European Commission

Implementing and Disseminating the European Tertiary Education Register

Final report

Contract EAC 2015-0280

Brussels, 2017

Benedetto Lepori, Andrea Bonaccorsi, Alessandro Daraio, Cinzia Daraio, Hebe Gunnes, Elisabeth Hovdhaugen, Michael Ploder, Monica Scannapieco, Daniel Wagner-Schuster

Education and Training

Implementing and Disseminating the European Tertiary Education Register

Contract EAC 2015-0280

Benedetto Lepori, Andrea Bonaccorsi, Alessandro Daraio, Cinzia Daraio, Hebe Gunnes, Elisabeth Hovdhaugen, Michael Ploder, Monica Scannapieco, Daniel Wagner-Schuster

Disclaimer

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

Abstract

This report presents the main elements of the European Tertiary Education Register (ETER). ETER is a project funded by the European Commission, Directorate General for Education Youth, Sport and Culture for the years 2013-2017, which has established a comprehensive register of educational institutions delivering degrees at ISCED levels 6 (bachelor), 7 (master) and 8 (PhD). ETER provides data on more than 2,500 higher education institutions (HEIs) in 36 European countries for the years 2011-2014, covering more than 22 million undergraduate students. Data at the level of individual HEIs include organizational characteristics and geographical information, staff, revenues and expenditures, students, graduates, research activities and can be downloaded from the public ETER website (www.eter-project.com).

This report presents the conceptual foundation of the ETER database and its methodological elements, the infrastructure which was developed to collect, validate and publish the ETER data, the main results in terms of coverage of HEI systems, of data availability and quality and data usage, some highlights on the structure of European higher education derived from ETER and, finally, a number of recommendations on the future development of the ETER system.

Executive summary

What is the European Tertiary Education Register?

The European Tertiary Education Register (ETER) is a database that provides a core set of data on Higher Education Institutions (HEIs) delivering degrees at the tertiary level in Europe.

ETER is a project funded by the European Commission's Directorate General for Education Youth, Sport and Culture (contracts EAC-2013-0308 and EAC-2015-280). The project began in August 2013 and ended in July 2017. It was a joint undertaking of five partners - USI, Università della Svizzera Italiana, Lugano, JOANNEUM RESEARCH, POLICIES, Graz, NIFU – Nordic Institute for Studies in Innovation, Research and Education, Oslo, University of Rome La Sapienza and University of Pisa – in close collaboration with EUROSTAT, with a network of national experts and with the National Statistical Authorities of the participating countries.

ETER currently provides information on nearly 3,000 HEIs in 36 European Research Area countries from the year 2011 (academic year 2011/2012) to 2014 (2014/2015), including EU-28 countries, EEA-EFTA countries and candidate countries. For a few countries, only descriptive and geographical information is available – these are the French part of Belgium, Montenegro, Slovenia and Romania.

Most ETER data can be freely downloaded from the project website (www.eter-project.com) and reused for analytical purposes, making ETER a truly common resource for policy-makers, administrators and scholars. A small part of ETER data is available only for research purposes under the signature of a non-disclosure agreement.

What is the rationale for ETER?

ETER represents an important contribution to the strategy for the modernization of European higher education, as a fundamental component of the Europe 2020 strategy. In this respect, higher education is facing fundamental challenges, like increasing the number of graduates, reaching international excellence, and contributing to economic development.

Reliable information is key for this process as it lays the groundwork for evidence-based policies: for example concerning the promotion of excellence, differentiation of higher education institutions, and the design of competitive funding policies. Information at the institutional level is also important to allow stakeholders to make sensible choices, for example concerning the selection of study locations, by comparing HEIs across dimensions of interest, like the type of subjects offered, quality of education, employability, and research quality.

ETER contributes to these goals in two main ways. First, it provides for the first time a reference list of HEIs in the European higher education area, including descriptive and geographical information, which can be used to describe the system and allow matching ETER with other data sources. Second, it provides a core set of statistical data on these HEIs, which are sufficiently comparable between European countries.

What are the key principles of ETER?

The key principle of ETER is to consider that Higher Education Institutions represent a meaningful unit of analysis. This implies for example that HEIs can be compared (albeit with some limitations) and that questions concerning their way of working, productivity and strategy can be answered. In methodological terms, this also means that HEI-level data (for example number of staff or counts of degrees) can be produced and analysed in a meaningful way.

ETER is therefore largely complementary with educational and R&D statistics provided by EUROSTAT, which focuses on the country or the region as unit of analysis. Therefore, ETER complements EUROSTAT data by providing a view of the diversity between

individual HEIs within countries and regions. Such a focus on Higher Education Institutions and their consideration as independent units in terms of quality of services (for example by students) is fully aligned with the idea of a European Higher Education Area, in which HEIs compete for students, researchers and funds in a market-like setting which is less and less regulated by national states (European Commission, 2011). Such an idea also informs the launch of projects like ETER – with the aim of providing comparable data on European HEIs – and U-MULTIRANK – with the aim of allowing students and stakeholders to compare systematically HEIs on different grounds.

More precisely, ETER focuses on providing HEI-level data on three dimensions:

- The inputs on HEI activities, specifically the amount and composition of financial resources, staffing and the enrolled students.
- The output of HEIs' educational activities, research and third-mission.
- A set of regulatory and institutional characteristics, like the legal status, the age of the institution and its geographical location, which are expected to influence its activities.

Which institutions are included in ETER?

ETER collects data on higher education institutions, defined as entities, which are recognisable as distinct organisations, which are nationally recognised as HEIs, and whose major activity is providing education at the tertiary level (ISCED 2011 level 5, 6, 7 and/or 8). R&D activities might be present, but are not a necessary condition for inclusion in the perimeter. Furthermore, HEIs covered by ETER should have at least 30 FTEs of staff or at least 200 enrolled students.

In practice, ETER provides a very complete coverage of HEIs graduating at least at the bachelor level (ISCED level 6) – in terms of enrolment, ETER coverage exceeds 90% of the EUROSTAT figures in most countries. At the European level, ETER included 95% of the students in 2013/2014 (ISCED 6), respectively 100% (ISCED 7) and 85% (ISCED 8). The coverage of institutions delivering only short degrees below the bachelor level (ISCED level 5) is much lower (34% of the student enrolments) due to the fragmented nature of this type of tertiary education, which is sometimes delivered by secondary education institutions or professional associations, and data is not readily available from the National Statistical Authorities.

Which data are provided by ETER?

ETER provides data on the following dimensions of HEIs:

- *Descriptors and regulatory characteristics*, like the foundation year, the institutional website, legal status, institutional type. These characteristics are informative of the HEI's position in the system.
- *Geographical information*, including the city, postcode, geographical coordinates of the main campus, the presence of branch campuses in other cities. This information is highly valuable to combine ETER with regional or city-level data.
- *Students and graduates* divided by level of study (diploma, bachelor, master), field of education, gender, nationality and mobility. This information fully characterizes the educational profile of the HEI. Additionally, data on Erasmus students are covered.
- Staff, divided between academic and non-academic. For the former, breakdowns by gender, nationality and field of education are provided, as well as the number of full professors. Staff data are informative of the human resources available to the HEI.
- *HEI expenditures*, divided between personnel, non-personnel and capital, and *revenues*, divided by stream (core funding, third-party, tuition fees) and source (public, private, international). These data characterize the financial resources used for activities and their origin.

- *Research and transfer activities,* including the number of PhD students and graduates and R&D expenditures. While data on research are less complete in ETER, they are readily available from international databases on scientific publications, European projects and patents, which can be easily combined with ETER.
- Finally, ETER includes a set of pre-defined indicators characterizing relevant dimensions of HEI activities, like the extent of subject specialization, international mobility, gender balance.

How has ETER data been collected? Is the data complete?

ETER is mostly a secondary data collection: with the exception of descriptors and geographical information, which have been partially collected from public sources, most data come from National Statistical Authorities in the participating countries and have been collected either in the framework of international education statistics, or for national statistical purposes.

ETER has developed for each variable standardized definitions and guidelines on how to harmonize national data, building extensively on the EUROSTAT definition in educational and R&D statistics, but extending them in areas not fully covered by EUROSTAT, particularly HEI finances and staff.

ETER has also developed a full data collection infrastructure, which allows collecting the data on a yearly basis from the National Statistical Authorities (NSA), validating and checking them and integrated in a relational database, which allows for safe storage of the data. The ETER database is also the basis for the ETER website, which allows users to select countries, HEIs and variables and download their data in a variety of formats.

On average, ETER data are 78% complete (data collection 2014). This average however conceals large differences between variables: descriptors and geographical information is nearly complete, in addition to nearly 90% completeness for student and graduate data, and only slightly lower for staff data. On the contrary, financial data are available only for about half of the HEIs in the database.

How is the data quality checked?

ETER data are subject to a systematic data validation and quality check, which begins from the moment the data is collected, and is organized into different stages. These checks focus on the internal quality of data, including their *format, accuracy* and the *consistency* between connected variables, but also performance of statistical checks on outlying data (for example costs per students) and on changes between years (for example in total revenues). Potentially suspect cases are then checked together with the NSAs and, in case, corrected. Deviant values, which are due to some substantive reasons, are annotated in the database to inform users.

Furthermore, ETER systematically collects metadata on deviations from standard definitions, on national specificities, as well on the mapping between national categories and ETER categories. These metadata are also available to users through the on-line interface.

What are ETER's uses?

ETER has been designed as a general public resource, which can be accessed free of charge and also be combined with other sources in order to analyse relevant problems. The potential uses therefore cover different scholarly and policy domains, like analysing the structure of European higher education (compared for example with the US), studying the impact of HEIs in regions and cities, analysing the efficiency of HEIs and the 'best' size to inform national consolidation policies.

Four main types of usages can be distinguished: (1) the direct usage of ETER data for policy analysis and reports, (2) their usage by scholars in order to analyse relevant problems in higher education, with potentially an indirect impact on policy-making, (3)

ETER as a reference list of HEIs for other studies (for example surveys) and (4) ETER as a direct source of data for other projects at the European and national level, which provide on-line visualizations and indicators on higher education.

What can we learn from ETER?

The main lesson of ETER is that the European higher education system is characterized by a high level of internal diversity, in terms of institutional characteristics, size, activities, resources, which broadly ranges from the research-intensive international universities to small-scale focused educational providers in domains like arts or humanities. In the simplest way, this diversity can be described in terms of two dimensions: the legal status of HEIs (public vs. private) and the right to award degrees, distinguishing between HEIs delivering only diplomas, those delivering at least masters and bachelors and those with the legal right to award the PhD. Despite this level of diversity, the core of European higher education is still a rather small number of universities, covering most subject domains: slightly more than 1,000 HEIs award doctorate degrees (over more than 2,700 HEIs in ETER) and account for 70% of total student enrolment.

Such diversity is of high policy relevance, as it implies that 'no size fits all', i.e. that there is not a single best model for European HEIs. On the contrary, public policies must be tailored and differentiated by HEI type and characteristics, while a major goal should be to maintain this diversity and to respond to diverse societal needs.

What is the future of ETER?

The main outcome of the two ETER contracts has been to develop a fully operational system, through which HEI data can be routinely collected on a yearly basis, validated and corrected, safely stored and made publicly available to a broader audience for further usage through a web interface. As documented in this report, the use of ETER is also growing rapidly, for both policy and scholarly purposes.

The main recommendation is therefore that the dataset should be maintained and updated through regular data collection. At the same time, there is room for further improvement in five critical areas, namely (1) extending coverage of the HEI system, (2) improving data completeness and data quality, (3) introducing additional variables, (4) making it easier to use ETER and (5) increasing visibility of ETER through targeted dissemination and communication activities. The report also provides recommendations on how these issues could be addressed in the future.

Table of Contents

Table 1. List of abbreviations

Abbreviation	Full Name
DG EAC	Directorate General Education and Culture
DG RTD	Directorate General for Research and Innovation
EC	European Commission
EEA	European Economic Space
EFTA	European Free Trade Agreement
ERA	European Research Area
ETER	European Tertiary Education Register
EU	European Union
EUMIDA	European Microdata Project
EUROSTAT	European Statistical Office
FOE	Fields of Education
FTE	Full Time Equivalents
FYROM	The Former Yugoslav Republic of Macedonia
HC	Head Count
HEI	Higher Education Institutions
ISCED	International Standard Classification of Educational Degrees
NE	National Experts
NIFU	Nordic Institute for Studies in Innovation, Research and Education
NSA	National Statistical Authority
OCED	Organisation for Economic Cooperation and Development in Europe
UAS	Universities of applied sciences
UOE	UNESCO OECD EUROSTAT handbook on education statistics
USI	Università della Svizzera italiana
a.a.	Academic year

Introducing the European Tertiary Education Register

The European Tertiary Education Register (ETER) is the first comprehensive database on European higher education institutions (HEIs). It is a publicly available data source constructed mostly from statistical data in the participating countries, which covers most of the Higher Education Institutions (HEIs) contributing at the bachelor, master and PhD level in the European Union member countries, EEA/EFTA countries and candidate countries (see box 1 for an overview).

The rationale for ETER is constituted by an increasing interest for more detailed data on European Higher Education, complementary to the national and regional statistics provided by EUROSTAT. On the one hand, scholars and decision-makers are becoming increasingly aware of the very large differences between HEIs in terms of their mission, characteristics and activities (Huisman *et al*, 2015; Van Vught, 2009). Under these conditions, averages at the country level do not provide reliable answers to important questions concerning, for example, the efficiency of HEIs (Daraio *et al*, 2015), the differences between HEI types (for example colleges vs. universities; Kyvik and Lepori, 2010) or the internationalization of European higher education (Lepori *et al*, 2014).

On the other hand, there is an increasing interest in comparing individual HEIs based on the services they deliver to students and their research activities. Manifestations of these tendencies are found in various rankings, from those focused on international research excellence to sectoral rankings and rankings focused on students (Waltman *et al*, 2012). Following these needs, the US implemented an integrated information system in the 1990s that covered all tertiary education institutions receiving federal aid, Integrated Postsecondary Education Data System (IPEDS; https://nces.ed.gov/ipeds/), on which the widely accepted Carnegie classifications of US HEIs is largely built (http://carnegieclassifications.iu.edu/).

Alongside the parallel running U-MULTIRANK project (http://www.umultirank.org/; van Vught and Ziegele, 2012), ETER represents the European Union's response to the need for advanced knowledge tools to improve the knowledge and transparency of European higher education (European Commission, 2017). Within this broader context, ETER fulfils two specific functions: first, providing a reference list of HEIs in Europe that graduate students at least at the bachelor level (ISCED 6) and, second, gathering and harmonizing the data on HEI input (revenues and expenditures), on personnel, students and graduates, which are collected at the national level by the National Statistical Authorities. Until a decade ago, this task was considered to be highly problematic in the European context, given the large diversity of national systems and data, which made the realisation of an integrated system difficult (Bonaccorsi *et al*, 2007). In 2011/2012, the European Commission financed a large-scale pilot project called European Micro Data (EUMIDA), which demonstrated the feasibility of collecting institutional-level data in Europe and produced the first register of HEIs in the continent (Eumida, 2009). The wide impact of this work demonstrated the relevance of these data for our understanding of European higher education (Bonaccorsi, 2014) and led the European Commission to launch a follow-up project.

The European Tertiary Education Register project built on these pilot activities, but aimed to move towards a fully operational system, where methodological standards are defined and enforced: data are collected on a yearly basis, subject to a systematic validation and data quality and, finally, made available to a broader set of users for analytical purposes. The study is a joint undertaking of five partners: USI – Università della Svizzera Italiana, Lugano, Center for Organizational Research, JOANNEUM RESEARCH, Graz, NIFU – Nordic Institute for Studies in Innovation, Research and Education, Oslo, University of Rome La Sapienza, Department of Computer, Control and Management Engineering Antonio Ruberti, Rome, Department of Electrical Engineering, University of Pisa – together with a network of experts in the concerned countries. ETER is supervised by the Directorate General of Education, Youth, Sport and Culture of the European Commission, in cooperation with EUROSTAT.

The first phase of the ETER project ran from July 2013 to July 2015 (contract EAC 2013-0308; see Lepori *et al*, 2015), while the second phase ran from August 2015 to July 2017 (contract EAC 2015-280). A call for tenders to continue ETER, including two further waves of data collection in 2018 and 2019, was published by the European Commission in summer 2017.

This final report presents the main building blocks of the ETER data system, including:

- The conceptual foundation of the ETER database and its methodological elements (page 16).
- The infrastructure, which has been developed to collect, validate and publish the ETER data (page 26).
- The main results in terms of coverage of the HEI system, of data availability and quality and, finally data usage (page 40).
- Some relevant highlights on the structure of European higher education derived from ETER (page 52)
- Finally, a number of recommendations on the future development of the ETER system (page 63).

This report is meant to summarize these items in a general and non-technical language for the reader interested to have an overview of ETER. It is accompanied by a set of annexes providing more in-depth and technical information, including particularly the ETER Handbook for data collection (Annex 1) and the ETER data quality report (Annex 2). These annexes are shortly described at page 66 of the report.

Box 1. ETER in a nutshell

2,764 individual Higher Education Institutions in 36 countries (academic year 2014/2015).

Countries covered: EU-28, EEA-EFTA countries (Iceland, Liechtenstein, Norway, Switzerland), candidate countries (the former Yugoslav Republic of Macedonia (fYRoM), Montenegro, Serbia and Turkey).

22,7 mio. undergraduate students and 0,7 mio. PhD students (a.a. 2014/2015).

85% of all tertiary education students in the participating countries

Data on organizational characteristics, staff, revenues, expenditures, students, graduates, research.

Average Completeness of the data around 78%.

Data available for the years 2011 to 2014.

Data can be downloaded from the public ETER website www.eter-project.com

The ETER goals, framework and methodology

Unlike *ad hoc* data collection for the purposes of a research contract or a policy study, which can be directly targeted to some specific policy or research questions, a system like ETER requires a high degree of standardization and consistency and must be designed for a broader set of potential uses and questions.

More specifically, ETER pursues two main goals, which are very similar to its US equivalent, i.e. IPEDS (https://nces.ed.gov/ipeds/):

- First, it aims to provide a census of higher education institutions in Europe. In addition to answering the question of how many HEIs are present on the continent, a *reference list* is fundamental for many purposes, like the ability to pair ETER with other data sources, designing samples of HEIs for in-depth research and collecting data from the HEIs themselves.
- Second, it aims to provide a core set of data on these HEIs, which allows for the characterization of their main inputs, dimensions of activities and outputs.

Therefore, ETER is largely complementary to the existing statistical system managed by EUROSTAT with respect to education, based on the UOE methodological manual (UOE, 2013), and R&D, based on the Frascati manual (OECD, 2015), which focuses on providing data at the national or regional level.

These goals can only be achieved by addressing three conceptual and methodological issues in a consistent way, i.e.

a) First, by defining the object of observation (Higher Education Institutions),
b) defining the perimeter of the system (which HEIs should be included and for which reasons) and
c) selecting the most suitable variables and indicators.

Such methodological choices result from a balance between three requirements (Barré, 2001):

- A conceptual understanding of the object of analysis in order to produce data, which reflects relevant questions.
- The compatibility with existing frameworks and methodologies, particularly from official statistics, in order to allow for comparisons and data integration.
- Pragmatic and resourcing considerations concerning the availability of data and the reasonable effort for their collection and standardization.

As will be shown in this chapter, concepts and definitions from the higher education literature (Bonaccorsi and Daraio, 2007) and from educational statistics constitute an important starting point for ETER's methodological choices. Specifically, ETER widely adopts many definitions and classifications from the UOE manual on educational statistics (UOE, 2013), in order to be able to reuse data collected for educational statistics and to achieve consistency with it.

None of these frameworks however is fully suited to the goals and needs of ETER and, therefore, some dimensions of ETER required *ad hoc* choices, which were extensively discussed with the European Commission and with the project task force, and then codified in the ETER handbook. The handbook is a core project output to achieve the reproducibility of the system in a long-term perspective (see section 0). Besides its data output, ETER therefore represents a core effort to advance methodological standardization concerning higher education institutions, particularly for dimensions such as personnel and finances, where educational statistics provides less detailed information.

Higher Education Institutions as units of analysis

The conceptual core of ETER is considering that Higher Education Institutions represent a meaningful unit of analysis. This implies for example that HEIs can be compared (albeit with some limitations) so that questions concerning their way of working, productivity and strategy can be answered. In methodological terms, this also means that HEI-level data (for example number of staff or counts of degrees) can be produced and analysed in a meaningful way.

Until the 80s, this was considered problematic both at the scholarly and at the political level. HEIs were considered in the same way as schools, functioning as parts of the public bureaucracy by delivering educational services that are directly subject to public regulation. Alternatively, universities were considered to be a loosely coupled collection of departments, without the ability of centrally managing processes like strategy, resourcing and delivery of education and research (Musselin, 2007). As a consequence, both studies and statistics on higher education focused on different levels, like comparing whole national higher education systems (Clark, 1983), analysing research groups or scientific disciplines (Becher and Trowler, 2001) and comparing educational curricula, the main unit of observation adopted by educational statistics (UOE, 2013).

The idea of Higher Education Institutions being considered as independent units in terms of quality of services (for example by students) was promoted by modernization agendas in different European countries related to the diffusion of New Public Management (Ferlie *et al,* 2008; Brunsson *et al,* 2012). It is consistent with the idea of a European Higher Education Area, in which HEIs compete for students, researchers and funds in a market-like setting, which is less and less regulated by national states (European Commission, 2011). Such ideas also inform the launch of projects like ETER – with the aim of providing comparable data on European HEIs – and U-MULTIRANK – with the aim of allowing students and stakeholders to systematically compare HEIs on different grounds.

Figure 1. ETER conceptual framework
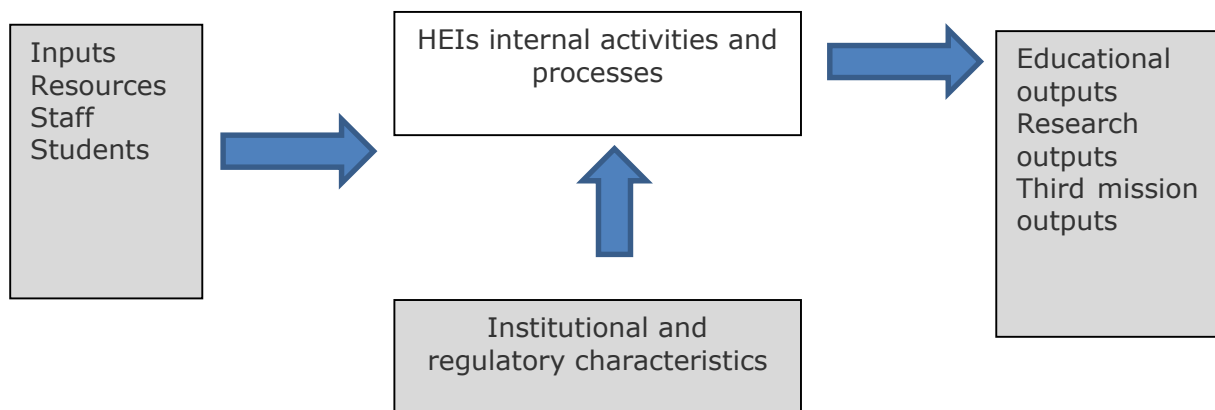
In this context, ETER does not aim to analyse how HEIs work in their internal processes, but focuses on a general and aggregated measurement of three dimensions:

- The inputs to HEI activities, specifically the amount and composition of financial resources, staffing and the enrolled students.
- The outputs of educational activities, research and third-mission.
- A set of regulatory and institutional characteristics, like the legal status, the age of the institution and its geographical location, which are expected to influence its activities.

Thus, ETER conceives HEIs as multi-input and multi-output entities, which jointly transform inputs into a range of outputs, possibly exploiting economies of scope, for

example between education and research. On the contrary, ETER will black-box internal processes and structures.

A limitation of the organizational approach of ETER is that it is only possible to a limited extent to take into account the internal heterogeneity of HEIs, for example between departments in different domains, as differences between the profile of activities, resourcing, and outputs might be very large (Lepori, 2007). Partially, this can be addressed by analysing the subject composition of HEIs since ETER provides data on the distribution of students and PhD students by subject domain[1].

A second issue concerns cases where the exact perimeter of HEIs cannot be defined unambiguously, because of their relationships with other entities. Two relevant cases are French universities, which have units in common with CNRS with mixed staff, budget and output, and university hospitals: the latter might be included within the university or legally independent units, but closely connected with universities[2]. The delimitation of university hospitals has a strong impact on staff data, resources and scientific output of universities. In order to detect such problems, ETER includes a dummy variable on whether a university hospital is present. Detailed information on coverage is provided in the metadata to the data collection.

In parallel to ETER, an effort has been made in the Research Infrastructure for Research and Innovation Policy Studies (risis.eu) to construct a register of public-sector research organizations by systematically tracking the linkages between types of organizations.[3] This information might be helpful in the future in order to deal more systematically with linkages between organizations within ETER.

Demographic events

An approach focused on HEIs as the main observational unit, which also foresees the collection of longitudinal data on different years requires to deal seriously with *organizational demography*, i.e. the fact that HEIs are founded, closed or merged together – events which are of increasing relevance in the wake of restructuring of European higher education (Pinheiro *et al*, 2016).

Building on EUROSTAT Business Units Register (*EUROSTAT, 2010*), ETER has therefore introduced a systematic approach to deal with organizational demography based on the following principles (see Figure 2)[4]:

- The introduction of unique identifiers for HEIs, which allows HEIs to be tracked even if the name changes.
- The distinction between demographic events, which imply the creation of a new entity, and change in organizational characteristics, particularly name changes – for example a college being renamed as a university.
- Rules for coding following types of demographic events and handling identifiers in ETER: the foundation of a new HEI, the closure of an existing HEI, the merger between HEIs and the split of HEIs in different components.
- A process to identify demographic changes as an integral part of the data collection process.

---

[1] See the ETER brief 1. What ETER tells us about subject specialization in European higher education https://www.eter-project.com/assets/pdf/ETER_brief_subjectmix.pdf.
[2] See https://www.cwts.nl/blog?article=n-q2w264.
[3] The register of public-sector research organizations (OrgReg) is available on-line at orgreg.joanneum.at.
[4] See the ETER Handbook for data collection, chapter 3.3.

Figure 2. Organizational demography in ETER

Besides its importance for data collection, the demography of HEIs is an important subject per se, as ETER is providing for the first time a European-wide overview of demographic changes in higher education, albeit for a limited period of time (2008-2014)[5].

Box. Demographic events

Merger. In 2013, the University of Lisbon and the Technical University of Lisbon were merged into the New University of Lisbon. Until 2012, they are present in ETER with their respective codes (PT0007, PT0009). They are no longer included in the 2013 collection, while the new university is included using the ID PT0140. In the demographic events for the University of Lisbon, a link is available to its ancestors.

Split. In 2012, the Teacher Training University of central Switzerland (CH0027) was split into three cantonal HEIs. In 2012, three new IDS are attributed to these three HEIs, whereas CH0027 is no longer present. In the demographic events for each of the three schools, a link is provided to their common ancestor.

Take-over. In 2012, the Tallinn Pedagogical College (EE0016) was taken over by the Tallinn University (EE0003). The corresponding IDs is no longer present in ETER after 2012. However, in the demographic events for the University of Tallin it is possible to trace the university back to the Pedagogical College.

The ETER perimeter

The second building block of ETER is a consistent definition of which entities should be considered as part of ETER. This is central to its function as a register of Higher Education Institutions in Europe.

Despite the wide use of the HEI concept, this task proved to be surprisingly complex and led to long discussions both in EUMIDA and in ETER for different reasons (Lepori and Bonaccorsi, 2013). First, there is no statistical definition for Higher Education Institutions, since the basic statistical unit in educational statistics by EUROSTAT is the educational program or curriculum. Educational programs are then classified based on their qualification level by adopting the International Standard Classification of Educational Degrees (ISCED). Tertiary education curricula (and students, respectively degrees) are classified at levels 5 (diploma), 6 (bachelor), 7 (master) and 8 (PhD) of the ISCED-2011 classification. The nearest approximation to HEIs would then be those educational institutions delivering degrees at ISCED levels 5 to 8 (hence the name of European Tertiary Education Register).

---

[5] See ETER brief 2. What ETER tells us about the history and demography of European HEIs https://www.eter-project.com/assets/pdf/ETER_brief_demography.pdf.

However, second, such a definition would not always correspond to what could be considered as a higher education *institution* – some tertiary education degrees are delivered within secondary schools (for example the French preparatory classes of Grandes Ecoles), while in countries like Switzerland and Germany professional degrees do not necessarily require attendance in a full-time education program similar to an equivalent school. Hence, the ETER requirement is that such degrees are delivered by an institution that is devoted to education at the tertiary level.

Third, the information available shows that most national higher education systems are characterized by a reasonably small number of HEIs delivering most of the degrees, including universities and colleges, and by a large queue of smaller educational providers, mostly providing short-cycle diplomas (at level ISCED 5). Not only does their sheer number generate a large burden for data collection, but also data availability for the smaller institutions outside the university sector is scattered – and in some cases it is not available from the NSA, but via other departments, and only basic data like student numbers are available.

The ETER approach to the perimeter is therefore based on the combination of a general definition of HEIs, plus some specific thresholds and rules for inclusion and exclusion[6]. In general, ETER collects data on *higher education institutions*, defined as entities, which are recognisable as distinct organisations, which are nationally recognised as HEIs, and whose major activity is providing education at the tertiary level (ISCED 2011 level 5, 6, 7 and/or 8). R&D activities might be present, but are not a necessary condition for inclusion in the perimeter.

Furthermore, a threshold in terms of size is set. HEIs covered by ETER should have at least 30 FTEs of staff or at least 200 enrolled students. Exceptions might apply for institutions of particular national importance. This threshold has been introduced in order to improve comparability of the perimeter.

We provide later in this report at page 40 details on the coverage of national systems by ETER and potential ways to improve coverage when it is insufficiently complete.

Variables and indicators selection

The selection of variables in ETER follows the conceptual framework of the whole project. ETER therefore plans to provide variables for the following dimensions:

- HEI regulatory, institutional and geographical characteristics, which are expected to influence their activities and their output.
- HEI inputs, particularly, financial resources, personnel and students.
- HEI outputs, distinguishing between education (degrees), research and third-mission.

At the same time, the choice of variables also follows considerations of feasibility and complementarity with other available sources. ETER is a database developed together with the National Statistical Authorities which focuses on integrating data that is already collected by the NSAs, which until recently was only available at the national level. Data from international databases concerning research output (publications) and technological output (patents) have not been integrated in ETER, but this should not be considered as a limitation since these data are available from other sources and can be easily matched with ETER[7]. Third-mission data and data on regional engagement and employability are not available from official statistical sources, but would require a dedicated data

---

[6] See the ETER handbook for data collection, chapter 3.2.

[7] Within the RISIS project, additional data for the ETER HEIs are being made available through OrgReg (orgreg.joanneum.at). These include various types of bibliometric indicators, numbers of EU-funded projects and numbers of patent applications. Since this information is partially proprietary, these data are however available only for research purposes and cannot be published on the open ETER website.

collection. Should some of these data become available from other projects, ETER now provides a platform for integrating them with basic statistical information.

A short description of variables is provided below. For full methodological details, please refer to the ETER handbook, chapter 5.

a*) Institutional descriptors* represent an important contribution of ETER to our knowledge of European higher education, as this information is not systematically collected elsewhere. They are mostly collected from public sources, like Websites and entries in Wikipedia.

Besides the official institutional name, its English translation and the acronym (if available), ETER includes some important categorizations, like the legal status and the classification of HEIs between universities, universities of applied sciences and other institutions. These can be combined with quantitative data in order to identify the role that different types of HEIs play in European higher education. ETER also comprises information on HEI history via the foundation year and, optionally, the ancestor year and legal status year. Finally, the URL of the website is included to ease the retrieval of information.

b) *Geographical information* is important in order to study the regional impact of Higher Education and to link with regional statistics. ETER includes the city and the postcode of the main seat, which are then transformed into regional coordinates. This allows easy localisation of HEIs within different spatial units, like cities, metropolitan areas and different regional classifications, making ETER an important source for regional analysis[8]. ETER also includes information on whether an HEI has campuses in different locations. An important future extension would be to also provide the localisation information for these campuses, as this would provide a more precise understanding of the regional coverage of higher education.

c) Data on *staff* are important for many purposes. They represent the most reliable measure of HEI size, which is more comparable than financial data between countries. Breakdowns by gender, nationality and scientific field provide relevant information on HEI profile, personnel composition and internationality. Additionally, the number of full professors is included. These breakdowns are not available in EUROSTAT, even at national level.

An important area of future development is a more fine-grained classification of academic staff by levels, which would be comparable across countries. National and institutional differences, however, made it impossible to build a consensus on a common set of definitions. The recent Eurydice study on academic staff might provide a helpful step in this direction (European Commission/EACEA/Eurydice, 2017).

d) *Data on education (students and degrees)* closely follow the definitions adopted by EUROSTAT in education statistics (UOE, 2013). ETER introduced the ISCED-2011 classification of education levels and the field of education and training classification (FET-2013). Breakdowns by student and degree for each level of education (ISCED5 to ISCED8) are provided by gender, nationality, mobility (based on the place of prior education) and educational field.

Three additional variables provide information on the lowest and highest degrees awarded by each institution and on whether the institution delivers mostly distance degrees

ETER also includes the number of incoming and outgoing Erasmus students by institution, which has been collected from data published by the Education, Audiovisual and Culture Executive Agency.

---

[8] See the ETER brief 4. What ETER tells us about the regional dimension of European higher education, https://eter-project.com/assets/pdf/ETER_regional_dimension.pdf.

e) Data on *research activities* are somewhat limited in ETER, which includes only information on PhD students and degrees (an important proxy of the extent of research, since they have to be considered in most cases as researchers) and on R&D expenditures. The latter are based on a breakdown of the use of time by academic staff, but are available only for a limited number of countries. The *research-active variable* (new to ETER) identifies those HEIs having an institutional research mission, even if they cannot award the doctorate. It is therefore meant to observe the extension of the HEI research mission beyond PhD-awarding HEIs (Lepori and Kyvik, 2010).

f) Data on *expenditures and revenues* are largely new to ETER, as the data provided by the education statistics are limited in this area. In this respect, ETER has introduced an important distinction of sources of revenues between core budget, third-party funds and student fees (which is relevant to characterize HEIs competitive position), which is then combined with a distinction by source (public, private, international) to yield a very fine-grained classification of revenue.

ETER also made an effort to harmonize the treatment of capital expenditures in European HEIs, a highly complex methodological task given the different accounting systems of European HEIs, which make comparability difficult.

Unfortunately, the level of completeness and comparability of these data are not yet fully satisfactory and further work in close cooperation with the NSAs is needed in order to harmonize national data.

g) Finally, ETER provides a set of pre-computed indicators built from the available data. They cover some substantive dimensions of HEI activities and characteristics, like gender balance[9], student mobility and internationalization[10], the structure of revenues of HEIs.

While the users themselves could compute these indicators, providing them directly makes the dataset more usable and avoids potential pitfalls, like computing indicators on data with a too many unclassified items.

Indicators and derived variables are an important area of future expansion of the ETER system, where the added value to users can be enhanced with moderate effort, since it does usually not require additional data collection.

---

[9] See ETER brief 6. What ETER tells us about gender balance in European Higher Education Institutions, available at https://eter-project.com/assets/pdf/ETER_gender.pdf.
[10] See ETER brief 3. What ETER tells us about student mobility in European higher education, available at https://www.eter-project.com/assets/pdf/ETER_student_mobility.pdf.

Box2. Who is who in ETER

The European Tertiary Education Register (ETER) is a project funded by the European Commission's Directorate General for Education, Youth, Sport and Culture (contract EAC-2015-0280), which aims to implement and disseminate a register of European Higher Education Institutions (HEI) and collect a comparable set of data for the HEIs within a defined perimeter. The project began in August 2015 and ended in July 2017. It followed a first contract in the years 2013 to 2015 (contract EAC 2013-0308).

The contract is a joint undertaking of five partners:
- USI, Università della Svizzera Italiana, Center for Organizational Research, Lugano (Benedetto Lepori, coordinator),
- JOANNEUM RESEARCH, POLICIES – Institute for Economic and Innovation Research, Graz (Michael Ploder, Daniel Wagner-Schuster),
- NIFU – Nordic Institute for Studies in Innovation, Research and Education, Oslo (Elisabeth Hovdhaugen, Hebe Gunnes),
- University of Rome La Sapienza, Department of Computer, Control and Management Engineering Antonio Ruberti, Rome (Cinzia Daraio, Tiziana Catarci, Leopold Simar, Monica Scannapieco).
- University of Pisa, Department of Electrical Engineering (Andrea Bonaccorsi, Alessandro Daraio,).

The project partners were supported by a number of national experts (Achilleas Mitsos, Greece; Krysztof Leja, Poland; Karel Sima, Czech Republic; Ben Jongbloed, Netherlands). The ETER contract is supervised by the Directorate General for Education, Youth, Sport and Culture of the European Commission, in cooperation with DG Research and Innovation and EUROSTAT, and by a task force composed of representatives from the National Statistical Authorities in the participating countries. The role of the task force is to discuss and make decisions concerning the design of the dataset, the selection of variables, and how to address methodological issues.

The ETER project is executed in close coordination with the National Statistical Authorities (NSA) in the participating countries: the NSAs provide important input concerning data and methodology and are the providers of most of the data included in ETER. They also gave consent for the publication of most ETER data. Even if it is not directly part of the European statistical system, the ETER project would have never been possible without this close cooperation with EUROSTAT and the National Statistical Authorities.

Table 2. Variables in the ETER database

Dimension	Variables
Identifiers	ETER ID National identifier (optional) Institution name (in own language) English institution name (if available) Acronym Year Demographic event (past) Affected HEIs (past) Remarks (past) Demographic event (future) Affected HEIs (future) Remarks (future)
Basic institutional descriptors	Country Code Legal status Institution category, national definition (in own language) Institution category, national definition (in English, if available) Institution category standardized Foreign campus Foundation year Legal status year Ancestor year University hospital Institutional website
Geographic information	Region of establishment, NUTS2 code Region of establishment, NUTS3 code Name of the city Postcode Multi-site institution Geographical coordinates
Educational activities	Highest degree delivered Lowest degree delivered Number of enrolled students at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility Total number of students enrolled at ISCED 5-7 Number of graduates at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility Total number of graduates at ISCED 5-7 Distance education institution Number of incoming Erasmus students Number of outgoing Erasmus students
Research activities	Research active institution Number of enrolled students at ISCED levels 8, by fields of education, gender, citizenship and mobility Number of graduates at ISCED levels 8 (doctorates), by fields of education, gender, citizenship and mobility
Expenditure	Current expenditure Personnel expenditure Non-personnel expenditure Expenditure unclassified Capital expenditure Accounting system of capital expenditures

	R&D expenditure
Revenue	Current revenue
	Core funding
	Basic government allocation
	Other core funding
	Third party funding
	Public third-party funding
	Private third-party funding
	Third-party funding from abroad
	Third party funding unclassified
	Tuition fees
	Student fees funding
	Revenues unclassified
	Non-recurring revenues
Staff	Number of academic staff in FTEs and headcounts
	Number of academic staff by fields of education, gender and citizenship in headcounts
	Number of non academic staff in FTEs and headcounts
	Number of professors by gender
	Inclusion of PhD students
	Number of total staff in FTE and HC
Indicators	Share of female among students and graduates (ISCED 6 and 7), respectively PhD students and graduates (ISCED 8).
	Share of female among academic staff and full professors.
	Share of foreigners among students and graduates (ISCED 6 and 7), respectively PhD students and graduates (ISCED 8).
	Share of foreigners among academic staff.
	Share of mobile students and graduates (ISCED 6 and 7), respectively PhD students and graduates (ISCED 8).
	Subject mix. Herfindahl index of the distribution by educational fields of students (ISCED 6 and 7), PhD students and academic staff.
	Degree orientation. PhD intensity (PhD degree/degree ISCED 5-7).
	Full professors as share of academic staff.
	Academic staff as share of total staff.
	Share in total revenues of the core budget, third-party funding and tuition fees.

The ETER data collection infrastructure

A major goal of the ETER contract was to establish and consolidate a standardized process for data collection, verification, storage and publication. The sheer amount of data to be collected (about 500,000 data cells per year) required the move from manual procedures to standardized processes, which allow the process to be deployed in a timely and manageable way. Moreover, the heterogeneity of data sources required the establishment of semi-automated data validation and quality procedures in order to detect mistakes and data issues, which might impair comparability.

While the process can be further streamlined, the ETER contract is now delivering a fully operational data collection infrastructure to the European Commission, which would allow data collection efforts to be repeated in the future with a reasonable effort.

In this chapter we describe the process and the main infrastructure elements. We also shortly describe the steps taken to promote ETER's usage.

The data collection, validation and publication process

Figure 3 provides an overview of the data collection process in ETER. The process is first managed bilaterally between the ETER core team and the respective national statistical authority contacts or contacts in the higher education ministry[11]. Data are collected through standardized excel templates, which are prefilled with the relevant information from previous years and include a number of checks and controls for consistency.

The process starts with the definition and update of the perimeter for each country. The country contact receives an excel sheet including the list of HEIs from the previous year's data collection and is asked to update the list by including name changes, new foundations, information on closures and demographic events. Together with the perimeter, a standardized description of the national system is provided and made available on the ETER website.

This information is integrated in the ETER database and is the basis to create data collection templates for each country including the correct list of HEIs for the reference year. The excel data collection template is already prefilled with information that is not expected to change by year, such as the foundation year, geographical position, etc. Such information should only be validated by the NSAs. The data collection sheet also includes a number of basic checks, particularly logical consistencies, such as sums of breakdowns equal to the total for each variable.

Collected data are then validated in order to correct simple mistakes and inconsistencies, and their quality needs to be thoroughly analysed, as there are many reasons why data might be not comparable by country. Data validation and quality might lead to corrections, but also to the addition of flags and methodological remarks, explaining why a value is deviant. After a first stage of validation, data are then uploaded to the ETER database (see chapter 0) and then subject to an extensive data quality process (see chapter 0).

The final step of the process lies in publishing the data on the website and in providing users with the required information in order to allow for sensible data usage (see chapters 0 and 0).

These steps are organized into a yearly cycle, starting in the summer of each year with the launch of the data collection process and concluding in late spring of the following year with the publication of data. ETER has therefore been designed for continuous annual data collection in order to produce a longitudinal database on the evolution of European higher education. This timeline is also broadly consistent with the Eurostat data

---

[11] The up-to-date list of ETER contacts is provided in annex to this report. (see chapter **Error! Reference source not found.**).

collection on education statistics. In the future this cycle could be somewhat shortened a by accepting the publishing of provisional or partial data[12].

The ETER data flow and database

The management of data and of data flows is a central question for a project like ETER, given the distributed nature of the data collection and the possibility that data are amended and corrected over time. Safely storing ETER data and managing different versions is therefore a condition that promotes the operational performance of the process.

The structure of the data flow is presented in Figure 3. It can be divided in three interconnected parts.

Figure 3. The ETER data flow
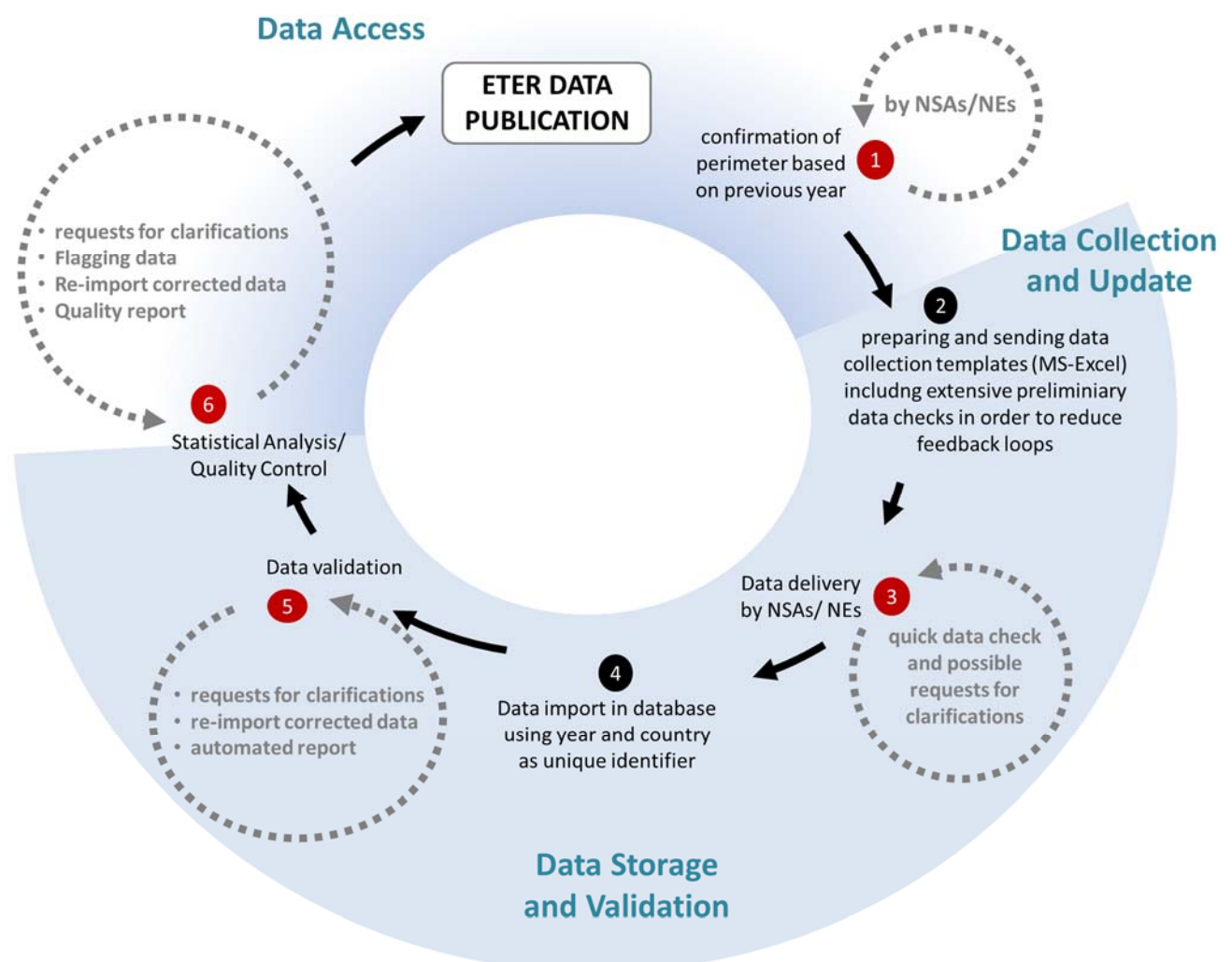
The data collection and update process is managed through excel files by country and year, which store the most recent versions of the data and integrate all variables for HEIs in a country. These files are updated, for example, when additional data are delivered or

---

[12] The ETER data is published at similar intervals as the Eurostat educational data. In considering timeliness, the reader should be careful that conventions for numbering of the years is different, i.e. ETER data for 2014 refer to the academic year 2014/2015 (for students), which is labelled as 2015 in EUROSTAT.

if clarifications create the need for corrections. Most of the data are provided by the NSAs, but some might be integrated by the consortium (particularly with respect to descriptors and geographical information). Excel data files might also be updated due to changes in the definitions or when new variables are added, these changes are also imported in the sheets for previous years in order to maintain consistency.

As soon as data are in a reasonable format and the main inconsistencies have been cleaned, they are uploaded to the ETER database. The database has different functions: it ensures safe storage and versioning of the data at a central place, it allows for data handling, for example exporting the data for data quality checks, and finally, it allows users to access the data. However, the general rule is that any changes to the data are first made in the excel files and then uploaded again to the database.

The ETER database is based on a MongoDB architecture, which uses an object-oriented approach in order to deal with different structured datasets, so called "documents". Different structured documents lead to a dynamic schema of the database itself, which leads to an iterative approach for the design and development of the ETER database. This is relevant since the database might change over time, for example with the addition of new variables.

This implies that the structure of the database will change slightly by every added or withdrawn field (bits and pieces of data). Thus, if new fields need to be added to a document, the field can be created without affecting all other documents in the system, without updating a central system catalogue and without taking the system offline. Also, if developers add more features, MongoDB continues to store the updated objects without the need for performing costly alteration operations, or worse – having to re design the schema from scratch.

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

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

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

For full documentation on the database, the reader should refer to the technical documentation provided in the annex of this report (see chapter **Error! Reference source not found.**).

Data quality procedure

ETER data collection is a secondary data collection, meaning that data are not collected directly from respondents, but the NSAs who rely on data sources for which ETER does not have any control.

As reported in the EUROSTAT Handbook on Methodology for the Modern Business Statistics[13], with reference to secondary data collection, three important dimensions to assess the 'fitness of use' of secondary data are used:

- assessing the metadata quality of the source,
- the data quality of the input data, and
- the data quality of the statistics produced.

---

[13] https://ec.europa.eu/eurostat/cros/content/collection-and-use-secondary-data-theme_en.

The first two dimensions are assessed in ETER through a standardized collection of metadata by country and variable, which is an integral part of the data collection process. The collected metadata are available on the ETER on-line database and represent key information to assess the usability of data. The third dimension is the central focus of the ETER data quality process, which checks systematically the collected data in terms of four data quality dimensions referring to different dimensions of the data[14]:

- *Format accuracy* evaluates the compliance of the value to the requested format, as defined in the data chapter of the ETER handbook, respectively in the definitions of each variable. This includes characteristics like being non-negative for all quantitative values, student and graduate data being integer variables, and so on.
- *Completeness* evaluates for each variable, dimension, and dataset the share of missing values (with the meaning relevant to completeness, i.e. unavailable or temporarily unavailable), which are present.
- *Consistency* verifies possible violations of semantic rules defined over the involved data, and specifically between different variables – for example HEIs not having the right to award the PhD, but for which the number of ISCED 8 graduates is non-null.
- *Timeliness* evaluates for each variable the time lag between the ETER Collection publication date and the Source Release date. Ideally, this lag should be reduced as much as possible.

By taking into account these characteristics, the ETER Data Quality Methodology has been designed and implemented as an articulated process that spans from the very first stages of the collection, of which the NSAs are in charge, to global checks and corrections done when the overall data collection is finished. This process occurs in three main stages.

In the early phases of data collection, the NSAs collect questionnaires, such as MS-Excel files. Some automatic checks have been implemented, including automatic rules within such files, i.e. checks of (i) incomplete data and blank cells, (ii) format accuracy compliance, (iii) mistakes in sums and inconsistencies between variables. In addition, a manual revision is performed on the collected files so that, if feedback to improve the collected data is identified, this feedback is conveyed to the NSAs for related corrections.

Later, a pre-validation of collected questionnaires is performed. As long as the questionnaires are provided by the countries, they are uploaded to the database and the re-exported for validation through an automated R-script, which checks country-by-country accuracy, completeness and consistency. Moreover, at this stage, a few key indicators are checked for extreme values, which might indicate data problems (for example expenditures per students). This stage produces a "Pre-validation quality report" by country highlighting potential problems with the data, which can then be checked by the NSAs.

The last stage is the final validation and correction stage. A global quality check is performed, which includes accuracy, completeness and consistency checks again and reporting on the identified violations. This quality check also includes far-reaching tests for the identification with statistical techniques of outliers concerning indicators like revenues per students, share of foreign students, and cross-year comparability checks to identify cases where the variation between years is very large and may point to data problems. When relevant, mainly to control for specific cases detected by previous methods, figures have been controlled against alternative data sources to solve possible doubts (for example cross referencing with published data on other sources).

Based on these checks, a number of actions are undertaken. This includes asking the NSAs to recheck the data and, if needed, make corrections. In the instances when this is

---

[14] For full details on the data quality process the reader should refer to the ETER data quality report provided as annex to this report.

not possible, because deviations are due to the specificities of the original data collection, data are flagged in the database and a remark is added so that users are aware of potential data limitations. Another category of flags and remarks covers the specificities of the data, which might impact the analysis – for example cases where the budget of an HEI suddenly increases because of a demographic event.

As a final step, a data quality report is produced, which systematically assesses and documents data quality (see chapter 0). We shortly report in the following chapter 0 on the main results of data quality and we discuss possible future evolutions of the data quality process.

Data publication and access

ETER is meant to provide a publicly available data source, where data can be searched by users or downloaded and combined with other data. Therefore, most ETER data can be freely reused in a similar format as the EUROSTAT statistical data, i.e. by requiring that the data source is mentioned and it must be explicitly mentioned when the data is modified. To achieve this goal, ETER has required that the NSAs consent to making the data publicly available.

As documented in the ETER data quality report, most NSAs informed the consortium that the delivered data were already public on national websites and therefore, no restrictions applied. Publication of data at the European level therefore largely reflects the on-going process of transparency in most European countries. In a few cases, the NSAs asked to HEIs consent to publish their respective data directly to the concerned HEIs.

For a small share of data, restrictions apply: these mostly concern financial data (Belgium-Flanders, Croatia, Italy, Latvia and Poland), as well as to a lesser extent staff (Belgium-Flanders) and data for a few HEIs that did not give consent for their data to be published. In these cases, ETER signed a Non-Disclosure Agreement with the respective NSAs.

For restricted data, access can be granted for research purposes and for carrying out statistical analyses, on the condition that direct identification of statistical units in the final product is not possible. Users need to register on the ETER website and accept these rules in order to download the data. In the public version of the dataset, data that were delivered, but restricted by a data disclosure agreement, were replaced by "c." For future direct access to the ETER database through the Advanced Programming Interface, suitable agreements will need to be defined depending on the final data usage.
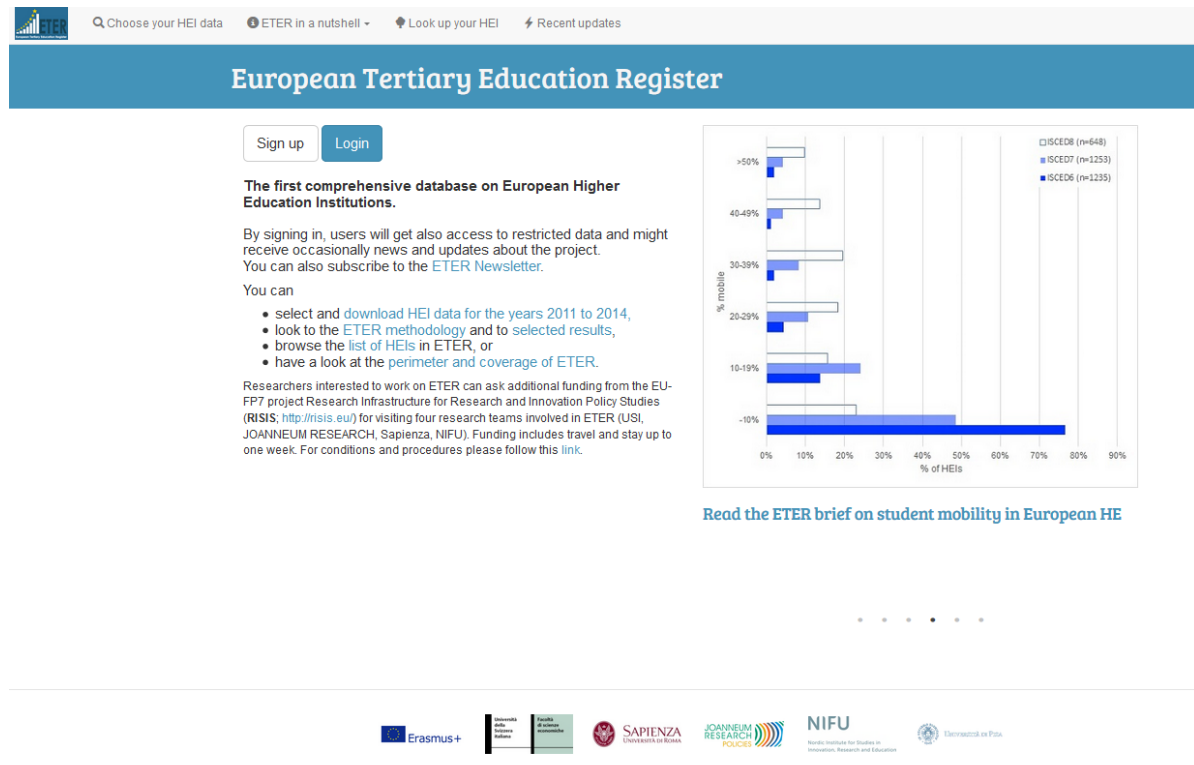
Moreover, ETER adopted usual statistical practices concerning data that are below some threshold, which could allow the identification of individuals, specifically for data on students and staff. To this aim, all cells below or equal to a count of 3 are set to "s" in the publicly available data. For breakdowns, the unclassified category is set to "s" in order to avoid the reconstruction of the concerned value by using the totals. The original data remain available for research purposes.

The ETER website and communication tools

The web application available under www.eter-project.com enables the user to retrieve data from the entire ETER data set in order to conduct research on European higher education sector micro data and provide the accompanying information required to make sensible use of data.

The website includes a short description of the ETER project and the performing consortium members. It also includes the possibility to register in order to have access to restricted data.

Figure 4. ETER homepage

Using the ETER web application, users have access to three main options:
- the user wants an overview of the register, its contents and documentation,
- the user wants to view and/or export data from the ETER micro data set, including country level metadata, or
- the user wants to look for specific HEIs in the register.

1) ETER overview.
The selection field "ETER in a nutshell" provides detailed information about the register, its contents and methodology:
- obtain a general overview about the project itself, its targets and contents.
- gather detailed information on the data collected, the origin of the data and special codes and flags used in the data collection.
- retrieve all relevant documents accompanying the project. These include a report about the main conceptual and methodological choices of ETER, the data collection handbook with an in depth description of the data collection methodology, a technical report with detailed information on methodology, data collection, data management and data quality, etc.
- access results of analyses (ETER briefs) based on ETER data.
- acquire information on publications based on the ETER data set.
- obtain answers to key questions by having a look at the frequently asked questions (FAQs) section.

2) View and export data and metadata
This section is the core of ETER website. By selecting the label "Choose your HEI data", users can choose the required higher education institutions by year and country.
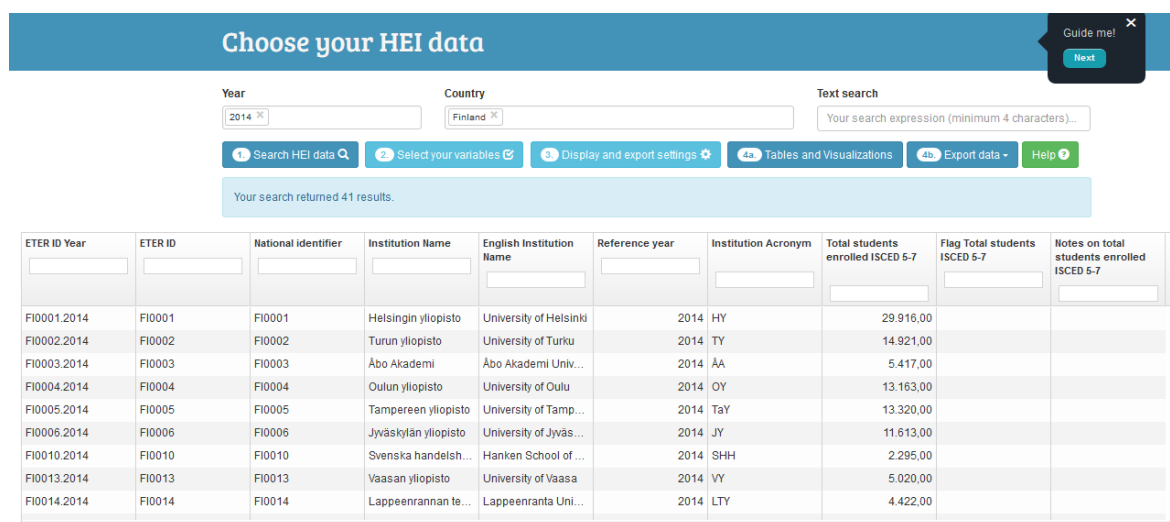
Additional possibilities allow the user to refine the selection and filter data. With the group of chosen institutions, the user can go a step forward and decide to download the requested data (Export data) or arrange and visualize them directly on the website (Tables and Visualizations). The default results mask includes some basic variables in order to get a first overview. ''Option Select your variables'' enables the user to select the required variables for export (exporting of all variables at once is also possible without selecting them – see export function), tables and visualizations.

After selecting variables, it is possible to change to the interactive *Tables and Visualizations space*. Using a table component, users are able to customize their data sets directly on the website. Users can drag and drop columns and arrange the data in order to retrieve the required data format. In interactive tables, different variables can be combined, displayed, and adapted to the specific user's needs. Users can additionally apply filters to variables and use the table grid in order to calculate shares. Additionally, graphical options for the selected data are available on the web interface. This is the case for selected indicators and allows the comparison of single institutions in a broader context.

The most important feature for data analysts is the *export function*, which is dedicated to usability in order to support data analysis and the dissemination objectives of ETER. Data export will be possible in .xlsx and .csv format, where users can choose the data to be exported by filtering and selecting variables or by exporting all variables at once and exporting the corresponding metadata.

The interface also includes an interactive help function providing guidance on the different options offered.

Figure 5. The ETER data search interface

3) Looking for specific HEIs

By accessing the "Look up your HEI" menu, the user is directed to a filtering mask where it is possible to insert the HEI name (or a portion) or the ETER code of the institutions of interest. The HEIs will be filtered in all available years, and by clicking on them an overview of all available data will be shown to provide quick consultation.

ETER dissemination activities

In order to promote the data's usage, ETER has organized a broad set of dissemination activities throughout the whole project. These include:

- The production of a set of ETER briefs.
- Dissemination events and dissemination workshops.
- The ETER mailing list.

The ETER briefs

The ETER briefs are meant to provide descriptive information, which shows the potential uses of the ETER database for analysing relevant policy issues. Each brief therefore includes the following information:

- The highlighting of a policy relevant issue in higher education, usually with references to EU policy documents.
- The presentation of key findings from the ETER dataset.
- Introducing users to the data and highlighting their added value with respect to existing data sources.

ETER briefs have been developed in a common four page format and have been available on the ETER website, as well as through the ETER newsletter.

Table 3 provides the list of policy briefs highlighting the policy issue and the relevance of ETER data. They show how the main contribution of ETER data, as compared for example with EUROSTAT educational statistics, is to provide a more disaggregated view looking not only to country differences, but also to differences among HEIs within countries. More details on these analyses are provided in chapter 0 of this report.

Table 3. ETER policy briefs

Topics	Date	Highlights
History and demography of European HEIs	June 2016	• European higher education is formed in layers created by different historical periods and with different policy goals. • Older universities, mostly founded before the 20th century, still enrol a large share of undergraduate students and most PhD students. • Universities of applied science only became significant in number of HEIs and students enrolled from the 1970s. • Mergers and consolidations into larger institutions occur rather frequently, while HEI closures are rare and usually limited to private HEIs.
Subject specialisation in European higher education	June 2016	• The core of the European higher education landscape, for both education and research, is composed of generalist institutions. • The rationale for and characteristics of specialised institutions differ substantially between countries, depending on specific national conditions. • Private specialised institutions often focus on Social Sciences and Business and Law, public specialised institutions often cover Humanities and Arts and Engineering.
Student mobility	November 2016	• Degree mobility at the bachelor level remains concentrated among a small number of HEIs (foreign campuses, foreigners' HEIs). • Erasmus credit mobility has its greatest impact in countries where total mobility at the Bachelor, Master and PhD levels is low. • Mobility at the master level is selective and includes several large HEIs. • Mobility at the PhD level occurs mainly towards countries with strong international reputations.
Universities and regional development*	March 2017	• Having a Higher Education Institution geographically close is a major advantage for the student population, for families and for firms. • There are large differences in the spread of Higher Education Institutions (HEIs) at the geographic level. • HEIs are an urban phenomenon: the largest number and density of HEIs are found in metropolitan areas and capital cities. • On average, one may find an HEI in each area with a diameter of 50 km. • However, there are as many as 580 provinces or small regions with no HEIs.
Size of HEIs	June	• Almost half of the HEIs in ETER are very small to small; 42% are

in European countries	2017	medium sized and only 12% are large to very large. • The size distribution of HEIs displays large differences between European countries and is also linked to structural characteristics that exist in their HE systems. • Size is strongly correlated with the institutional profile in terms of institutional category, legal status and coverage of educational levels and fields.	
Gender balance of academic staff	August 2017	• The majority of European countries and Higher Education Institutions (HEIs) have achieved gender balance (40-60%) among academic staff. • A high proportion of female academic staff, yet female full professors are in the minority. • Strong variation in the proportion of female full professors at institutions within countries. • Institutions with a high proportion of female full professors are often small, and found in fields of science such as nursing, education or art.	

ETER dissemination activities

During the course of the project, ETER organised a large number of dissemination events and activities targeted at three types of audiences: the European Commission and international organizations, national policymakers and higher education managers, scholars in research policy and higher education. Some of these events were in collaboration with the FP7-RISIS project, as ETER is officially recognized as one of the major infrastructure contributors on science studies for this project (risis.eu).
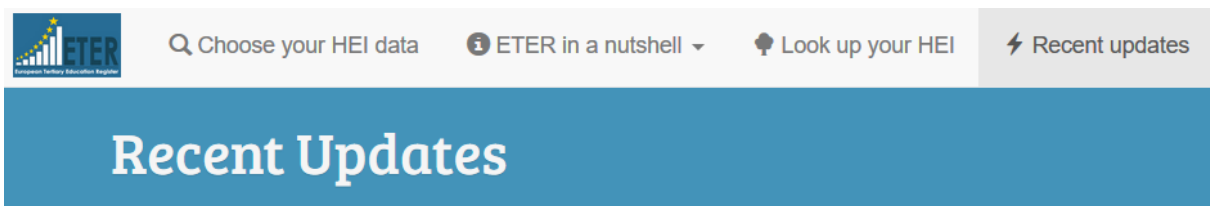
Table 4. List of dissemination events

Date	Location	Event	Remarks
05.09.2015	Berlin	Science and Technology Innovation Conference	ETER poster and presentation
26.11.2015	Paris	OECD DSTI ETER presentation	Presentation of ETER by B. Lepori and discussion of usage by OECD.
16.03.2016	Paris	OECD NESTI workshop	Presentation by A. Bonaccorsi
22.04.2016	Brussels	RISE expert group presentation	Presentation by A. Bonaccorsi
23.05.2016	Paris	Dissemination workshop at the French Ministry of Research	Presentation by A. Bonaccorsi
02.06.2016	Luxembourg	Working Group Education and Training Statistics June 2-3, 2016	Presentation of ETER by B. Lepori
14.06.2016	Warsaw	Warsaw School of Economics, in collaboration with Ministry of Research and Ministry of Economy	Presentation by A. Bonaccorsi
23.06.2016	Berlin	Dissemination meeting ZWB Berlin	Presentation by A. Bonaccorsi
11.09.2016	Brussels	JRC General Conference on Human Capital and Regional Development	Invited talk on ETER
17.11.2016	Copenhagen	Dissemination meeting at Copenhagen Business School	Presentation by A. Bonaccorsi

15.12.2016	Paris	OECD Working Party on Technology and Innovation Policy (TIP)	Presentation by A. Bonaccorsi, discussion on integration of ETER with the knowledge triangle project by OECD
25.01.2017	Athens	Athens University, in collaboration with Greek Ministry of Research	Presentation by A. Bonaccorsi
31.03.2017	Brussels	Meeting with Austrian Ministry of Research, OECD delegates and representatives of the Austrian research community	Presentation by A. Bonaccorsi and by Joanneum Research
01.06.2017	Brussels	ETER final dissemination workshop	EC, international organizations

ETER Newsletter

The ETER newsletter has been implemented through the MailChimp tool (https://mailchimp.com/), a tool for sending mass e-mails, which provides reach features for creating the layout of messages, customizable content and up-to-date statistics on message views. The messaging tool is interfaced with the ETER website, where outgoing messages are listed and their content can be consulted.

Figure 6. The Recent updates section of the ETER website

Ten messages were delivered from February to September 2017. Six presented the ETER policy briefs, while the four remaining messages included important information on ETER (a welcome to users, information on the possibility of visiting ETER teams, announcement of the ETER data collection 2014, the new ETER tender).

The mailing list has been populated by various sources: including contacts provided by the European Commission, registered users of the website, users who participated in the

dissemination workshops (mostly from ministries and national agencies), contacts from the ETER team of individuals involved in higher education management (both academically and in practice). The list therefore is largely representative of public interest in ETER data. It has been expanded progressively to 771 individuals.

As shown by Figure 7, the opening rate of the ETER messages has been quite stable and, for most messages sent, exceeded the 30% threshold, a remarkable level for mass mailing.

Figure 7. Messages sent and opened by the mailing list
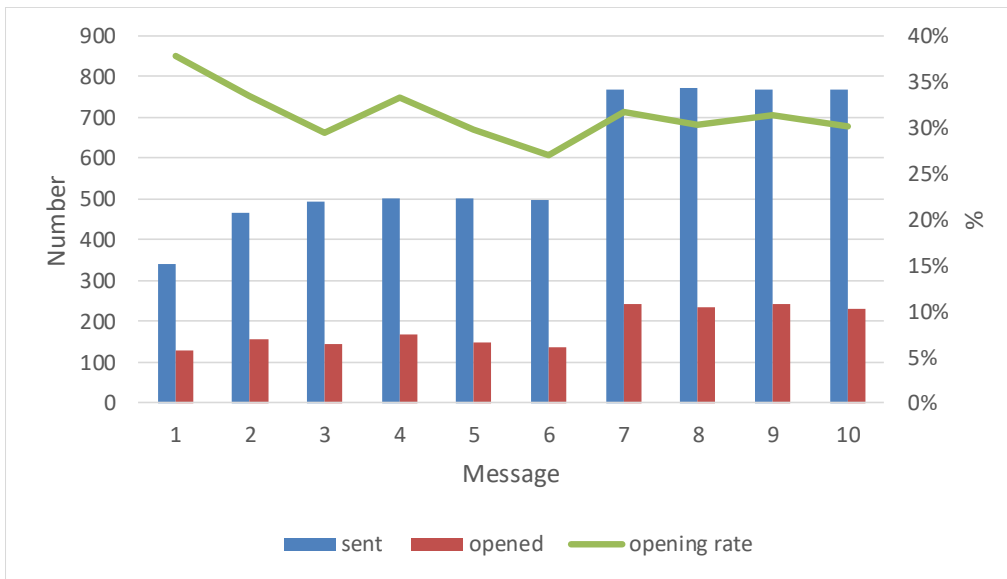
The location of subscribers who are opening ETER messages (Figure 8) shows that ETER dissemination has reached nearly all European countries, with high attention coming from the US and subscribers throughout the world.

Figure 8. Locations where messages were opened

Top locations by opens

🇺🇸	USA	96	17.9%
🇬🇧	United Kingdom	87	16.2%
🇨🇭	Switzerland	48	9.0%
🇵🇱	Poland	35	6.5%
🇮🇹	Italy	35	6.5%

Website usage

Use of the ETER website increased significantly over time. The best indicator to measure the usage of a website for statistical data is the number of users and their development over time. From the start of the website in mid-June, 308 users had already registered

with ETER by August 19. Taking into account the comparatively small community of potential users, this number is seen as a satisfying achievement. An additional positive signal about ETER's web use is the gradually increasing number of users. After a very intense starting phase, there are still more than 10 new observable users every month. The increase in the number of users was also supported by the newsletter's release in February 2017.

Figure 9. Cumulative number of registered users
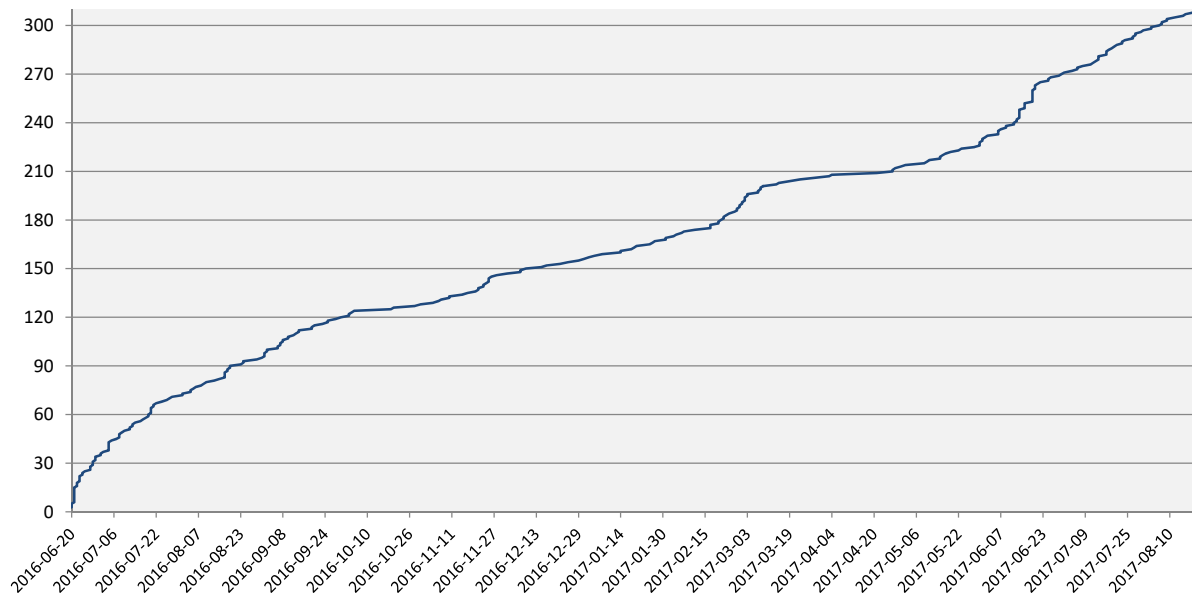
A detailed look at the list of users shows a very heterogeneous picture in different respects. Besides the fact that users are spread all over Europe, researchers from other continents also show their interest in the ETER microdata. Additionally, users differ in their sector of work. Alongside a large part of users from the academic sector, there is also large interest from the non-academic research sector and the public sector, which includes the European Commission as well as ministries and national statistical authorities. This is also confirmed by the website's statistics. A look at the number of sessions by country shows how users are spread around the world (see Figure 10). Besides a high rate of access from countries where ETER core team members are located (Austria, Italy, Switzerland and Norway), users in the United States (7.5 % of all sessions), Belgium (6 %), Germany (5.1 %), Spain (4.8 %), France and the United Kingdom (4.5 % each) show high interest in the ETER dataset.

Figure 10. Distribution of sessions on the ETER website
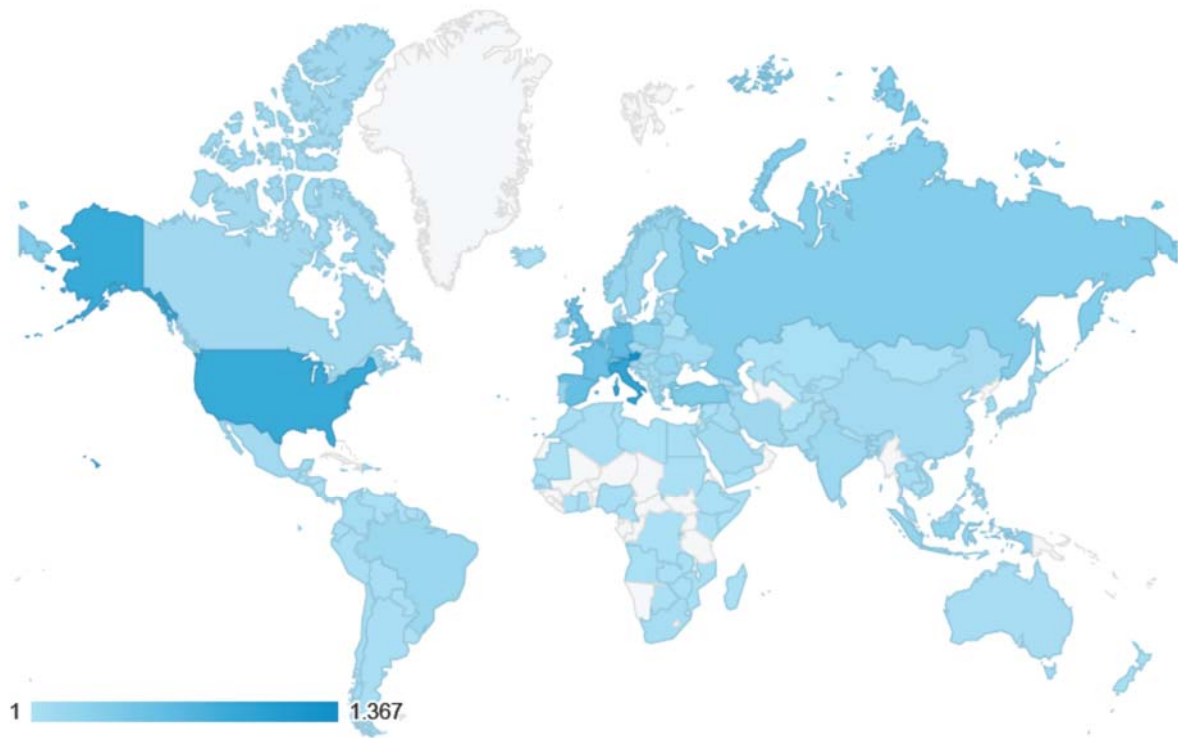
The website's statistics also show sustained interest in ETER by the number of users visiting the web application. On average, 674 users access the ETER website each month (May-July, users with at least one visit). From the launch of the website, there have been 11,584 (until July 31th) sessions of users and 8,709 users with at least one visit. The number of sessions is still quite stable after an enthusiastic start; the increasing share of returning visitors (29.9 % compared to 27.4 % in mid-April 2017 and 24.6 % in mid-November 2016) is a good indicator that the web application was successfully built in a way that is dedicated to the user's needs. This means at the same time that 70.1 % of the visitors are new to the website.

The website statistics show that, as expected, the main function for users is the possibility to search and export microdata on higher education institutions. A significantly smaller share of users access the menu "Look up your HEI", while in the "ETER in a Nutshell" section, especially general information and the ETER briefs are often chosen.

Table 5: No. of Users and Sessions of the ETER web application

Year	Month	Users	Sessions
2016	June	731	944
	July	583	722
	August	484	644
	September	529	683
	October	602	738
	November	715	1.058
	December	583	737
2017	January	626	805
	February	618	771
	March	677	872
	April	540	725
	May	692	1.002
	June	716	1.055
	July	613	828
	Total	**8.709**	**11.584**

Results of the data collection

Data in ETER are available for the years 2011 (academic year 2011/2012), 2012, 2013 and 2014[15]. For these four years, ETER grew into a widely accepted database with an impressive amount of data – the current ETER database contains more than 2 million data cells.

This chapter presents an overview of the results of the data collection, focusing on the following aspects: First, we discuss the coverage in terms of countries and HEIs, i.e. to which extent ETER covers all relevant HEIs in a country. Second, we present some general descriptive statistics on the content of the database and its level of completeness; reasons for missing data are also discussed. Third, we shortly present the results of data quality and their implication for data usage. The last part of this chapter provides evidence on the current use of ETER data for policy and scholarly purposes.

Coverage of countries and HEI systems

The ETER perimeter includes 36 countries, i.e. EU-28, EEA-EFTA countries (Liechtenstein, Iceland, Norway and Switzerland) and candidate countries (Former Yugoslav Republic of Macedonia FYROM, Montenegro, Serbia and Turkey). For all the included countries, the HEI perimeter and the HEI and geographical descriptors are included in the database for all years, so that ETER indeed provides a register of European higher education.

However, for a limited number of countries, no statistical data are available, mostly due to workload and organizational problems at the NSAs in the concerned countries: these are the French part of Belgium, Montenegro, Slovenia and Romania. Additionally, for FYROM and Turkey only basic data has been collected by the ETER consortium from the NSA website. Finally, for Luxembourg only data for 2011 are available and for Denmark 2014 data have not been delivered.

When taking into account the fact that ETER was a voluntary data collection, in addition to the regular delivery to EUROSTAT for the NSAs, the outcome in terms of country coverage can be considered as very good.

---

[15] Data for the year 2008 were collected by the EUMIDA project and have been made available through the RISIS project in a format compatible with ETER (orgreg.joanneum.at). They are available only for research purposes.

Figure 11. Country coverage of ETER

ETER data availability
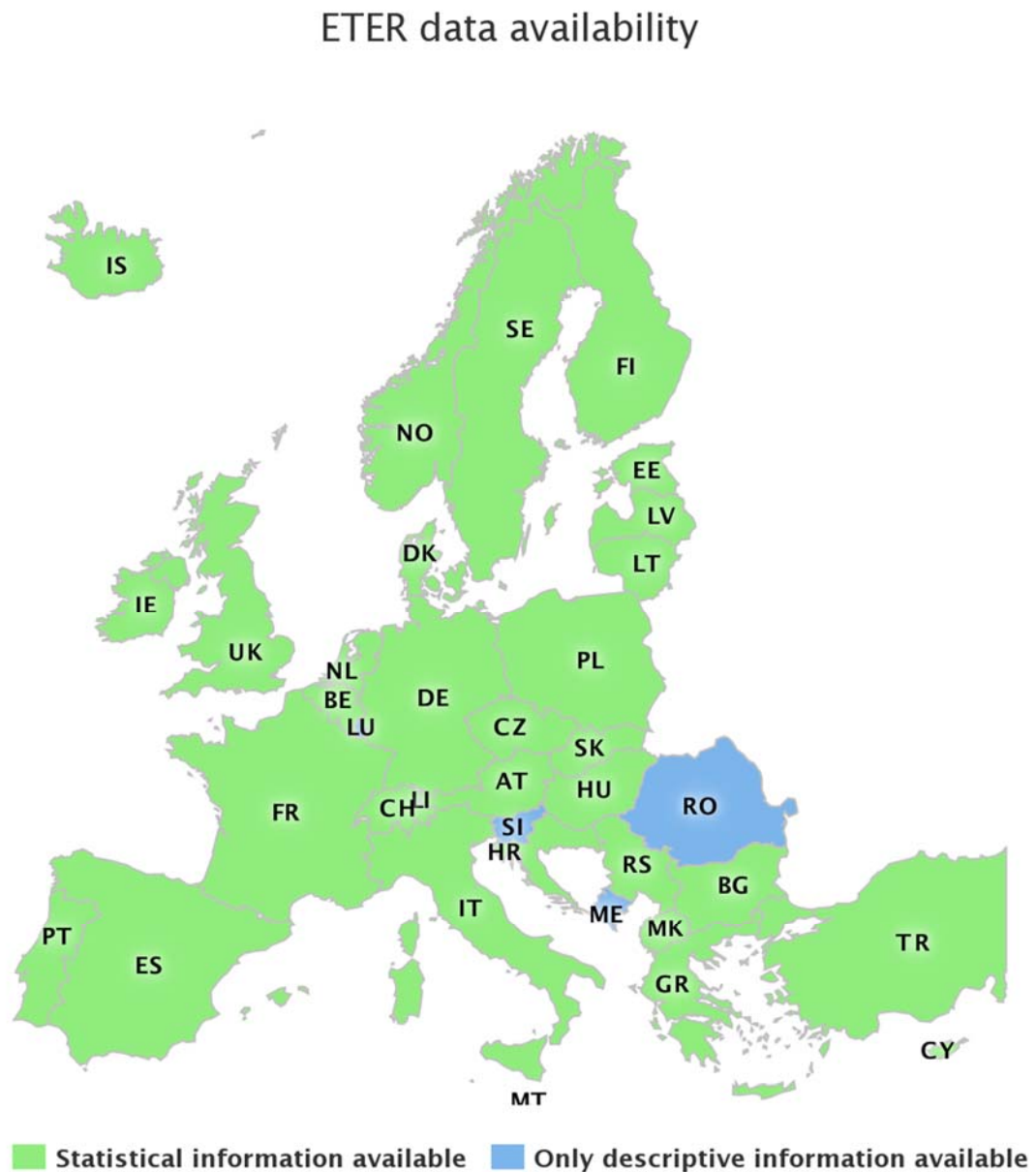
Highcharts © Natural Earth

As introduced in a previous chapter, the perimeter of HEIs to be included was subject to debate. Therefore, during the contract, the ETER coverage was benchmarked with the whole perimeter of tertiary education, as defined in the UOE data collection. For each country, a detailed comparison between ETER coverage and the whole sector of tertiary education by using the so-called ISCED mappings of educational systems was made[16]. ISCED mappings provide a fine-grained identification and classification of national educational curricula, providing information on their characteristics, classification, number of students, and level of qualification. For our purpose, the curricula in ISCED mappings were attributed to groups of HEIs and checked against curricula covered by ETER. This allowed a fine-grained correlation between the curriculum-based approach of

---

[16] http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx.

UOE and the institution-based approach of ETER in order to better understand the sources of differences in the numbers of students and/or graduates.

These mappings are published on the ETER website and allow the user to assess, which groups of tertiary education providers are included and excluded in each country[17]. The main conclusions of these systematic comparisons are the following:

- ETER coverage is much larger than universities, and includes almost all colleges and large numbers of specialized schools like art and music schools (only 40% of the ETER HEIs have the right to award PhD degrees).

- In terms of the number of students, the ETER figures by country are very close to EUROSTAT national aggregates at ISCED levels 6 (bachelor), 7 (master) and 8 (PhD), for most countries, showing that the perimeter is largely the same (see Figure 12). For these three levels, ETER included in 2013/2014 95% of the students (ISCED 6), respectively 100% (ISCED 7) and 85% (ISCED 8).

- France, Germany and Switzerland represent the only country cases where ETER coverage at ISCED levels 6 and 7 is relatively low when compared with EUROSTAT, due to the existence of a large professional sector delivering degrees at ISCED levels 6 and 7 and, for France, to the exclusion of some relevant HE sector (like most business schools).

- ETER coverage at ISCED level 5 is limited to HEIs delivering also degrees at ISCED level 6 and 7. Therefore, coverage at this level is relatively low when compared with EUROSTAT (34% of the students in 2013/2014), but varies strongly by country depending on whether ISCED 5 education is provided by separate institutions or not.

Figure 12. Coverage of ETER as compared with EUROSTAT, 2013/2014
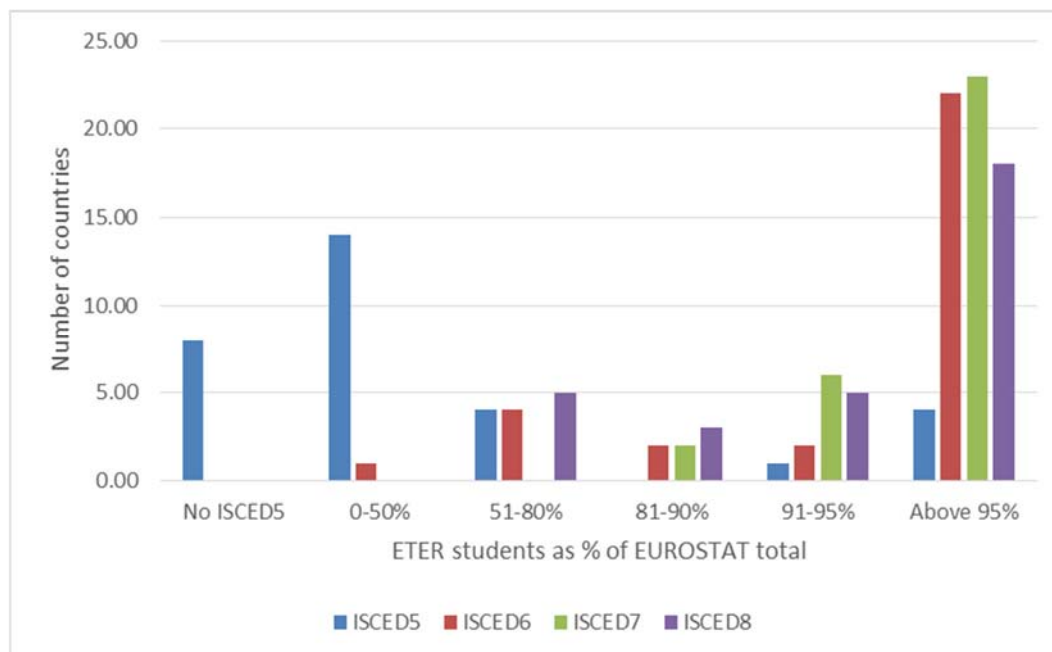
This analysis showed that there are different reasons for exclusion (by the NSAs) of tertiary education providers at ISCED level 5: first, in general, data collection for short-term education is less complete and structured than for higher education, which is usually subject to more policy attention. Second, short-term education is organized on a regional basis, particularly in federal countries like Germany and Switzerland, and therefore data is not readily available at the national level. Third, the average size of

---

[17] See https://www.eter-project.com/about/perimeter.

institutions delivering ISCED 5 degrees is much smaller and, therefore, data collection is more difficult when compared with higher education.

In terms of future extension of data collection, the consortium considered that, while some additional inclusions in ETER might be relevant, extending the current ETER setting to all tertiary education would not be feasible in terms of data availability and the data collection effort, as it would probably amount to multiplying by a factor of 5 the amount of data (for an additional coverage of 15% of students). One possible option that could be envisaged is to selectively extend the coverage to some institutions delivering only ISCED 5 degrees including only descriptors and geographical information, since these data are basically collected once, as well as a few core variables, like the number of students and degrees by level. The feasibility of this option must first be carefully assessed.

The ETER data: general presentation

Table 6 provides some general descriptive information on the ETER database. The database covers 36 countries and includes roughly 2,700 HEIs for each individual year – with fluctuations due to demographic events, such as new foundations or the accreditation of existing HEIs within national systems. These HEIs enrolled more than 22 million students in 2014 and delivered nearly 5 million diplomas at ISCED levels 5 to 7 – the lower numbers for 2011 and 2012 are due to the missing data for Turkey. These HEIs also enrolled about 700,000 PhD students, which constitute a key component of their research training activities. The same HEIs employed 1.7 million staff in full time equivalents, i.e. nearly one for every 12 undergraduate students – this ratio being underestimated since data for staff are not fully complete – and had revenues over 150 billion euros in purchasing power parities (PPP), the figure is also underestimated due to missing values.

Table 6. General descriptive information on ETER

	2011	2012	2013	2014
Countries covered	36	36	36	36
Number of HEIs	2,683	2,684	2,712	2,764
Students ISCED 5-7	16,667,140	16,257,747	22,178,999	22,725,593
Graduates ISCED 5-7	4,099,981	4,081,889	4,935,728	4,860,344
Students ISCED 8	582,865	582,334	659,635	696,248
Graduates ISCED8	48,810	70,298	56,899	47,987
Staff FTE	1,735,054	1,689,554	1,781,018	1,771,770
Current revenues in bio. PPPs euros	136.53	150.07	153.52	153.77

This very coarse data displays the sheer size of the system that is represented by the ETER dataset.

Further, Table 7 depicts what could be considered the "typical" European HEI, by using median values. As we shall discuss more in detail in chapter 0, these figures have only limited validity since higher education is characterized by a very large internal diversity – the main advantage of ETER being the possibility of analysing such differences quantitatively both within and across countries.

However, to give an order of magnitude, a "typical" European HEI would enrol slightly less than 3,000 undergraduate students and slightly more than 200 PhD students, have 300 units of staff and a budget of nearly 40 million euros in PPPs. This gives a preliminary indication of how European higher education is populated by a large number of small and medium institutions, alongside very few large HEIs[18].

---

[18] See *ETER brief 5. What ETER tells us about the size distribution of Higher Education Institutions in Europe, available at*

Table 7. Median characteristics of HEIs in ETER

	2011	2012	2013	2014
Students ISCED 5-7	2,660.0	2,596.0	2,728.0	2,667.5
Graduates ISCED 5-7	690.0	666.0	691.0	664.5
Students ISCED 8	251.0	258.0	215.5	214.0
Graduates ISCED8	40.0	50.0	33.0	32.0
Staff FTE	327.0	330.1	320.0	301.7
Current revenues in bio. PPPs euros	34.0	37.8	39.4	37.5

Data quality

Introduction

In ETER, data quality was mostly aimed at assessing the internal quality of the (secondary) data collected, as ETER has little control on the external validity of the original sources and, therefore, many substantial issues impacting the comparability cannot be resolved at this stage, but would require changes to the original data collection practices. Nevertheless, these quality checks also provided important information to detect far-reaching comparability issues.

The ETER database reaches a very high level of *format accuracy*, i.e. there are practically no values that do not comply with the format for that variable. This includes rules governing foundation year, which must be four digits, financial values that cannot be negative, and student numbers as integers, etc.

This control is performed systematically during the preliminary validation phase. After which, it is performed again on the complete dataset. The identified deviant cases are either directly corrected or reported back to the national experts for checking and correction. Only a few cases of format inaccuracy are remaining in the dataset, which are explained by national specificities – for example in NL the number of ISCED 8 students is not an integer since figures reported are in FTE based on their contractual percentage. All these cases have been flagged and explained in the remarks and metadata.

For a dataset like ETER, a central dimension of accuracy consists of checking whether no cells are left blank and that rules concerning missing values are respected, particularly the correct distinction between "0" values and missing values (respectively not applicable values).

Data consistency was implemented through a large number of checks, mostly of dependencies between variables[19]. This includes logical relationships between variables – if an HEI does not have the right to award a PhD, there should be no PhD graduates, the sum of individual breakdowns are equal to their totals, and finally, relationships between valued variables, for example the value of academic staff in FTEs should be lower than in HC.

Most reported consistency problems have been corrected in agreement with the NSAs. However, there are a few remaining cases, which have been flagged and documented in the dataset. This includes small-scale differences due to rounding (particularly for financial variables), breakdowns of student/graduate data being non consistent due to the use of different data sources with respect to their totals, and finally, cases due to national specificities, for example HEIs enrolling students where the degrees are delivered by another HEI. These cases have been documented and flagged.

Timeliness of data publication has been assessed by comparing the publication date in ETER with the corresponding national source's date of publication. Reducing this time-lag

---

https://eter-project.com/assets/pdf/ETER_size_distribution.pdf.

[19] The reader should refer to the ETER data quality report, annex to this report, for a full description of these analyses and their outcomes (see chapter 0).

would be important in order to provide the most actual release of data. Published in June each year, the ETER data collection has an average lag of 12-18 months with respect to individual national releases. The choice of a more recent reference year (i.e. 2012 instead of 2011 for the 2014 data collection) to reduce the lag would have implied the exclusion of a number of countries because of temporary unavailability of data. While the schedule of data collection cannot be easily changed since it is synchronized with the UOE data collection, in principle it would be possible to reduce the gap between the beginning of the ETER data collection and the publication of data (currently 9 months), by accepting the publication of partial and provisional data, which might be updated at a later stage.

A further important check for outliers occurred, i.e. values that are significantly different (in a statistical sense) from the expected distribution of values. This was made first by defining simple thresholds, like a value for the expenditures per student above which data are considered suspect, and, in a more refined way, through advanced statistical techniques. The identified outliers have then been manually checked in order to identify whether this represented a data problem.

The outlier analysis detected a number of data mistakes, such as financial figures being reported in thousands instead of units, or columns or rows being swapped in the data. Most identified outliers however were of substantive nature, and pointed to the specificities of a particular data point. A good example is the Dutch distance university, whose expenditures per student are very high since the HEI mostly offers courses for students enrolled in other Dutch universities and, therefore, the number of students formally enrolled does not reflect the teaching burden. All such cases have been flagged and commented.

Multiannual checks are based on the same principle, i.e. to identify cases, which are suspect since their level of variation (relative *and* absolute) is very large in respect to what can be reasonably expected – a university with 50,000 students losing half of its enrolment within one year being a typical case. These checks uncovered a number of cases where data for a single year were not reliable and had to be corrected. Other cases pointed to a change in the definition and data collection methodology, for example changing reference dates for staff counts or rules to classify national degrees to the ISCED classification or, even, a change in the time limit to be considered a student, which strongly affected data in one country. Other cases were due to demographic events – the budget increasing suddenly because of a merger – or to HEI specificities, like an HEI being progressively closed down. Finally, a number of strong fluctuations, particularly concerning financial data, remained unexplained.

The latter remark emphasizes that this set of quality check was essential to increase the internal quality of ETER data especially at the beginning of the contract, leading to a large number of corrections, but over the years the cases identified increasingly were already known from previous years or had no simple explanation or couldn't be corrected because of underlying problems in the primary data. Therefore, while these kind of checks must remain, as they are a core dimension of the ETER system, a move towards a more substantive engagement with the comparability of data is advised for the future.

Completeness of Data

Data completeness was a major concern in ETER. From the EUMIDA pilot study it was known that data routinely collected by the NSAs within the framework of EUROSTAT educational statistics were much more available than other types of data, which have been less standardized by EUROSTAT, like financial and staff data.

ETER results broadly confirmed this pattern, but brought a number of improvements, notably concerning staff data – both for the aggregate variables and for a number of breakdowns of high policy relevance, like academic staff by citizenship (Lepori *et al*, 2014).

At an aggregate level, the whole ETER database reached a completeness level of 73% in 2011 and 2012, which increased to 76% in 2013 and 78% in 2014. This score excludes

countries where only descriptive information is available, i.e. the French region of Belgium, Montenegro, Slovenia and Romania.

However, this average conceals large variations by countries and variables.

A large group of countries within the analysed dataset have a very high level of completeness (over 85%) including BE (Flanders), BG, CH, CY, DE, ES, IE, IT, LI, LU, MT, PT, SE and UK, a second group with an high level (50%-85%) including AT, CZ, DK, EE, FI, FR, GR, HR, HU, IS, LT, LV, NL, NO, PL, SK and a third group with a fair data availability (below 50%) including MK, RS, TR.

Variation is equally very large by variable (see Table 8).

Table 8. Completeness by variable and reference year for some key variable
Remark. This table underestimates the improvement in availability over the project since in many cases new data were made available also for the former years. For more detailed information see the Data Quality Report (Annex 7.2).

Variable name	2011	2012	2013	2014
Legal status	0.99	1.00	1.00	1.00
Foundation year	0.98	0.99	0.99	0.99
Name of the city	0.99	1.00	1.00	1.00
Total Current expenditure	0.48	0.52	0.58	0.57
Total Current revenues	0.48	0.52	0.54	0.51
Total academic staff (FTE)	0.69	0.66	0.70	0.70
Academic staff by sex	0.61	0.59	0.69	0.74
Academic staff by citizenship	0.38	0.38	0.51	0.55
Academic staff by FoE	0.27	0.29	0.30	0.35
Students enrolled at ISCED 6 by sex	0.88	0.85	0.95	0.94
Students enrolled at ISCED 6 by citizenship	0.82	0.80	0.84	0.81
Students enrolled at ISCED 6 by mobility	0.52	0.52	0.54	0.52
Total students enrolled ISCED 5-7	0.91	0.89	0.98	0.97
Total graduates ISCED 5-7	0.87	0.87	0.97	0.94
Total students enrolled at ISCED 8	0.73	0.70	0.87	0.92
Total graduates at ISCED 8	0.77	0.66	0.84	0.75
R&D Expenditure	0.18	0.16	0.17	0.15

Descriptor availability is nearly 100%, with the sole exception of three variables, i.e. the acronym (many HEIs simply don't have one) and legal status and ancestor year, which are optional. The same applies for geographical information, which is almost complete.

Staff data is the area where ETER made the most significant progress. The most important variable, i.e. academic staff in FTEs, is available for 70% of HEIs. Four countries delivered data in headcounts only (FR, GR, LV, TR), while academic staff data are not available for only three countries: AT, EE, IS. When considering HEIs with staff data available in FTEs or HCs, these enrol 97% of undergraduate students, as most of the missing cases are very small HEIs. There has been a general improvement over the project with respect to EUMIDA in terms of the availability of breakdowns by gender and, even to a large extent, citizenship, information which is of increasing policy relevance. The availability of staff breakdown by field of education remains very low.

The area of main concern in terms of data availability (and comparability) is financial data. Despite improvements over time, there is still a large number of countries, mostly from Central and Eastern Europe, which did not deliver financial data at all, while in two large countries (France and Poland) data are available for less than half of the HEIs (which include however the large universities). This implies that availability of financial data weighted by the number of students (or PhD students) is higher than figures based on the number of HEIs. For countries that delivered data, the availability of breakdowns

has generally improved and is fairly complete over time, with the notable exception of capital expenditures, whose availability is lower.

Student and graduate data is fairly complete, which is expected since these data are part of the EUROSTAT-UOE data collection. An exception remains the breakdown by mobility, which has not yet been introduced in some countries – a situation expected to improve over time as more countries introduce the new EUROSTAT standards. The availability of breakdowns is also lower for countries with data collected by the consortium.

Finally, R&D expenditures have the lowest availability over all ETER variables. This is due to the fact that, in order to compute this variable at the institutional level, a survey to determine the use of academic staff time is required, following the Frascati manual methodology (OECD, 2015). This is not the case in many countries, including some of the largest European countries like France, Germany and Italy.

Exhaustive discussions have been conducted during the project on how to improve data availability, notably for staff and finances. It turns out that there are three broad groups of missing data:

- Data for which European guidelines exists, but are not yet implemented by all countries. In these cases, the situation will improve progressively over time, as evidenced by student breakdowns of field and mobility.
- Data that is available in some form in some countries, but with a lack of standardization – particularly staff and financial data – a complicated issue given the diversity of personnel and the financial arrangements of European HEIs. The consortium therefore strongly suggests an investment in resources for future ETER data collections in order that the ETER consortium can spend resources in helping NSAs to provide and standardize the data.
- Finally, data that are not currently collected and for which little expectation for improvement – specifically R&D data.

An important implication of the scattered data availability is that analyses using ETER data have to be carefully designed on the grounds of data availability for individual variables, in order to avoid situations where the number of cases becomes too low when combining variables.

Comparability of data

Data comparability is a particularly difficult issue for higher education, given the diversity of national systems, HEIs and data collection practices. The concept itself is surprisingly subtle and elusive: that data are comparable means in general that users are confident that the results of some comparisons between objects (in our case HEIs) are meaningful for their specific purposes. Comparability does therefore not mean that data are exactly the same, which would be not possible, but that potential differences are considered not relevant for some kind of scholarly or political conclusion. In this respect, ETER is not different from official statistics, which shares similar problems, particularly at the international level (Godin, 2005).

In practical terms, it is useful to distinguish between different sources of comparability problems, as the potential remedies are different.

A first source of comparability issues are simply measurement errors or issues with the data collection, while methods and definitions are comparable. For example, figures on the number of PhD students in Germany are underestimated, since not all are enrolled at universities during the full duration of their thesis.

A second source are differences in the underlying definitions. For example, some countries might set a maximum duration of enrolment to count students, differences in this respect might strongly affect the comparability of the figures.

More complex issues derive from differences in the national statistical or accounting systems between countries, for example between countries adopting a cash accounting

system and countries adopting a commercial accrual system, which strongly affects how capital expenditures are reported.

Finally, other issues related to more basic differences on how national systems and/or individual HEIs are organized, implying that the same figures might have a completely different meaning in each context. For example, in some countries a PhD student is a student following some courses and writing a dissertation, while in other countries PhD students are paid research fellows who work in a research team. Also, a statistical category like 'professor' might have a very different meaning in a research university, when compared to a college.

This discussion also implies that there is not a single strategy for dealing with comparability, but rather the need to resort to a combination of approaches.

A first approach is a standardization of definitions, by providing common definitions at the European level and detailed examples on how to implement data collection. This is the approach followed by educational statistics (UOE, 2013) and R&D statistics (OECD, 2015), which was largely adopted in ETER and recorded in the ETER handbook, which provides detailed methodological guidance on how to collect data for this project. Standardization is also important in order to be able to report methodological deviations in metadata. For ETER, the major limit of this approach is that no primary data collection is foreseen and, therefore, secondary standardization is limited by the available data.

A second approach is mapping national data sources and categories to the ETER definitions in order to achieve cross-country comparability. This approach was, for example, essential for the breakdown of HEI revenues, where each country uses its own classification system. Ex-post standardization is therefore a choice approach for administrative data, which have been collected for other purposes. It was adopted in ETER to some extent, but there is important potential in this direction by working more directly with the NSAs on handling the primary data.

A third approach is based on imputation through statistical techniques, when data clearly do not comply with definitions. This approach has not been used until now in ETER, where all data are identical to those delivered by the NSAs, but could be adopted in future to correct some deviations reported in the metadata, like data reported in headcounts rather than FTEs, the inclusion of healthcare in HEI revenue data or the use of contractual equivalents for PhD students. The availability of a large multi-annual dataset where regularities between variables can be analysed statistically suggests that such techniques could be quite helpful to improve the quality and comparability of data for final users.

In the ETER data, the following major systematic *comparability problems* were identified:

a) Despite more precise guidelines for financial data, the exact mapping of income categories with ETER breakdowns is still not perfectly clear in several countries, implying possible comparability problems (i.e. HEIs with a negative core budget in DE), which are largely due at national specificities in the primary financial data. In-depth analysis of the financial data by the ETER team might solve some of these issues, but this requires the collaboration of the NSAs and intensive effort, including probably site visits from the ETER team.

b) Concerning staff, specificities about the inclusion and classification of staff across countries and within countries among HEI categories (typically university vs. colleges) may impact full comparability. Figures reported might be incomplete excluding some categories (i.e. atypical staff in UK). A major issue remains the scattered availability of FTE and HC measures of staff, which strongly limits the usability of data. Since *some* staff data are available in most countries, the use of imputation techniques to standardize them would be particularly helpful.

c) The classification of students and graduates according to the new ISCED levels it is still not perfectly stable in every country, but it has been possible to solve with an ad hoc

concordance table for ETER. In some cases figures reported exclude minor categories or programmes. Also, breakdowns of students and graduates by mobility status are not fully comparable among countries. Finally, problems due to the jeopardised application of the new ISCED classifications -by level of education and by field of education- were finally solved in 2014 with the complete adoption in all countries. This had a positive impact, especially for the comparability of data by field. In these areas, close monitoring of national practices remains important to avoid new comparability problems.

ETER usage and added value

In a broad sense, four main categories of ETER usage can be distinguished, each with different preconditions and requirements for the data infrastructure (Table 9).

Table 9. Categories of ETER usage

Category	Prerequisites	Status
ETER data for policy analyses	Yes, but requires some in-depth work Usually combined with other sources	Some first examples emerging.
ETER data for scholarly analyses	Is improving over time Strongly enhanced by data combination.	Is emerging, will develop rapidly in the coming years.
ETER as a reference perimeter	For projects focusing on a subset of HEIs.	Some good examples, little prerequisites.
ETER as an automated data source	Other projects can directly interface with the ETER database and automatically extract data	Some first requests received, requires a suitable legal framework for access.

ETER potentially covers many different policy domains, since it has been designed as a general public resource, which can be freely exploited and combined with other sources in order to analyse relevant problems. This is clearly illustrated by the cases presented in Table 10 and the analysis presented in the following chapter 0. This list however is far from complete, as ETER is an open resource, and therefore, the ETER team is not immediately notified of most users or uses.

Table 10. Selection of use cases

Policy domain	Type of usage	Description
Academic staff	Policy analysis	The Eurydice report *Modernisation of Higher Education in Europe: Academic Staff - 2017*[20] explores the current realities for academic staff within the changing higher education landscape in Europe. The report is based mainly on qualitative data gathered by the Eurydice Network, covering higher education systems in 35 countries. In addition, quantitative data from Eurostat and the European Education Tertiary Register (ETER) are also used, as well as information gathered from surveys developed for this report to academic staff Trade Unions and Quality Assurance agencies. Despite the analysis being at the country level, ETER data on full professors where

---

[20]

https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Publications:Modernisation_of_Higher_Education_in_Europe:_Academic_Staff_%E2%80%93_2017

		used, since this variable is not available in EUROSTAT.
Compari ng US and Europe	Policy analysis/ scholarly analysis	As part the impact assessment of public policies at OECD, data from ETER were combined with IPEDS data and bibliometric information to compare the European and US higher education. Panov C., Scott R. (2015) How to meet the demand for more and better research, education and innovation? An empirical analysis of universities in Europe and the United States, DSTI/STP/TIP(2015)15/REV1 This work is being further extended and developed within the RISIS project to a wide-ranging comparative studies of excellence in both systems. Lepori B., Geuna A., Mira A., Money matters, but why? Scaling properties of US and European Universities, paper to be presented at the Atlanta Conference on Science and Innovation Policy, October 2017.
Quality insuranc e	Register function	The European Quality Assurance Register for Higher Education (EQAR; https://www.eqar.eu/) *aims at constituting a register of evaluation of higher education curricula, EQAR plans to use ETER as a register and statistical basis for* the collection of data on curricula.
Regional studies	Policy analysis	Use of ETER data to identify the presence and size of universities in European cities (students, academic staff, number of universities) as indicators for the DG Education and Culture- JRC Cultural and Creative Cities Monitor, 2017[21].
Impact analysis	Policy analysis	The RISE group book 'Europe's future: Open Innovation, Open Science, Open to the World' makes use of ETER data for the analysis of the HEIs on regional development in Europe.
Rankings	Data source	Data for the U-MULTIRANK tool have been completed with ETER data, particularly for some descriptors like foundation years, as well as for missing quantitative information. ETER also allowed for cross-checking some of the data collected from the universities.
Educatio nal statistics	Data source	ETER data have been directly interfaced with a blog of the Finnish Ministry of Education about statistical indicators, to provide comparative evidence across Europe. Indicators can therefore be generated on-line from the ETER platform. https://app.powerbi.com/view?r=eyJrIjoiNmI0ZmYwMzYtMDIxNC00M2NkLWI zMzQtMTljM2NkMzQ1NWFiIiwidCI6IjkxMDczODlkLTQ0YjgtNDcxNi05ZGEyLW M0ZTNhY2YwMzBkYiIsImMiOjh9

Limitations and future strategies for expanding the ETER usage

The information presented in the previous sections documents that ETER is quickly becoming the reference dataset on higher education institutions in Europe. The amount of usage has rapidly increased and is expected to increase further in the coming years.

At the same time, a micro-data database like ETER is challenging for users for different reasons: the sheer amount of data, which requires the use of statistical software for extractions and analyses; its multi-level nature with variables nested within HEIs nested

---

[21] https://composite-indicators.jrc.ec.europa.eu/cultural-creative-cities-monitor/media/c3monitor2017.pdf

within countries, which requires a suitable analytical design; finally, the limitations in terms of availability and comparability of data, which requires a high degree of familiarity with the database in order for it to be used in an adequate way.

This limitation, which was clearly expressed by some ETER users, is particularly relevant for non-expert users, who do not have the resources for an in-depth analysis of the database, and therefore are somewhat limited in the scope of usage.

The consortium suggests that a suitable strategy to deal with this issue is the construction of a core set of data and indicators with the following characteristics:

- Including only a core set of variables, as the experience of usage shows that most users will focus on a few aggregated variables, like total staff or total students and do not need many of the more detailed breakdowns.
- Also including some pre-computed indicators of interests, like indicators of subject composition, therefore avoiding complex calculations by users.
- Improving completeness by imputation for some key variables, particularly for academic staff, where data is available in FTEs or HCs.
- Excluding countries and variables for which the availability of data is too low for analytical purposes.

These data could be made available on the same interface as the main ETER database, for example as a specific set of variables – a similar approach as the one adopted for the main statistical tables section of the EUROSTAT data interface. Therefore, users could easily and smoothly switch to the main dataset and include additional variables if needed.

The sub-dataset could therefore include 20-30 variables by HEIs (against 250 for ETER) and should reach a higher level of completeness (well above 90%) and of comparability.

An helicopter view on the diversity of European Higher Education Institutions

As already highlighted, the main contribution of ETER to our understanding of European Higher Education lies in the possibility to analyse differences between individual HEIs, which is not possible with national or regional-level data provided by EUROSTAT.

Indeed, higher education literature has highlighted that HEIs differ widely in terms of their size, the balance between education and research, the scope of subject domains covered (Huisman et al, 2015), with the two extremes being represented by the large research-oriented generalist universities, like Oxford or Munich, and by the small specialized schools focused on training in fields like arts and architecture. Such diversity, until now, has made attempts to develop a classification of European HEIs, like the well-known US Carnegie classification (http://carnegieclassifications.iu.edu/) difficult (see Lepori et al, 2017).

The implication of such diversity is that speaking of an "average" European HEI is not meaningful and that even the question "How many HEIs do we have in Europe?" depends on the specific perspective adopted. Such issues are also relevant from a policy perspective, as policy interventions might be quite different depending on the subpopulation of HEIs concerned: just to provide two examples, first, the "best" size or critical mass of an HEI will be widely different for an art school and for a research-oriented international university; second, spreading higher education across Europe does not necessarily require that all regions host a research university, but different functions might be distributed differently across European regions.

In this chapter, we build on work done in the last ten years by the ETER team and other scholars, as well as on the data analysis performed in the ETER briefs, in order to highlight some aspects of this diversity and their policy relevance.

A layered view of European higher education

The EU's agenda for modernizing higher education highlights the need for wide diversity of Higher Education Institutions (Jongbloed and de Boer, 2012) and renewed the EU Agenda for Higher Education published in May 2017. Maintaining diversity is considered by most scholars as a suitable strategy to respond in an effective way to the diverse requests of higher education, including international research competitions, education at various educational levels and knowledge transfer to the local society.

An important dimension of such diversity is regulation: the presence of HEIs with different legal statuses and institutional mandates also fosters the differentiation of profiles and functions, as demonstrated by the US system (Cohen, 2007). A central question in the higher education policy debate indeed concerns the best means to promote differentiation of institutional types and activity profiles.

Figure 13 provides an overview of ETER HEIs in terms of two dimensions: the legal status, i.e. publicly-regulated HEIs vs. private HEIs, and the legal right to award degrees, by distinguishing those HEIs delivering only short-term diplomas (ISCED5), those delivering degrees up to the bachelor or master level (ISCED6 and 7) and those HEIs with the right to award the PhD (ISCED 8), a distinctive sign of the presence of a research mandate.

This figure shows that there is a considerable level of institutional diversity in European higher education, since public institutions delivering the PhD constitute only 42% of the HEIs in ETER. However, when looking at activities, these institutions – mostly public universities – account for the largest share of enrolments.

By design, ETER comprises only a very small number of HEIs delivering only short-cycle diplomas. As introduced in chapter 0, there are probably many thousands of tertiary education providers at this level. However, the contribution of these institution's educational activities is relatively limited – according to EUROSTAT, in ETER countries there were 3.5 million students at ISCED5 level, of which about 1.5 million are within

institutions that also grant higher degrees and therefore are included in ETER. This marks a clear difference with the US system, where colleges account for about one-third of student enrolment.

About half of the ETER HEIs deliver diplomas at the bachelor or master level; these include a large number of specialised schools in arts and humanities, as well as public and private colleges, particularly in those countries having a binary higher education system (Kyvik and Lepori, 2010). With more than 3 million student enrolled, their contribution to education is substantial, particularly for public colleges – some of them being among the largest institutions in their countries, for example in the Netherlands, Norway and Switzerland.

In ETER, about one in every five HEIs that award a PhD degree are private, yet these enrol a small share of students (ISCED5-7) and a negligible number of PhD students, displaying their limited contribution to research activities. This again sharply contrasts with the US, where private non-profit universities are among the best research universities in the country.

Figure 13. Institutional diversity in ETER

	Legal status	
	Public HEIs	**Private HEIs**
Diploma	15 HEIs 6,259 students	5 HEIs 4,672 students
Bachelor or master	673 HEIs 2,377,640 students	514 HEIs 729,321 students
PhD	1,049 HEIs 18,488,490 students 695,477 PhDs	224 HEIs 1,052,341 students 19,533 PhDs

(Highest degree delivered)

The summary of this first preliminary assessment is that institutional diversity in European higher education has been developed in two directions. The first direction is the creation of colleges as providers of education at the bachelor and master level; this direction has been particularly successful in a number of countries, which have created as so-called "binary systems", like the Netherlands, Norway, Switzerland, Finland (and to a lesser extent Germany). The second direction has been the entry of private educational providers, a phenomenon common to many countries, but more present in some central and eastern European countries.

Yet, despite these developments, which in Europe started only began in 70s, the European higher education system remains dominated by a core of about 1,000 universities, which account for the largest share of education (in terms of diplomas) and for almost all research activities.

No size fits all

HEI size represents an important policy concern, which is of particular interest as university mergers are on the rise in Europe. A common assumption driving mergers in many countries is that fewer larger institutions are associated with positive effects on costs and increases in visibility. Another factor is that consolidation may lead to an improved quality of teaching and research. However, there is little evidence of these effects (Pinheiro et al, 2016), while studies on the relationship between HEI size and research performance provide divergent results (von Tunzelmann et al, 2003).

ETER shows that the European higher education landscape is extremely diversified[22]. Using student population (all levels ISCED 5-7) to define size classes (Daraio et al., 2011), 18% are very small institutions (≤500 students); 26% are small (>500 to ≤2000); and medium institutions have the highest incidence, representing 42% of total HEIs (>2000 to ≤20 000). Only 1 in 10 institutions are large with more than 20,000 enrolments and an additional 2% (40 universities) are very large with more than 50,000 students (Figure 1). In 2014, the largest university – Anadolu University in Turkey, which offers both traditional and distance education – enrolled more students than the total number of students enrolled in most other ETER countries.

At the same time, ETER shows strong differences between groups of HEIs in terms of their size. For example, universities are more than three times larger than an average institution in ETER (the median size is 10,074 against 2,704), while specialised HEIs are expectedly much smaller than their generalist counterparts. This also implies that differences in the average size of HEIs by country reflect the structure and composition of the national system – for example where universities are the majority of HEIs, the average size is larger than in countries with many colleges and other HEI categories.

To illustrate this effect,

---

[22] See *ETER brief 5. What ETER tells us about the size distribution of Higher Education Institutions in Europe, available at* https://eter-project.com/assets/pdf/ETER_size_distribution.pdf.

Figure 14 shows the average size (circle) by country and type of HEIs. In all countries, universities are larger than UASs and other HEIs. In Switzerland, the Netherlands and Greece, the difference in size between universities and UASs is smaller in terms of student population, as UASs cover a large share of student enrolments in these countries, and went through a process of consolidation through mergers. We also see that comparing the average size by country can be misleading: Spain, where there are only universities, has the highest average HEI size among all countries (except Turkey) because of the composition effect. However, when comparing only universities, their size in Spain is smaller than in Denmark and Greece and fairly comparable with Italy and France, where a large number of small HEIs in other categories lowers the overall average.

Figure 14. HEI size by country and type of HEI
Source: ETER brief
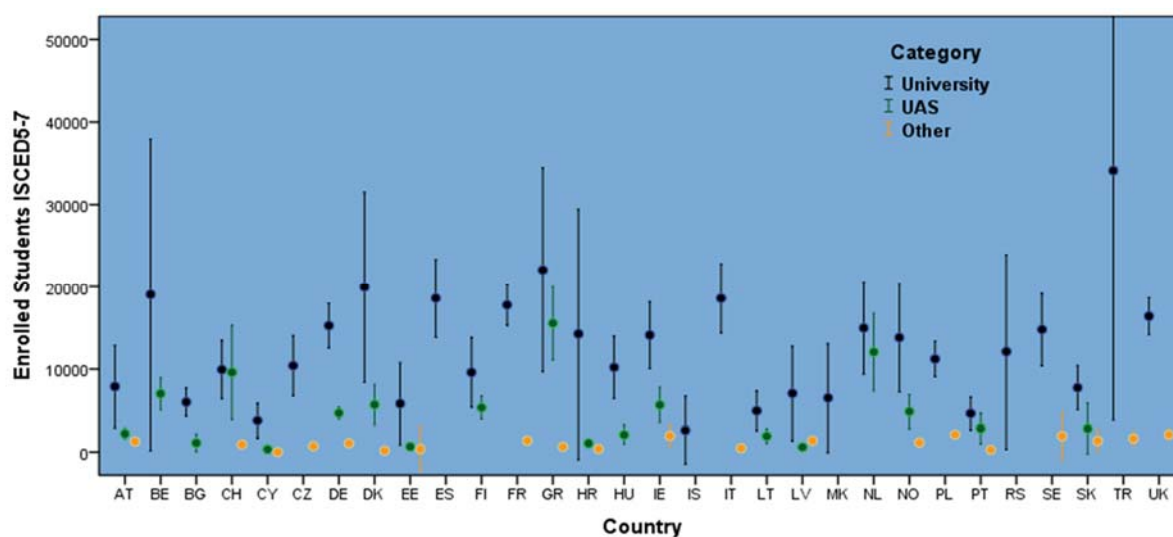
We draw two important implications for policy analysis. First, there is no "one size fits all". The optimal size for a generalist research university is likely to be very different than for a professional specialised school in arts. Second, policy conclusions derived from simple averages of HEI size might be misleading, as they will be systematically affected by the system's composition.

In this context, besides providing data on HEI size measured by different variables (students and staff), the added value of ETER is that it is possible to combine the size measures with information on the HEI legal status (public vs. private), institutional type (university vs. colleges) and subject composition (generalists vs. specialized), in order to ground policies tailored to specific HEI groups.

Generalist and specialist HEIs

A major source of diversity in higher education systems is represented by the scope of subject domains covered (Meek *et al*, 1996). Traditionally, ancient universities evolved over centuries by progressively expanding the range of their subjects, until covering most scientific domains both in teaching and education. But in some fields specialized HEIs have emerged, because of the specificity of the discipline (arts and humanities), of the policy will to promote technological development (technical schools) or of the high demand for education (business school). As a matter of fact, there are good reasons for both approaches as each institution must pursue excellence in line with its mission and strategic priorities. Specialized institutions are commonly known as important providers of professional education, able to target both the markets and students needs with high accuracy. At the same time, these institutions are often very small, which may lead to issues of critical mass. Impact, research performance and visibility seem to be easier to achieve for larger institutions, which also tend to be at the top of research-based rankings.

By extending previous studies (Lepori *et al*, 2010), ETER allows for the first time a systematic analysis of the role of generalist and specialist institutions in European higher education, respective of their importance by mission (education vs. research), type (university vs. college) and country[23].

---

[23] See the ETER brief 1. What ETER tells us about subject specialization in European higher education https://www.eter-project.com/assets/pdf/ETER_brief_subjectmix.pdf.

Based on the distribution of students by ten fields of educational statistics, ETER HEIs have been divided between those that are specialized in a single field, focused (2-3 important fields) and generalist (covering most fields of education). As shown by Figure 15, HEIs can be divided almost equally between these three categories, but the generalist institutions account for more than 70% of enrolled students and for more than 80% of PhD students. Particularly concerning research, the core of European Higher Education is therefore constituted by generalist institutions, the few exceptions being represented by some large technical or medical schools.

*Figure 15. HEIs by level of subject specialization*
Source: ETER brief.
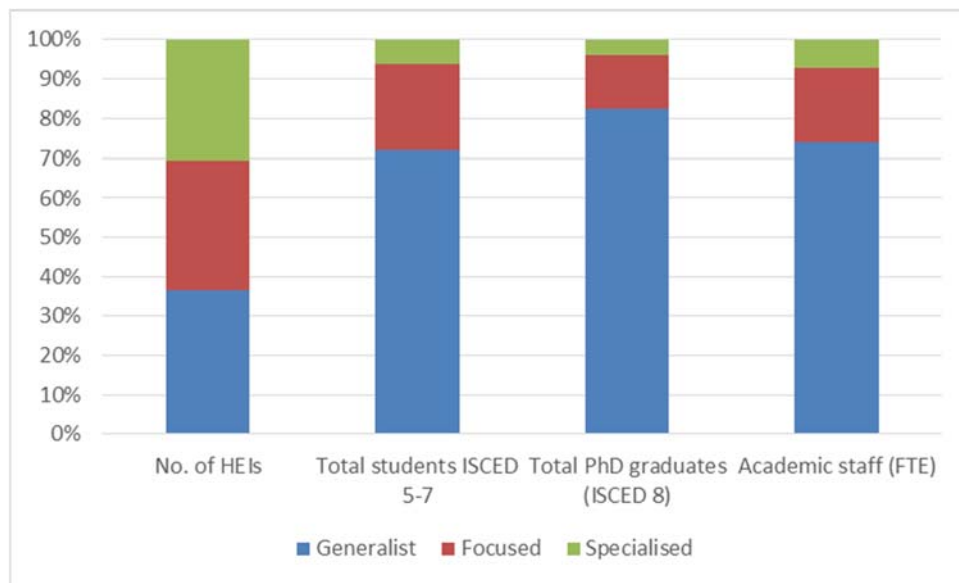
On the contrary, specialized institutions are found in distinctive domains, like arts and humanities, business and law and engineering. These findings therefore highlight the division of work and importance of diversity within higher education systems. Their distribution within European countries is also very heterogeneous, pointing to the importance of public policies for their emergence.

As shown by Figure 4, Germany, France and Italy have a large number of specialized institutions. The map also shows that in some regions, e.g. Warsaw, Paris and Prague, a high number of specialized HEIs are clustered, while in other regions a small number of HEIs (or even one single institution) enrol a large number of students (e.g. the Polytechnic Universities of Milan and Turin or the BI - Norwegian School of Management in Oslo).

Figure 16. Number of specialized institutions and their enrolled students per region
Source: ETER brief.
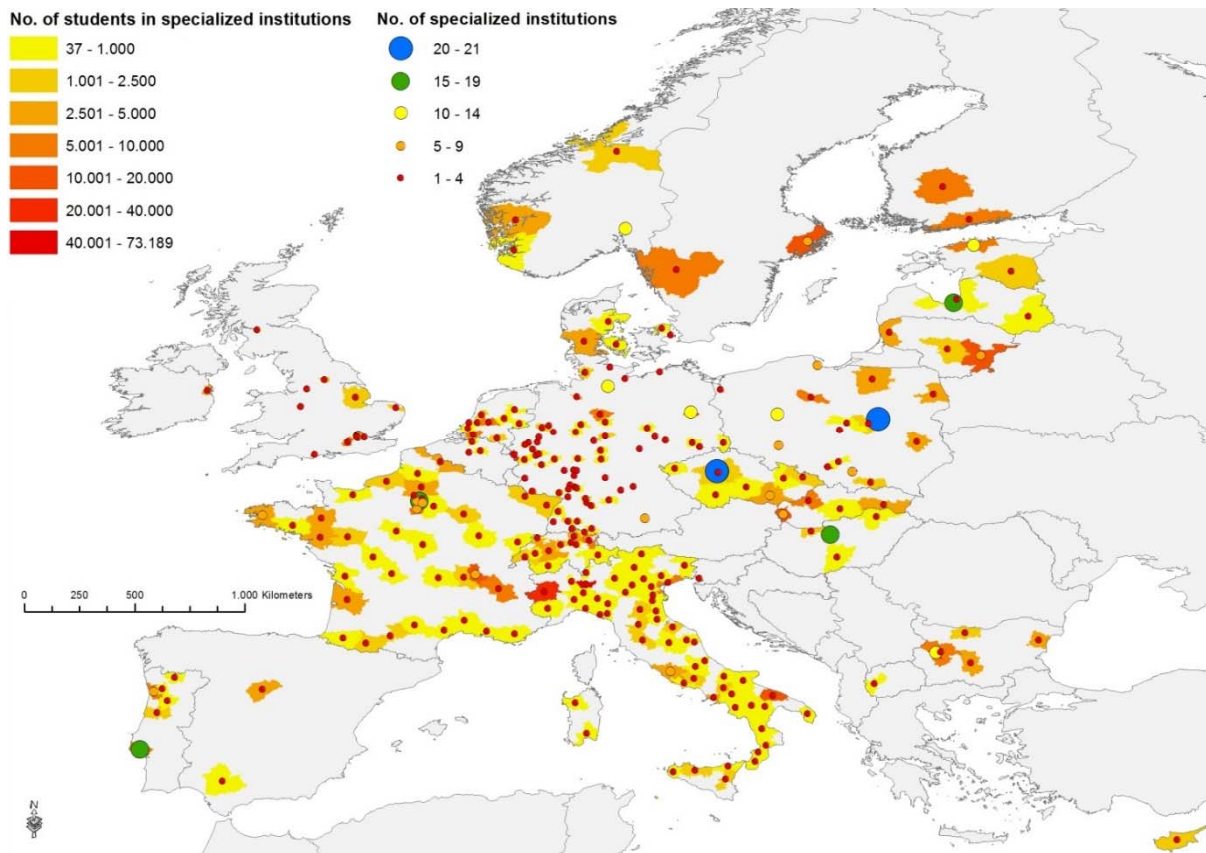
The distribution of HEIs across regional spaces

The regional spread of higher education institutions is an important dimension of the EU's agenda for the modernization of higher education and, more broadly, to the goals of the Europe 2020 strategy. On the one hand, higher education institutions (HEIs) within or in proximity to densely populated areas, allow students to access higher education without having to move to other regions and therefore, represent an important contribution towards achieving the target of 40% of young people completing higher education or equivalent studies by 2020. On the other hand, having universities geographically close is a major advantage for firms and social actors that make use of research inputs, as these partnerships provide direct knowledge flows through information exchanges and face-to-face collaboration. Therefore, the presence of one or more HEIs in a region is an asset that helps foster knowledge flows between education, research and business in order to promote the balanced development of European regions.

ETER represents a useful tool for policymakers to improve the alignment between regional development strategies and higher education policies. ETER allows for more fine-grained analyses of the localization and other associated factors (legal status, mission, subject focus, educational vs. research orientation) of individual higher education institutions in Europe when compared with EUROSTAT, which provides regional level data. For example, ETER may allow for the closer inspection of regional smart specialization strategies and their relationships with the geographical distribution of HEIs, particularly when combined with data on research output and technology transfer.

Figure 17 shows two features of the regional distribution of HEIs in Europe. On the one hand, policies for enhancing regional access have been moderately successful, since many European regions now include at least an HEI, but regional differences remain

important. Overall, it appears that some countries have a more geographically distributed higher education system: among the Nordic countries, Netherlands; Ireland and Italy[24].

On average, a student in Europe can find a HEI by traveling 50 km, but, 181 million people – one-third of the European population – live in regions where students may find a HEI only by traveling more than 50 km. These areas are found in large and dispersed countries (Norway, Sweden) but also in largely populated countries such as the UK, Germany, France and Spain, and in some parts of Eastern Europe. The number of regions without HEIs needs however to be relativized. Many of these regions are small and not very far from places where HEIs are located, while some of them might comprise campuses of HEIs located in other regions.

On the other hand, HEIs are concentrated in a limited number of metropolitan areas. The 36 regions with more than 10 HEIs include one-third of all HEIs in Europe. Most regions with more than 10 HEIs are capital cities, the list being championed by Paris, Warsaw, Berlin and Lisbon. This group also includes a few large cities, like Hamburg, Munich and Lyon. London is a special case since, for administrative reasons, it is divided between different NUTS 3 regions.

Concentration is even stronger when considering only PhD-awarding institutions (mostly universities), which account for the bulk of the research effort in higher education. Universities offering PhD degrees are only present in large numbers in capital cities and in a small number of university towns, or medium-sized cities with more than one university. This dynamic creates delicate policy issues, in particular for peripheral areas where there is a need for both the creation of human capital and knowledge spill over from research activities. The ability for ETER data to provide views of the spatial distribution of HEIs by their characteristics and activities—particularly education vs. research—represents an important asset for policy analysis.

---

[24] See the ETER brief 4. What ETER tells us about the regional dimension of European higher education, https://eter-project.com/assets/pdf/ETER_regional_dimension.pdf.

Figure 17. Number of HEIs by region (EUROSTAT, NUTS regions level 3)
Source: ETER brief.

N. of: HEI
- (0,m) (1217)
- (1) (334)
- (2) (169)
- (3) (124)
- (4, 10) (71)
- > 10 (36)
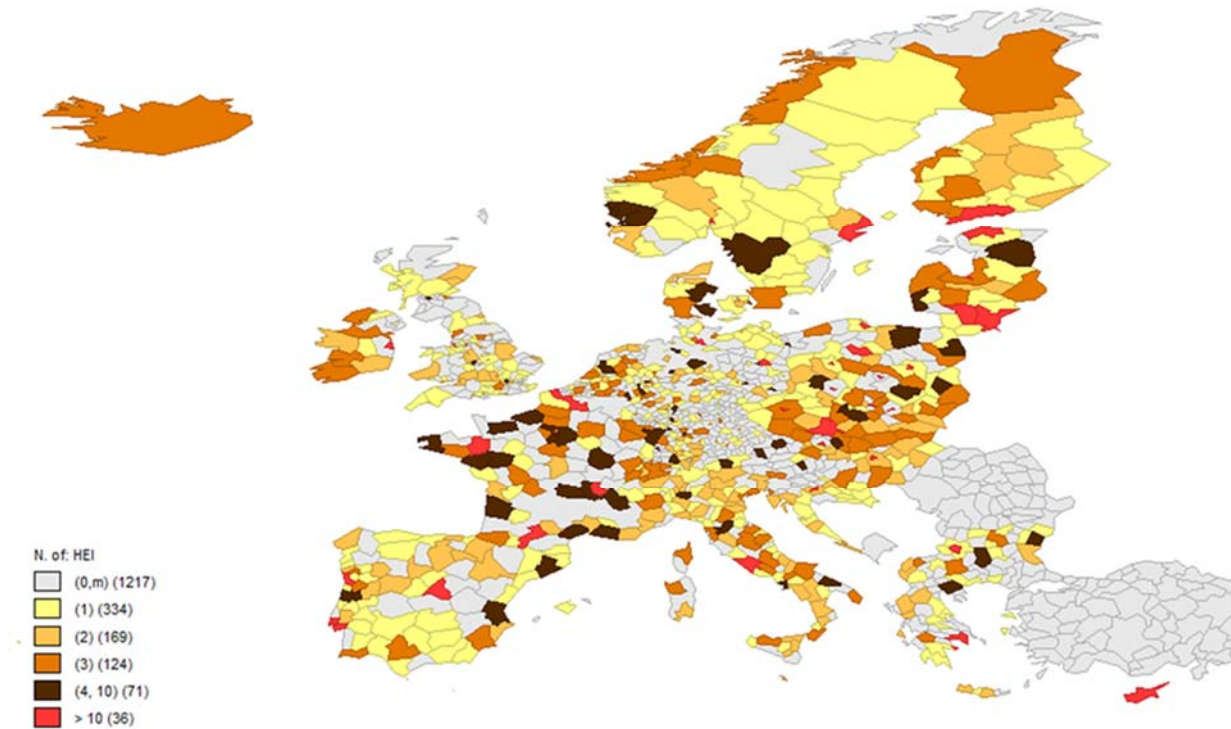
Gender balance and student mobility

Finally, the two remaining ETER briefs highlight two specific dimensions of diversity impacting relevant policy issues, i.e. achieving gender balance in European HEIs and promoting student mobility in the European Higher Education Area.

Gender equality and gender mainstreaming are a core objective of the European Research Area's strategy to avoid talent loss, to diversify the views and approaches in research, and to foster excellence. The *Conclusions on advancing gender equality in the European Research Area*, adopted by the European Council in 2015, reaffirm the commitment of the European Union to achieve gender balance. Yet, data at the national level, published by the She Figures (European Commission, 2015) confirm that, while gender balance – defined as reaching at least a 40% ratio of females – has been achieved for students, it remains an issue among academic staff, particularly for full professors.

Moving beyond national averages, ETER provides a disaggregated view of gender balance at the institutional level, displaying large differences between HEIs within the same country and identifying groups of HEIs, which made more progress towards this goal[25]. More than half of the HEIs included in the data set have achieved gender balance among academic staff, whereas only a minority (13%) of HEIs achieved gender balance among female full professors. As shown by Figure 18, the share of female full professors, defined as those individuals at the top of the academic hierarchy, displays wide variation both between and within countries.

---

[25] See ETER brief 6. What ETER tells us about gender balance in European Higher Education Institutions, available at *https://eter-project.com/assets/pdf/ETER_gender.pdf.*

In most countries, just a handful of HEIs achieved gender balance for female full professors. In eight countries (BG, CH, HR, LT LV, RO, RS and UK), more than one out of every five HEIs reached the 40% threshold.

A closer inspection shows that ETER includes 35 HEIs in which at least three out of four academic staff members are female. Only a few of them have more than 100 academic staff and most of them are in fields with a large presence of women, including health – particularly nursing and education – and both regular and pre-school teacher training. This is evident by looking at the institutions with the highest proportion of female academic staff. There are several institutions in Latvia on the list, which can likely be linked to a generally high proportion of female academic staff in Latvian HEIs.

Figure 18. HEIs by share of female full professors
Source: ETER brief.
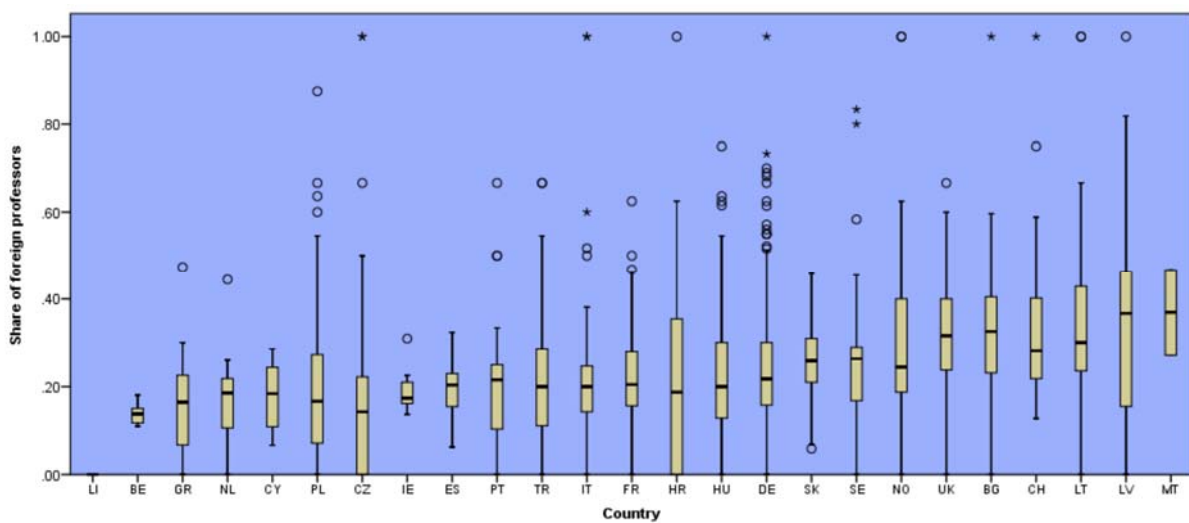
Improving student mobility is another core goal of the European Higher Education Area and a major policy priority in the EU's agenda for modernizing higher education. In the words of the European Council, it is "one of the fundamental ways in which young people can strengthen their future employability, as well as their intercultural awareness, personal development, creativity and active citizenship". At the institutional level, attracting students from other countries is expected to improve the quality of Higher Education Institutions and their ability to recruit talented researchers, as well as improve education and research quality. The European Union set a target of having 20% of graduates study abroad for a period by 2020. However, according to preliminary data, the amount of outward mobility in EU-28 countries was a mere 2.9% in 2013, with only Cyprus and Luxembourg exceeding the EU benchmark.

Despite some data limitations, ETER provides a disaggregated view at the HEI level for student mobility and, particularly, for *degree mobility*, i.e. students who moved for the purpose of studying abroad (Figure 19)[26]. At the bachelor level (ISCED 6), 77% of the HEIs had less than 10% of mobile students in 2013 and heightened levels of mobility are limited to a small number of cases, mostly branch or franchise campuses. A few large HEIs however enrol a substantial number of mobile students, the list being led by the University of Vienna with more than 10,000 mobile degree students at the bachelor level, followed by Manchester and Innsbruck.

---

[26] See *ETER brief 3. What ETER tells us about student mobility in European higher education, available at*
https://www.eter-project.com/assets/pdf/ETER_student_mobility.pdf.

At the master level, and even more so at the PhD level, differences between HEIs are larger. At the master level, the level of mobility is very high in the UK, fostered by linguistic and historical reasons, but also by the deregulation of tuition fees, which makes non-EU students an interesting market. Swiss HEIs have high shares of mobile students, also due to linguistic reasons, but overall one-quarter of all HEIs for which data is available in ETER enrol more than 20% of degree mobility students.

Finally, mobility at the PhD level is a widespread phenomenon – more than one PhD student out of three is mobile – which displays significant differences between countries and HEIs. In half of the 1200 HEIs in question, at least one out of every four PhD students will obtain their PhD in a country that is different from where they earned their secondary degree. The largest numbers of mobile PhD students in Europe are found in top-ranked research universities, with Cambridge, Oxford and ETH Zurich exceeding 2,000 mobile PhDs, followed by Vienna, UCL, Imperial College and KU Leuven. These figures emphasize the strong asymmetry of PhD student mobility towards top-ranked international universities, which function as European-wide training places for researchers

Figure 19. Degree mobility by educational level
Number of HEIs reaching a certain level of mobility for each degree level. Source: ETER brief.
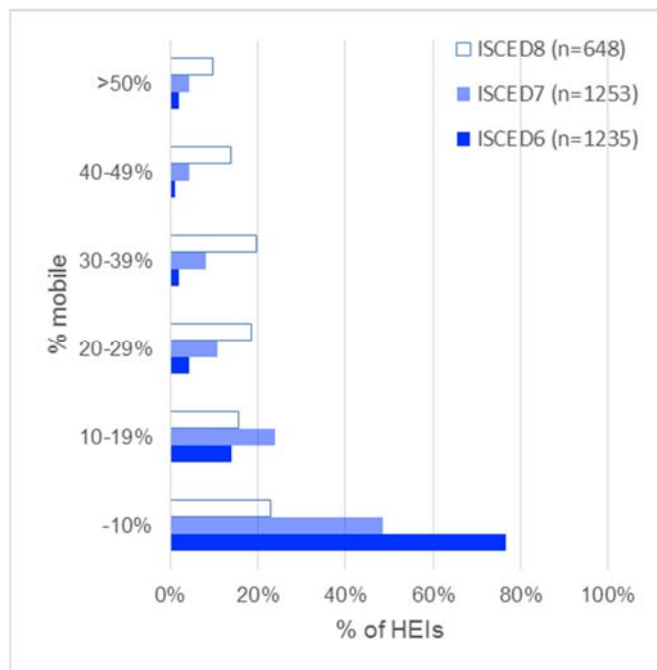
Conclusions and recommendations

As documented in this report, the main outcome of the ETER contract has been to move beyond a feasibility study to a fully operational data collection system, through which HEI data can be routinely collected on a yearly basis, validated and corrected, safely stored and made publicly available to a broader audience for further usage through a web interface. The ETER consortium also undertook a sizeable effort to promote the usage of ETER through dissemination events, a dedicated mailing list and examples of the ETER's value through the so-called ETER briefs.

ETER now represents by and large the core reference dataset on European higher education, providing for the first time a reference list of HEIs – those graduating at least at the ISCED 6 (bachelor level); it also provides a core set of data, which were previously available only at the level of individual countries, including geographical information, staff and finance data, as well as rich data on students and graduates. This reference list can then be easily matched with other data sources for a more in-depth analyses, for example concerning research activities. As documented in this report, the usage of ETER is also rapidly growing, both on the policy and on the scholarly side

The main recommendation of this report is therefore that the dataset should be maintained and updated through regular data collection. Updating will be important not only to provide more accurate data, but also since a longer time series will provide a more thorough understanding of the changes over time in European higher education. The latter is highly important also for policy purposes, since higher education systems have a lasting heritage from the past, but what matters is the direction and pace of change over time, in order to attain new policy goals.

Though, the ETER dataset is not perfect. Despite harmonization efforts, it still largely reflects particularities and idiosyncrasies of higher education data, linked to national specificities, but also to different levels of standardization, for example when comparing students with financial data. Since ETER does not manage its own data collection, this process has to go in parallel with efforts to improve the availability and quality of the primary data, particularly in those areas, which were not previously standardized by EUROSTAT like staff and financial data. Since the first pilot in the EUMIDA feasibility study, this process has come a long way: standards for additional variables have been defined, the NSAs are starting to collect more data given their policy relevance (data on the internationalization of academic staff being a case in point), and it has become widely accepted that HEI-level data should be publicly accessible.

Given this context, a pragmatic approach is needed for the future development of ETER. It is strongly advised to set clear and realistic priorities, which are also shared and supported by NSAs when adapting the data, which is required to extend the system.

More specifically, we provide recommendations on how to address four critical issues, which emerged in the current ETER contract, i.e. extending coverage of the HEI system, improving data completeness and data quality, introducing additional variables and making the usage of ETER easier.

Extending the coverage of the HEI system. As discussed in chapter 0, ETER coverage has long been subject of discussions between divergent perspectives: providing the largest possible coverage of tertiary education; avoiding a large number of observations; while taking into account data availability at the national level. In this respect, the current ETER perimeter represents a pragmatic compromise. While a few additional HEIs might be included in the future, the consortium considers that substantially extending the number of institutions within the current system will not be feasible: on the one hand, it would increase the burden for data collection beyond what the NSAs are willing to invest and, second, it would simply translate into a lower availability of data. An option would be to distinguish between a core set of HEIs, for which all ETER variables are collected (about 250 variables per HEI) are collected, and an extended set, for which only few data are collected – including descriptors, geographical information, and maybe, the number

of students. This option should be explored carefully in terms of the definition of an extended perimeter and of its feasibility.

Improving data completeness and data quality should be a major focus of the future development of ETER: current limitations at this level limit the scope ETER usage – for example an analyses on financial data can be performed on only about 20 countries – and the process requires a high level of proficiency and investment from users in order to get acquainted with the specificities of the data, therefore limiting usage. While the current ETER contract largely focuses on internal data quality (including annotating problems), the focus in the future should be to work together with the NSAs to find pragmatic solutions to map and possibly impute national data to achieve better comparability. This work would first focus on a number of large countries, which constitute the core of ETER, and focus on staff and financial variables, where most of the problems are currently encountered.

Introducing additional variables. The nature of ETER as a generic information system on higher education constantly generates requests for integrating additional variables in order to answer relevant policy and scholarly questions. The inclusion of additional variables raises feasibility problems in terms of the data collection burden, particularly for the NSAs. To address these issues, the consortium recommends a differentiated strategy based on three pillars:

- First, the selective introduction of additional variables to the ETER data collection. This should be limited to a few items, since it generates an additional data collection burden for the NSAs. Two possible examples of this extension are the localisation of HEI campuses located in cities other than the main seat and a classification of academic staff by level (provided a common classification can be agreed).
- Second, the use of ETER to publish HEI-level data available from other sources. This was already the case for data on Erasmus students, but important data could also be provided in the future by projects like U-MULTIRANK and EUROGRADUATES. While these projects provide much more detailed and disaggregated data (which sometimes cannot be published), ETER could offer a suitable platform to publish HEI-level aggregated data.
- Third, matching ETER with other existing datasets through the use of common identifiers, so that users can easily combine different sets of data for analytical purposes. This path has been pursued within the RISIS project (risis.eu), where ETER has been matched with databases on scientific publications and participations in EU-FP projects and patents, but could also be extended to a large set of databases in the future, thus strongly enhancing the interoperability of HEI datasets in Europe.

Making usage easier. While ETER usage is rapidly growing, it remains that ETER is a complex dataset, by its disaggregated nature, but also by the presence of missing data and of data specificities. The current approach has been to carefully document these particularities in the database, therefore displacing to users the burden of carefully analysing the data. While this may be a good approach for expert users, it limits the scope of ETER for less advanced users, which need to be able to produce some useful figures in a reasonable amount of time (for example for policy or information purposes). The consortium therefore recommends placing a particular emphasis in the future on 'intermediate products' for ETER users. Illustrative examples include the following: data explanations on the model of the 'statistics explained' webpages published by EUROSTAT; pre-prepared analysis and tables with extract of data (similar to the main indicators of EUROSTAT). An on-going approach is to plan for a core ETER dataset, which includes only those cases and variables with data that is sufficiently complete, possibly

by using statistical imputation techniques to address the known data problems. We emphasize that, since these products are intermediate between the ETER full dataset and users, they need to be designed together with a user group in order to carefully match their needs.

Increasing the visibility of ETER through targeted dissemination and communication activities. As documented in this report, ETER has become widely known and four main types of usage are emerging – as data source in policy documents and for scholarly reports, as a reference perimeter for other project, finally as primary data source for projects and websites on higher education. We suggest that, besides general dissemination activities, efforts to increase visibility should be now targeted to lead users, including international organizations and national ministries that could multiply the visibility of ETER by including it in their data offers and in their website. An important goal would therefore to define with these key stakeholders institutional agreement on the usage of ETER data and to provide them more direct support in the exploitation and analysis of the data.

Annexes

These annexes can be downloaded from the ETER website (https://www.eter-project.com/about/documents).

ETER handbook for data collection

The ETER handbook for data collection is the main methodological document of ETER. It provides definitions and detailed guidelines on variables and data collection and, therefore, is the basis for the whole ETER data collection.

Data quality report

The data quality report provides a detailed presentation of the ETER quality process and the results of find round of data quality checks performed in June 2017 on the whole dataset.

Full dataset

The full ETER dataset can be downloaded from the ETER website (https://www.eter-project.com/search) by first searching HEIs without inserting country and year and then selecting the option 'download all data' in the export table.

The display and export tab allows selecting different formats, including excel, csv and machine-ready formats for Stata and SPSS.

References

Barré, Rémi 2001. Sense and nonsense of S&T productivity indicators. The contribution of European Socio-Economic Research to the Benchmarking of RTD policies in Europe. Proceedings of a conference organised by the European Commission Directorate General for Research, Brussels, March 15-16, 2001.

Becher, Tony and Paul R. Trowler. 2001. Academic Tribes and Territories. Intellectual enquiry and the culture of disciplines. Ballmoor, Buckingham / Philadelphia, PA: The Society for Research into Higher Education and Open University Press.

Bonaccorsi, A. 2014. Knoweldge, Diversity and Performance in European Higher Education. A Changing Landscape. Cheltenam: Edward Elgar.

Bonaccorsi, Andrea and Cinzia Daraio. 2007. Universities and Strategic Knowledge Creation. Specialization and Performance in Europe. Cheltenham: Edwar Elgar.

Bonaccorsi, Andrea, Cinzia Daraio, Benedetto Lepori and Stig Slipersaeter 2007. Indicators on individual higher education institutions: addressing data problems and comparability issues. Research Evaluation, 16, 66-78.

Brunsson, Nils, Andreas Rasche and David Seidl 2012. The dynamics of standardization: Three perspectives on standards in organization studies. Organ.Stud., 33, 613-632.

Clark, Burton R. 1983. The higher education system. Academic organization in cross-national perspective. Berkeley: University of California Press.

Cohen, Arthur M. 2007. The shaping of American higher education: Emergence and growth of the contemporary system. John Wiley & Sons.

Daraio, Cinzia, Andrea Bonaccorsi and Léopold Simar 2015. Efficiency and economies of scale and specialization in European universities: A directional distance approach. Journal of Informetrics, 9, 430-448.

Eumida. 2009. EUMIDA Handbook - Collection of institutional-level data on tertiary educational institutions at the European level.

European Commission. 2017. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on a renewed EU agenda for higher education. Brussels: European Commission, COM(2017) 247 final.

European Commission. 2011. Supporting growth and jobs – an agenda for the modernisation of Europe's higher education systems. Brussels: European Commission, SEC(2011) 1063 final.

European Commission/EACEA/Eurydice. 2017. Modernisation of Higher Education in Europe: Academic Staff – 2017. Eurydice Report. Luxembourg: Publications Office of the European Union.

EUROSTAT. 2010. Business Registers. Recommendation manual. Luxembourg: EUROSTAT.

Ferlie, Ewan, Christine Musselin and Gianluca Andresani 2008. The steering of higher education systems: A public management perspective. Higher Education, 56, 325-348. Godin, Benoit. 2005. Measurement and Statistics on Science and Technology. London: Routledge.

Huisman, Jeroen, Benedetto Lepori, Marco Seeber, Nicoline Frølich and Lisa Scordato 2015. Measuring institutional diversity across higher education systems. Research Evaluation, 24 (4), 369-379.

Jongbloed, Ben and Harry de Boer 2012. Higher Education Funding Reforms in Europe and the 2006 Modernisation Agenda. The Modernisation of European Universities: Cross-National Academic Perspectives, 1, 127.

Kyvik, Svein and B. Lepori. 2010. Research in the non-university higher education sector in Europe. Dordrecht: Springer.

Lepori, B., M. Seeber and A. Bonaccorsi 2014. Competition for talent. Country and organizational-level effects in the internationalization of European higher education institutions. Research Policy.

Lepori, B. and A. Bonaccorsi 2013. The Socio-Political Construction of a European Census of Higher Education Institutions: Design, Methodological and Comparability Issues Minerva, 51(3), 271-293.

Lepori, B. and Svein Kyvik 2010. The research mission of Universities of Applied Science and the future configuration of Higher Education systems in Europe. Higher Education Policy, 23, 295-316.

Lepori, B., Carole Probst and Lukas Baschung 2010. Patterns of subject mix of higher education institutions: a first empirical analysis from the AQUAMETH database. Minerva, 48(1), 73-99.

Lepori, Benedetto, Aldo Geuna and Valerio Veglio. 2017. A typology of European Universities. Differentiation and Resource Distribution. Brighton: University of Sussex, SPRU Working Paper Series.

Lepori, Benedetto, Andrea Bonaccorsi, A. Daraio, Cinzia Daraio, H. Gunnes, E. Hovdhaugen, M. Ploder, Monica Scannapieco and D. Wagner-Schuster. 2015. Establishing a European Tertiary Education Register. Final Report. Brussels: European Commission.

Lepori, Benedetto 2007. Patterns of diversity in the Swiss higher education system. In Universities and Strategic Knowledge Creation. Specialization and Performance in Europe, ed. A. Bonaccorsi and C Daraio, pp. 209-240. Cheltenham: Edwar Elgar.

Meek, V. Lynn, Leo Goedegebuure, Osmo Kivinen and Risto Rinne. 1996. The Mockers and Mocked: Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education. Oxford: Pergamon.

Musselin, Christine 2007. Are Universities Specific Organisations? In Towards a Multiversity? Universities between Global Trends and National Traditions, ed. G. Krücken, A Kosmützky and M Torka, pp. 63-84. Bielefeld: transcript.

OECD. 2015. Frascati Manual 2015. Guidelines for Collecting and Reporting Data on Research and Experimental Development. Paris: OECD.

Pinheiro, Rómulo, Lars Geschwind and Timo Aarrevaara 2016. Mergers in higher education. European Journal of Higher Education, 6, 2-6.

UOE. 2013. UOE data collection on education systems. Volume 1. Manual. Concepts, definitions, classifications. Montreal, Paris, Luxembourg: UNESCO, OECD, Eurostat.

van Vught, F. A. and F. Ziegele. 2012. Multidimensional Ranking. The Design and Development of U-Multirank. . Dordrecht: Springer.

Van Vught, Frans. 2009. Mapping the higher education landscape. Towards a european classification of higher education. Milton Keynes, UK: Springer.

von Tunzelmann, Nick, Marina Ranga, Ben Martin and Aldo Geuna. 2003. The effects of size on research performance: a SPRU review.

Waltman, Ludo, Clara Calero-Medina, Joost Kosten, et al 2012. The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. J.Am.Soc.Inf.Sci.Technol., 63, 2419-2432.

Publications Office