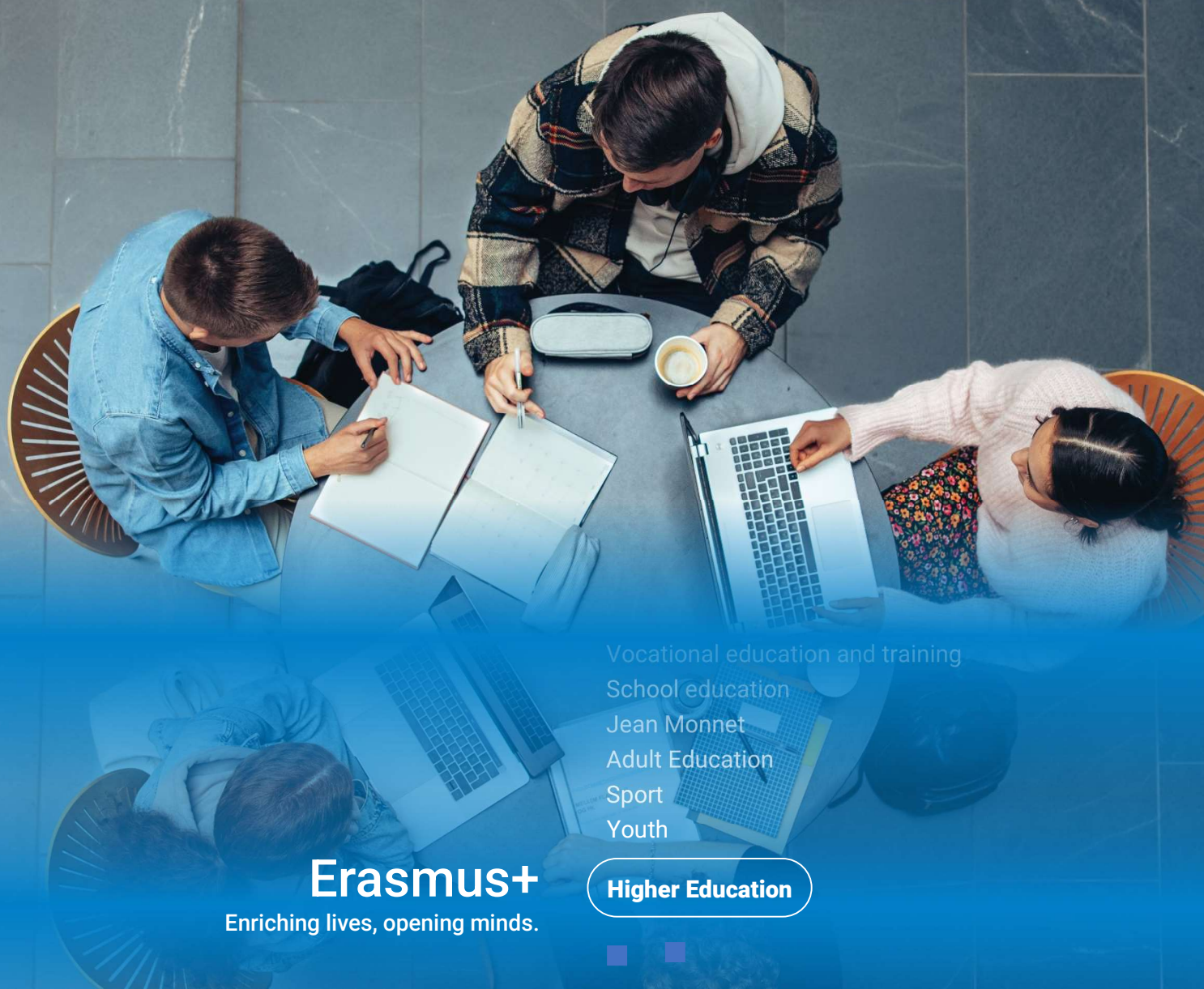




EUROPEAN UNION

The European Higher Education Sector Observatory

Handbook for the European Tertiary Education Register data collection



Vocational education and training
School education
Jean Monnet
Adult Education
Sport
Youth

Erasmus+

Enriching lives, opening minds.

Higher Education



This document is published by the European Education and Culture Executive Agency (EACEA, Platforms, Studies and Analysis).

Please cite this publication as:

European Commission / EACEA (2025), Handbook for the European Tertiary Education Register data collection – 2025. The European Higher Education Sector Observatory technical report. Luxembourg: Publications Office of the European Union.

European Education and Culture Executive Agency

Platforms, Studies and Analysis Boulevard Simon Bolivar 34 (Unit A6)
1049 Brussels
BELGIUM
E-mail: EACEA-HE-Observatory@ec.europa.eu
Website: <https://national-policies.eacea.ec.europa.eu/eheso>

Text completed in March 2025.

The European Higher Education Sector Observatory, and this publication in particular, are possible thanks to funding from Erasmus+, the European programme for education, training, youth and sport. It is an initiative by the European Commission's Directorate-General for Education, Youth, Sport, and Culture. It is operated by a consortium under a service contract with the European Education and Culture Executive Agency (EACEA). The opinions expressed are those of the contractor only and do not represent the contracting authority's official position.

Reproduction is authorised provided the source is acknowledged.

Luxembourg: Publications Office of the European Union, 2025

© European Union, 2025

The Commission's reuse policy is implemented under Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39, ELI: <http://data.europa.eu/eli/dec/2011/833/oj>).

Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed, provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders. The European Union does not own the copyright in relation to the following elements:

Cover image: © Jacob Lund / <https://stock.adobe.com>.

The European Higher Education Sector Observatory

Handbook for the European Tertiary Education Register data collection

2025

Contents

Abbreviations.....	6
Document summary	7
1 Introduction.....	8
2 Basic principles of the data collection.....	10
2.1 Higher Education Institutions as the statistical unit.....	10
2.2 A focus on Higher Education Institutions within tertiary education	11
2.3 HEIs as multi-functional institutions and profiling.....	12
2.4 Building on existing methodological frameworks	12
2.5 EHESO-ETER as a secondary data collection.....	13
3 Perimeter and list of entities.....	14
3.1 Country coverage	14
3.2 Overall structure of the perimeter	14
3.3 Inclusion of educational institution in the main data collection.....	15
3.3.1 Exclusion cases	15
3.4 Demographic events	18
3.5 Types of instructional providers.....	20
4 Classifications	21
4.1 ISCED Fields of Education and Training classification.....	21
4.2 Levels of education (ISCED)	22
4.3 Geographical classification	23
4.3.1 NUTS regions.....	23
4.3.2 Postcodes	23
4.4 Reference period for data collection.....	24
4.5 Special codes.....	25
4.6 Data flags and notes	25
4.7 Monetary units.....	27
4.7.1 ISO country and currency codes in EHESO-ETER.....	27
4.8 Treatment of small values	28
5 Variables	30
5.1 HEI-level data collection	30
5.2 Identifiers and demographic information.....	31
5.2.1 ETER identifier.....	31
5.2.2 Reference year	31
5.2.3 ETER ID – Year.....	31
5.2.4 National ID.....	32

5.2.5	Legal name of the institution	32
5.2.6	English translation	32
5.2.7	Institution acronym	32
5.2.8	Demographic event ID	32
5.3	Basic institutional descriptors.....	33
5.3.1	Country of establishment.....	33
5.3.2	Legal status.....	34
5.3.3	Institution category (national classification).....	34
5.3.4	Institution category (standardised).....	34
5.3.5	Year of establishment	35
5.3.6	University hospital	36
5.3.7	Institutional website	37
5.4	Geographic information	37
5.4.1	Region of the main seat - NUTS code	37
5.4.2	Postal code	38
5.4.3	Name of city	38
5.4.4	Geographic coordinates.....	38
5.5	Multi-site institutions	38
5.6	Educational activities and outputs.....	39
5.6.1	Lowest and highest degree delivered	40
5.6.2	Students	40
5.6.3	Graduates.....	42
5.6.4	Distance education institution.....	44
5.7	Research activities	44
5.7.1	Research active institution	44
5.7.2	ISCED 8 students	45
5.7.3	ISCED 8 graduates.....	46
5.7.4	Inclusion in R&D statistics	48
5.8	Revenues and expenditures.....	48
5.8.1	General definitions and principles.....	48
5.8.2	Structure of the EHESO-ETER financial information	49
5.8.3	Expenditures	50
5.8.4	Revenues.....	52
5.9	Personnel	56
5.9.1	Classification of HEI personnel	57
5.9.2	Measurement units	58
5.9.3	Breakdowns of academic personnel.....	58

5.9.4	Employed PhD candidates	59
5.9.5	Academic personnel.....	59
5.9.6	Academic personnel by seniority level (HC).....	60
5.9.7	Research and teaching assistants	62
5.9.8	Support and administrative personnel.....	62
5.9.9	Total number of personnel	63
5.10	Supplementary data collection by institutional types.....	63
6	Data validation and metadata	64
6.1	Internal validity checks	64
6.2	External validity.....	65
6.2.1	EUROSTAT checks	65
6.3	Outlying values	66
6.4	Multiannual checks	67
6.4.1	Country-level checks	68
6.4.2	HEI-level checks.....	68
6.5	Metadata.....	69
6.5.1	General Metadata.....	70
6.5.2	In-depth information.....	71
7	Data collection and validation process.....	74
7.1	Overview.....	74
7.2	Perimeter and descriptors validation.....	74
7.3	Main data collection and data collection by types	75
7.3.1	Procedures and options for data deliverers	76
7.4	Data validation process	77
7.5	Open data policy and data publication	78

Abbreviations

Abbreviation	Definition
AY	Academic year
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 Association
ETER	European Tertiary Education Register
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
EHESO	European Higher Education Sector Observatory
HE	Higher Education
EHESS	European Higher Education Sector Scoreboard
HEI	Higher Education Institution
ISCED	International Standard Classification of Education Degrees
NE	National Experts
NIFU	Nordic Institute for Studies in Innovation, Research and Education
NSA	National Statistical Authority
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Cooperation and Development
UAS	Universities of applied sciences
UOE	UNESCO-UIS/OECD/EUROSTAT joint data collection on education statistics

Document summary

This document provides a description of the data collection methodology for the ETER data collection, i.e. the main data collection on individual Higher Education Institutions undertaken with National Authorities; this data collection is part of the broader data provided for the European Higher Education Sector Observatory (EHESO). The handbook 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 April 2025 and has been adopted for the ETER data collection 2022 (as well as for updating the former data collected).

1 Introduction

This handbook provides methodological guidance for the ETER data collection within the European Higher Education Sector Observatory (EHESO), referred in the following as the EHESO-ETER data collection.

The EHESO-ETER data collection has been developed through a series of contracts funded by the European Commission and collects data from National Statistical Authorities and higher education ministries (NSAs) on individual Higher Education Institutions. EHESO-ETER largely follows established methodologies in educational and research statistics, specifically the UNESCO-UIS/OECD/EUROSTAT (UOE) joint data collection on formal education manual (UOE manual)¹ and the OECD Frascati Manual on research and experimental development (R&D) statistics². Close cooperation is established with EUROSTAT and OECD concerning data collection methodology.

From 2024 onwards, the EHESO-ETER data collection has become part of the broader framework of the European Higher Education Sector Observatory (EHESO³), an initiative by the European Commission's Directorate-General for Education, Youth, Sport, and Culture funded from Erasmus+, the European programme for education, training, youth and sport.

The EHESO will leverage the strengths of existing EU data tools and capacities, including the European Tertiary Education Register (ETER), U-Multirank (multidimensional institutional benchmarking), Erasmus+ database (student and staff mobilities, transnational cooperations of HEIs), Database of External Quality Assurance Results (DEQAR), Eurostudent (Social dimension of Higher Education), Eurograduate (higher education graduate outcomes), Eurydice higher education policy data (Bologna Process Implementation Reports, Mobility Scoreboard, National Education systems descriptions), and other relevant data sources.

The EHESO will make it possible to compare, analyse and showcase the higher education sector's performance across various fields. Institutions and governments will be able to strengthen their evidence base on key topics such as inclusion, learning outcomes, progress on digital, green and entrepreneurial skills, technology transfer, employability, students and labour market needs. This process will also reinforce the institutions' role in innovation ecosystems, and transnational cooperation in the higher education sector.

The consortium implementing the activities of the service contract is coordinated by PPMI, in Lithuania, and includes the Centre for Higher Education (CHE) in Germany, the Centre for Higher Education Policy Studies (CHEPS) at the University of Twente in the Netherlands, the Austrian Institute of Technology (AIT) and JOANNEUM RESEARCH in Austria. The project is supported by the Nordic Institute for Studies in Innovation, Research and Education (NIFU) in Norway and the Centre for Science and Technology Studies (CWTS) from Leiden University in the Netherlands, as well as a number of individual experts.

The EHESO-ETER data collection represents the backbone of the EHESO data, as it provides a consistent list of Higher Education Institutions (HEIs) and basic statistical data such as numbers of students, graduates, personnel, and financial data, as well as basic descriptors on institutional characteristics and geography. EHESO-ETER data will be progressively enriched

¹ [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=UNESCO_OECD_Eurostat_\(UOE\)_joint_data_collection_%E2%80%93_methodology](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=UNESCO_OECD_Eurostat_(UOE)_joint_data_collection_%E2%80%93_methodology)

² <https://www.oecd.org/sti/inno/frascati-manual.htm>

³ https://www.eacea.ec.europa.eu/news-events/news/launching-eheso-european-higher-education-sector-observatory-2024-05-13_en

with data from other sources and projects both at the country and the HEI level. As variables and definitions remain largely the same, the EHESO data also integrates previous ETER data collection from the year 2011 onwards to allow for longitudinal analysis.

This handbook is specifically targeted at data collectors in the EHESO team and at the NSAs and provides instructions on how to collect and prepare the data for integration in the EHESO. Specifically, it is organised in the following sections:

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

2 Basic principles of the data collection

EHESO-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 allow also a better understanding of the relationships and complementarity 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 that builds on the statistical definition of tertiary education, defined as levels 5 to 8 of the ISCED classification, and takes into account the specificities of collecting data at the institutional level.
- An understanding of HEIs as multifunctional organisations displaying a diversity of profiles according to different dimensions.
- A methodological approach that builds on existing statistical definitions (specifically from the UOE joint data collection on education statistics), while adapting definitions to the specific needs of EHESO-ETER and 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.

2.1 Higher Education Institutions as the statistical unit

The basic design of choice of the EHESO-ETER data collection is to focus on 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 where the primary statistical units are educational programmes with data aggregated at the country or regional level. As discussed later in this handbook, this is generating differences in how the data are collected and aggregated in respect to the UOE main data collection.

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. Universities have also increasingly started competing at the European and international levels, as evidenced by emerging policies such as the European Universities Initiative⁴.

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

This principle has far-reaching consequences for the EHESO-ETER data collection.

⁴ https://ec.europa.eu/education/education-in-the-eu/european-education-area/european-universities-initiative_en.

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 3, identifying the units is not always straightforward and includes some elements of conventionality.

Second, data and indicators are referred to the whole higher education institutions' level rather than to its constitutive units, such as faculties and departments or campuses. EHESO-ETER has been designed for analyses for which an institutional level is a meaningful unit of analysis; therefore it is not suited for studies of differentiation within institutions.

Third, this principle has implications for defining variables, particularly for what concerns financial variables, where EHESO-ETER focuses on the HEI's resources and expenditures rather than on national investment. Even if basic definitions might be the same, the different unit of analysis generates different measures compared with educational statistics at the country level.

2.2 A focus on Higher Education Institutions within tertiary education

The overall data collection perimeter for EHESO-ETER is identical to tertiary education as defined in the UOE manual, i.e. the whole of educational activities classified at levels 5 to 8 of the International Standard Classification of Education Degrees (ISCED). However, the focus on institutions rather than on programmes generates a number of differences with UOE data collection for substantive and practical reasons.

At the substantive level, the program-based view and the institutional-based view largely overlap, in the sense that most programmes in tertiary education are offered by specific institutions (Higher Education Institutions; HEIs) that focus on tertiary level of education, such as universities and Universities of Applied Sciences.

However, in many countries, short cycle vocational education (at ISCED5 level) is decentralized to a large number of small regional providers, which however account only for a minority of total students and degrees at the tertiary level (on the whole ISCED5 students account for less of 10% of total tertiary enrolments in EU-27).

While the main EHESO-ETER data collection currently includes about 3 000 Higher Education Institutions, the number of those 'other' educational providers exceeds 10 000 entities. Moreover, the experience in the ETER data collection shows that for vocational educational providers less data are available. The feasibility of collecting data for these providers is, therefore, problematic.

In some other cases, tertiary education degrees that are offered by institutions whose primary mission is to provide secondary education, such as the so-called preparatory classes to tertiary education in France and the French-speaking part of Belgium. And, finally, in countries where vocational education is largely delivered at the workplace, such as in Austria, Germany and Switzerland, vocational tertiary education degrees are awarded to individuals, which are not enrolled in a formal study program, but follow ad hoc courses to prepare to vocational exams, frequently offered by industrial associations. The degrees (and sometimes students) can be attributed to tertiary education, but they are not delivered by an institution whose main activity is education.

To address these issues, while keeping compatibility with statistical definitions of tertiary education, EHESO-ETER has now introduced a nested structure of data collection at two levels:

- Collecting simplified data aggregated by the *types* of educational providers at the tertiary level, with the goal of achieving a coverage of tertiary education in terms of students and graduates as near as possible to EUROSTAT data.
- Collecting full data at the level of individual Higher Education Institutions for a subset of HEIs, which are sufficiently large and for which institutional-level data are available at

national level. By experience, this amounts to between 80% and 95% of the whole students in tertiary education depending on the country.

The corresponding definitions and rules are described in detail in chapter 3, while the data collected are described in chapter 5.

2.3 HEIs as multi-functional institutions and profiling

EHESO-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.

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 EHESO-ETER framework overcomes the traditional distinction between educational and R&D statistics and does not attempt to separate these activities, for example by distinguishing between teachers and researchers, but instead provides distinct indicators concerning educational and research production. EHESO-ETER builds in that respect on the literature on the differentiation of HEI profiles to identify a set of core dimensions for the characterisation of individual HEIs. 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. Additionally, EHESO-ETER provides core information on the level and composition of resources available for HEI activities, notably the composition of academic personnel and institutional revenues.

Within the EHESO, this approach is operationalized in the EHESO benchmarking tool, which grounds largely on the EHESO-ETER data collection and allows users to select and compare similar institutions based on their profile.

2.4 Building on existing methodological frameworks

A core design choice of EHESO-ETER is to rely on, as far as possible, existing methodologies, specifically from education and R&D 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 and guarantees the possibility of comparison with international educational statistics.

More specifically, the definition of EHESO-ETER variables largely builds on the UNESCO-UIS/OECD/EUROSTAT (UOE) manual⁵ for the data collection on education. EHESO-ETER definitions concerning students and degrees are fully aligned with the UOE with few adaptations related to different statistical units.

EHESO-ETER definitions concerning HEI personnel also largely comply with UOE definitions, but more detail is provided concerning the breakdowns of personnel by level, as well as citizenships.

⁵ [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=UNESCO_OECD_Eurostat_\(UOE\)_joint_data_collection_%E2%80%93_methodology](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=UNESCO_OECD_Eurostat_(UOE)_joint_data_collection_%E2%80%93_methodology)

As of financial data, UOE methodology has been designed to measure the educational investment from the state and society, while EHESO-ETER is focused on the resources available at the institutional level, and accordingly different definitions and measures are given. As for the definition of Research & Development (R&D) expenditures, EHESO-ETER follows definitions and rules in the Frascati Manual (OECD, 2015), also adopted by the EUROSTAT for R&D statistics.

Finally, EHESO-ETER uses its own definitions and methodologies for variables of non-statistical nature (most institutional descriptors and geographical information) and for financial data.

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 programmes) are counted only once, while in EHESO-ETER, they will be counted in each individual institution. Accordingly, the sum of students enrolled in EHESO-ETER HEIs might exceed national totals in the UOE data collection.

2.5 EHESO-ETER as a secondary data collection

A final core characteristic of EHESO-ETER is that it 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 undergoes extensive validation checks (see chapter 6), but the project team assumes no responsibility for mistakes in the original data. Observed discrepancies are reported back to the producing entity and any changes are discussed with them.

Most EHESO-ETER data are derived from the data sources behind UOE data collection on educational statistics. Primary data are the same as those in the country and regional education and training statistics⁶ 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 EHESO-ETER team from official sources, such as the NSAs' websites, respectively from public open data portals. Differences in coverage between UOE data and ETER data are mostly due to the lack of institutional-level data for vocational education.

The only exception is some descriptive information, such as foundation years and websites, as well as most geographical information, which are collected by the project team from public sources, such as institutional websites and Wikipedia. NSAs are given the possibility of validating and checking these data.

⁶ <https://ec.europa.eu/eurostat/web/education-and-training/data/database>.

3 Perimeter and list of entities

A central issue for EHESO-ETER is establishing the list of institutions (HEIs) to be included in the data collection, the so-called *perimeter*. This perimeter is stored in the RISIS Organizational Register (OrgReg), the register application developed by the RISIS EU-FP infrastructure initiative, whose function is providing a list of higher education institutions in a country, including some essential information like history, demography, and linkages between institutions, and maintaining unique identifiers (e.g. AT0001), which are the basis for the EHESO statistical data. OrgReg is described in the OrgReg methodological manual⁷.

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 programmes 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.

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.

3.1 Country coverage

The EHESO-ETER data collection includes currently the UK, EEA/EFTA countries (Iceland, Liechtenstein, Switzerland and Norway), as well as candidate and potential candidate countries to the European Union (Albania, Bosnia and Herzegovina, Kosovo, Montenegro, Serbia, Türkiye, The Republic of North Macedonia), as well as Andorra.

The extension to the three countries, which were awarded the candidate status in 2023, i.e. Georgia, Moldova and Ukraine, will be further considered in future data collections.

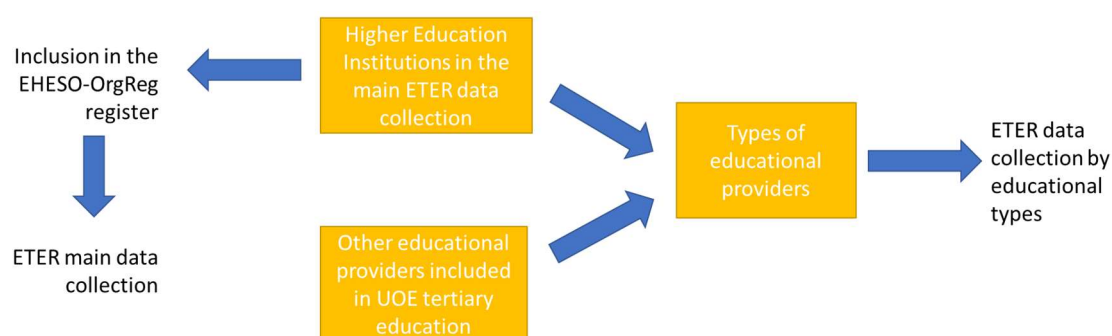
3.2 Overall structure of the perimeter

Figure 1 below describes the logic of the EHESO-ETER perimeter and its relationships with the UOE definition of tertiary education.

Following the UOE manual, *instructional educational institutions* are those that provide *educational programmes* for students as their main purpose (e.g. schools, colleges, and universities) that fall within the scope of education statistics, i.e. formal education. Such institutions are normally accredited or sanctioned by the relevant national education authorities or equivalent authorities. Educational institutions may also be operated by private organizations, such as religious bodies, special interest groups or private educational and training enterprises, both for profit and non-profit.

⁷ <https://zenodo.org/records/6396703>

Figure 1: Overall structure of the EHESO-ETER perimeter



EHESO-ETER collects data at the level of individual institutions for a subset of the educational institutions included in the UOE scope, following the criteria described below in section 3.3; the remaining instructional providers are then grouped in types as described below in section 3.5 and a core set of aggregated data is collected in a complementary data collection.

3.3 Inclusion of educational institution in the main data collection

The main EHESO-ETER data collection includes those educational providers (conventionally labelled as *higher education institutions*), defined as instructional educational providers that:

- are recognisable as distinct organisations, meaning 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. 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.
- are nationally recognised as HEIs. 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 recognised HEIs.
- the institution's 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. These institutions can have a large share of research activity, but it should not be their only purpose.
- exceed certain size thresholds. 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.

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.

Figure 2 below provides a decision tree to apply these criteria.

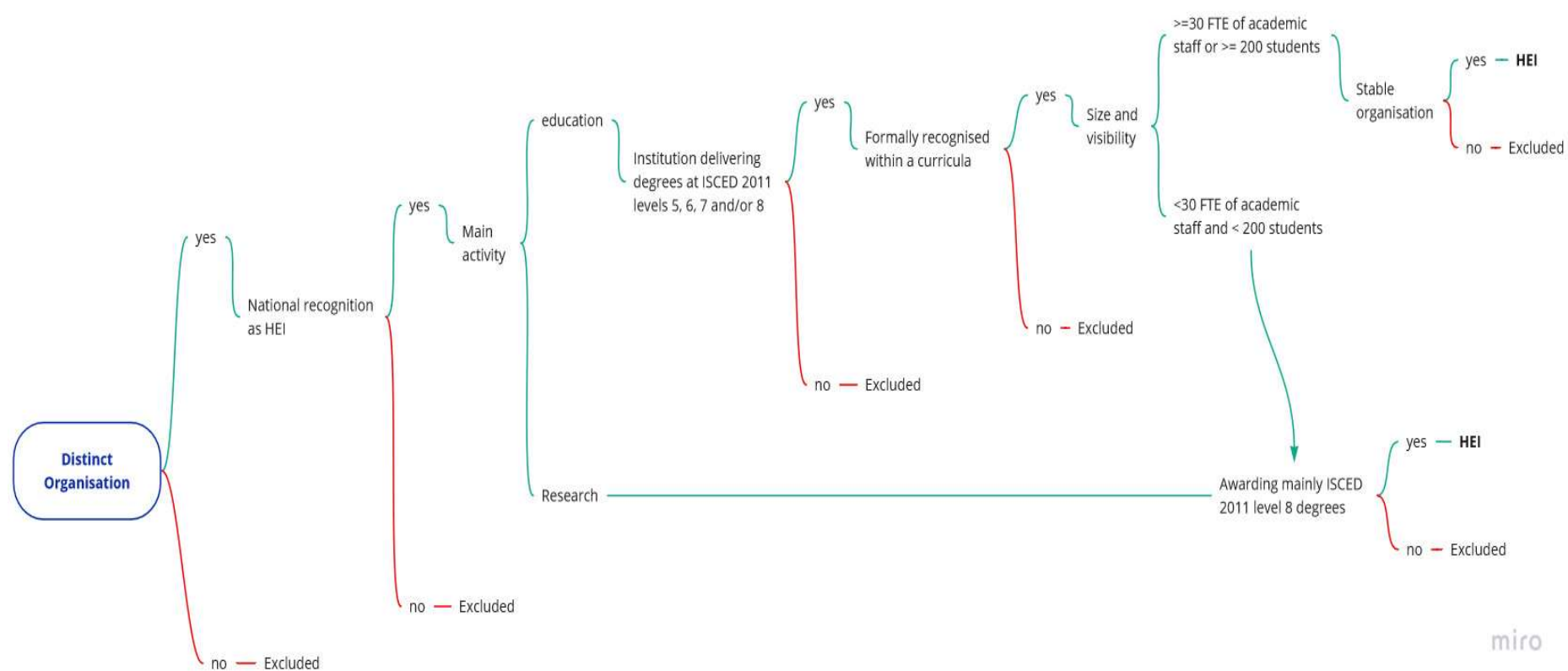
3.3.1 Exclusion cases

The following cases lead to exclusion from the EHESO-ETER perimeter:

- Institutions offering only services for education (but no curricula) are not included in the EHESO-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 secondary-level Institutions of Vocational Education and Training and 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 vocational 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 2: Decision tree for identifying institutions as HEIs



3.4 Demographic events

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. Proper tracking of demographic events is fundamental for keeping continuity of data series, for example in case of mergers.

Demographic events are recorded in the OrgReg register and then automatically copied (synchronized) in the EHESO-ETER data centre.

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 or changes of names, 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.

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. The 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 types:

- Events that imply the creation of entities (*birth*) or the closure of existing entities (*death*). Since they are unique to each entity, they are handled directly by attributing the corresponding foundation and closure years.
- Events that 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 that 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).

Table 1: 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.	New independent identifier (ID) is introduced in the case of births.
Death. The complete closure of activities of an existing institution.	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.	New independent identifier (ID) introduced analogously to births. Identifiers of antecedent HEIs are reserved and not reused.
Split of an existing institution into two or more independent institutions (two or more new institutions, new name, accreditation, maybe legal status).	New independent identifiers (ID) introduced analogously to births. Identifiers of antecedent HEI are reserved and not reused.
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 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. 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.	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. 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.

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 2).

Table 2: Examples of demographic events

Demographic event ID	Parent ID	Child ID	Year of demographic event	Type of demographic event	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

3.5 Types of instructional providers

Instructional providers are further regrouped into types, meaning providers, which have a similar status, legal mandate and educational offerings. The identification of types is therefore meant to reflect the structure of national educational systems as defined for example by its founding act, as for example described in the Eurydice reports on national systems⁸.

Examples of types are 'universities', 'colleges', 'academy of arts', etc.

The list of types for each country participating in the EHESO-ETER data collection is agreed by the project team with NSAs or national higher education ministries and is specific to each country. A key source to this aim are the so-called ISCED mappings within UOE data collection, which list by country the main educational programmes, their ISCED level and, in most cases, information on the related providers⁹.

HEIs in the EHESO-ETER main data collection are assigned to types through to a specific variable in the data collection, which is also used to aggregate institutional-level data at the level of whole types.

Other educational providers are assigned to types in the framework of the specific data collection by types as described in chapter 5.10.

⁸ <https://eurydice.eacea.ec.europa.eu/national-education-systems>

⁹ <https://isced.uis.unesco.org/data-mapping/>

4 Classifications

This chapter introduces the basic conventions and classifications that are used for the EHESO-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 EHESO-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.

4.1 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-F¹⁰) should be used at the first level (broad fields; see Table 3).

Please note that some older data are available based on the previous Fields of Education classification 1997; in that case, data for field 04 (Business, administration and law) and field 06 (ICT) are set to 'm' and following remark is included 'FOE-1997 used'.

Table 3: Fields of Education and training classification

Code	Name	Subfields
00	General programmes and qualifications	001 Basic programmes and qualifications 002 Literacy and numeracy 003 Personal skills
01	Education	011 Education
02	Humanities and Arts	021 Arts 022 Humanities 023 Languages
03	Social sciences	031 Social and behavioural science 032 Journalism and information
04	Business and law	041 Business and administration 042 Law
05	Natural sciences, mathematics and statistics	051 Biological and related sciences 052 Environment

¹⁰ <http://uis.unesco.org/en/topic/international-standard-classification-education-isced>

		053 Physical sciences 054 Mathematics and statistics
06	Information and communication technologies	061 Information & Communication Technologies
07	Engineering, manufacturing and construction	071 Engineering and engineering trades 072 Manufacturing and processing 073 Architecture and construction
08	Agriculture, forestry, fisheries and veterinary	081 Agriculture 082 Forestry 083 Fisheries 084 Veterinary
09	Health and welfare	091 Health 092 Welfare
10	Services	101 Personal services 102 Safety services 103 Security services 104 Transport services

4.2 Levels of education (ISCED)

Data on students and degrees will be divided by the level of education of the programme to which they are enrolled, using the International Standard Classification of Education (ISCED) in its 2011 version.

The table below provides the main categories and descriptions for the relevant levels for the EHESO-ETER data collection. For full details, please refer to the ISCED-2011 handbook¹¹.

Table 4: Levels of education

	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

¹¹ <http://uis.unesco.org/en/topic/international-standard-classification-education-isced>

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 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.	Duration: 2-3 years Entry requirements: ISCED 6 Usually: second degree at the tertiary level Direct access to ISCED 8 level
ISCED 7 Master or equivalent level long degrees	ISCED 7 long 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 programmes is presented separately in ETER (not included in ISCED 7) wherever possible, given their different characteristics.	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 programmes (not only courses).

4.3 Geographical classification

4.3.1 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 NUTS3¹². In EHESO-ETER, NUTS regions are derived directly from institutional addresses 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 EHESO-ETER.

4.3.2 Postcodes

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

¹² <https://ec.europa.eu/eurostat/web/nuts/>

Table 5: 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, Türkiye	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

4.4 Reference period for data collection

In principle, EHESO-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 6 below. Specific national departures from these practices should be detailed in the metadata.

Table 6: Reference periods

Variable	Reference period/date
Descriptors and geographical information	Last day of the calendar year (31 st of December)
Expenditures	Calendar year (1 st January – 31 st of December)
Revenues	Calendar year (1 st January – 31 st 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 beginning of the academic year.
Graduate 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, the year '2012' refers to the academic year 2011/2012 for students, 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 (for the latter, there are differences between countries).

4.5 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.
- Code 'm' refers to the fact that the data in question is missing.

Note that there is a difference between 'm' and '0', best described by the example of student and graduate data: If a degree (ISCED-level) is not offered, data should be '0' (before 2023, it was 'a'). If the degree is offered but data is not available, it should be 'm'. The information to interpret the '0' correctly can be taken from other variables, in this case lowest and highest degree delivered.

An 'unclassified' category is provided for all breakdowns (students, personnel, revenues, expenditures). This column should provide the number of cases that cannot be classified (or otherwise it should be '0'); when all data for a breakdown are missing, this column should also be 'm'.

There are two additional special codes shown only in the public dataset, which should not be inserted by data deliverers but will be generated automatically by the system (see chapters 4.8 on the treatment of small values and 7.5 on the data collection process for details):

- Code 'c' is shown in the public dataset for confidential data.
- Code 's' is shown in the public dataset for cell values smaller or equal to 3 to keep the anonymity of individuals.

In general, no blank cells are allowed in the data collection, except for 'Notes' and 'Flags'.

4.6 Data flags and notes

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 to quickly identify the reasons for the flags.

The EHESO-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 (see Table 7).

Table 7: 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 note 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 notes/metadata	There are specific conditions that imply that the value of a cell should be interpreted differently or not directly compared with others. The implications are explained in the note section and/or the metadata.
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 data centre.
c	confidential	When data are available but restricted to public 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 note field explains the source of the anomaly.

4.7 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 EHESO-ETER, this applies specifically for Latvia and Lithuania, where Euro was adopted on 01.01.2014 and 01.01.2015, respectively, as well as for Croatia (from 01.01.2023) and Bulgaria (from 01.01.2025). This implies that data in NC will show break in series due to change in the currency.

ISO_4217 codes for the currency are included in a separate field in the data collection (see Table 8) and reported in the country-level metadata in EHESO-ETER.

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

- *Euros* at nominal exchange rates from national currency to euro for non-euro countries based on national averages as published by EUROSTAT (<https://ec.europa.eu/eurostat/web/exchange-and-interest-rates>).
- *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/overview>).

4.7.1 ISO country and currency codes in EHESO-ETER

See Table 8 for a full list of ISO country and currency codes in the countries included in the data collection¹³. Some special cases to be mentioned: For the United Kingdom, the code UK is adopted. For Greece, the code GR is used in EHESO-ETER (compared to EL in EUROSTAT).

Table 8: 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
BA	Bosnia and Herzegovina	Convertible mark	BAM
BE	Belgium	Euro	EUR
BG	Bulgaria	Lev (pl. leva), Euro from 01.01.2025	BGN/EUR
CH	Switzerland	Swiss Franc	CHF
CY	Cyprus	Euro	EUR
CZ	Czechia	Czech Koruna (pl. koruny)	CZK
DE	Germany	Euro	EUR
DK	Denmark	Danish Krone (pl. kroner)	DKK
EE	Estonia	Euro	EUR

¹³ See http://www.iso.org/iso/country_codes

GR	Greece	Euro	EUR
ES	Spain	Euro	EUR
FI	Finland	Euro	EUR
FR	France	Euro	EUR
HR	Croatia	Kuna (inv.), Euro from 01.01.2023	HRK/EUR
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), Euro from 01.01.2015	LTL/EUR
LU	Luxembourg	Euro	EUR
LV	Latvia	Lats (pl. lati) · Euro from 01.01.2014	LVL/EUR
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	Türkiye	Turkish Lira (inv.)	TRY
UK	United Kingdom	Pound Sterling (pl. pounds)	GBP
XK	Kosovo	Euro	EUR

4.8 Treatment of small values

Small values (non-null and below 3 or below 5 depending on the country) might pose some issues in terms of data privacy, for example when an HEI enrolls only one female student at the ISCED5 level, this might allow her identification. Small values are treated in different ways depending on the country:

- Some countries allow for the publication of small values in the EHESO-ETER public dataset.
- Other countries allow for their publication in the restricted data (accessible upon registration), but these values are shown with the code 's' in the public dataset (implemented through color-coding by data collection coordinator, see section 7.5).

- Other countries do not allow for the publication of small values also in the restricted dataset, and accordingly they are replaced with 'm' on the EHESO-ETER dataset (typically done by the data deliverer, with support of data managers).

The handling of small values requires specific attention when dealing with breakdowns (such as by gender or by field of education and training) to avoid reconstruction of the original value. Specifically, the following rules apply:

- When a total is replaced with 'm' or 's', all its breakdowns are also replaced.
- Within breakdowns, at least two values have to be replaced, while keeping the others and the total. If there is only one small value within a breakdown, the unclassified category should be replaced.
- If only unclassified breakdown is a small value, it should not be replaced.

See Table 9 for some examples.

Table 9: Examples for treatment of small values

Breakdown	Original values	Replacement
Gender	Women: 2 Men: 10 Unclassified: 0 Total: 12	Women: s/m Men: 10 Unclassified: s/m Total: 12
Age	Under 20 years old: 2 Between 20 and 21 years old: 10 Between 22 and 24 years old: 8 Between 25 and 29 years old: 12 Over 29 years old: 7 Age group unclassified: 0 Total: 39	Under 20 years old: s/m Between 20 and 21 years old: 10 Between 22 and 24 years old: 8 Between 25 and 29 years old: 12 Over 29 years old: 7 Age group unclassified: s/m Total: 39
	Under 20 years old: 2 Between 20 and 21 years old: 10 Between 22 and 24 years old: 8 Between 25 and 29 years old: 12 Over 29 years old: 3 Age group unclassified: 0 Total: 39	Under 20 years old: s/m Between 20 and 21 years old: 10 Between 22 and 24 years old: 8 Between 25 and 29 years old: 12 Over 29 years old: s/m Age group unclassified: 0 Total: 39
	Under 20 years old: 2 Between 20 and 21 years old: 10 Between 22 and 24 years old: 8 Between 25 and 29 years old: 12 Over 29 years old: 0 Age group unclassified: 11 Total: 39	Under 20 years old: s/m Between 20 and 21 years old: 10 Between 22 and 24 years old: 8 Between 25 and 29 years old: 12 Over 29 years old: 0 Age group unclassified: s/m Total: 39

5 Variables

This chapter provides detailed information on the variables collected through the EHESO-ETER data collection, organised in the two levels described in chapter 7. Within the EHESO, these data are complemented with data from other data sources, such as international databases, EU data, etc. These data and their methodology is further described in the EHESO *Handbook for external institutional data sources*. Data are complemented with metadata described further in chapter 6.

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 4 for a detailed description of the classification schemes used in collecting these data.

5.1 HEI-level data collection

Table 10 lists the dimensions in which EHESO-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 10: Dimensions of EHESO-ETER variables

Dimension	Description
Identifiers and demographic information	These variables identify the HEI in the EHESO, matching information with other datasets and information on demographic events.
Institutional descriptors	Descriptive information on the HEI, such as legal status, foundation year, history, national label.
Geographic information	Address and geographical coordinates of the main and satellite campus.
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.
Research activities	Includes information on research activity, the number of ISCED 8 (PhD) students and graduates, R&D expenditures and researcher mobility.
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.
Revenues	Includes detailed breakdowns on HEI revenues (core budget, third-party funding, student fees funding, current and non-recurring revenues).
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.

5.2 Identifiers and demographic information

These variables provide unambiguous identification of institutions, information on demographic events, and matching information with other datasets.

All these variables are stored in the OrgReg register and copied automatically in the EHESO-ETER data. These are pre-validated in the ETER perimeter validation phase (see section 7.2 below).

Table 11: Variables included in identifiers and demographic information

Variable	Source
ETER ID	Pre-assigned
Reference year	Pre-assigned
ETER ID – Year	Pre-assigned
National identifier (optional)	OrgReg-NSA (validation)
Institution name (in own language)	OrgReg-NSA (validation)
English institution name (if available)	OrgReg-NSA (validation)
Acronym	OrgReg-NSA (validation)
Demographic event ID	OrgReg-NSA (validation)

5.2.1 ETER identifier

	Type of variable	Data source
ETER ID	ISO code (two-digit country code and four digits numeric)	OrgReg

Country code (ISO 3166) + integer (four digits code, e.g. AT0001). Identifiers are assigned in 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.

5.2.2 Reference year

	Type of variable	Data source
Reference year	four digits numeric	Pre-assigned

MMMM (e.g. 2011). The year to which the data in EHESO-ETER refer. For rules concerning the reference period, see chapter 4.4.

5.2.3 ETER ID – Year

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

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.

5.2.4 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 'a' as in 'not applicable'.

5.2.5 Legal name of the institution

	Type of variable	Data source
Legal name of the institution	text, using ISO/IEC 8859 encoding	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 (i.e. the name shown in EHESO-ETER relates to a specific year); changes of names across years without a change of IDs is possible when they imply a relabelling of the same institution (keeping its location, most of the activities, personnel, etc.). This also includes name changes related to changes in status (Hogeschool > University).

5.2.6 English translation

	Type of variable	Data source
Official English translation	text, using ISO/IEC 8859 encoding	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.

5.2.7 Institution acronym

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

The acronym of the institution as it 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 'm' as in 'missing'.

5.2.8 Demographic event ID

This variable is derived from OrgReg and informs whether the HEI has been affected by a demographic event as this is likely to affect also the data (for example in case of a take-over). It is provided for all HEIs affected by an event in the corresponding year; for example, if institutions CH0033 and CH0034 were merged to CH0035 on 01.01.2016, the event is recorded in 2015 for CH0033 and CH0034, i.e. the last year the figure in ETER, and for 2016 for CH0035, i.e. the first year in ETER. Details of the event can be retrieved directly from OrgReg.

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

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

5.3 Basic institutional descriptors

Table 12: Variables included in basic institutional descriptors

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

5.3.1 Country of establishment

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

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.

Following the practice of the UOE manual, 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).

5.3.2 Legal status

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

Following the UOE manual, 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.

5.3.3 Institution category (national classification)

	Type of variable	Data source
Institution category (national classification)	text	NSA

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. These categories are also used to aggregate ETER data by institutional type (see section 5.10).

5.3.4 Institution category (standardised)

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

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 or 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.

5.3.5 Year of establishment

The foundation year is highly relevant 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)	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 are collected if applicable:

	Type of variable	Data source
Legal status year	integer (four digits)	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)	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 'a' ('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.

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).

5.3.6 University hospital

Variables 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 criteria¹⁴:

- 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)	OrgReg, NSA might correct these data if relevant

Note: affiliated hospitals are only those for which the link with university as described in OrgReg is either ‘membership’ or ‘affiliated’; associated hospitals, without a strong connection with the HEI, are excluded.

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

The identifier in OrgReg. 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	OrgReg, NSA might correct these data if relevant

In English when available, otherwise national language.

¹⁴ See Elizondo, Andrea Reyes, Clara Calero-Medina, and Martijn S. Visser. ‘The three-step workflow: A pragmatic approach to allocating academic hospitals’ affiliations for bibliometric purposes.’ *Journal of Data and Information Science* 7.1 (2022): 20-36.

<https://sciendo.com/article/10.2478/jdis-2022-0006>

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

Including details on the nature of linkage.

5.3.7 Institutional website

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

Registration of the official website. *Format* [www.website_name](#). When no website is available, the code 'm' should be used.

5.4 Geographic information

These variables provide information on the HEI location, including city, postcode, and geographical coordinates; all of them are stored in OrgReg and validated by NSAs in the EHESO-ETER perimeter validation.

Table 13: Variables included in geographical information

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

5.4.1 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)	OrgReg

This variable provides 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 4.3.1.

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 localised according to the location of their main seat, independent of where the students live (and the existence of decentralized student support services).

5.4.2 Postal code

	Type of variable	Data source
Postal code	text	OrgReg

The postal code of the official address of the HEI. Different postal code systems throughout Europe are described in chapter 4.3.2. 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).

5.4.3 Name of city

	Type of variable	Data source
Name of city	text	OrgReg

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

5.4.4 Geographic coordinates

	Type of variable	Data source
Geographic coordinates	numeric	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.

5.5 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)	OrgReg

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

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

5.6 Educational activities and outputs

Table 14: Variables included in educational activities and outputs

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

This set of variables provides information on educational activities focusing on the enrolled students, and on educational outputs through graduations. It focuses on ISCED levels 5, 6 and 7, while PhD students (ISCED8) are included under 'research activities' (see section 5.7 below). All these variables conform to definitions in the UOE manual. Note that ISCED 7 long is a separate category and should not be included in ISCED 7 to avoid double counting. However, in some countries a separation of ISCED 7 (Master's) and ISCED 7 long degrees is not possible, which should be flagged with 'd' for 'definition differs' and provided with a respective note.

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' (before 2023: 'a').

5.6.1 Lowest and highest degree delivered

	Type of variable	Data source
Lowest and highest degree delivered	nominal	NSA

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.

This variable refers to the legal right of awarding a degree, for example based on national accreditation. It is still possible that in a year no degrees are awarded at one level.

5.6.2 Students

The number of students enrolled per institution at ISCED level 5, 6, 7 levels (with separate figures for ISCED 7 long 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 headcount of students.

In general, EHESO-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 equals 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 ISCED 7 long students and degrees cannot be separated, they should be included with the other ISCED 7 students and degrees, a respective flag and note is required in this case to inform the user (and the code 'm' is used for ISCED 7 long cells).

5.6.2.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. The unclassified category should also include non-binary students. If this breakdown is not possible, please use the appropriate special code (see chapter 4.5).

5.6.2.2 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 4.5).

5.6.2.3 Mobility

	Type of variable	Data source
Mobility	numeric	NSA

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 note should be added ('unclassified are distance education students based in other countries'). If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.6.2.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Students should be divided by fields of education using ISCED-F fields of education and training 2013 (see section 4.1). If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.6.2.5 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 whose 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 load. If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.6.2.6 *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 (most likely 'm', see chapter 4.5).

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

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

5.6.2.7 *Total number of students*

	Type of variable	Data source
Total number of students	numeric	NSA

The sum of students by ISCED level.

5.6.3 Graduates

Numbers of graduates per institution at ISCED level 5, 6, 7 levels by fields of education (with separate figures for ISCED 7 long 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 calculated automatically in the data collection sheet but can be overwritten, if necessary, by the data deliverer.

When long ISCED 7 graduates cannot be separated, they should be included with the other ISCED 7 graduates, whereas a respective flag and note is required in this case to inform the user (and the code 'm' is used for ISCED 7 long cells).

5.6.3.1 *Gender*

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. The unclassified category should also include non-binary students. If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.6.3.2 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 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 (most likely 'm', see chapter 4.5).

5.6.3.3 Mobility

	Type of variable	Data source
Mobility	numeric	NSA

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 (most likely 'm', see chapter 4.5).

5.6.3.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Graduates should be divided by fields of education using ISCED-F fields of education and training 2013 (see section 4.1). If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.6.3.5 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 (most likely 'm', see chapter 4.5).

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

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

5.6.3.6 Total number of graduates

	Type of variable	Data source
Total number of graduates	numeric	NSA

The sum of graduates by ISCED level.

5.6.4 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 programmes).
- 1 = a substantial share of distance education. An example is Anadolu University in Turkey.
- 2 = mostly distance education (>90% of the student in distance programmes). Typical examples are Fernuniversität Hagen or the UK Open University.

5.7 Research activities

This section includes some core data on research activities collected from official statistics; additional information on publication output and research projects is provided in the EHESO from international databases.

Table 15: Variables included in research activities

Variable	Data source
Research active institutions	NSA
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

5.7.1 Research active institution

	Type of variable	Data source
Research active institutions	Binary (0=non-research active; 1=research active)	NSA

Research-active institutions are those that have institutionalised research activities. Criteria for inclusion are 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.

5.7.2 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 to reduce the burden for data collection.

5.7.2.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. Non-binary students should be included under unclassified. If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.7.2.2 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 (most likely 'm', see chapter 4.5).

5.7.2.3 Mobility

	Type of variable	Data source
Mobility	numeric	NSA

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 (most likely 'm', see chapter 4.5).

5.7.2.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Students should be divided by fields of education using ISCED-F fields of education and training 2013 (see section 4.1). If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.7.2.5 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 load. If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.7.2.6 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 (most likely 'm', see chapter 4.5).

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

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

5.7.2.7 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.

5.7.3 ISCED 8 graduates

This variable asks for the number of degrees delivered at ISCED 8 qualification. It excludes intermediate-stage ISCED 8 programmes 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 to reduce the burden for data collection.

5.7.3.1 Gender

	Type of variable	Data source
Gender	numeric	NSA

A breakdown between men and women. Non-binary graduates should be included under unclassified. If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.7.3.2 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 (most likely 'm', see chapter 4.5).

5.7.3.3 Mobility

	Type of variable	Data source
Mobility	numeric	NSA

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). If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.7.3.4 Fields of Education

	Type of variable	Data source
Fields of Education	numeric	NSA

Graduates should be divided by fields of education using ISCED-F fields of education and training 2013 (see section 4.1). If this breakdown is not possible, please use the appropriate special code (most likely 'm', see chapter 4.5).

5.7.3.5 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 (most likely 'm', see chapter 4.5).

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

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

5.7.3.6 Total number of graduates

	Type of variable	Data source
Total number of graduates	numeric	NSA

The sum of graduates at level ISCED 8.

5.7.4 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 set to null, the corresponding R&D expenditures variable should be 'm'.

5.8 Revenues and expenditures

5.8.1 General definitions and principles

Financial data in EHESO-ETER refer to the HEI's own budget using concepts similar to the ones adopted for companies. This follows the idea that most HEIs in Europe now have a budget and some autonomy in managing their revenues and expenditures. The focus on HEI's budget is a major difference compared to the UOE data collection, which is mostly concerned with measuring national investment in tertiary education. Accordingly, ETER has developed its own definitions in this area.

More specifically, financial data in EHESO-ETER deal with two dimensions:

- *Revenues*, i.e. the amount of money available to the HEI from different revenue sources, 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 a HEI to manage its activities, for example, for paying its employees, external services, general expenditures and maintenance costs, as well as capital costs.

While some correspondence between revenues and expenditures is generally expected, 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. Differences might be also due to the treatment of capital costs, as explained below. However, systematic differences over many years will be checked as they might signal methodological problems.

Accounting systems. EHESO-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 (see section 4.7).

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.

5.8.2 Structure of the EHESO-ETER financial information

The structure of the EHESO-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 despite differences in HEIs' accounting systems.

Table 16: Classification of revenues and expenditures

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

This takes into account differences in how non-recurring expenditures and revenues, particularly those associated with investments, are treated: 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 and revenues also include investment contributions from the State.

To illustrate these differences, Table 17 compares of expenditures and revenues for two HEIs with the same level of current expenditures and different accounting systems. 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 17: 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

5.8.3 Expenditures

Table 18: 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

5.8.3.1 Current expenditure

	Type of variable	Data source
Current expenditure	numeric	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).

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. Specifically, all patient care costs and other general expenses of academic hospitals, even if the education authorities pay such expenses, are excluded. 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.

5.8.3.2 Personnel expenditure

	Type of variable	Data source
Personnel expenditure	numeric	NSA

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

5.8.3.3 Non-personnel expenditure

	Type of variable	Data source
Non-personnel expenditure	numeric	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.

5.8.3.4 Expenditure unclassified

	Type of variable	Data source
Expenditure unclassified	numeric	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.

5.8.3.5 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').

5.8.3.6 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.

5.8.3.7 R&D Expenditure

	Type of variable	Data source
R&D Expenditure	numeric	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').

5.8.4 Revenues

Table 19: 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

5.8.4.1 Current Revenue

	Type of variable	Data source
Current Revenue	numeric	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 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). Table 20 maps these classifications to the variables collected in ETER.

Table 20: 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)	

5.8.4.2 Core funding

	Type of variable	Data source
Core funding	numeric	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, EHESO-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'.

5.8.4.3 Basic government allocation

	Type of variable	Data source
Basic government allocation	numeric,	NSA

Basic 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.

5.8.4.4 Other core funding

	Type of variable	Data source
Other core funding	numeric	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'.

5.8.4.5 Third-party funding

	Type of variable	Data source
Third-party funding	numeric	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 or 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'.

5.8.4.6 Public third-party funding

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

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

5.8.4.7 Private third-party funding

	Type of variable	Data source
Private third-party funding	numeric	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, as well as households.

5.8.4.8 Third-party funding from abroad

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

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

5.8.4.9 Third-party funding undivided

	Type of variable	Data source
Third-party funding undivided	numeric	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).

5.8.4.10 Student fees funding

	Type of variable	Data source
Student fees funding	numeric	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.

5.8.4.11 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.

5.8.4.12 Revenue unclassified

	Type of variable	Data source
Revenue unclassified	numeric	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'.

5.8.4.13 Non-recurring revenue

	Type of variable	Data source
Non-recurring revenue	numeric	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'.

5.9 Personnel

These variables provide information on the personnel employed by higher education institutions. Conforming to the UOE manual, HEI personnel includes all persons employed by the HEI and whose activities are required for HEI operations.

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. Any exclusions (for example based on an employment threshold) should be indicated in the metadata.

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.

While following the UOE definition, the EHESO-ETER data collection comprises additional breakdowns of academic personnel by seniority categories and citizenships, as well collects data also for research and teaching assistants and non-academic personnel.

Table 21: 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

5.9.1 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. 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 programmes 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 and of employed PhD candidates.

5.9.2 Measurement units

Two measurement units are adopted for HEI personnel, namely headcounts (HC) and full-time equivalents (FTE).

5.9.2.1 Numbers of personnel in headcounts (HC)

The UOE manual foresees 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.

5.9.2.2 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.

5.9.3 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.

Breakdowns are required only in headcounts. If they are not available, data should be 'm'.

5.9.4 Employed PhD candidates

Employed PhD candidates are an important component of HEI personnel: these individuals are enrolled as students at ISCED8 level, for obtaining a doctoral degree, and are frequently employed full-time or part-time with different tasks in research and educational support.

There is variation in how these tasks are defined and, accordingly, how employed PhD candidates are treated in personnel statistics. In many cases, the individuals work in research projects funded by their supervisor and enjoy of considerable autonomy in the conduct of research. In other cases, their support tasks are additional and independent from the PhD thesis. Accordingly, some countries include employed PhD candidates in academic personnel, others in Research and Teaching Assistants. Moreover, not in all countries employed PhD candidates can be singled out as personnel category.

To account for this situation, the EHESO-ETER data collection includes two variables in personnel data collection for employed PhD candidates, under academic personnel and under RTAs, as well as a specific flags. Notes should be left to provide additional information (use the 'Notes on academic personnel' field).

Table 22: Inclusion of employed PhD candidates in personnel data

	Employed PhD candidates included in academic personnel	Employed PhD candidates included in Research and Teaching Assistants
Included in academic personnel as a separate category	Numeric value	0
Included in academic personnel but cannot be separated.	code 'm' and flag 'i'*	0
Included in RTA as a separate category	0	Numeric value
Included in RTA, but cannot be separated	0	code 'm' and flag 'i'*
No information	code 'm' and flag 'i'*	code 'm' and flag 'i'*

*details to be provided in the specific remarks field.

5.9.5 Academic personnel

5.9.5.1 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.

5.9.5.2 *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).

5.9.5.3 *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. Non-binary personnel should be included in the unclassified category. If this breakdown is not possible, this variable should be coded as 'm'.

5.9.5.4 *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, this variable should be coded as 'm'.

5.9.5.5 *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 4.2). If this breakdown is not possible, this variable should be coded as 'm'.

5.9.6 *Academic personnel by seniority level (HC)*

Academic personnel is classified by seniority level following the classification proposed by OECD-INES in cooperation with EUROSTAT.¹⁵ 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.

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

¹⁵ EDU/EDPC/INES/WP(2022)5.

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), and post-doctoral researcher.

	Type of variable	Data source
Number of employed PhD candidates included in academic personnel (HC) by gender	numeric	NSA

If employed PhD candidates are not included in academic personnel, this variable should be set to null. If there are included in other categories of academic personnel, such as the junior or the other category, it should be set to 'm' and a note should be added.

	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 as well as 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'.

5.9.7 Research and teaching assistants

5.9.7.1 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.

5.9.7.2 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).

5.9.7.3 Employed PhD candidates in Research teaching Assistants (HC)

	Type of variable	Data source
Employed PhD candidates in Research teaching Assistants (HC)	numeric	NSA

The total number of employed PhD candidates that are included in research and teaching assistants in HC. If employed PhD candidates are included in academic personnel, this variable should be null. If their number is unknown, this variable should be 'm'.

5.9.7.4 Other Research and Teaching Assistants (HC)

	Type of variable	Data source
Other Research and Teaching Assistants (HC)	numeric	NSA

The number of Research and Teaching Assistants who are not of employed PhD candidates. If RTA cannot be broken down between employed PhD candidates and other, this variable should be 'm'.

5.9.8 Support and administrative personnel

5.9.8.1 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.

5.9.8.2 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).

5.9.9 Total number of personnel

5.9.9.1 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.

5.9.9.2 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.

5.10 Supplementary data collection by institutional types

This data collection completes the main EHESO-ETER data collection with few additional variables covering also the educational institutions not included in the main EHESO-ETER data collection. It has two main functions:

- To allow more precise reconciliation of national totals for key variables with EUROSTAT data.
- To provide a quantitative summary view of the national tertiary education system and their structure by types of institutions, thereby complementing the country fiches provided by the Eurydice network¹⁶.

Following variables are collected:

- The number of educational providers belonging to a type.
- The total number of students enrolled at ISCED levels 5, 6, 7 and 8 in the reference year.
- The number of graduates at ISCED levels 5, 6, 7 and 8 in the reference year.
- Total academic staff in Headcounts in the reference year.

All definitions conform to those in the EHESO-ETER main data collection. As described further in section 7.3, data for the HEIs included in the EHESO-ETER main data collection are automatically computed using the English institutional category variable (national level, see 5.3.3). The institutional type data can be accessed via API or directly via the following link (download starts automatically): <https://www.eter-project.com/api/4.0/institution-type-data/export>.

¹⁶ <https://eurydice.eacea.ec.europa.eu/>

6 Data validation and metadata

The EHESO-ETER data collection includes several steps of data validation and documentation through metadata that conform to well-established practices at EUROSTAT and National Statistical Agencies. Additionally, ETER introduced over the year statistical approaches for more refined validation steps such as longitudinal analyses¹⁷.

Validation has different purposes:

- First, to detect mistakes in the EHESO-ETER data collection process such as swapping lines and rows in the data collection files, wrong currency units, etc.; these mistakes lead to corrections the EHESO-ETER data.
- Second, to detect problems and inconsistencies in the primary data at NSAs. These issues might lead to correction in the primary data, but more frequently to annotating and flagging the data (for example when they are generated by differences in definitions and data collection practices).
- Third, to document extensively through metadata (at the country level), and flags and remarks (at the HEI level) deviations and issues encountered; this is information is essential for the purposes of data analysis.
- Fourth, to document the overall quality of the EHESO-ETER dataset based on a set of standard indicators.

As described, further in chapter 7, these checks are implemented in a cascading approach where the simple checks are performed already in the data collection sheets and more complex data checks are performed in the data centre. This ensures that simpler mistakes are corrected as soon as possible in order to avoid unnecessary feedback loops.

A full list of these checks is maintained by the EHESO-ETER team in a separate file.

6.1 Internal validity checks

This set of checks verifies the internal validity of the EHESO-ETER data as referred to the definitions provided by this handbook. This includes following set of checks:

- *Format accuracy*, i.e. verifying that data entered have the right foreseen by the handbook and that no logically impossible values are found, e.g. letters or other codes for numeric variables and codes not foreseen in the variable definition for categorical variables.
- *Blank cells checking*. No blank cells are allowed in the dataset except for notes and flags. Blanks should be recoded with a special code (see 4.5) or with zero.
- *Consistency checks*. 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 level should be 0). When logical inconsistencies are justified (for example HEIs enrolling students, but not awarding the corresponding degree), a flag ('ic' in the example) and a note should be added.

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). A special flag 'rd' is foreseen for cases where inconsistencies are generated by rounding.

¹⁷ The approach developed in the past ETER contracts is documented in Bruni, Renato, Cinzia Daraio, and Simone Di Leo. "Flexible Techniques to Detect Typical Hidden Errors in Large Longitudinal Datasets." *Symmetry* 16.5 (2024): 529.

These checks are implemented already in the data collection files as highlighted cells (see further in chapter 7.4) in order to allow for immediate detection and correction. They are repeated when importing the data in the data centre.

6.2 External validity

External validity refers to checks comparing the EHESO-ETER data with external data sources; large disagreement might be indicative of data problems or of underlying differences in definitions.

External validity is undertaken at different levels in the ETER data collection process:

- A first level of external checks is systematically performed during the data collection process in order to compare EHESO-ETER data, aggregated at the national level, with the same data as published by EUROSTAT. Since most definitions are identical and most EHESO-ETER data are derived from the same data sources, these checks are highly useful in order to detect any issues in the data collection process.
- A second level of external checks is mobilized case by case by the EHESO-ETER team when other checks casts doubts on specific data points, such as revenue or personnel data for individual HEIs. The suspect data points may be crosschecked with information from national databases, institutional websites and reports, as well as public sources such as Wikipedia. Given that these checks are very time-consuming, they are undertaken only for individual cases detected with other methods.
- A third level of external checks are feedback by users, which remark potential data problems when using EHESO-ETER data. These might be revealed by statistical analysis, by comparison with other sources and by detailed knowledge of specific national or institutional contexts.

6.2.1 EUROSTAT checks

External validation with EUROSTAT is made with reference to two groups of variables:

- Totals for main variables such as students, graduates and personnel. These are compared with the data provided in the additional data collection by institutional type (see section 5.10) since this is expected to provide a more complete coverage.
- Indicators for major breakdowns such as gender and mobility. These compare the published indicators by EUROSTAT with those computed in EHESO-ETER from original data.

The former are revealing systematic mistakes (such as the use of wrong units) and differences in the data collection methodology. They might also be related to systematic differences in EHESO-ETER vs. EUROSTAT coverage. The latter are particularly useful to detect problems in the data collection process such as column swaps.

Table 23: Eurostat checks

Data category	Totals checks	Breakdowns checks
Students	Total students by ISCED level (5, 6, 7). ISCED7 includes also ISCED7 long.	Share of female students. Share of mobile students. Share of students by FET. All by ISCED level.
Graduates	Total graduates by ISCED level (5, 6, 7).	Share of female graduates. Share of mobile graduates.

	ISCED7 includes also ISCED7 long.	Share of graduates by FET. All by ISCED level.
Academic personnel	Total academic personnel HC.	Share of female academic personnel.
Research	Total students and graduates at ISCED8.	Share of female/mobile/by FET students and graduates at ISCED8.

Remark: depending on data availability, checks might be based on previous years' EUROSTAT data.

6.3 Outlying values

Outlier detection identifies values outside the normal range of variation for a set of ratios between variables, which might indicate possible data problems. The analysis is implemented for a set of ratios between basic variables, such as students per staff, and of indicators, such as the share of foreign students.

While it is acknowledged that critical values of the ratios can be affected by explanatory variables, such as the HEI's legal status, the approach adopted in the EHESO is a simple one, where the top 1% and bottom 1% of the cases for each ratio are singled out and examined. When a factual explanation of a case can be provided, the corresponding observation is added a specific flag to drop the case from the following round of checks.

The list of ratios is listed below in Table 24.

Table 24: Outlier checks

Type	Ratio	Remarks	Flag field
Basic ratio	students ISCED 5-7 / academic staff FTE		
	academic staff FTE / total staff FTE	Only lower bound	
	personnel exp (PPP) / total staff FTE		
	personnel exp (PPP) / total exp (PPP)		
	total exp (PPP) / total rev (PPP)		
	graduates ISCED 5-7 / students ISCED 5-7		
	graduates ISCED 8 / students ISCED 8		
Indicators	Share of foreigners students ISCED 5-7	Only upper bound	
	Share of mobile students ISCED 5-7	Only upper bound	
	Share of mobile students ISCED 5-7/ Share of foreigners students ISCED 5-7		
	Share of foreigners graduates ISCED 5-7	Only upper bound	
	Share of mobile graduates ISCED 5-7	Only upper bound	
	Share of foreign PhD students	Only upper bound	
	Share of mobile PhD students	Only upper bound	
	Share of mobile PhD students/ Share of foreign PhD students		
	Share of foreign PhD graduates	Only upper bound	
	Share of mobile PhD graduates	Only upper bound	
	Share of mature students ISCED 5-7		
	Share of mature graduates ISCED 5-7		
	Herfindahl index students ISCED 5-7		
	PhD intensity		

Outlier detection is applied to the last three years of date in order to detect whether the outlier is specific to a single year or recurrent.

The list of outliers is produced from the EHESO-ETER database. The output will include:

- The ratio considered
- The eterid
- The reference year
- Total student number ISCED 5-7 as a measure of size.
- The numerator (when applicable)
- The denominator (when applicable)
- The ratio value
- The corresponding flag field and remarks.

Outliers are first checked by the EHESO-ETER core team in order to:

- Analyse whether there is a factual explanation.
- Identify and correct simple mistakes (like column swaps).
- Identify deviant cases that need to be reported to national experts for checking.

In order to avoid checking always the same case, it is essential that detected cases are systematically flagged, particularly when an outlier is explained by substantive reasons, such as students to staff ratio in the case of distance universities.

Following flags should be used:

- Flag p (provisional) if the cases is clearly deviant, but cannot be explained at this stage. This is informative that the case was already detected.
- Flag o (outlier) when the case is deviant, but there is a clear substantive explanation, like distance universities or large share of foreigners in small countries.

6.4 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.

Multiannual checks are performed on a small set of variables, which are expected to show a high level of stability over time (see Table 25) and at two analytical levels, i.e. whole countries and individual HEIs.

Table 25: 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)

6.4.1 Country-level checks

Large aggregates such as total students enrolled by country are expected to be very stable over time and accordingly, these checks provide a first-level control as they might reveal systematic mistakes and changes of methodology for whole countries. Some changes might also be related to changes in data coverage.

These checks are performed in a descriptive manner by providing time series of totals for each country in the data validation reports.

6.4.2 HEI-level checks

HEI-level checks rely on statistical analysis to identify potential deviant cases to be further examined.

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:

- 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:

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

Hence we have $GM \geq S1$.

- Compute the discontinuity measure (DM) as: $DM = \frac{DV}{PGM}$
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.
- Compute the variance of the annual delta measure (VM) as: $VM = \frac{VD}{PGM}$
Where VD is the variance of the deltas and PGM is the powered geometric mean.
- On the relevant sample (with $GM \geq S1$), identify the HEIs that have:
 - the discontinuity measure (DM) greater than a predetermined percentage (e.g. 10%) of highest values (S2). If the calculated DM value is $\geq S2$ then set Alarm 1=TRUE;
 - 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 $\geq S3$ then set Alarm 2=TRUE.
- 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.

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).

Based on this inspection, country data managers have to decide whether cases should be subject to further scrutiny and explanations requested to NSAs. Typical examples of non-relevant cases are HEIs with small numbers of students or personnel, and cases already explained, for example by demographic changes. In addition, cases with a regular linear trend of growth occur regularly and are non-problematic.

On the contrary, particularly suspect cases are large variations for large institutions, and deviant values for a single year within an overall stable time series.

6.5 Metadata

According to the ISO standard, *metadata* are defined as data that defines and describes other data and processes. The EHESO-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 EHESO-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 should be reported in the 'notes'-fields of the different data fields.

Table 26: 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

Note: The metadata system has been under internal evaluation and will be simplified in future data collections. Since EUROSTAT is currently also working on a revision of the metadata system, the EHESO project team will coordinate its metadata actions with EUROSTAT.

6.5.1 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 EHESO-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 EHESO-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 EHESO-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 as described in section 4.4. Please be careful to update this information for the current year of data collection.

	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 EHESO-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.

6.5.1.1 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.

6.5.2 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.

6.5.2.1 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 EHESO-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 EHESO-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 EHESO-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 EHESO-ETER breakdown and not included at all, if existing, together with any other relevant remark and additional information.

6.5.2.2 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 EHESO-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 EHESO-ETER breakdown and not included at all, if existing, together with any other relevant remark and additional information.

6.5.2.3 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 EHESO-ETER.

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

6.5.2.4 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.

6.5.2.5 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.

6.5.2.6 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.

7 Data collection and validation process

7.1 Overview

The EHESO-ETER data collection and validation process is organized as a collective effort of the project team and the national statistical authorities of participating countries. The roles are defined as follows:

- **Data collection coordinator:** For the EHESO-ETER data collection, this role is taken over by JOANNEUM RESEARCH. The coordinator is responsible for providing templates, guidelines and doing final data checks. JOANNEUM RESEARCH also hosts and manages the EHESO data centre and OrgReg database.
- **Data managers:** These include all EHESO-ETER project team members responsible for managing the data collections for one or several countries. They are responsible for coordinating the data collection in their respective countries and doing first data checks (in Excel). In some countries, **national experts** serve as a bridge between data managers and data deliverers.
- **Data deliverers (data owners):** These include all data delivering institutions, which usually are National Statistical Offices and (in a few cases) Other National Authorities. They are responsible for validating the perimeter and filling in the data collection sheets.

All files and additional documentation in the EHESO-ETER data collection process are shared with data managers and data deliverers via **Nextcloud**, a self-hosted cloud service of JOANNEUM RESEARCH. Open links to the country folders are provided by the data managers when starting the perimeter validation process.

The following chapters will give an overview of tasks and responsibilities in each step of the data collection and validation process. The process starts with validating the perimeter (chapter 7.2), then goes over to the main data collection phase (chapter 7.3) and the data validation process (chapter 7.4). The final chapter covers disclosure agreements and data publication (chapter 7.5).

7.2 Perimeter and descriptors validation

The perimeter and descriptors validation process starts from the previous round of data collection and is first crosschecked with institutions and demographic events in the register of public organisations (OrgReg), in order to provide a list of suggested institutions to the data deliverers. The goal of the perimeter validation process is to move from a suggested list of institutions to a validated list for the specific year for each country. More specifically, the process looks as follows:

- The **data collection coordinator** creates the perimeter files (i.e. lists of higher education institutions) from the previous round of data collection and crosschecks information in OrgReg.
- The data collection coordinator then shares the perimeter files with the **data managers** responsible for each country together with decision guidelines for the inclusion or exclusion of higher education institutions. Data managers coordinate the validation process with their contact persons within each country (**data deliverers**). In some countries, data managers receive support from **national experts** (NEs).
- **Data deliverers** check the perimeter file and mark an 'x' for all institutions to be included in the new round of data collection. They also provide information on demographic events (see chapter 3.4) and leave additional notes for HEIs that are not included in EHESO-ETER

(as a reference). They should also fill in missing information (additionally included indicators, see below) for HEIs new in EHESO-ETER.

- After the perimeter is validated, the **data collection coordinator** (with the help of data managers) implements all necessary changes (e.g. name updates, demographic events) in OrgReg in order to keep the register up-to-date. In the end, the data collection sheets are created based on the perimeter files.

Besides a list of higher education institutions and options to fill in demographic events, the perimeter files also include additional data, specifically most institutional descriptors (see 5.3). These are taken from EHESO-ETER and are included because they stay mostly constant over time. For institutions already in EHESO-ETER, the data is prefilled. For new institutions included in the EHESO-ETER data collection, these data should be filled in.

After the perimeter is validated and all required information is filled in, it is uploaded to the data centre. Indicators included in the perimeter file as well as indicators synchronized from OrgReg are prefilled. All other data is coded as 'm' and should be filled in/replaced during the main data collection. The data collection coordinator then uses the web application to export data collection sheets directly from the data centre.

7.3 Main data collection and data collection by types

As described above, the data collection files are created automatically from the EHESO data centre and saved in the respective country folder on Nextcloud by the data collection coordinator. All obligatory cells are **prefilled with 'm'**. Data deliverers should replace 'm' with the actual data or leave 'm' in case of missing data.

Please note the difference between 'm' and zero: Data deliverers should replace 'm' with 0 if data is actually zero and not missing (e.g. if ISCED-level is not offered, see information on warning colours below).

Data collection sheets include the following tabs (see chapter 5 on details concerning the data collected), which should be filled in/checked by the **data deliverer**:

Table 27: Tabs and data in the data collection sheet

Tab in Excel	Type of data	Required action
Metadata_collected	metadata taken from previous data collection	data should be updated/completed
In-depth_information	metadata taken from previous data collection	data should be updated/completed
Basic_institutional_descriptors	descriptors prefilled from previous data collection and additional indicators synchronized from OrgReg	In case of changes, data should be marked green and data managers should be informed
Geographic_information	descriptors synchronized from OrgReg	In case of changes, data should be marked green and data managers should be informed
Expenditures	collection data prefilled with 'm'	data should be inserted
Revenues	collection data prefilled with 'm'	data should be inserted
Personnel	collection data prefilled with 'm'	data should be inserted

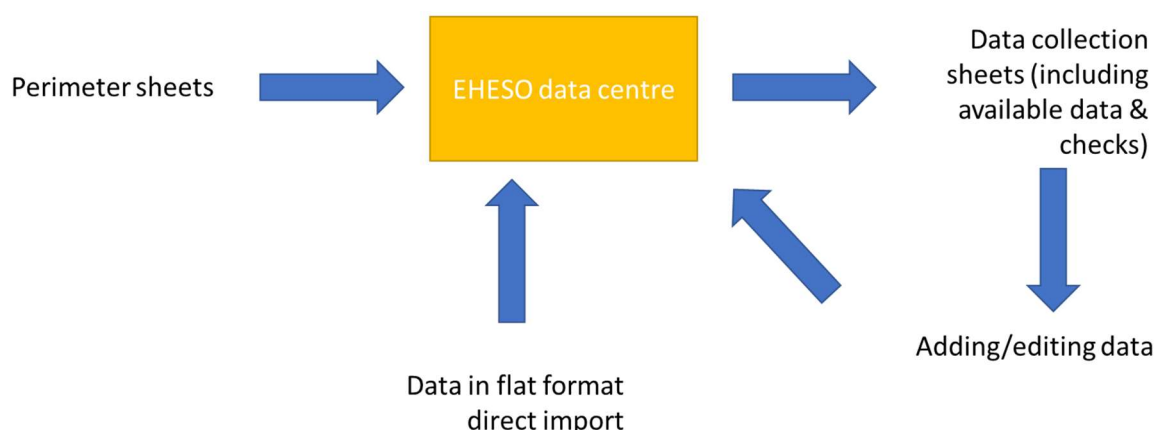
Education – Students	collection data prefilled with ‘m’	data should be inserted
Education – Graduates	collection data prefilled with ‘m’	data should be inserted
Research	collection data prefilled with ‘m’	data should be inserted
InstTypes (new!)	totals calculated from main data collection (other tabs)	data should be checked, missing types and potentially notes should be added (see details below)

Institutional type data is collected together with the main data collection (i.e. in the same Excel sheet). Totals by type are calculated automatically based on the inserted data¹⁸. **Data deliverers** should add missing types and the respective sums for data not included in EHESO-ETER in order to fill the gap between the EHESO-ETER and EUROSTAT data collections (gap in coverage). In case of methodological particularities, respective notes should be added. For details on the variables collected as well as more background information, see chapter 5.10.

7.3.1 Procedures and options for data deliverers

Figure 3 presents an overview of the whole data collection process and of the available options.

Figure 3: Overview of data collection process



Once the perimeter sheets have been prepared and imported in the data centre, data deliverers have **two options** to deliver data:

1. *Export the collection sheet for the country and reference year, complete data and deliver the data collection sheet for import via Nextcloud.*

This is the main way of delivering the data. Automatic sum calculation and checks in the Excel sheet (more on this in chapter 7.4) help the data deliver to identify potential issues immediately.

2. *Provide data in a flat format using simple .xlsx templates.*

¹⁸ Formulas work correctly as long as the order of institutions is the same in each tab of the Excel sheet. If institutions are reordered to insert data more easily, simply go back to the initial order after data is inserted.

The benefit of the flat (or 'additional') import option is that only selected variables and/or selected institutions can be imported at a time and that output from NSA databases can be reused directly. This option is also very useful if a single variable needs to be corrected. To be correctly imported, delivered data must include the 'ETER ID – Year' (in the format 'BAS.ETERIDYEAR') in the first column, followed by the selected variables to be imported. Rows contain all institutions for which the respective variables are available.

If this option is of interest, data deliverers should contact the data manager in order to discuss whether the flat option may be suitable. The data manager will then provide templates and guidelines on how to use it.

As data is prefilled with 'm' in the data centre as well as in the main data collection sheet, delivering and importing data to the data centre in batches (i.e. for a group of variables and/or institutions) is possible through either option. Re-exporting the whole data sheet after using the flat import option allows for running the validation checks in Excel; this is strongly advised when data are imported in batches.

7.4 Data validation process

The data validation process is a crucial instrument for ensuring high data quality. The performed checks are described in chapter 6).

The process consists of three steps:

1. *Data validation checks in the data collection files, giving immediate feedback to the data deliverer.*

The data collection files (Excel files) include a first wave of basic data validation checks by using conditional formatting. In case of incomplete values, format accuracy issues (e.g. wrong data formats or logical inconsistencies) or mistakes in the comparison between sums of breakdowns and totals, the **data deliverer** is notified to check these cases again. Such alarms should lead either to corrections or to flagging of the data. If data is correct, data collection files of previous years should be checked as issues typically reoccur. The same flags and notes can be used.

The following warning colours are used in the data validation checks:

- **Yellow:** Obligatory data are missing
- **Orange:** Accuracy problem (e.g. wrong format, logical inconsistencies, etc.)
- **Red:** Mistakes in sums

These checks have proven to reduce the number of feedback-loops required in the data validation process significantly. Note that these checks are not available when using the flat import option, which is why re-exporting the data in the main data collection sheet is advised.

Data managers are responsible to follow up on any unaddressed/unflagged warnings in the Excel files before notifying the **data collection coordinator** (JOANNEUM RESEARCH) to conduct final checks and import the data to the data centre.

2. *Data validation during data upload, allowing for more extensive data checks.*

Once data is imported to the data centre, a more detailed and automated data validation process starts. Using data validation scripts, more sophisticated tests can be implemented compared to the checks done in Excel. This process has so far been implemented by the **data collection coordinator** and will be moved more towards **data managers** in future data collections.

The uploaded data will go into a transformation table, where two types of checks are implemented:

- Automatic checks executed on each field (e.g. by field types), which exclude impossible formats and codes.
- Dynamic check scripts, which control for missing values, logical inconsistencies and mistakes in sums. The data collection coordinator maintains these scripts and can adapt them anytime.

Detected issues will be reported to the responsible **data manager**, while clean data will be uploaded into the data centre. Data managers will contact data deliverers in case of unsolved issues. As a rule, detected cases need either to be solved (i.e. not be detected in a new data validation run) or flagged, including a note to explain the case.

3. Extensive data checks including longitudinal analysis and check with external data.

The extensive data checks are executed at a later stage country by country, when the full dataset has been collected for a country. These checks focus on the following areas:

- **Multiannual (longitudinal) checks:** these look at validity over time and breaks in time series.
- **Outlier detection:** these look whether values are classified as outliers and hence checked.
- **Comparison to EUROSTAT:** these external data checks compare coverage of selected indicators in the UOE and EHESO-ETER data collections (always done one year later, as typically EUROSTAT data is published later than EHESO-ETER data).

Whenever data is updated or corrected, it undergoes the same data validation process. It is further important to note that there are also **other sources** of data checks and updates, including for example problems detected by data users. Therefore, it is important to use the **most current version** of the data by exporting the data collection sheet from the data centre before making adjustments. If updated data is available, data deliverers should contact the respective data manager to provide the correct sheet. Old data collection files on Nextcloud are kept only as a reference and should not be reused.

7.5 Open data policy and data publication

EHESO follows the European Union open data policy as implemented by the Commission Decision of 12 December 2011 on the reuse of Commission documents¹⁹ and as adopted also by EUROSTAT²⁰. More specifically, data and metadata provided by the EHESO are delivered under the Creative Commons BY 4.0 license terms²¹ and can be reused without any payment or written licence provided that:

- the source is indicated as EHESO;
- when re-use involves modifications to the data or text, this must be stated clearly to the end user of the information.

It is expected that EHESO-ETER data conform to these licensing conditions, as most data are also available from national sources and subject to open data policies. However, there might be some exceptions where restrictions are required, including specific HEIs (for example private universities) and variables (for example financial information). Restrictions might be also required to protect anonymity for students and graduates, when numbers are low (smaller or equal to 3). All restrictions to the EU open data policy need to be duly justified.

¹⁹ https://commission.europa.eu/legal-notice_en#copyright-notice

²⁰ <https://ec.europa.eu/eurostat/about-us/policies/copyright>

²¹ <https://creativecommons.org/licenses/by/4.0/>

Restrictions can be set at two levels:

- restricted data are only accessible to registered users, and can be used only in aggregated form (for example statistical analyses) and for research purposes, excluding commercial purposes, pursuing the code of conduct developed by the RISIS research infrastructure, which must be explicitly accepted upon registration²². These data are replaced in the publicly available data by code 'c'.
- Confidential data are not accessible to EHESO users. This data should not be delivered.

Before publication, deliverers, and specifically NSAs, need to confirm the licensing conditions. This consent can be given in the following ways:

- Signing disclosure agreements (signed only by the data deliverer).
- Providing consent via Email.

Either way, the following **information** needs to be covered:

- The country and the data covered by the agreement.
- **Time reference:** which years are covered by the agreement
- Required **restrictions** and their level (restricted, confidential) such as
 - Restrictions for selected institutions
 - Restrictions for selected variables
 - Restrictions for small values to avoid identification of individuals (smaller or equal to 3)

The process of implementing these restrictions for the public dataset is done as follows:

- **Data deliverers** fill in the respective data in the data collection sheet as it can be shown to registered users. In the disclosure agreement (in whichever form), data deliverers specify which restrictions need to be met for the public dataset.
- **Data managers** are responsible for checking the agreements and pointing out restrictions to the data collection coordinator.
- The **data collection coordinator**, when doing final data checks, implements the restrictions by using color-coding in Excel that is translated to special codes 'c' and 's' in the public dataset.
 - **Purple colour** is used for confidential data for specific institutions or variables.
 - **Light blue colour** is used for small values (smaller or equal 3), as described
- The color-coding is shown also in the data collection export when using the Admin menu (only available to the data collection coordinator and data managers). In case that data is revised to be imported again, attention should be paid not to overwrite existing color-coding, unless data is replaced entirely. Then the steps outlined above need to be done again.
- Note that registered users that sign the terms of use see all data. If data should be strictly private, it should not be delivered and instead coded with 'm' with a flag 'c'.

After a reasonable number of countries are included in the EHESO-ETER data centre, with all data checks and adjustments made, data for the respective year is ready for publication. Typically, **data publication** of each round of data collection is planned for spring of the following year (i.e., for the data collection in fall 2024, data publication is planned for March 2025). The data collection coordinator then sets the respective year from 'collecting' to 'public'. Additional data updates are possible at any time by contacting the responsible data manager.

²² <https://www.risis2.eu/risiis-code-of-conduct/>

About the European Higher Education Sector Observatory

The European Higher Education Sector Observatory (EHESO) is a user-centred data and information platform on the higher education sector in Europe. The Observatory enables the comparison, analysis, and showcasing of the sector's performance across multiple levels, actors, and themes.

