility and training cooperation supported by the EU Framework Programmes	EUPRO

All variables refer to the count of active projects during the reference year, except for the mobility and training indicator which refers to the project starting in the reference year. If an HEI has not been identified in EUPRO, these variables are set to '0'; hence, there are no missing data for these variables.

6.10.1 Project participation and coordination

	Type of variable	Data source
EU-FP participation by field of education and training and by subprogramme	numeric	EUPRO
EU-FP coordination by field of education and training and by subprogramme	numeric	EUPRO

[11] https://rcf.risis2.eu/dataset/4/metadata.

The count of the projects active in the reference year in which an HEI is involved as a consortium partner. This can be in the role of a normal partner (a partner in a FP project is defined to receive dedicated funding for a specific task defined in the project; subcontracting partners are not included) or as a coordinating organisation of the project (with the responsibility to manage the project and distribute funds to partners based on the description of work).

Participations and coordination are broken down by Fields of Education and Training using the correspondence table as in Table 22 below. Since some projects are assigned to more than one topic, the sum of breakdown might exceed the total (full counting).

Table 22 Breakdown by FET of EU-FP participation and coordination

FET	EUPRO subject codes	Remarks
00 Generic	NA	
01 Education	105 Educational sciences	
02 Arts and Humanities	31 Humanities	
03 Social Sciences, Journalism and Information	29 Social sciences	excluding 89, 91, 105
04 Business, Administration and Law	91 Economics and Business; 89 Law	
05 Natural Sciences, Mathematics and Statistics	23 Natural Sciences	excluding 47
06 Information and Communication Technologies	47 Computer and Information sciences	
07 Engineering, Manufacturing and Construction	25 Engineering and Technology	
08 Agriculture, Forestry, Fisheries and Veterinary	27 Agricultural Sciences	
09 Health and Welfare	21 medical and Health Sciences	
10 Services	NA	

Further, the number of participation and coordination is broken down by EU-FP subprograms as follows:

- Research and Innovation Action (RIA).
- Innovation Action (IA).
- Coordination and Support Action (CSA).
- European Research Council Grants (ERC).

Due to changes in the EU-FP program structure, this breakdown is available only from 2014 onwards.

6.10.2 Industrial collaboration

	Type of variable	Data source
EU-FP participation with at least one company partner	numeric	EUPRO

The number of active projects in the reference year in which the HEI is involved and in which at least one partner is a company. A company is defined as a private for-profit organisation without any public mission. Note that, for instance, research organisations that are established as companies legally but have a public mission and receive public funding are not treated as a company.

6.10.3 Regional collaboration

	Type of variable	Data source
EU-FP participation with at least one regional non-academic partner	numeric	EUPRO

The number of active projects in the reference year in which the HEI is involved and in which at least one partner is in the same region, using the NUTS3 adapted classification. The latter has been developed within RISIS (risis2.eu) and in the EU-funded project KNOWMAK (knowmak.eu), providing a more appropriate regional breakdown for analysing – broadly speaking – R&D activities since it doesn't intersect functional metropolitan areas by artificial administrative borders (for details see https://zenodo.org/record/3752861 and https://zenodo.org/record/3746455#.YwTYDRzP0ZU).

6.10.4 Researchers' mobility and research training cooperation

ETER includes three variables to characterise researchers' mobility and research training cooperation supported by the European Framework Programs.

	Type of variable	Data source
EU-FP Researchers mobility	numeric	RISIS/AIT

The number of research training fellowships hosted by the HEI. This includes all MCA fellowships under EU FP7 and MSCA individual fellowships under H2020.

	Type of variable	Data source
EU-FP Personnel mobility	numeric	RISIS/AIT

The number of mobile personnel hosted by the HEI. This includes the FP9 MCA International Research Staff Exchange Scheme (IRSES) and the H2020 MSCA Research and Innovation Staff Exchange (RISE).

	Type of variable	Data source
EU-FP Research training cooperation	numeric	RISIS/AIT

This includes Innovative Training Networks under FP9 and H2020.

6.11 Revenues and expenditures

6.11.1 General definitions and principles

Financial data in ETER deal with two distinct aggregates:

- *Revenues*, i.e. the amount of money received by the HEI, either as a general allocation from the state/public authorities or other funding for research, education and other services (which could be coming from public and/or private sources).
- *Expenditures*, i.e. the amount of money spent by an HEI in order to manage its activities, for example, for paying its employees, external services, general expenditures and maintenance costs, as well as capital costs.

Therefore, it is generally assumed that HEIs have their own budget and that there might be a difference between the amount of revenues and expenditures received during the financial year. Some differences might be due to different schedules in revenues and expenditures (for example, delayed payment for services) or the fact that HEIs incur profit or loss. Finally, differences might be due to the treatment of capital costs, as explained below. However, substantial and systematic differences over many years will be checked as they might signal methodological problems.
This distinction is a major difference with UOE data collection, which has a single perimeter for expenditures and provides information on sources of funds (thereby implicitly assuming revenues = expenditures). This also implies that adjustments for changes in fund balances reported in UOE data collection are not included in ETER.

- *Accounting systems*. ETER acknowledges the variety of accounting systems of HEIs in Europe, some having the traditional cash-based approach of public administration (where all revenues and expenditures are accounted in the current year and the HEI cannot build reserves and make profits and losses), others having a commercial accounting system based on the accrual principle. The major

difference between the two methods concerns how capital expenditures are accounted for (see below).

- *Currency*. All financial data should be provided in the national currency at current prices. All data are in currency units (not thousands). In the ETER database, they will be converted into Euros (at average yearly exchange rates) and in Purchasing Power Parities (PPP) in euro.[12]
- *Reference periods*. It is generally assumed that all financial variables refer to the calendar year (01.01-31.12). Other definitions of the reference period should be noted in the metadata.

6.11.2 Structure of the ETER financial information

The structure of the ETER financial information, both on the revenues and expenditures side, is based on the distinction between operating and non-operating accounts. The former includes revenues and expenditures related to the university's everyday operations, while the latter include extraordinary costs and revenues, like significant investments and non-recurrent donations as endowment capital. This distinction is meant to make data on the current revenues and expenditures more comparable over time and across HEIs and limit comparability problems due to differences in accounting systems.

Table 23 Classification of revenues and expenditures

Expenditure			
Current expenditure			**Non-current expenditure**
Personnel expenditures	Non-personnel expenditure	Expenditure unclassified, including depreciation (accrual basis)	Capital expenditure (cash basis)
Revenues			
Current revenues			**Non-recurring revenues**
Core budget	Third-party funding	Students fees	Revenues unclassified	Contributions for investments, donation (capital)

The main aim of this classification is to improve the comparability between HEIs having different accounting systems, particularly by distinguishing current expenditures and revenues on the one side, non-current expenditures and revenues on the other side.

[12] Until 2020, Eurostat data on PPPs for EU-28 countries is used; from 2021 onwards, only data for EU-27 countries is available and used in ETER.

This takes into account lasting differences in this respect between those universities having an accrual accounting system and those having a cash accounting system:

- In an *accrual accounting system*, capital costs are not part of the operating statement, which includes, however, depreciations of past investments.
- In a *cash accounting system*, capital costs are part of the operating statement in the year they have been disbursed; revenues might also include large investment contributions from the State.

The table overleaf (22) provides a comparison of the expenditures and respective revenue categories for the two accounting systems for two HEIs with the same level of current expenditures. Both HEIs undertook an investment of 500 units in the first year, one having an accrual accounting system and one a cash accounting system.

In both cases, total expenditures and total revenues are the same over the five-year period, but the repartition by year is different. Moreover, the levels of current revenues will differ since the first HEIs is financing their investments from the current revenues, whereas the second one receives them as a specific endowment. Therefore, excluding capital costs and depreciation, this reporting model ensures comparability between current expenditures and removes the most significant differences in revenues due to large one-time contributions for investments.

Table 24 Example: Comparison of expenditure and revenue categories in HEIs with different accounting systems

Accrual basis						
	Current expenditures	Capital costs/ depreciation	Total expenditures	Current revenues	Contributions for investments	Total revenues
2010	1000	100	1100	1100	0	1100
2011	1000	100	1100	1100	0	1100
2012	1000	100	1100	1100	0	1100
2013	1000	100	1100	1100	0	1100
2014	1000	100	1100	1100	0	1100
Cash basis						
	Current expenditures	Capital costs/ depreciation	Total expenditures	Current revenues	Contributions for investments	Total revenues
2010	1000	500	1500	1000	500	1500
2011	1000	0	1000	1000	0	1000
2012	1000	0	1000	1000	0	1000
2013	1000	0	1000	1000	0	1000
2014	1000	0	1000	1000	0	1000

6.11.3 Expenditures

Table 25 Variables included in expenditures

Variable	Data source
Current expenditure	NSA
Personnel expenditure	NSA
Non-personnel expenditure	NSA
Capital expenditure	NSA
Accounting system of capital expenditure	NSA
R&D expenditure	NSA

6.11.3.1 Current expenditure

	Type of variable	Data source
Current expenditure	numeric, national currency	NSA

This variable includes all current expenditures of the considered HEI for its activities, including salaries, payment for services, repairs and maintenance expenditures, electricity, and other financial costs. It excludes depreciation and capital costs even if they are included in the operational account (cash accounting principle).

This variable measures the total level of expenditure of higher education institutions in the reference period. The perimeter corresponds to expenditures inside institutions as defined in the UOE manual covering all types of goods and services. It includes expenditures for educational core goods and services, R&D expenditures, and other non-instructional goods and services, such as meals and transportation, when directly made by HEI institutions.

Special case: Expenditures for hospitals. Consistent with the UOE manual, expenditures by or on teaching hospitals (sometimes referred to as academic hospitals or university hospitals) is excluded from the educational expenditure. In particular, all patient care costs and other general expenses of academic hospitals, even if the education authorities pay such expenses. However, expenditure by or on teaching hospitals that is directly and specifically related to the training of medical students, and expenditure on R&D at teaching hospitals to the extent that is included in the OECD/DSTI data collection on R&D are included.

Please note: the sum of personnel, non-personnel and unclassified expenditure should equal operating expenditure.

6.11.3.2 Personnel expenditure

	Type of variable	Data source
Personnel expenditure	numeric, national currency	NSA

This variable includes salaries and social expenses such as payroll tax (employer's contribution), insurance, and pensions. It requires having employees.

6.11.3.3 Non-personnel expenditure

	Type of variable	Data source
Non-personnel expenditure	numeric, national currency	NSA

Includes other current expenditure, i.e. all current expenditure on goods and services other than compensation of personnel. This includes electricity, costs for equipment included in the current expenditures (for example, small equipment), services paid outside the HEIs, financial costs, rents, repairs and maintenance of infrastructure.

6.11.3.4 Expenditure unclassified

	Type of variable	Data source
Expenditure unclassified	numeric, national currency	NSA

If any current expenditures are not assigned to personnel or non-personnel current expenditure, they should be indicated as unclassified. When all expenditures can be assigned to the provided categories, this cell should contain '0'. When a breakdown is not available, this column should be 'm' (including a flag and a note), and only total expenditures be filled in.

6.11.3.5 Capital expenditure

	Type of variable	Data source
Capital expenditure	numeric, national currency	NSA

Capital expenditure is expenditure on assets that last longer than one year. They include spending on construction, renovation and major repair of buildings, and expenditures on new or replacement equipment. It is understood that most countries report small outlays for equipment below a certain cost threshold, as current rather than capital spending. This variable should refer to large investments like new facilities and buildings. Accounting of capital expenditure is based on different principles depending on the accounting system:

- In *accrual accounting*, it should correspond to the depreciation as recorded in the operating statement.
- In *cash accounting*, it should correspond to the investment cost as recorded in the operating statement.

Please note: If capital expenditure cannot be separated from current expenditures, this variable should be 'm' (including a flag and a note), and the current expenditure variable flagged ('d').

6.11.3.6 Accounting system of capital expenditure

	Type of variable	Data source
Accounting system of capital expenditure	nominal (0=not included in university account, 1=cash accounting, 2=accrual accounting).	NSA

This variable indicates the basic method of computing capital expenditures within a HEI:

- *Not included in university account*, when the universities facilities are owned and financed by the State directly (or by third parties).
- *Cash accounting*, meaning that capital expenditures are accounted for in the year the expenses are incurred (following provisions of UOE manual).
- *Accrual account basis*, meaning that capital expenditures are depreciated across years in the HEI accounts and only these are included in the operating accounts.

This variable is collected at the HEI level since the method might differ between HEIs.

6.11.3.7 R&D Expenditure

	Type of variable	Data source
R&D Expenditure	numeric, national currency	NSA

Data on R&D expenditures can be obtained either by surveys or in the form of administrative data, as described in detail in the Frascati Manual. The calculation of research expenditure of higher education institutions is particularly complex since, in most cases, research and educational activities are not clearly separated (both in real activities and in accounting). The corresponding methods are based on a breakdown of worked hours of HEI personnel, which are then converted into a breakdown of cost by use of data on salaries. Data can be collected from time surveys of HEI personnel or administrative information, frequently using R&D coefficients to convert data to R&D costs.

If data are not available at the HEI level, this variable should be missing. Metadata on the methodology of the R&D statistics should be included in the metadata worksheet. For non-research-active institutions, this variable should be 0 (and not 'm' for 'missing').

6.11.4 Revenues

Table 26 Variables included in revenues

Variable	Data source
Current Revenue	NSA
Core funding	NSA
Basic government allocation	NSA
Other core funding	NSA
Third-party funding	NSA
Public third-party funding	NSA
Private third-party funding	NSA
Third-party funding from abroad	NSA
Third-party funding undivided	NSA
Student fees funding	NSA
Tuition fees	NSA/NE
Non-recurring revenues	NSA

6.11.4.1 Current Revenue

	Type of variable	Data source
Current Revenue	numeric, national currency	NSA

This variable measures the amount of money received by the considered HEI in the reference period, excluding non-recurring revenues, which are recorded in a separate variable.

Two different types of breakdowns of revenues are possible. First, UOE manual provides a breakdown of the source of funds by distinguishing between public, private and households. The second type of breakdown is the distinction, *which is not used in UOE*, between core funding (resources for the general operations of the HEI) and third-party funds, i.e. resources devoted to specific activities (usually at the level of researchers or organisational subunit). The two definitions do not match each other, and, therefore, care needs to be taken in order to achieve consistency.

Table 27 Revenues classification

	Government	Private	Households	Abroad
Core funding	Basic government allocation	Other revenues (e.g., donations, interests)		
Third-party funding	Public third-party funds	Private funds (contracts and payments for education)		Third-party funds from abroad
Student's fees			Fees (student only)	

6.11.4.2 Core funding

	Type of variable	Data source
Core funding	numeric, national currency	NSA

Core funding is defined as funding available for the operations of the whole institution, which is not earmarked to specific activities and whose internal allocation can be decided freely by the institution itself. Thus, the main criterion to separate core funding and third-party funds is that the former is managed at the level of the whole HEI, and there is discretion to which activities to devote to them. In contrast, third-party funds are usually attributed and managed directly by organisational subunits.

In most institutions, the main component of the core budget is the government-based allocation (either from a national or regional government). Other components are financial revenues, donations at the institutional level, as well as income from the rent of premises and other income like sales from commercial activities and paid services to students. When possible, ETER asks for a distinction between the two.

Core funding is divided into two distinct and non-overlapping categories, i.e. basic government allocation and other core funding. If the breakdown is not available, the total should be entered in the aggregated variable, and the subcategories should receive the code 'm'.

6.11.4.3 Basic government allocation

	Type of variable	Data source
Basic government allocation	numeric, national currency	NSA

Core government allocation is defined as the amount of money transferred to the HEI by the government for its primary operations. In most cases, the use of this money is at the

discretion of the HEI. Both contributions from the national and federal governments should be included.

6.11.4.4 Other core funding

	Type of variable	Data source
Other core funding	numeric, national currency	NSA

This variable includes all other components of the HEI core budget, like interests and endowments, donations at the institutional level, as well as income from the rent of premises and other non-identified income.

When only basic government allocation is available and is equal to the total core budget, other core budgets should be coded as 'm'.

6.11.4.5 Third-party funding

	Type of variable	Data source
Third-party funding	numeric, national currency	NSA

Third-party funding is funding earmarked for specific activities and institutional units. In most cases, it is also limited in time. Most of these funds come from public sources, but they constitute a different stream than core funding. Third-party funding specifically includes:

- Grants from national and international funding agencies for research activities, such as research councils. This includes, for example, grants from agencies like the Norwegian Research Council or DFG in Germany, European Union framework programmes, international programmes like Eureka or COST.
- Funds from charities and non-profit organisations for specific research and educational purposes (like the Wellcome Trust of the Bill Gates foundation).
- Contracts from public bodies, non-profit organisations and private companies for specific research and services.
- Fees/payments from companies for educational services and research and service grants from companies.

In practical terms, the third-party funding variable will be the sum of the categories above. Detailed explanations on which items have been included should be provided in the metadata.

Third-party funding is divided into two distinct and non-overlapping categories depending on whether funds come from public and private sources. If the breakdown is not available,

the total should be entered in the aggregated variable, and the subcategories should receive the code 'm'.

6.11.4.6 Public third-party funding

	Type of variable	Data source
Public third-party funding	numeric, national currency	NSA

Public third-party funding includes third-party funding provided by public entities, such as government bodies and research-funding organisations. It includes, for example, funding from national research councils, public administration contracts, and grants from national sources.

6.11.4.7 Private third-party funding

	Type of variable	Data source
Private third-party funding	numeric, national currency	NSA

It includes funding by private entities on contract research and contract education, including private businesses and non-profit organisations, religious organisations, charitable organisations, business and labour associations, and well as households.

6.11.4.8Third-party funding from abroad

	Type of variable	Data source
Third-party funding from abroad	numeric, national currency	NSA

It includes funding from abroad, like funding from international research programs and companies abroad. The amount should therefore correspond to funding from abroad in the EUROSTAT definition.

6.11.4.9 Third-party funding undivided

	Type of variable	Data source
Third-party funding undivided	numeric, national currency	NSA

This variable includes third-party funding that cannot be separated between the above categories, particularly revenues items where the distinction between public and private is not possible (revenue items like 'other contracts for education'). Therefore, this variable should be 'm' (including a flag and note).

6.11.4.10 *Student fees funding*

	Type of variable	Data source
Student fees funding	numeric, national currency	NSA

The amount of money the institutions get from student fees paid by households and students to higher education institutions for participation in educational programmes (see UOE manual). Conforming to the manual, these include:

- tuition fees,
- other fees charged for educational services and other services, and
- fees paid for other welfare services furnished to students by the institution.

Payments are reported as net amounts after subtracting any scholarships or other forms of financial aid (such as reductions in tuition fees or waivers of fees) provided to students by the educational institutions themselves. Other financial aid to students from governments or privates is not netted out.

6.11.4.11 *Tuition fees*

	Type of variable	Data source
Tuition fees	nominal (0=no fees, 1=partial fees, 2=fees for all students)	NSA (preferred)/National Experts (optional)

This variable registers if the institution charges tuition fees for all, some or none of the students. It takes three values:

- no fees,
- fees for some students (e.g., Sweden has tuition fees for foreign students but not for national), and
- fees for all students.

6.11.4.12 *Revenue unclassified*

	Type of variable	Data source
Revenue unclassified	numeric, national currency	NSA

All revenues, which were not assigned to the above revenues, should be indicated as unclassified. When all revenues can be assigned to the provided categories, this cell should contain a '0'; when a breakdown is not available at all, this column should be 'm'.

6.11.4.13 *Non-recurrent revenue*

	Type of variable	Data source
Non-recurrent revenue	numeric, national currency	NSA

This variable includes all revenues, which have an extraordinary and non-repeating character. Therefore, this kind of revenue is not included in the operating statement in an accrual accounting system. Therefore, singling out extraordinary revenues is meant to avoid significant changes from year to year and limit comparability problems between accounting systems (as extraordinary revenues are treated in different ways in cash accounting and accrual accounting systems). The most relevant cases of extraordinary revenues are:

- Specific contributions for investment when, for example, the State provides a large amount of money for a facility, which is then capitalized and spent over the year when the construction progresses, are excluded. This should be the case only in cash accounting systems (as in the accrual system, such contributions will be included in the revenues).
- Large capital endowments or large non-recurrent donations (for example, a charity providing a large endowment for a HEI).

Please note: If non-recurrent revenues cannot be separated from current revenues, this variable should be 'm' (including a flag and note) and the current revenues variable flagged with 'd'. If the HEI has accrual accounting, this variable should be 0.

6.12 Personnel

These variables provide information on the personnel employed by higher education institutions. Conforming to the UOE manual, HEI personnel comprises all persons employed by the HEI and whose activities are required for HEI operations. In the 2022 edition, these variables have been extended by including the breakdown of academic personnel by seniority level developed by OECD-INES and EUROSTAT (see below section 6.12.9).

This definition is extensive as it includes all persons:

- Involved in student instructions and R&D activities.
- Providing professional support to students, including both academic support and counselling and social support.
- Involved in the management and administration of the HEI.
- Providing services to the HEI like social services, food and maintenance.

Inclusion criteria are not dependent on the level or duration of employment. Therefore, also temporary personnel should be included generally. However, exceptions might be made (particularly for data on headcount) for personnel employed with a very small percentage. Temporary replacement personnel should be included as well.

Exclusions and special cases. Generally:

- Persons working for HEI subcontractors are only included if the personnel working for the HEI can be distinguished from personnel devoted to other services. This might include, for example, personnel working for outsourced services like the canteen or a technology transfer office, serving only the HEI. Other subcontracted work is excluded.
- Student aids (master's students or PhD students) are included if they have a contractual relationship with the HEI. This applies particularly to PhD students.
- Personnel temporarily not at work (e.g. due to illness or accident, or paternal leave) is included.
- Retired personnel is not included.

6.12.1 Classification of HEI personnel

The classification of HEI personnel closely follows the UOE manual and distinguishes between three categories based on primary or major functions performed by personnel:

- academic personnel, with a further breakdown into senior, intermediate and junior levels,
- teaching and research assistants,
- support and administration personnel.

The sum of personnel categories should be equal to the total personnel.

Academic personnel. Following the UOE manual, academic personnel includes:

- Personnel employed at the tertiary level of education whose primary assignment is instruction and/or research,
- Personnel who hold an academic rank with such titles as a professor, associate professor, assistant professor, instructor, lecturer, researcher or the equivalent of any of these academic ranks,
- Personnel with other titles (e.g. dean, director, associate dean, assistant dean, chair or head of a department), if their principal activity is instruction or research.

Teaching/research assistants. This category includes personnel employed on a full or part-time basis for the primary purpose of assisting academic personnel in classroom or laboratory instruction or in the conduct of research and receiving payment (in cash or in-

kind) for their activity. Personnel in these positions are usually the employed graduate students or other personnel who hold such titles as teaching assistant, teaching associate, teaching fellow, research assistant, or equivalent personnel with other titles.

Support and administration personnel. Support and administration personnel include the following categories:

- Academic support covers all personnel whose primary responsibility is to support the academic programs of students. This personnel classification includes the following types of personnel: guidance counsellors, librarians, educational media specialists, and attendance officers.
- HEI-level management personnel such as the director, administrative director, and head of service.
- HEI-level administrative personnel, including both central-level and department-level personnel.
- Personnel engaged in maintenance and operations, including special services like IT.
- Undergraduate students employed for teaching assistance or research.

Please note: given that it is a source of comparability differences between countries, metadata is provided on the inclusion of medical personnel, particularly whether personnel of the university hospital is included in the data.

6.12.2 Measurement units

Two measurement units are adopted for HEI personnel:

1) *Numbers of personnel in headcounts (HC).*

There are three different ways of computing headcounts. One approach is based on apportioning: personnel having different functions or active in different fields is distributed between functions proportionally to their engagement so that the total HC equals the number of individuals working in the institution. This is the preferred method. It implies that some breakdowns will be fractional even if based on HC.

A second method is based on functions: individuals are counted more than once when they have different functions. This method implies that the HC figures will exceed the number of individuals working in the institutions and, therefore, there will be inconsistencies between breakdowns and totals.

A third method is based on single attribution: individuals are attributed only once to the function or field where they are mostly engaged. With this method, the total HC equals the number of individuals working in the institution.

The choice made has to be indicated in the metadata.

Following UOE conventions, Headcounts are computed on the last day of the calendar year (i.e. only persons who are employed at that date are included). Departures from these rules should be remarked in the metadata.

2) *Personnel activities in full-time equivalents (FTE).*

Full-time equivalents are defined as the number of actual working hours of HEI personnel during a reference period (usually the calendar year) divided by the total number of hours conventionally worked in the same period by a full-time employed individual.

For most purposes, FTE data are the most important measure of HEI personnel, as they measure the total effort performed by personnel in HEI activities. HC data are relevant when observing personnel characteristics referring to individuals, like their gender, nationality and field.

6.12.3 Breakdowns of academic personnel

The following breakdowns are required:

- *Citizenship:* breakdown of academic personnel between national and foreign personnel using citizenship as a criterion. This breakdown is relevant to analyse the internationalisation of HEI.
- *Gender:* breakdown of academic personnel and professors between men and women (with an 'unclassified' option also available). This breakdown is relevant to analyse the gender balance of HEI.
- *Fields of education*: breakdown of academic personnel by fields of education. This breakdown provides the opportunity to analyse differences in the composition of fields of education between institutions and may contribute to explaining observed differences.
- *Seniority level*: breakdown of academic personnel by seniority between senior, intermediate and junior.

Data should be collected in both full-time equivalents (FTE) and headcounts (HC). However, breakdowns by citizenship, gender and fields of education are required only for

Headcounts. Additionally, data are collected for a specific subcategory of academic personnel, i.e. full professors.

Table 28 Variables included in personnel data

Variable	Data source
Total academic personnel (FTE and HC)	NSA
Academic personnel in HC by gender, citizenship, fields of education	NSA
Academic personnel in HC by seniority level and gender	NSA
Teaching and research assistants (FTE and HC)	NSA
Support and administration personnel (FTE and HC)	NSA
Total personnel (FTE and HC)	NSA

6.12.4 Total number of academic personnel (FTE)

	Type of variable	Data source
Total number of academic personnel (FTE)	numeric	NSA

The total number of academic personnel employed by the HEI during the reference period in full-time equivalents.

6.12.5 Total number of academic personnel (HC)

	Type of variable	Data source
Total number of academic personnel (HC)	numeric	NSA

The total number of academic personnel employed by the HEI during the reference period in Headcounts (referring to the last day of the calendar year).

6.12.6 Academic personnel by gender (HC)

	Type of variable	Data source
Academic personnel by gender (HC)	numeric	NSA

A breakdown between men and women in HC. If this breakdown is not possible, please use the appropriate special code (chapter 5.5).

6.12.7 Academic personnel by citizenship (HC)

	Type of variable	Data source
Academic personnel by citizenship (HC)	numeric	NSA

A breakdown between national and foreign personnel in HC. Employees are considered foreign if they do not have the citizenship of the country for which data is collected.
If this breakdown is not possible, please use the appropriate special code (chapter 5.5).

6.12.8 Academic personnel by Field of education (HC)

	Type of variable	Data source
Academic personnel by Field of education (HC)	numeric	NSA

Academic personnel in HC should be divided by the fields of education they work in by using the Field of Education and Training classification (see chapter 5). If this breakdown is not possible, please use the appropriate special code (chapter 5.5).

6.12.9 Academic personnel by seniority level (HC)

Academic personnel is classified by seniority level following the classification proposed by OECD-INES in cooperation with EUROSTAT.[13] While academic titles might be used as proxies, seniority levels refer to the functions assumed within the HEI and, therefore, might correspond to different academic titles by country and/or sector. In principle, countries should use the same correspondence table between national categories and seniority levels as in the OECD pilot and remark any modifications.

Given its importance for the study of gender differences in careers, for each level, a breakdown is requested by gender, i.e. women, men, and unclassified.

Remark: given that there is currently no consensus on the classification of employed PhD candidates, these are not singled out in a separate category but, depending on how countries interpret their functions, might be included in junior academic staff or in research and teaching assistants variable. A specific question is included in the metadata.

[13] EDU/EDPC/INES/WP(2022)5.

	Type of variable	Data source
Number of senior academic personnel (HC) by gender	numeric	NSA

Senior academic staff is defined as the highest grades/posts for academic staff pursuing an academic career in either instruction or research. Staff allocated to this category must hold similar qualifications, pay range and level of responsibilities, although the nature of their responsibilities may differ. It is possible to have one grade/post per career track if relevant (i.e. if the tracks are clearly separate).
Examples are full professor, Professeur titulaire et corps assimilés (FRA), director of research.

Remark: for the years 2011-2019, the variable 'number of full professors' has been collected with a very similar definition. Hence, values have been copied automatically in the new senior academic personnel variable (but might be revised when a full breakdown by seniority level is provided).

	Type of variable	Data source
Number of intermediate academic personnel (HC) by gender	numeric	NSA

Academic staff pursuing an academic career working in positions not as senior as the top position but more senior than entry-level positions.
Examples are associate professor, Maître de conférences (FRA), and senior researcher.

	Type of variable	Data source
Number of junior academic personnel (HC) by gender	numeric	NSA

Entry grades/posts into which an individual would normally be recruited to begin their academic career. Staff allocated to this category must hold similar qualifications, pay range and level of responsibilities, although the nature of their responsibilities may differ. It is possible to have one grade/post per career track if relevant (i.e. if the tracks are clearly separate).
Examples include assistant professor, lecturer, Professeurs agrégés (FRA), junior researcher, and post-doctoral researcher.

	Type of variable	Data source
Number of other academic personnel by gender	numeric	NSA

Instructional and research personnel who are not considered to be on the academic career track. This excludes doctoral candidates and teaching and research assistants. Examples include adjunct professors or fellows.

	Type of variable	Data source
Number of unclassified academic personnel by gender	numeric	NSA

The number of academic personnel that cannot be classified by seniority level. The gender breakdown is requested for consistency reasons.
Remark: if the seniority breakdown is not available, this variable should be coded as 'm'.

6.12.10 Total number of research and teaching assistants (FTE)

	Type of variable	Data source
Total number of research and teaching assistants (FTE)	numeric	NSA

The total number of research and teaching assistants employed by the HEI during the reference period in full-time equivalents.

6.12.11 Total number of research and teaching assistants (HC)

	Type of variable	Data source
Total number of research and teaching assistants (HC)	numeric	NSA

The total number of research and teaching assistants employed by the HEI during the reference period in headcounts (referring to the last day of the calendar year).

6.12.12 Total number of support and administrative personnel (FTE)

	Type of variable	Data source
Total number of support and administrative personnel (FTE)	numeric	NSA

The total number of support and administrative personnel employed by the HEI during the reference period in full-time equivalents.

6.12.13 Total number of support and administrative personnel (HC)

	Type of variable	Data source
Total number of support and administrative personnel (HC)	numeric	NSA

The total number of support and administrative personnel employed by the HEI during the reference period in headcounts (referring to the last day of the calendar year).

6.12.14 Total number of personnel (FTE)

	Type of variable	Data source
Total number of personnel (FTE)	numeric	NSA

The total number of personnel employed by the HEI during the reference period in full-time equivalents.

6.12.15 Total number of personnel (HC)

	Type of variable	Data source
Total number of personnel (HC)	numeric	NSA

The total number of personnel employed by the HEI during the reference period in headcounts.

6.13 Erasmus mobility data

Table 29 Variables included in Erasmus mobility data

Variable	Data source
Erasmus Code 2014-2020	Erasmus Executive Agency
Erasmus Code 2021-2027	Erasmus Executive Agency
Erasmus incoming students by ISCED level	DG EAC Open Data Portal
Erasmus outgoing students by ISCED level	DG EAC Open Data Portal
Erasmus incoming personnel	EUPRO
Erasmus outgoing personnel	EUPRO

These variables provide information on the student and personnel mobility supported by the Erasmus program. They are based on data from the Education, Audiovisual and Culture Executive Agency of DG EAC.

The data source for the years 2011-2013 (the academic years 2011/2012 to 2013/2014) are micro-data published on the open data portal of the European Commission: http://data.europa.eu/euodp/en/data/dataset/erasmus-mobility-statistics-2013-14

Data for 2014-2016 have been delivered directly by the EAC Executive Agency and refer to the completed mobility.

From 2017 (2017/2018 academic year) onwards, data for student mobility are again incorporated from the open data portal of the European Commission: https://data.europa.eu/data/datasets?query=erasmus&locale=en&dataScope=eu&country=eu&page=1

Remark: For Switzerland, data refer to the Swiss mobility scheme replacing Erasmus and are delivered directly by Movetia, the national agency for the promotion of exchanges and mobitliy in the education system (https://www.movetia.ch/en/).

Data for personnel mobility are incorporated from the EUPRO database (https://rcf.risis2.eu/dataset/4/metadata).

The call year has been used as a reference year. This implies some shifts, as selected mobility are generally divided between years (e.g. 2015 mobility takes place mostly in the academic years 2015/2016 and 2016/2017).

6.13.1 Erasmus charter code 2014-2020

	Type of variable	Data source
Erasmus charter code 2014-2020	text	Erasmus Executive Agency

The Erasmus charter code is valid for the period 2014-2020 of the ETER HEI as attributed by the Erasmus Executive Agency.
For ETER HEIs that could not be identified in the Erasmus program, the code 'a' is used since they probably did not participate. This also applies to all HEIs in the countries, which did not participate in the program.

6.13.2 Erasmus charter code 2021-2027

	Type of variable	Data source
Erasmus charter code 2021-2027	text	Erasmus Executive Agency

The Erasmus charter code is valid for the period 2021-2027 of the ETER HEI as attributed by the Erasmus Executive Agency.

For ETER HEIs that could not be identified in the Erasmus program, the code 'a' is used since they probably did not participate. This also applies to all HEIs in the countries which did not participate in the program.

6.13.3 Erasmus incoming students

	Type of variable	Data source
Number of incoming Erasmus students at ISCED 5 level	numeric	DG EAC Open Data Portal
Number of incoming Erasmus students at ISCED 6 level.	numeric	DG EAC Open Data Portal
Number of incoming Erasmus students at ISCED 7 level	numeric	DG EAC Open Data Portal
Number of incoming Erasmus students at ISCED 8 level	numeric	DG EAC Open Data Portal
Number of incoming Erasmus students at unclassified ISCED level	numeric	DG EAC Open Data Portal
Total number of incoming Erasmus students	numeric	DG EAC Open Data Portal

The number of Erasmus students hosted by the HEI during the respective academic year. This includes students' mobility at ISCED levels 5 to 8, disaggregated by ISCED level, as well as totals for all levels.

When the HEI has no Erasmus code, the code 'a' is used. When the HEI does not offer a level of education, the code 'a' is used.

6.13.4 Erasmus outgoing students

	Type of variable	Data source
Number of outgoing Erasmus students at ISCED 5 level	numeric	DG EAC Open Data Portal
Number of outgoing Erasmus students at ISCED 6 level.	numeric	DG EAC Open Data Portal
Number of outgoing Erasmus students at ISCED 7 level	numeric	DG EAC Open Data Portal
Number of outgoing Erasmus students at ISCED 8 level	numeric	DG EAC Open Data Portal
Number of outgoing Erasmus students at unclassified ISCED level	numeric	DG EAC Open Data Portal
Total number of outgoing Erasmus students	numeric	DG EAC Open Data Portal

The number of Erasmus students sent out by the HEI during the respective academic year. This includes students' mobility at ISCED levels 5 to 8, disaggregated by ISCED level, as well as totals for ISCED 5-8.

When the HEI has no Erasmus code, the code 'a' is used. When the HEI does not offer a level of education, the code 'a' is used.

6.13.5 Erasmus personnel mobility

	Type of variable	Data source
Erasmus incoming staff	numeric	EUPRO
Erasmus outgoing staff	numeric	EUPRO

The number of support mobility of personnel by the Erasmus program.

6.14 Quality assurance variables from DEQAR

These variables provide information on the number and characteristics of quality assurance reports to which HEIs in ETER have been subjected. It is based on information collected from national quality assurance agencies in the Database of External Quality Assurance results (DEQAR; https://www.eqar.eu/qa-results/search/by-institution/). DEQAR has been implemented by the European Quality Assurance Register for Higher Education (EQAR), i.e. the official register of such agencies, listing those that work in-line with an

agreed common framework to ensure the quality of higher education institutions and study programmes.

Importantly, the absence of reports does not imply that the HEI has not been subject to quality assurance, as the responsible agency might not be a member of EQAR or reports have not been uploaded. For detailed information on country-level quality assurance, the reader should refer to the DEQAR knowledge base (https://www.eqar.eu/kb/country-information/).

All variables for this dimension are computed by ETER from the DEQAR database. Therefore, these variables are not included in the NSA data collection.

Table 30 Variables included in DEQAR data

Variable	Data source
DEQAR identifier	DEQAR
Institutional external quality assurance	DEQAR
Programme external quality assurance – ISCED 5	DEQAR
Programme external quality assurance – ISCED 6	DEQAR
Programme external quality assurance – ISCED 7	DEQAR
Programme external quality assurance – ISCED 8	DEQAR
Joint programme quality assurance	DEQAR
Cross-border external QA	DEQAR

6.14.1 DEQAR Identifier

	Type of variable	Data source
DEQAR identifier	string, multiple identifies are separated by a semicolon (;)	DEQAR

DEQARINST(+four digits) (e.g. DEQARINST0584). The identifiers in the DEQAR database associated with the ETERID. There may be multiple identifiers since, in some cases, quality assurance is referred to the department or faculty level. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.2 Institutional external quality assurance

	Type of variable	Data source
Institutional external quality assurance	nominal (0=not external quality assured (including negative decisions in DEQAR), 1=externally quality assured)	DEQAR

Whether the HEI was externally quality assured at the institutional level with at least a report valid on the last day of the ETER reference year. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.3 Program external quality assurance ISCED5

	Type of variable	Data source
Program external quality assurance ISCED5	numeric	DEQAR

Number of externally quality assured programs per ISCED5 (short cycle) valid on the last day of the ETER reference year. Excluding programs with negative decisions in DEQAR. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.4 Program external quality assurance ISCED6

	Type of variable	Data source
Program external quality assurance ISCED6	numeric	DEQAR

Number of externally quality assured programs per ISCED6 (bachelor and equivalent) valid on the last day of the ETER reference year. Excluding programs with negative decisions in DEQAR. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.5 Program external quality assurance ISCED7

	Type of variable	Data source
Program external quality assurance ISCED7	numeric	DEQAR

Number of externally quality assured programs per ISCED7 (master or equivalent) valid on the last day of the ETER reference year. Excluding programs with negative decisions in DEQAR. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.6 Program external quality assurance ISCED8

	Type of variable	Data source
Program external quality assurance ISCED8	numeric	DEQAR

Number of externally quality assured programs per ISCED8 (PhD) valid on the last day of the ETER reference year. Excluding programs with negative decisions in DEQAR. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.7 Joint program external quality assurance

	Type of variable	Data source
Joint program external quality assurance	numeric	DEQAR

Number of externally quality assured programs that involved joint programs, i.e. programs in an integrated curriculum coordinated and offered jointly by different higher education institutions from EHEA countries, leading to double/multiple degrees or a joint degree (all levels) valid on the last day of the ETER reference year. Most of these programs will also be counted in the ISCED-level variables. Excluding programs with negative decisions in DEQAR. If the institution is not reported in DEQAR, this variable should be 'm'.

6.14.8 Cross-border external quality assurance

	Type of variable	Data source
Cross-border external quality assurance	numeric	DEQAR

Number of externally quality assured programs, where the assessment was conducted by an agency outside the HEI country of establishment, valid on the last day of the ETER reference year. Excluding programs with negative decisions in DEQAR. Most of these programs will also be counted in the ISCED-level variables. If the institution is not reported in DEQAR, this variable should be 'm'.

7 Indicators in ETER

Please note that the indicators chapter is provisional; a revision is foreseen later in the project based on additional variables and breakdowns.

In addition to the variables described in chapter 6, the ETER database includes indicators that characterise HEIs across important dimensions, like gender balance, mobility and internationalisation, and research orientation. The indicators highlight traits and characteristics in HEIs that are often unavailable through other sources.

Moreover, ETER includes *HEI classification indicators* that divide the ETER HEIs in groups with similar basic characteristics, such as size, educational intensity, etc. These are meant to allow users to construct groups of HEIs, for example for comparative purposes.

The choice of directly integrating indicators in the database is motivated by robustness considerations: the indicators include those that the ETER consortium considers sufficiently robust and available for a sufficient number of countries. Moreover, the methodology for calculating the indicators has been carefully considered, taking into account data specificities.

This chapter presents the rationale for selecting indicators, the current list of indicators and the rules for their calculation.

7.1 General approach

There are several reasons for selecting the specific set of indicators used in ETER. Firstly, the indicators presented contribute valuable information on HEIs, as they can be used to characterise relevant dimensions of HEIs. In addition, the indicators are selected based on the availability of data in the database and the robustness of the indicator itself.

All indicators are available at the institutional level, and it is possible to compare different types of institutions within a country or between countries. It is further possible to select a certain type of institution, either by size or by type (for example, private/public or university/university of applied science) and compare this subgroup. This makes the ETER indicators unique compared to official statistics from EUROSTAT, collected at the national level. The indicators discussed in this document are all derived from the variables present in the ETER database.

The indicator must be available for a significant number of HEIs and countries in order to provide meaningful results. Currently, ETER has good coverage of variables on students (ISCED levels 6-8), graduates (ISCED levels 6-8) and personnel (except for the breakdown of personnel by field). However, coverage of ISCED level 5 is quite limited; therefore, no indicators are provided at this level.

Approximately two-thirds of the countries have delivered data on revenues and expenditure, and a half for R&D expenditure. Therefore, in the initial phase of ETER, we suggest only using data from the areas where the coverage is good for the calculation of indicators. This implies that we currently have a few indicators related to research and more indicators related to gender balance and teaching.

Economic indicators, like costs per student, are problematic methodologically for a few different reasons: first, costs per student vary greatly by subject domain and HEI mission and, therefore, comparing aggregated HEI-level values might lead to misleading interpretations in terms of efficiency. Second, even when financial data are compared across countries using purchasing power parities, these do not fully account for the different cost structures in higher education.

When counting individual HEIs rather than countries, the situation does not change substantially: breakdowns of students by gender, citizenships and field of study are available in most cases, slightly less so for mobility. However, the situation is more problematic when considering breakdowns of personnel data (even when aggregates are available in most cases) and for revenues and R&D expenditures.

7.2 List of indicators

Table 31 overleaf provides the list of indicators currently included in ETER. Most of these indicators are related to students, graduates, PhD-students, PhD-graduates and personnel in HEIs since these are the types of variables where most countries have delivered data.

Table 31 List of indicators in the ETER database by dimension

Dimension	Variable	N.
Gender (share of women)	Students (ISCED 6 and 7)	2
	Graduates (ISCED 6 and 7)	2
	PhD-students	1
	PhD-graduates	1
	Academic personnel	1
	Full professors	1
Mature students (>24 years old)	Students (ISCED 6 and 7)	2
	Graduates (ISCED 6 and 7)	2
Nationality	Students (ISCED 6 and 7)	2
	Graduates (ISCED 6 and 7)	2
	PhD-students	1
	PhD-graduates	1
	Academic personnel	1
Degree mobility	Students (ISCED 6 and 7)	2
	Graduates (ISCED 6 and 7)	2
	PhD-students	1
	PhD-graduates	1
Credit mobility	Incoming Erasmus students	1
	Outgoing Erasmus students	1
Researchers' mobility	Researchers' mobility intensity	1
Subject mix (Herfindahl index)	Students (ISCED 5-7)	1
	PhD-graduates	1
	Personnel	1
STEM orientation	Share of students and graduates in science and technology	2
Degree orientation	Master orientation	1
	PhD-intensity	1
Personnel	Full professors, as a share of academic personnel (HC)	1
	Academic personnel (FTE), as a share of total personnel	1
Revenues	Total core budget as a share of total current revenues	1
	Total third-party funding as a share of total current revenues	1
	Student fees funding as a share of total current revenues	1
HEI classification indicators	Institutional control	1
	PhD status	1
	HEI size	1

	Education intensity	1
	Master degree orientation	1
	PhD intensity	1
	Subject concentration	1
	STEM specialisation	1

7.3 Calculation of indicators

7.3.1 Overall rules and special cases

While specific formulas for each indicator are provided below, a number of general rules apply.

First, unclassified values are not included in the calculation of indicators. For example, if the gender is unknown, the share of women will be calculated as the share of women among men and women, where any unclassified person (missing) is excluded. A similar case is when the nationality is unknown. In these cases, the share of foreign personnel will only be calculated as the share of foreign nationals among foreign nationals and nationals. However, indicators computed from breakdowns where the share of unclassified cases exceeds 25% will be flagged with 'ic'. Any flags from the parent variables are copied over for the indicators in order to make users aware of possible data problems.

Second, for countries that have reported restricted data, the indicators using that kind of data will be set to restricted as well (special code 'c') and coded as 'c' in the publicly available dataset. Additionally, indicators that are calculated using small values coded into 's' are also coded into 's' in order to prevent recalculation.

Flags related to data will also be applied to the indicators using that data as follows:

- When data used for calculating an indicator is missing (code 'm'), the indicator will be set to missing.
- When a total is '0', the value for the indicator referring to a breakdown will be set to 'a' (not applicable). For example, this applies when a HEI has '0' students at a specific ISCED level.

All indicators are calculated for individual years and HEIs.

7.3.2 Gender

Indicators concerning gender are computed as follows:

$$share\ of\ women = \frac{number\ of\ women}{number\ of\ women + number\ of\ men}$$

'Unclassified' is not included in the denominator.

These indicators are computed for the following variables in the ETER database:

- Share of women students ISCED 6. Students at level ISCED 6.
- Share of women students ISCED 7. Students at level ISCED 7 (including ISCED 7 long degrees).
- Share of women students ISCED 8. Students at level ISCED 8. For HEIs not delivering the PhD, this indicator is set to 'a' (not applicable).
- Share of women graduates ISCED 6. Graduates at level ISCED 6.
- Share of women graduates ISCED 7. Graduates at level ISCED 7 (including ISCED 7 long degrees).
- Share of women graduates ISCED 8. Graduates at level ISCED 8. For HEIs not delivering the PhD, this indicator is set to 'a' (not applicable).
- Share of women academic staff. Academic personnel in Headcounts.
- Share of women full professors. Professors in Headcounts.

These indicators are highly relevant to analyse gender balance and, specifically, how it differs by activity and level and by type of HEIs.

7.3.3 Mature students and graduates

By definition, mature students and graduates are those above 24 years; their share is an important indicator of the participation of adult learners to higher education. They are computed as follows:

$$share\ of\ mature\ students/graduates = \frac{number\ of\ students\ /\ graduates\ above\ 24\ years}{total\ number\ of\ students\ /\ graduates}$$

'Unclassified' is not included in the denominator.

These indicators are computed separately for ISCED5 and ISCED7 (including 7 long) students and graduates.

7.3.4 Share of foreign students, graduates and personnel

Indicators concerning citizenship are computed as follows:

$$share\ of\ foreign\ students/graduates/personnel = \frac{number\ of\ foreign\ citizens}{number\ of\ foreign\ citizens + number\ of\ nationals}$$

'Unclassified' is not included in the denominator.

These indicators are computed for the following variables in the ETER database:

- Share of foreign students ISCED 6. Students at level ISCED 6.
- Share of foreign students ISCED 7. Students at level ISCED 7 (including ISCED 7 long degrees).
- Share of foreign students ISCED 8. Students at level ISCED 8. For HEIs not delivering the PhD, this indicator is set to 'a' (not applicable).
- Share of foreign graduates ISCED 6. Graduates at level ISCED 6.
- Share of foreign graduates ISCED 7. Graduates at level ISCED 7 (including ISCED 7 long degrees).
- Share of foreign graduates ISCED 8. Graduates at level ISCED 8. For HEIs not delivering the PhD, this indicator is set to 'a' (not applicable).
- Share of foreign academic staff. Academic personnel in Headcounts.

These indicators are relevant to analyse the internationalisation process of HEIs and its determinants depending on level and activity. Please note that the availability of indicators on foreign academic personnel is rather limited.

7.3.5 Share of mobile students and graduates

Indicators concerning mobility are based on the EUROSTAT definition of place of prior education (providing access to tertiary education). They provide a more precise measure of student mobility, which allows for excluding foreign citizens born in the country.

$$share\ of\ mobile\ students = \frac{number\ of\ mobile\ students}{number\ of\ mobile\ students + number\ of\ resident\ students}$$

'Unclassified' is not included in the denominator.

These indicators are computed for the following variables in the ETER database:

- Share of mobile students ISCED 6. Students at level ISCED 6.

- Share of mobile students ISCED 7. Students at level ISCED 7 (including ISCED 7 long degrees).
- Share of mobile students ISCED 8. Students at level ISCED 8. For HEIs not delivering the PhD, this indicator is set to 'a' (not applicable).
- Share of mobile graduates ISCED 6. Graduates at level ISCED 6.
- Share of mobile graduates ISCED 7. Graduates at level ISCED 7 (including ISCED 7 long degrees).
- Share of mobile graduates ISCED 8. Graduates at level ISCED 8. For HEIs not delivering the PhD, this indicator is set to 'a' (not applicable).

These indicators are relevant to analyse the internationalisation process of HEIs and their determinants depending on level and activity.

7.3.6 Credit mobility intensity

These two indicators measure the number of students involved in credit mobility supported by the Erasmus+ programme as compared to the overall volume of education.
They are computed as follows:

$$Credit\ mobility\ incoming = \frac{Number\ of\ incoming\ Erasmus\ students\ ISCED\ 5-7}{Total\ number\ of\ students\ ISCED\ 5-7}$$

$$Credit\ mobility\ outgoing = \frac{Number\ of\ outgoing\ Erasmus\ students\ ISCED\ 5-7}{Total\ number\ of\ students\ ISCED\ 5-7}$$

If the Erasmus code is 'a', these two indicators should equally be code as 'a'.

7.3.7 Researchers' mobility intensity

This indicator measures the (incoming) mobility of researchers supported by the EU-FP program. It is computed as follows:

$$Researchers'mobility\ intensity = \frac{Number\ of\ EU-FP\ researchers'mobility}{FTE\ academic\ staff}$$

7.3.8 Subject concentration in education and graduate education

These indicators characterise the level of specialisation of HEIs education by field. The indicator is calculated as a Herfindahl index of the distribution of students and graduates by field of education.

$$Herfindahl = \frac{1}{n^2} * \sum_{1}^{11} n_j^2$$

Where n_j^2 is the number of students or graduates in field j, and n is the total number of students or graduates for that level within the HEI. The sum runs over the 11 Fields of Education and Training. The index runs from 1, when all students are in the same field, to 0.1 when the students are equally distributed across fields in a HEI.

- Herfindahl index students ISCED 5-7. The indicator is calculated using the distribution of the total number of students at levels ISCED 5-7 by field of education and training (ISCED-F).
- Herfindahl index PhD graduates. The indicator is computed using the distribution of PhD graduates (ISCED 8) by field of education and training (ISCED-F). Graduates are preferred to students, as data are considered more reliable. Since PhD education is closely associated with research, this can also be considered as an indicator of subject concentration in research.
- Herfindahl index academic staff. The indicator is calculated using the distribution of the total number of academic personnel by field of education and training (ISCED-F).

When data are computed using ISCED-1997 FOET (8 fields), a flag is added ('d') and a remark is included.

7.3.9 STEM orientation

These indicators provide a proxy of the orientation of education towards science and technology based on the classification of students in the Fields of Education and Training.

More precisely, they are computed as follows:

$$STEM\ orientation\ (students) = \frac{\begin{array}{c} number\ of\ students\ ISCED\ 5-7\ in \\ fields\ 05\ (Natural\ sciences, mathematics\ and\ statistics), \\ 06\ (Information\ and\ communication\ technologies) \\ and\ 07\ (Engineering, manufacturing\ and\ construction) \end{array}}{Total\ number\ of\ students\ ISCED\ 5-7}$$

$$STEM\ orientation\ (graduates) = \frac{\begin{array}{c} number\ of\ graduates\ ISCED\ 5-7\ in \\ fields\ 05\ (Natural\ sciences, mathematics\ and\ statistics), \\ 06\ (Information\ and\ communication\ technologies) \\ and\ 07\ (Engineering, manufacturing\ and\ construction) \end{array}}{Total\ number\ of\ graduates\ ISCED\ 5-7}$$

7.3.10 Master degree orientation

This is an indicator frequently used to characterise the level of orientation of an HEI to master-level education, which is usually associated with a stronger research orientation.

It is computed as follows:

$$Master\ degree\ orientation = \frac{number\ of\ graduates\ at\ levels\ ISCED\ 7 \text{ and } 7 \text{ long}}{number\ of\ graduates\ at\ ISCED\ levels\ 5,6,7\ and\ 7\ long}$$

When the HEI does not deliver master's degrees, this indicator is set to 'a'.

7.3.11 PhD intensity

This indicator is frequently used to characterise the HEI's level of orientation towards research in relation to the volume of educational activities at the undergraduate level.

It is computed as follows:

$$PhD\ intensity = \frac{number\ of\ graduates\ at\ level\ ISCED\ 8}{number\ of\ graduates\ at\ ISCED\ levels\ 5,6,7\ and\ 7\ long}$$

When the HEI does not deliver the doctorate or when an HEI awards only PhD degrees, this indicator is set to 'a'.

7.3.12 Indicators related to personnel composition

These indicators provide information on the personnel structure of HEIs, which is relevant to analyse their internal organisation.

Two indicators are included in ETER.

$$ShareAcadStaff = \frac{number\ of\ academic\ staff\ (FTE)}{number\ of\ academic\ staff\ (FTE) + number\ of\ administrative\ staff\ (FTE)}$$

Is a relevant indicator to analyse the weight of administration for HEIs.

$$ShareProf = \frac{number\ of\ professors\ (HC)}{number\ of\ academic\ staff\ (HC)}$$

It is relevant to analyse the academic personnel structure, also depending on the country and type of HEI.

7.3.13 Indicators related to revenues

Three complementary indicators are provided for the share of revenue sources of HEIs. Cases, where the share of unclassified revenues exceeds 25% will be flagged.

$$ShareCore = \frac{Total\ core\ budget}{Total\ current\ revenues}$$

$$ShareThirdParty = \frac{Total\ \ third\ party\ funding}{Total\ \text{current}\ revenues}$$

$$ShareFees = \frac{Student\ \text{fees}\ funding}{Total\ \text{current}\ revenues}$$

7.3.14 Classification indicators

Classification indicators in ETER allow classifying in groups HEIs included in ETER according to some core dimensions. There are useful for the identification of groups of HEIs. Moreover, classification indicators can be integrated in datasets, such as those from EUROGRADUATES and EUROSTUDENTS, where HEI-level individual data cannot be disclosed because of requirements of anonymity.

The basis of the indicators are the main dimensions identified in recent HEI classifications, as summarised in Lepori, Benedetto (2022), The heterogeneity of European Higher Education Institutions: a configurational approach. *Studies in Higher Education*, 47(9), 1827-1843. https://www.tandfonline.com/doi/abs/10.1080/03075079.2021.1968368.

Classification indicators build on available indicators in ETER and categorise them using the distributions of variables so that the size of each data cell remains large enough to guarantee anonymisation.

The table overleaf provides the list of indicators and how they are computed and categorised for continuous variables.

Since most variables are highly skewed, the choice of cut points takes into account the quantiles (to have good numerosity by class), adapted however to provide a better

coverage of the large institutions – implying that generally large classes will include less HEIs.

Missing data are completed by interpolation at two conditions:

- at least ¾ of the yearly values should be available.
- categories are stable across years.

This procedure reduces somewhat the missing values in the dataset, particularly when a single year is missing in the dataset.

Dimensions	Variable	Definition and remarks	Computation in ETER	Categorization
Institutional variables	Institutional control	Dummy, 0 if the institution is under public control or is mostly financed by the state, 1 if it is private and mostly funded by private sources.	Recoded from the legal status variable by including private government-dependent in public.	0 Public control 1 Private control
	PhD status	Dummy, 1 if the HEI has the legal right to award the PhD, 0 otherwise.	Recoded from the highest degree variable, 1 if highest degree = 3, 0 otherwise	0 No PhD 1 PhD
HEI size	HEI size	Number of academic personnel FTE	Already available. When only HC are available, FTEs estimated from HC to minimise missing data.	1 below 100 FTEs 2 below 500 FTEs 3 below 1,500 FTEs 4 more or equal 1,500 FTEs
Educational activities	Education intensity	Number of diploma, bachelor and master students ISCED5-7 divided by academic staff		1 below 10 2 below 25 3 below 50 4 50 or above
	Master degree orientation	Number of master students divided by the number of diploma, bachelor and master students.	Already available in ETER	1 below 0.30 2 below 0.50 3 below 0.70 4 more or equal to 0.70

Research activities	PhD intensity	Number of ~~PhD students divided by students~~		~~1 below 0.005~~ 2 below 0.03 3 below 0.08 4 more or equal to 0.08
Subject scope	Subject concentration	The index is computed as the sum of the squares of the share of bachelor and master students in each of the ten subject fields of educational statistics (Herfindahl concentration index). It ranges between 1 (all students in a single field) to 0.1 (students equally distributed between fields).	Already available in ETER	1 below 0.30 2 below 0.65 3 more or equal to 0.65
	Relative specialisation in natural sciences and engineering	Share of bachelor and master students in the corresponding fields normalised by the average share in the whole sample.		1 below 0.15 2 below 0.35 3 below 0.60 4 more or equal 0.60
	Main educational field	The field with the largest number of students ISCED5-7	Computed	Using ISCED-F classification 0 = generic 1 = education 2 = arts and humanities 3 = social sciences 4 = business, administration and law 5 = natural sciences

				6 = information and communication technologies 7 = engineering 8 = agriculture 9 = health and welfare 10 = services 11 = unknown

8 The ETER data quality process

8.1 General introduction and relevance of data quality[14]

Data quality is a relevant issue in any data collection, but it is often underestimated in practical data collection procedures except for the Official Statistics field, in the context of which data undergo a well-defined correction and control procedure.

Data quality is a multidimensional concept including several distinct dimensions (Batini & Scannapieco, 2006), and a number of different definitions have been proposed in various contexts (Scannapieco & Catarci, 2002). It can be intended as 'the distance between the data views presented by an information system and the same data in the real world'. Such a definition can be seen as an 'operational definition', although evaluating data quality based on comparison with the real world is a very difficult task.

Data accuracy is a crucial dimension of data quality: it measures the closeness between a value v and a value v', considered the correct representation of the real-life phenomenon that v is intended to represent. However, quality is more than simply data accuracy. Other significant dimensions, such as, for example, completeness, consistency, and timeliness (i.e. degree of up-to-datedness), play an essential role in the definition of the data quality concept.

Some international standards for defining data quality concepts and related dimensions have been proposed. ISO 25012 introduces and defines three possible levels (views) of data quality to be considered individually, namely:

- Internal Data Quality, related to values and formats of data (e.g., consistency, completeness);
- External Data Quality, related to characteristics of the software and hardware used to store and access data (e.g., response time, portability);
- Data Quality in Use, related to the final user of data (e.g., effectiveness, level of satisfaction).

Internal quality control is a major issue for HEI-level international data collection, as, on the one hand, the lower level of disaggregation of data makes them more sensitive and

[14] This chapter includes contributions from Renato Bruni, Giuseppe Catalano e Giorgio Matteucci from Sapienza University in Rome.

increases the chances of mistakes. On the other hand, there are many different sources of comparability problems between HEIs and countries.

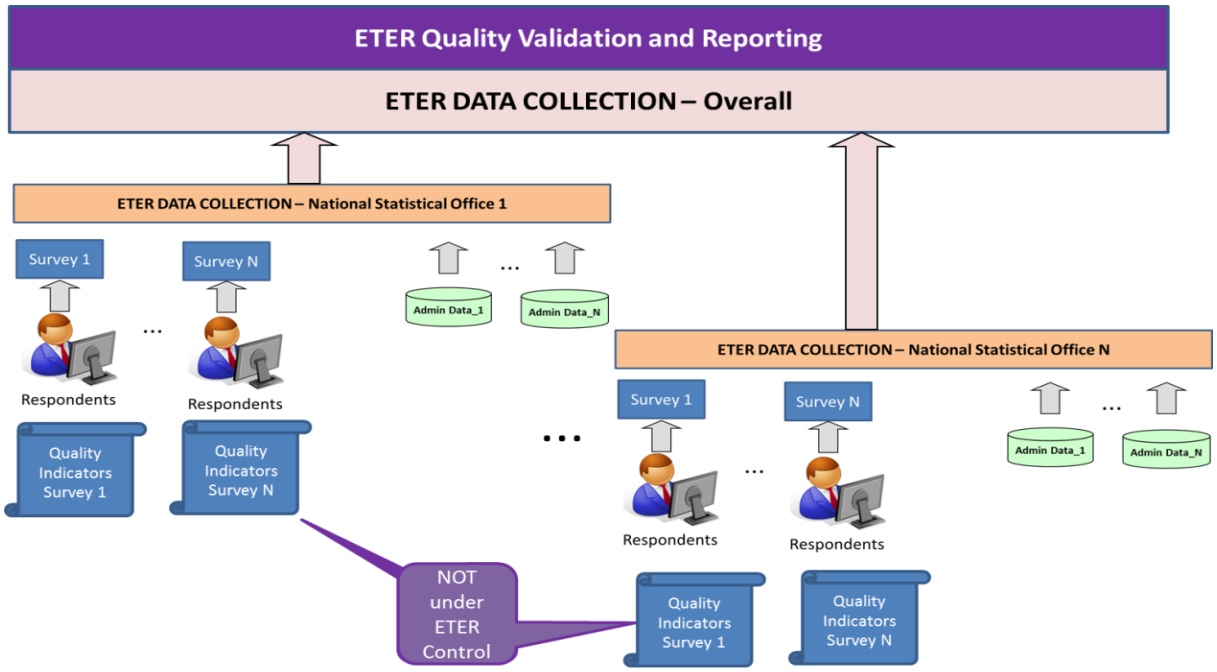
In addition, ETER data collection is performed by integrating data already collected by statistical institutions by means of different surveys or administrative data. This means that the 'survey design' phase of a classical statistical production process cannot be applied in this case to the 'overall' process, and, consequently, the proposed quality indicators cannot be reused directly in ETER quality validation and reporting.

For these reasons, quality validation and reporting on data collected by ETER are performed as follows:

- A dedicated quality validation activity is performed on collected data, independent of the survey-specific quality characterisation, but instead based on a newly proposed quality characterisation described below in section ☐.
- Where data available at National Statistical Offices can be accompanied by statistical process quality reporting (like prescribed by EUROSTAT (2007)), such quality reporting is part of a detailed quality characterisation of the 'source data' to the ETER process.

The overall process is depicted in Figure 8.

Figure 8. Overall ETER data collection and quality management



This chapter describes in detail the ETER quality validation and reporting process. Firstly, a set of operational steps is detailed to make this process feasible; secondly, the process itself is described in terms of design choices and quality methods used.

The ETER quality validation and reporting process consist of the following phases:

- Quality metadata collection contextual to the data collection, in which NSAs provide the set of quality metadata as specified in the next section.
- Quality indicators and validation checks are based on the computation of 'Quality indicators' as defined in section 8.2. They are performed within the data collection phase on a country basis and on the whole dataset.
- Multi-annual checks.
- Ratios to detect comparability problems.
- Investigation of the comparability dimension on the basis of the previous analysis.
- Checks with external data sources to assess the overall coverage against official statistics and national aggregates or to explain/correct problems detected in the previous steps.

8.2 Quality indicators

In order to characterise the quality of data collected in ETER, we decided to evaluate a set of dimensions as listed in Table 32. These dimensions belong to the ISO 25012 standard related to Internal Data Quality. They are a subset of those proposed by EUROSTAT, resulting from the considerations described above related to the specificity of the ETER collection.

Table 32 Quality Dimensions

Dimension	Purposes
Accuracy (Format)	To evaluate the conformity of the provided values to the specified format in the collected data sets.
Completeness	To evaluate the number and meaning of missing values that are present in the collected data sets.
Consistency	To verify possible violations of semantic rules defined over the involved data, and specifically between different variables.
Timeliness	To evaluate the lapse of time between the ETER collection date and the source release date.

Quality indicators allow for a quantitative assessment of the quality dimensions listed in Table 32. They will be compiled based on quality metadata collected from the NSA/NE as described in the previous chapter of this handbook.

Each quality dimension can be evaluated according to different granularity levels, namely:

- Variables, e.g., personnel expenditures, third-party funding, etc.
- Dimensions, e.g. expenditures, income, etc.
- Data set, e.g. related to the whole bunch of data provided by a single NSA.

Hence, for each dimension, the following indicators have been defined.

8.2.1 Format Accuracy

		Value
Variable Format accuracy Indicator	Provided Format= Requested Format	TRUE, FALSE

For each variable, *format accuracy* evaluates the compliance of the value to the requested format, as defined in the data chapter of the ETER handbook.

8.2.2 Completeness

		Value Range
Completeness Variable Indicator	Number of missing variable values/total number of variable values	[0, 1]

		Value Range
Completeness Dimension Indicator	Number of missing variable values for all the variables of the dimension/total number of values for all the variables of the dimension	[0, 1]

		Value Range
Completeness Data set Indicator	Number of missing variable values for all the variables of the data set/total number of values for all the variables of the data set	[0, 1]

For each variable, dimension and data set, *completeness* evaluates the number of missing values (with the meaning relevant to the completeness, i.e. unavailable or temporarily unavailable) that are present.

8.2.3 Consistency

Consistency verifies a possible violation of rules associating different variables, such as that the sum of subcategories is equal to the total, or identifying logical relationships between variable, such as the number of ISCED8 graduates being 0 if the HEI does not have the right of awarding such degrees.

The first round of consistency checks is performed automatically within the data collection tool: whenever one of the following rules is violated, the respondent is required to provide an explanation. Under the column 'notes', the respondent is asked to describe the cause of departure.

Additionally, an annual comparison of national aggregates is implemented in the set of internal data quality mechanisms. All quantitative variables are aggregated within a country; whenever a variable changes from the previous year to the current year by more than 20%, it will be automatically detected. Although country comparison at the aggregate level may lead to some distortions (e.g. for small countries), comparing national aggregates across years is a further useful consistency check.

Table 33 Consistency indicators

Consistency indicator		Value
1	Total Expenditure=SUM(personnel expenditure, non-personnel expenditure, capital expenditure, unclassified expenditures)	**TRUE, FALSE**
2	Total expenditure>0	**TRUE, FALSE**
3	Total Income=SUM(core budget, third party funding, tuition fees, revenues unclassified)	**TRUE, FALSE**
4	Total Income>0	**TRUE, FALSE**
5	Personnel Total (HC and FTE)=SUM(academic personnel, non-academic personnel)	**TRUE, FALSE**
6	Personnel Total>0	**TRUE, FALSE**
7	Academic personnel total=SUM(female academic personnel, male academic personnel, unclassified)	**TRUE, FALSE**
8	Academic personnel total=SUM(national academic personnel, foreign academic personnel, unclassified)	**TRUE, FALSE**

9	Academic personnel total=SUM(academic personnel by field of education)	**TRUE, FALSE**
10	Academic personnel total-full professors>0	**TRUE, FALSE**
11	Full professors=SUM(female full professors, male full professors, unclassified)	**TRUE, FALSE**
12	*If* lowest degree delivered=ISCED 8 *then* Enrolled Students, Graduates ISCED 5-7 =0 *If* lowest degree delivered=ISCED 7 *then* Enrolled Students, Graduates ISCED 5-6 =0 *If* lowest degree delivered=ISCED 6 *then* Enrolled Students, Graduates ISCED 5 =0	**TRUE, FALSE**
13	*If* highest degree delivered=ISCED 5 *then* Enrolled Students, Graduates ISCED 6-8 =0 *If* highest degree delivered=ISCED 6 *then* Enrolled Students, Graduates ISCED 7-8 =0 *If* highest degree delivered=ISCED 7 *then* Enrolled Students, Graduates ISCED 8 =0	**TRUE, FALSE**
14	Student Total=SUM(female students, male students, unclassified) (for each ISCED level)	**TRUE, FALSE**
15	Student Total=SUM(national students, foreigner students, unclassified) (for each ISCED level)	**TRUE, FALSE**
16	Student Total=SUM(resident students, mobile students, unclassified) (for each ISCED level)	**TRUE, FALSE**
17	Student Total=SUM(students by fields of education) (for each ISCED level)	**TRUE, FALSE**
18	SUM(Total students enrolled ISCED 5-7, Total students ISCED 8)>0	**TRUE, FALSE**
19	Graduates Total=SUM(female graduates, male graduates, unclassified) (for each ISCED level)	**TRUE, FALSE**
20	Graduates Total=SUM(national graduates, foreigner graduates, unclassified) (for each ISCED level)	**TRUE, FALSE**
21	Graduates Total=SUM(resident graduates, mobile graduates, unclassified) (for each ISCED level)	**TRUE, FALSE**
22	SUM(Total graduates ISCED 5-7, Graduates ISCED 8)>0	**TRUE, FALSE**
23	*If* Number of students=0 *then* number of graduates=0 (for each ISCED level)	**TRUE, FALSE**
24	*If* Non research active *then* R&D expenditure 0	**TRUE, FALSE**
25	Total expenditure-R&D expenditure>0	**TRUE, FALSE**
26	Ancestor year ≤ foundation year ≤ legal status year	**TRUE, FALSE**

8.2.4 Timeliness

		Value
Timeliness Variable Indicator	ETER Collection Date- Source Release Date< Tolerance Threshold	• TRUE, if ETER Collection Date- Source Release Date lower than Tolerance Threshold • FALSE, otherwise

For each variable, timeliness evaluates the time lag between the ETER publication date and the Source Release date.

8.2.5 Accuracy, completeness and consistency checks implemented

In this section, the detailed types of checks performed to investigate accuracy, completeness and consistency are illustrated. Table 34 reports the types of checks in the first column, a description of them in the second column and a short outline of the implemented procedure to run it in the last column.

Table 34 Types of checks implemented

Type of checks	Description	Procedure
Accuracy checks	Accuracy checks verify that data entered have the right format foreseen by the handbook and that no logically impossible values are found.	Accuracy checks are performed in the data collection sheet and on delivered data. Simple mistakes are corrected directly, whereas unclear cases are reported back to NSAs/NEs for clarification.
Completeness checks	No blank cells are allowed in the dataset except for remarks. Blanks should be recoded as missing, confidential, not applicable or "0". This control is extremely important for the final quality of the database.	Blank cells are highlighted automatically. Clear cases are recoded directly, and ambiguous cases (for example, between missing and 0) are reported back to national experts and NSAs for clarification.
Consistency checks	a) These checks control for logical consistency between different variables (for example, when the highest degree delivered is at ISCED 7 level, all values for students and graduates at ISCED 8	Deviant values are identified and checked. In case there are specific reasons, an explanation

	level should be 0). Rules in this respect are stipulated in the handbook. b) Further, these checks control whether the sums of breakdowns by subcategories equal the total and numerical relationships between values (for example, R&D expenditures lower than total expenditures).	is added to the metadata for that specific HEI.
Check of missing data	An analysis of missing data is performed (including also issues of breakdowns by subcategories).	When it is expected that data should be available, possibly with some limitations, this is requested to NE/NSAs.
Control of metadata completeness	Metadata are systematically controlled for completeness, taking into account also issues emerging from the checks on the data.	When metadata are missing or incomplete, further information is requested.
Expert checks	Expert checks based on knowledge of national systems, as well as information available on the Web and EUMIDA data, are performed in order to ensure that provided data are realistic.	Potential problematic cases are notified back to national experts and NSAs. When these are related to methodological issues, the corresponding remarks are integrated into the metadata.

8.3 Comparability dimensions

The peculiarity of ETER data collection, which is performed by integrating data already collected by statistical institutions using different surveys or administrative data in about 40 countries, also requires the evaluation of another dimension related to the comparability of data across countries.

Based on the specific metadata collected for each variable ('Content of the variable and departure from definition') and relevant notes filled in at a record level for sub-categories of HEIs, a common data glossary will be produced at the end of each data collection wave. For a subset of variables, specific questions will be addressed to NSAs in order to highlight several issues that emerged as relevant.

Therefore, even if the challenge of comparability would require a longer time and cannot be completely addressed within the current contract, ETER will contribute:

- to highlight in a transparent way the comparability problems and differences among different countries' statistics definitions;

- to provide common definitions collected in a data glossary, which NSAs could eventually follow for future data collections;
- to assess the magnitude of departures and data distortion and decide when integration is possible with specific flags and when it is not possible at all.

8.3.1 Outlier Detection

An outlier can indicate an observation, a processing error, or a special element of the observed population that needs to be treated differently from the bulk in the subsequent processes.

In ETER, the outlier detection phase is part of the quality control and validation process and has been performed with two different objectives: (i) identification of possible errors in data collected; (ii) a better understanding of the collected data in order to perform proper analysis on them. Relevant outliers should be flagged and accompanied by a specific explanation in the notes and eventually rechecked. The complete list of ratios that will be explored for outlier detection is included in the table below.

The check is applied on an extended set of ratios between variables either pertaining to the same dimension (i.e. basic funding over total income) or different dimensions (i.e. personnel expenditure on total personnel). The list of ratios is reported in Table 35.

Table 35 Noticeable Ratios

Ratio	Rule
Students ISCED 5-7 / Academic Personnel FTE	Lower bound threshold: 0.27 Upper bound threshold: 2.53
Academic Personnel FTE / Total personnel FTE	Lower bound threshold: 0.71 Upper bound threshold: 1.54
Personnel expenditures / Total personnel FTE	Lower bound threshold: 0.42 Upper bound threshold: 1.64
Personnel expenditures / Total expenditures	Lower bound threshold: 0.71 Upper bound threshold: 1.3
Total expenditures / Total revenues	Lower bound threshold: 0.85 Upper bound threshold: 1.11
Basic government allocation / Total revenues	Lower bound threshold: 0.0032 Upper bound threshold: 1.9
Graduates ISCED 5-7 / Students ISCED 5-7	Lower bound threshold: 0.49 Upper bound threshold: 2.23
Graduates ISCED 8 / Students ISCED 8	Lower bound threshold: 0 Upper bound threshold: 2.85

8.3.2 Multiannual checks

The availability of data across different years raises the issue of the longitudinal consistency of data collected (impact of demographic events, revision of variable categories and definitions, etc.). On the other hand, the availability of several yearly editions of data offers an additional possibility for quality control. Indeed, multi-annual checks can help to detect suspect cases where the level of variation from year to year is substantial (as compared with what is expected, respectively, with the average change in the whole sample). This type of check is particularly useful in detecting and reporting mistakes of respondents and/or changes in the methodology for data collection.

However, the limited number of years available does not enable application-specific methods for time series analysis that require much more data. In addition, the ETER dataset is composed of different typologies of variables (e.g. structural descriptors rather than quantitative variables) with a different propensity to change over time.

For these reasons, the methodological approach developed for the multiannual checks consists of multiple procedures and is based on the use of different techniques:

- *Manual check* of the impact of demographic events (take-over, spin-off) on concerned institutions' figures and respective flagging (the code 'b' for the breakdown in time series was already foreseen);
- *Analytic control* of descriptors and status variables which supposed to be stable over time, i.e. legal status, foundation year, geographical information, lowest/highest degree awarded, etc.;
- *Comparison of national aggregates* over time for a selected number of quantitative variables already during the validation phase, with alarm if the variation is over a pre-defined threshold;
- *Use of measures of statistical dispersion* (interquartile range comparison over time) to assess the overall stability of the distribution of quantitative variables;
- *Statistical analysis* to highlight the HEIs with annual growth outstanding from the overall distribution (outlier) with an adaptation of the methodology already used for yearly outlier detection.

The approach proposed for multiannual checks is a flexible one. It is based on *thresholds* and *parameters* that can be modified based on expert knowledge or fixed on the basis of the empirical distributions of the observed data.

It relies on the following checks:

- *Check of the discontinuity*: it is useful to identify large jumps and, therefore, to capture the volatility of variations over time. It is based on the computation of the variance of the deltas (i.e. the variations) and on a normalisation of it that uses the power of the geometric mean (PGM). The power is used to implement the level of scale-invariant chosen, that is, the level of variation of the check with respect to the size of the variable. For this purpose, a scale invariance parameter (SI) is used to account for intermediary cases between fully scale-invariant and fully scale-variant cases. The SI parameter is fixed at 0.5 (intermediary case) but can be changed manually from 1 (fully scale-invariant) to 0 (fully scale-variant).
- *Check of the variance of deltas*: it identifies fluctuating trends and therefore captures the broader changes between one year and the next available year. It is based on the calculations of the sum of the positive deltas and the sum of the negative deltas. After that, the proposed algorithm makes the product of the two sums of deltas and normalises it using the power of the geometric mean, again to implement the desired level of scale invariance.

The methodology works as follows:

i) Start by considering only HEIs above a certain relevance threshold (S1). S1 is set to exclude a predetermined percentage of cases (e.g. 2%). S1 is defined calculating the geometric mean (GM) of the yearly data available for the variable under control for each HEI:

$$\mathrm{GM} = (\prod_{i=1}^{n} x^i)^{1/n}$$

Hence we have GM > = S1.

ii) Compute the discontinuity measure (DM) as:

- $\mathrm{DM} = \frac{DV}{PGM}$

Where $DV = |SPD * SND|$, SPD is the sum of the positive deltas, where deltas are the differences of the values of the observed variable observed in two different years (note that the deltas are computed on the closest years for which the data are available), SND is the sum of the negative deltas, and therefore DV is the absolute value of the product of these two; PGM is the powered geometric mean.

iii) Compute the variance of the annual delta measure (VM) as:

- $\mathrm{VM} = \frac{VD}{PGM}$

Where VD is the variance of the deltas and PGM is the powered geometric mean.

iv) On the relevant sample (with GM > = S1), identify the HEIs that have:

a. the discontinuity measure (DM) greater than a predetermined percentage (e.g. 10%) of highest values (S2). If the calculated DM value is > = S2 then set Alarm 1=TRUE;

b. the variance of the annual delta measure (VM) greater than a predetermined percentage (e.g. 10%) of highest values (S3). If the calculated value is > = S3 then set Alarm 2=TRUE.

v) Finally, identify the outliers (i.e. the units that have to be checked) as those HEIs that have GM>S1 (hence are relevant units) and have Alarm1 and Alarm2 positive. These HEIs present the values of the considered variable too discontinuous and too variable over time.

To clarify the approach, we describe in the following the case of enrolled students ISCED 5-7 (the same approach has been applied to the other relevant variables listed in Table 36):

1) Fix S1=120 (minimum relevance S1=120 excludes 2% of the smallest HEIs)
2) Fix S2= 13.06 (alarm level S2=13.06 highlights the 10% with the highest variance of delta)
3) Fix S3=1.47 (alarm level S3=1.47 highlights the 10% with the highest degree of discontinuity)
4) Fix the scale invariance parameter SI=0.5 (it may vary between 0 and 1: 0 means fully scale variant, 1 fully scale-invariant)
5) compute the geometric mean of the yearly values available for each HEI ($\mathrm{GM} = (\prod_{i=1}^{n} x^i)^{1/n}$)
6) implement the comparison If (GM >= S1) then the institution is relevant; so in the following of the check, you should discard those observations that do not meet this threshold S1
7) compute the powered geometric mean (PGM)(in this, you use the SI parameter): $\mathrm{PGM} = \mathrm{GM}^{(1+SI)}$
8) compute the deltas (deltas are the differences of the closest yearly data available for the controlled variable)
9) compute the sum of the positive deltas (SPD) and the sum of the negative deltas (SND)

10) calculate the discontinuity value (DV) as the absolute value of the product of SPD times SND: $DV = |SPD * SND|$
11) compute the discontinuity measure (DM) dividing the discontinuity value by the PGM: $\mathrm{DM} = \frac{DV}{PGM}$
12) implement the comparison Alarm1: If (DM >= S2), then raise Alarm 1
13) compute the variance of deltas (VD)
14) compute the variance of deltas measure (VM) dividing the variance of deltas by the PGM: $\mathrm{VM} = \frac{VD}{PGM}$
15) implement the comparison Alarm2: If (VM >= S3) then raise Alarm 2
16) Finally, identify the outliers that have to be checked as those that have GM>S1 (are relevant units) and have Alarm1 and Alarm2 positive (hence are too discontinuous and too variable over time).

The statistical multiannual approach is applied to a list of the most relevant (sensible) variables (see Table 36), covering the most important ETER dimensions. The comparison is made between the current years against the previous available years. Whenever an *outlier*, that is, a value that *exceeds the threshold* value, is detected, the whole longitudinal set (starting from 2011) is manually analysed to better interpret its nature. In this case, for each record identified as an outlier, the following data are inspected (with some variations depending on the nature of the variable): yearly data, identification of the HEIs, the values of the thresholds, flags and notes of the variables checked, and other relevant variables (according to the type of variable analysed).

Therefore, the goal of the statistical method is essentially related to the identification of cases to be controlled and possible comparability problems emerging from the data. Further, an expert-based evaluation is performed by the consortium working team.

Table 36 List of variables considered for the multiannual checks

Variable
Total expenditure (PPP)
Total revenues (PPP)
Total academic personnel (FTE)
Total academic personnel (HC)
Number of administrative personnel (FTE)
Total personnel (FTE)
Total personnel (HC)
Total students enrolled (by ISCED level)
Total graduates (by ISCED level)

8.4 Quality Reporting

The purpose of quality reporting is to summarise the data quality validation procedures and to report the quality metadata calculated in each wave of data collection. The overall data quality status of each wave of data collected is summarised in a technical data quality report prepared by the ETER data quality managers, which accompanies the data. Quality problems will also be highlighted at a record level with a system of flags and notes that will be integrated into the dataset by the project team.

9 The ETER data and publication infrastructure

The ETER project provides a complex technical infrastructure for data collection, which allows standardisation and systematisation of the data collection process. This infrastructure includes the following:

- The software of the technical ETER infrastructure. It consists of:
  - A database to collect, organise and retrieve data.
  - A website using a content management system (CMS) to support dissemination activities.
  - A web application in order to work with the data.
- The administrative part of the ETER infrastructure. It includes hosting of servers and applications as well as their maintenance, which is done by JOANNEUM RESEARCH. Maintenance covers keeping the system up-to-date and secure and monitoring applications and databases for the ETER environment.

The following sections cover the three parts of the ETER software in detail.

9.1 The ETER database

The ETER database is based on a MongoDB architecture, which uses an object-oriented approach to deal with different structured datasets, so-called 'documents'. Different structured documents lead to a dynamic schema of the database itself, which leads to an iterative and agile approach for the design and development of the ETER database. This implies that the database structure will change slightly by every added or withdrawn field (bits and pieces of data). Thus, if new fields need to be added to a document, the field can be created without affecting all other documents in the system, without updating a central system catalogue and without taking the system offline. In addition, if developers add more features, MongoDB continues to store the updated objects without the need for performing costly alteration operations.

The ETER database consists of some master data and the dynamic data collected via Excel files. Important parts of the master data are the 'field' documents. A field document is a meta-description of a column within the HEI data collection Excel files. The field description defines:

- The type of data that is stored in this field.
- The display labels for the represented column.
- The possible flags available for this field.
- A list of display and export formats.
- The link of the Excel-Column ID and the data path within the database object.

The dynamic data are split into subdocuments in order to improve performance and data handling. Each HEI base document is linked to its subdocuments via a 1:1 connection.

9.2 The ETER CMS system

The ETER web infrastructure is designed to address a broad range of interested users. Therefore, the infrastructure does not only host a very detailed database but also supports dissemination – it has been designed in a hybrid approach, including a website with a content management system and an application for the work on and with data. The CMS is used to create, store and publish content on the ETER website. It enables the ETER team to work collaboratively since several users can contribute to content creation and sharing. Thus, the planned dissemination activities in the ETER contract can be implemented in a flexible way. The content management system is the landing page of the ETER domain. It is easily extendable and enables the consortium to place all dissemination activities in one place (e.g., videos of webinars and training activities, Twitter feed, reports, newsletters, and country reports).

9.3 The ETER web application

The aim of the ETER web application is to enable users (both the project team and external users) to work with the data and satisfy different user needs. It is designed in a way to make ETER users feel comfortable while using the app. The following requirements have been met with the ETER web application:

9.3.1 User administration

The ETER web application enables the project team to manage data and features by user groups:

- *Non-registered users* can access all dissemination content on the website and access data via the web application. However, non-registered users cannot see data which have been either marked as confidential by NSAs or small data which are coded according to personalisation issues.
- *'User'(s)* are all persons who registered to ETER and agreed to the terms of usage. Persons included in this group can access all dissemination content and the full ETER dataset without restrictions.
- *'CoreTeam'* includes members of the ETER project team. Additional to the permissions for user group 'user', core team members can export data collection files, see the change history and are able to access also data currently collected but not yet published.

- *'Admin'* - additional to the permissions of core team members, administrators have access to the user administration, code administration, year and country administration, fields administration and data import.

9.3.2 Data management

The web application includes many features which are necessary for managing the ETER dataset:

- *Code administration:* The ETER dataset includes an extensive coding and flagging system in order to be able to switch between codes and text as well as storing both codes and values (e.g. in cases where values are confidential for non-registered users but accessible for registered users). ETER administrators have access to the code administration menu and can edit the coding system.
- *Year administration:* During the data collection phase, the ETER project team works with the web application to prepare the data publication. In order to do this, the administrators can use status 'collecting' for a year. Years having this status can only be seen by user groups 'admin' and 'CoreTeam'. When the status is changed to 'Public', all users can see the respective data collection years.
- *Country administration:* Country information in ETER includes names of countries, country codes, exchange rates and purchasing power parity (PPP). This information is managed in the country administration menu by the administrators.
- *Fields administration:* Within the ETER project, a flexible approach has been chosen in order to be able to add, remove and change variables quickly. So-called 'fields' are metadata about variables and include every information about them (e.g. names, field type, access rights, possible codes, a formula in case of calculation). These metadata can be edited using an Excel file and imported by the administrators.
- *Data import:* The ETER infrastructure team has implemented several import mechanisms in order to be flexible for different types of data updates:
    - *Import data by year and country*. This import type is used for getting data into the database in the first place and when updating data for a specific country and year.
    - *Import data by ETER ID*year*. This import type is especially helpful for importing additional data and imports of several years and countries at once. It is, for example, possible to import new variables for all countries and all years in one batch.
    - *Import data by ETER ID*year range*. This import type proved especially valuable when unchanging data needed to be amended. Examples are names, acronyms, foundation years, websites or geographical information.

Instead of using an import of format *ETER ID*year*, the data can be updated for all concerned years with only one input row.

- *Export of data collection files:* ETER project team members can export data collection files directly from the web application. This is helpful when updates are needed, e.g. as a result of the data quality process.

9.3.3 Searching and extracting data

The primary purpose of starting ETER was the provision of microdata on the level of higher education institutions. Analytics of the online ETER usage and a survey on the usage of ETER (launched in summer 2021) showed that searching and extracting data are still the features of ETER that are used by far the most often. ETER provides a large set of possibilities for searching the data:

- Users can simply download the whole dataset at once with one click. This is the preferred option for users who are experienced in data analysis and want to dig deep into the data themselves.
- Users can export data for all countries and all years for selected variables. ETER has a wide range of variables. In order to narrow the selection down, some users may want to export data for all years and countries, but only a subset of variables.
- Select predefined groups of variables. Users may want to work with specific groups of variables, e.g. data about personnel, students or finances. In this case, users can select all variables in a group with only one click. The ETER team also provided a core set of data, which includes the most important totals.
- Select all variables at once. This option is the best if users want to export all variables at once, but for selected years and countries.
- Finally, users can also compile a fully individual dataset by selecting single variables from the list of variables.

The ETER web application has the target to support its users as much as possible in searching, screening and exporting data - users can choose between different settings and different export formats:

- Replace variable codes (for nominal variables), special codes (for missing variables) and flags with their full labels in order to support statistical analyses and graphical illustrations. A full list of codes and their meanings is provided online.
- Choose a preferred export header to use either variable labels (full name) or variable names (systematic variable naming). In addition, it is possible to export the data with both header options.
- Choose a preferred export format where the following possibilities are available: csv (; separated), Microsoft Excel (.xlsx) and machine-ready datasets. In machine-

ready datasets, special codes are replaced with fixed values to allow the respective statistical software to recognise missing values and save time before starting data analysis. Two different types of machine-ready exports are provided, targeting the statistical programmes SPSS and STATA. Detailed information on the replacements can be found online.

9.4 Data publication and conditions for access

Most data in the ETER database are publicly available once the national statistical authorities have given publication permission. Thus, two different types of access are provided in ETER:

- Open public access, where small numbers and all data for which a national statistical authority restricted access, are coded with 'c'. These data can be freely downloaded from the ETER website and used for different purposes (under the condition that the source is mentioned).
- Restricted access where registered users receive access to the entire data set for research purposes under the condition that individual data are not disclosed. In order to receive access to all data, interested users have to sign a non-disclosure agreement. This agreement allows the usage of restricted data for research purposes under the condition that no individual data points are disclosed publicly.

Because some data are delivered but restricted for public access, data have to be recoded for the publicly available data set. The same is valid for values larger than 0 but below or equal to 3 when this is requested by a country. The following codes are used in this respect:

- When data are available but restricted to public access, they are recoded to 'c' in the publicly available database.
- Values larger than 0 and below or equal to 3 will be recoded to 's'. To prevent that these data can be inferred, some additional rules have to be followed:
    - If a total is larger than 0 but below or equal to 3, all breakdowns are coded to 's'.
    - If one or more breakdowns are larger than 0 but below or equal to 3, the unclassified category is also recoded to 's'.
    - If the value of the 'unclassified' category is larger than 0 but below or equal to 3, it is not coded to 's' as long as the corresponding values of this breakdown are larger than 3.
        - If the corresponding values are larger than 0 but below or equal to 3, they will be recoded to 's' together with the 'unclassified' category.
        - If corresponding values are either larger than 3 or 0, one additional 0 in the breakdowns has to be recoded to 's' in order to prevent recalculation.

- If one of the categories, 'Total academic personnel' or 'Total administrative personnel', includes values larger than 0 but below or equal to 3, both categories, including breakdowns, have to be recoded to 's' to preserve the value of total personnel.
- FTE values in personnel data are not recoded as long as the corresponding headcounts are larger than 3.

The ETER website (www.eter-project.eu) can be used without registration. However, in order to explore the full dataset (including confidential and small values as described above), users need to register and agree to the ETER terms of use. The access will then be activated immediately, and users can access the whole dataset without restrictions.

10 References

Bonaccorsi, A. & Daraio, C. (2007a). *Universities and Strategic Knowledge Creation. Specialization and Performance in Europe.* Edwar Elgar.

Bonaccorsi, A. & Daraio, C. (2007b). Universities as strategic knowledge creators: some preliminary evidence, in Bonaccorsi, A., Daraio, C. (Eds.), *Universities and Strategic Knowledge Creation. Specialization and Performance in Europe*. Edwar Elgar. pp. 31-81.

Borden, V.M. & McCormick, A.C. (2020). Accounting for diverse missions: can classification systems contribute to meaningful assessments of institutional performance? *Tertiary Education and Management, 26*, 255-264. DOI: 10.1007/s11233-019-09045-w

Daraio, C., Bonaccorsi, A., Geuna, A., Lepori, B., Bach, L., Bogetoft, P., Cardoso, M.F., Castro-Martinez, E., Crespi, G., Lucio, I.F.d., Fried, H., Garcia-Aracil, A., Inzelt, A., Jongbloed, B., Kempkes, G., Llerena, P., Matt, M., Olivares, M., Pohl, C., Raty, T., Rosa, M.J., Sarrico, C.S., Simar, L., Slipersaeter, S., Teixeira, P.N., & Van den Eeckaut, P. (2011). The European university landscape. *Research policy, 40*, 148-164.

de Boer, H., Enders, J., & Leisyte, L. (2007). Public Sector Reform in Dutch Higher Education: the Organisational Transformation of the University. Public Administration, 85, 27-46. DOI: 10.1111/j.1467-9299.2007.00632.x

Fumasoli, T. & Lepori, B. (2011). Patterns of strategies in Swiss Higher Education Institutions. *Higher Education, 61*, 157-178. http://www.jstor.org/stable/41477813

Hazelkorn, E. (2015). *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence.* Springer.

Huisman, J., Lepori, B., Seeber, M., Frølich, N., & Scordato, L. (2015). Measuring institutional diversity across higher education systems. *Research Evaluation, 24* (4), 369-379. DOI: 10.1093/reseval/rvv021

Huisman, J., Meek, L., & Wood, F., (2007). Institutional diversity in higher education: A cross-national and longitudinal analysis. *Higher Education Quarterly, 61*, 563-577. DOI: 10.1111/j.1468-2273.2007.00372.x

Kyvik, S. (2004). Structural changes in higher education systems in Western Europe. *Higher education in Europe, 29*, 393-409. DOI: 10.1080/0379772042000331679

Lepori, B. (2021). The heterogeneity of European Higher Education Institutions. A configurational approach. *Studies in Higher Education*. DOI: 10.1080/03075079.2021.1968368

Lepori, B. (2020). Non-university Education and Professional Institutions, in Shin, J.C., Teixeira, P. (Eds.), *Encyclopedia of International Higher Education Systems and Institutions*. Springer.

Lepori, B., Geuna, A., & Mira, A. (2019). Scientific output scales with resources. A comparison of US and European universities. *PloS ONE 14*. DOI: 10.1371/journal.pone.0223415

McCormick, A.C. & Zhao, C.M., 2005. Rethinking and reframing the Carnegie classification. *Change, 37*(5), 50-57.

OECD (2015). *Frascati Manual 2015. Guidelines for Collecting and Reporting Data on Research and Experimental Development.* OECD.

van Vught, F. (2007). *Diversity and Differentiation in Higher Education Systems.*

Van Vught, F., Bartelse, J., Bohmert, D., File, J., Gaethgens, C., Hansen, S., Kaiser, F., Peter, R., Reichert, S., Taylor, J., West, P., & van de Wende, M. (2008). *Mapping diversity. Developing a European classification of higher education institutions.*