Establishing a European Tertiary Education Register

Annex I. Technical report on the data collection

Contract EAC 2013-038

Brussels, 2016

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

Education and Culture

Establishing a European Tertiary Education Register

Annex I. Technical report on the data collection

Table of Contents

Table of figures

List of tables

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
OECD	Organisation for Economic Cooperation and Development
UAS	Universities of applied sciences
UOE	UNESCO OECD EUROSTAT handbook on education statistics
USI	Università della Svizzera italiana

1 Introduction

The European Tertiary Education Register (ETER) is a study contracted by the European Commission's Directorate General Education and Culture (contract EAC-2013-0308), which aims to compile a register of European Higher Education Institutions (HEI) and collect a comparable set of data for the HEIs in the perimeter.

The study is a joint undertaking of four partners, Università della Svizzera italiana (USI), Lugano, Centre for Organizational Research, Joanneum Research, Graz, Nordic Institute for Studies in Innovation (NIFU), Research and Education, Oslo, University of Rome, La Sapienza, Department of Computer, Control and Management Engineering, Antonio Ruberti, Rome – together with a network of experts in the concerned countries. It is supervised by the Directorate General of Education and Culture from the European Commission, in cooperation with DG Research and Innovation, and with EUROSTAT.

It follows the European MIcroData project (EUMIDA), a large-scale study supported by the European Commission from 2009 to 2011, which demonstrated the feasibility of collecting European-level data on individual HEIs.

The study began on August 1st, 2013 and lasted for 24 months until July 31, 2015.

This report provides technical information on the ETER database as of June 2015. It complements the ETER final report, which is more oriented to the general policy-relevant public. The report covers the following items[1]:

- In chapter 2 we provide an overview of the ETER database concerning its geographical and temporal coverage, and detail the variables included; we also provide a synthetic comparison between ETER and EUROSTAT data and information on data publication and confidentiality.
- In chapter 3 the coverage of the ETER dataset is analysed, including definitions of the perimeter, exclusion and coverage in respect to the entire field of tertiary education.
- In chapter 4 we present the general status and analysis of the data collection process, while also highlighting the problems met, with specific reference to countries that did not deliver data.
- Chapter 5 presents the outcome of the data validation and quality control processes on the ETER dataset and examines the accuracy, completeness, consistency, and timeliness of the data, as well as emerging comparability problems.
- Chapter 6 presents the indicators selected by ETER in order to characterize HEIs and their activities, which have been included in the dataset.

[1] While ETER represents a truly collective effort by all core team partners and national experts, we acknowledge the specific contributions to individual chapters: chapters 2, 3 and 4 have been mostly written by USI with contributions by Joanneum concerning the data collection process and the database and by Sapienza on data source analysis. Chapter 5 has been written by Sapienza, while chapter 6 has been prepared by NIFU.

2 An overview of the ETER database

2.1 General information

The European Tertiary Education Register is a database that collects information on Higher Education Institutions (HEIs) in Europe, concerning their basic characteristics and geographical position, educational activities, staff, finances and research activities.

The current ETER coverage includes 31 countries: EU-28 countries – excluding the French part of Belgium, Romania and Slovenia - as well as Iceland, Liechtenstein, Norway, Switzerland and the Former Yugoslav Republic of Macedonia. Data are provided for the year 2011 (academic year 2011/2012) and the year 2012 (academic year 2012/2013). For a few countries, only data for 2011 are currently available (see Table 3 for a general overview of the country status).

The combined sample between the 2011 and 2012 data collections includes 2293 HEIs with about 16.6 million undergraduate students and about one-half million PhD students. For some countries (the French part of Belgium, Romania, Slovenia, Montenegro, Serbia and Turkey) only a definition of the HEI perimeter and a list of HEIs for the year 2011 are available.

ETER data has mostly been provided by the National Statistical Authorities (NSAs), based on national statistical databases or higher education information systems. They have been complemented by descriptors and geographical information mostly collected by the ETER consortium and validated by NSAs.

ETER should be considered as complementary to education and R&D statistics collected by EUROSTAT, as well as with the MULTIRANK project. EUROSTAT has provided national aggregates on educational activities at the tertiary level since 1998 – e.g. the number of students and of degrees – for both EU and some non-European countries. ETER covers a subset of tertiary education activities, mostly those institutions who at minimum, graduate students at the bachelor level (ISCED-2011 level 6). By providing institutional-level data, ETER allows for the investigation of the diversity of HEI characteristics and the distribution of functions between HEIs (for example different orientations towards education or research). In terms of the number of students, ETER includes in the considered countries, 85% of all students at the tertiary level, with reference to EUROSTAT national-level data.

Table 2. General characteristics of ETER as compared with EUROSTAT educational and R&D statistics

	ETER	EUROSTAT	MULTIRANK
Analytical level	Individual institutions	Countries	Individual institutions
Number of countries covered	31	37	74
HEIs included in the dataset	2293**	not applicable	850
Years covered	2011 and 2012	1998-2012	2012-2013
Organizational descriptors	yes	no	yes*
Geographical information	yes	no	yes*
Student's and graduates	yes	yes	yes*
Staff data	yes	limited	yes*
Revenues and expenditures	yes	yes	yes*
Research information	yes	yes	yes*
Institutional indicators	yes	no	yes
Metadata	yes	yes	yes*
Data flags	yes	yes	no
Data source	NSA	NSA	HEIs
* data are collected, but not published			
**combined sample 2011 and 2012			

ETER is similar to the MULTIRANK project (http://www.u-multirank.eu) in its focus on individual HEIs as a unit of analysis. As far as their goal to compare HEIs worldwide, MULTIRANK has a much broader geographical coverage, but mostly focuses on the largest and most reputed HEIs, while ETER aims to provide a reasonably compete census of HEIs graduating at least at ISCED level 6 in ERA countries.
There is also a large overlap in the data collected, even if MULTIRANK provides a broader set of data concerning research and technology transfer, as well as regional engagement. The two projects substantially differ however in their specific goals: ETER is a basic statistical facility, with the aim of publishing micro-data on individual HEIs, which are then made available to a broader range of users. ETER data can therefore be downloaded freely from the project website. On the contrary, MULTIRANK aims to construct a set of indicators to allow customers of higher education (for example students) to compare HEIs. Accordingly, individual data are not published, but HEIs can be compared on the project website using different indicators. The source of data is also different: ETER includes mostly official data provided and validated by the National Statistical Authorities, while MULTIRANK largely relies on data self-declared by the institutions themselves.

Table 3. Overview of ETER data by country

Country	Data 2011	Data 2012	Remarks
Austria	1	1	No financial and staff data available; no breakdown of students and graduates by FOE and nationality.
Belgium (French)	0	0	Delays due to administrative issues, procedure for participation is ongoing.
Belgium (Flemish)	1	1	Data almost complete (except a few breakdowns).
Bulgaria	1	1	All financial data, including R&D expenditures are missing.
Croatia	1	1	Very limited data available. All financial data missing. Students only aggregated for ISCED 5-7 and no breakdowns.
Cyprus	1	1	Data are fairly complete, except for revenues, for which only the aggregated figures are available.
Czech Republic	1	1	Very little financial data available. For students and graduates breakdown by FOE was not delivered.
Denmark	1	1	Data quite complete. Data for the analysis refer to 2011.
Estonia	1	1	Only descriptors and students and graduates data provided, no financial or staff data.
Finland	1	1	Data delivered.
Former Yugoslav Republic of Macedonia	1	0	NSA is not be able to deliver. A core of data (excluding all financial information) was been retrived by consortium using information available on the Internet (2011 only).
France	1	1	Data on students and graduates, as well as most of financial data delivered. Additional data expected by end of March 2015
Germany	1	1	Data delivered and fairly complete.
Greece	1	1	All financial data, including R&D expenditures are missing.
Hungary	1	0	
Iceland	1	0	Data on staff have been delivered, students and descriptors provided by the consortium; only 2011 data.
Ireland	1	1	Data nearly complete, breakdown of students available only on FOE-1997, private funding missing, no postocode available since the country has no postcode system.
Italy	1	1	Data are quite complete.
Latvia	1	1	Missing all financial data.
Liechtenstein	1	1	Missing all financial data.
Lithuania	1	1	All breakdowns by fields of education of students and graduates are missing.
Luxembourg	1	1	Data fairly complete.
Malta	1	1	Data almost complete.
Montenegro	0	0	Data not delivered. The NSA does not hold HEI-level data.
Netherlands	1	1	Data on PhD students and degrees largely missing, as well as all financial data.
Norway	1	1	Data delivered. A few breakdowns missing.
Poland	1	1	No financial data available, no breakdown of staff data, breakdown of students and graduates by FOE not available.
Portugal	1	1	No financial data available, breakdown of students and graduates by fields of education not available.
Romania	0	0	Data cannot be delivered by NSA, consortium is looking for a solution through a national expert, but this can be done only at a later stage.
Serbia	0	0	Did not participate in the data collection. National Statistical Authority does not hold data.
Slovakia	1	1	
Slovenia	0	0	Data collection on-going.
Spain	1	1	All financial data missing.
Sweden	1	1	Data very complete, FOE-1997 used for students and graduates.
Switzerland	1	1	No specific issues, data fairly complete.
Turkey	0	0	Data not delivered. Council for Higher Education did not reply.
United Kingdom	1	1	Data available with limitations, since most recent data are paying in the UK.
Total	31	28	

2.2 Variables and indicators

ETER includes the following main groups of variables:

- *Institutional descriptors and geographical information on the included HEIs.*
- *Data on students and graduates, including breakdowns by ISCED-2011 level, gender, nationality, mobility and fields of education.*
- *Data on research, including students and graduates at the PhD level, as well as R&D expenditures.*
- *Financial data: expenditures and revenues of the HEI.*
- *Staff data, including some breakdowns by gender, nationality and field.*

The ETER database also includes a small number of indicators that have been selected to characterize individual HEIs and their profile of activity. These indicators are further presented and analysed in chapter 5 of this report.

When compared with the data provided by educational and R&D statistics at EUROSTAT, ETER includes very similar variables and breakdowns for students and graduates, since ETER readily adopted the definitions from the UOE manual on educational statistics. There are however two main differences: first, HEI-level data are provided rather than

national aggregates and, second, the ISCED-2011 classification (and FOE classification of educational fields) was already introduced (whereas it will be used in EUROSTAT data only from next year).
ETER provides substantial additional information concerning the other dimensions: descriptors are of the utmost importance in order to characterize types of HEIs and their history, while geographical information allows for an analysis of the distribution of HEI activities across the European space. ETER also provides more detailed information on expenditures and revenues, including an important breakdown of revenues by core budget and third-party funds, which is not foreseen in educational statistics. Additional data have also been collected concerning staff, including the number of professors and breakdowns by gender and nationality.

Table 4. ETER list of variables

Dimension	Variables	Included in EUROSTAT	Availability
Identifiers	ETER ID National identifier (optional) Institution name (in own language) English institution name (if available) Year	No	Complete (100%)
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	No	Almost complete (100%)
Geographic information	Region of establishment, NUTS2 code Region of establishment, NUTS3 code Name of the city Postcode Multi-site institution Geographical coordinates	No	Almost complete (100%)
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	National aggregates for students and graduates (not yet published with the ISCED-2011 classification). Variables for degree level and distance education.	80-90% (lower for mobility)

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 R&D expenditures	National aggregates for ISCED 8 students and graduates. R&D expenditures. Research-active variable not provided.	80-90% for ISCED 8 students and graduates, 40% for R&D expenditures.
Expenditures	Personnel expenditure Non-personnel expenditure Capital expenditure Accounting of capital expenditures	Yes, except the categorical variable for the accounting system.	Around 50%
Revenues	Core budget Third party funding Private funding Tuition fees Student fees funding	Total revenues available, the breakdown by core and third-party not available.	Around 50%
Staff	Number of academic staff in FTEs and headcounts Number of academic staff by fields of education, gender and citizenship in headcounts Number of administrative staff in FTEs and headcounts Number of professors by gender Inclusion of PhD students Number of total staff in FTE and HC	Only total teaching staff available as national aggregate; other variables not available.	Around 60% for the total, much lower for citizenship and staff by field.

Table 3 provides a disaggregated view of the availability of data by country and by group of variable. Overall, we can summarize the situation as follows (see chapter 5.2.2 for more detailed information on completeness):

- Descriptors are generally available for all countries, with the exception of a few cases where information on foundation years was not available.
- Financial data (revenues, expenditures, R&D expenditures) are available for about half- of the countries. These include some of the large European countries like Germany, France and Italy. Therefore, the completeness in terms of the number of HEIs is larger.
- Staff data are generally available for most countries; major exceptions are currently France and the UK. However, the breakdown of academic staff between nationals and foreigners is available for a much smaller number of countries and data on staff by field of education is available only for very few countries.
- Student and graduate data are available for most countries, including the breakdown by gender, nationality, and fields of education. The breakdown by mobile students is

less widely available. The situation is similar for PhD students, but some countries are still missing.

During the second round of data collection, a number of gaps could be filled, including financial data for France, PhD students for Spain and breakdowns of students by fields for a few countries. It is reasonable to expect that more gaps can be filled in the future, especially in the domains where new classifications have being introduced. The consortium considers that there is scope for improvement concerning financial variables, at least for HEIs in the public sector.

The consortium also considers staff data as a major area for future improvement. There are currently no established definitions and classifications of academic staff and data are lacking for some critical variables, particularly for the breakdown of staff by nationality and scientific domains.

Table 5. Availability of data by country

The information combines 2011 and 2012 data depending on availability.

	Descriptors			Expenditure data			Revenues				Staff						Students					Graduates					Research: phd students					0
	Institutional descriptors	Foundation year	Geographic information	Personnel	Non-personnel	Capital	Core	Third-party	Fees	Private	HC	FTE	By FOE	By gender	Nationality	Professors	Total	By gender	By FOE	Mobile	Nationality	Total	By gender	By FOE	Mobile	Nationality	Total	By gender	By FOE	Mobile	Nationality	R&D expendi
Austria	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	0
Belgium (Flemish)	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Bulgaria	1	1	1	0	0	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0
Croatia	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Cyprus	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Czech Republic	1	1	1	1	1	0	1	0	0	0	1	1	0	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0
Denmark	1	1	1	1	1	1	1	0.5	1	0.5	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Estonia	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Finland	1	1	1	1	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0.5
Former Yugoslav Re	1	1	1	0	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	0	0	0
France	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Germany	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Greece	1	1	1	0	0	0	0	0	0	0	1	0	0	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1
Hungary	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Iceland	1	1	1	0	0	0	0	0	0	0	0.5	0.5	0	0.5	0.5	0	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0
Ireland	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0
Latvia	1	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1
Liechtenstein	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Lithuania	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	0	1	1	0	1	0	1	1	0	1	0	1
Luxembourg	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	0	0	0	1	0	1	1	1	1
Malta	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
Montenegro																																
Netherlands	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0
Norway	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	1
Poland	1	1	1	1	0	0	1	0	0	0	0	1	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	0.5
Portugal	1	1	1	1	0	0	0.5	0.5	0.5	0.5	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	0
Romania																																
Serbia																																
Slovakia	1	1	1	1	1	1	1	0	1	0	1	1	0	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0
Slovenia																																
Spain	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
Sweden	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Switzerland	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Turkey																																
United Kingdom	1	1	1	0	0	0	0	0	0	0	0.5	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
	31	31	31	21	19	16	18.5	14	16.5	14	24	22.5	7	25.5	12.5	24	31	28	26	17	26	29	27	25	15	24	30	29	23	16	23	13

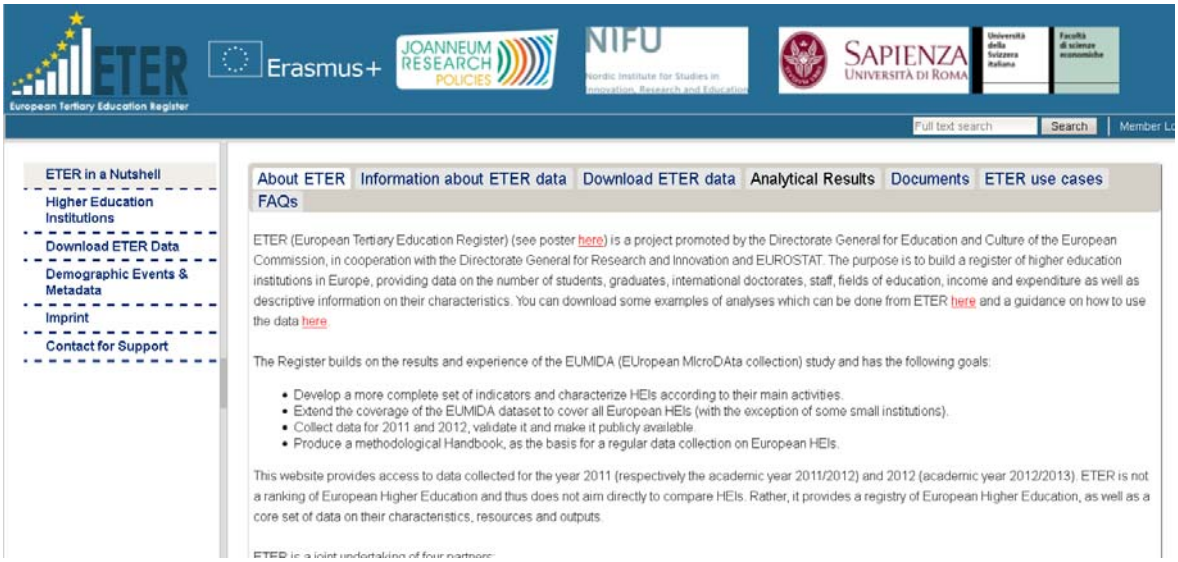
2.3 Data publication

In order to increase transparency in European higher education, the ETER project aims to make data publicly available to interested users. This requires addressing a number of issues:

- Providing an interface where data can be searched and downloaded in a suitable format.
- Providing users with sufficient information on methodology in order to enable a consistent use of data.
- Addressing questions of data confidentiality.

ETER data can be accessed and searched on-line from the project website http://eter.joanneum.at/imdas-eter/ (see Figure 1).

Figure 1. Entry page to ETER



The website has been designed as a unique entry point for all information in ETER. It provides users with the following information:

- Basic information on the project and methodological information in order to use the data.
- Downloading the list of HEIs in the dataset and retrieving demographic information on time changes.
- Searching and downloading subsets of the data or downloading the whole dataset in. xlsx or .csv format.
- Downloading metadata information.

The website provides open access for all users, as well as a restricted access (password-protected) through which users can also download data that are restricted to scholarly usage.

Data for both the year 2011 and 2012 are available on the website.

The website also includes general information on the project, full methodological documentation, instructions and examples of usage. More specifically, the following information is available:

- General information on the project and a set of Frequently Asked Questions.
- The methodological handbook, the full list of ETER variables and the variable codes (including special codes and data flags).
- Examples of usage of the dataset and an explanation of how to download and analyse the data (in PowerPoint format).

A new release of the website will be prepared for the publication of the second wave of data collection, which will also include additional information and the presentation of ETER analyses.

2.4 Data disclosure

In order to make the data public, ETER needed to clarify the status of the data and possible restrictions with the original data providers, i.e. National Statistical Authorities, who are the owners of most of the primary data.

It is generally intended that most data on public higher education institutions should be publicly available, but from past experiences, it is known that some restrictions may apply, especially concerning financial data and private HEIs. A related concern is to avoid data that are not yet available at the national level, but are made accessible at the European level. Some coordination between the national and European publication of data then becomes important.

To address these issues, NSAs were asked to explicitly agree to the publication of the data they delivered. They were given the option of signing a data disclosure agreement, which allows restricted access for some data to only be used for research purposes, on the condition that the direct identification of statistical units in the final publication is not possible.

As shown by Table 6, 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, NSAs asked for consent to publish data directly to the concerned HEIs.

For the countries that signed an agreement, two types of restrictions emerged:

- On one hand, financial data (expenditures, revenues, R&D expenditures) is restricted in a few countries. This might apply to other countries that deliver data in the future.
- On the other hand, restrictions for a few individual HEIs apply in two cases: private HEIs in some countries (due to their legal status) and a few HEIs which did not give consent to deliver data in countries where NSAs had to ask permission directly, as there was no legal basis (Bulgaria, Lithuania).

In the public version of the dataset, data that was delivered, but restricted by a data disclosure agreement, were replaced by "c."

Moreover, ETER adopted usual good statistical practices concerning data that are below some threshold, which would 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 restricted purposes.

Table 6. Status of data disclosure

	Data disclosure agreement signed	Specific issues by variable	Specific issue by HEI
Austria	no	All data are already public, no data disclosure agreement required.	Confirmed by NSA.
Belgium (Flemish)	yes	Staff and all financial data are confidential.	
Bulgaria	yes	None.	Data restricted for 3 HEIs which did not give consent.
Croatia	yes	All financial data (revenues, expenditures, R&D expenditures) are restricted.	
Cyprus	no	All data are public	
Czech Republic	yes	None.	
Denmark	no	All data are public	Confirmation by national expert.
Estonia	no	All data are public.	Confirmed by NSA.
Finland	yes	All data are public	
Former Yugoslav Republic of Macedonia	no	All data are public.	All data were retrieved from public sources by the consortium.
France	no	All data are public	
Germany	yes	All data are public	
Greece	yes	R&D expenditures are restricted only for research purposes.	
Hungary	no	All data are public	
Iceland	no	Staff data are restricted access.	Agreement to be signed.
Ireland	yes	No restrictions, all data are public.	
Italy	yes	Financial data are confidential.	
Latvia	yes	Data on expenditures and revenues are restricted access only for scientific purposes.	One HEI did not provide consent to publication of data.
Liechtenstein	yes	None. All data are public.	
Lithuania	yes	None.	All data are public except for 4 HEIs which did not provide consent to trasmission of data.
Luxembourg	no	All data are public.	Confirmed by NSA.
Malta	no	All data are public.	Confirmed by NSA.
Netherlands	yes	All data are public.	
Norway	no	All data are already public, no data disclosure agreement required.	
Poland	yes	Financial data (including R&D) are public only for the public HEIs.	
Portugal	no	All financial data (revenues, expenditures, R&D expenditures) are restricted. Some restrictions for staff data	Agreement to be signed.
Romania	no		Data not yet delivered.
Slovakia	no	All data are public.	Confirmed by NSA.
Slovenia	no		Data not yet delivered.
Spain	no	All data are public.	
Sweden	yes	No restrictions, all data are public.	
Switzerland	yes	None. All data are public.	
United Kingdom	no	All data are public.	Confirmation by NSA.

For final users, the conditions of access to ETER data are stipulated on the website (see Figure 2). Public data can be freely downloaded and reused for any type of analysis. Due reference has to be given to the data source and data providers assume no liability for the data and their usage. As usual for statistical data, it must be mentioned when users modify data points for any reason, and also, users are not permitted to reproduce or make publicly available the entire dataset.

Figure 2. Conditions for usage of ETER data (source. ETER website)

- *Under which conditions can I use the ETER data?*
  This website has been developed by the ETER consortium, under a contract with the European Commission, to enhance public access to information about higher education in Europe. The website and the information are the property of the European Union. The European Commission's goal is to keep this information accurate and up to date. If errors are brought to its attention, the Commission will endeavour to correct them. However, the Commission accepts no responsibility or liability whatsoever with regard to the information on this site.

  The data available on this platform are publicly available and, accordingly, can be used to make analyses of European higher education and for descriptive and policy purposes. This includes both the data and accompanying information like flags and metadata. The ETER consortium, on behalf of the European Commission, has acquired permission from National Statistical Authorities to disseminate these data.

  Use is permitted on the following conditions:

  - Reference to the ETER project has to be made as follows: "Data source: ETER project. Download date XXX".
  - In scientific publications and reports, the following acknowledgment should be included: "Data have been provided by the European Tertiary Education Register (ETER), funded by the European Commission under the contract EAC-2013-0308".
  - The user has to provide to the ETER consortium a copy of reports and scientific publications issues from these data.
  - When re-use involves modifications to the data, this must be stated clearly to the end user.
  - While users can download the entire dataset or parts of it for analytical purposes, they are not allowed to make available the dataset (or parts of it) on an on-line support.
  - The European Commission, the ETER consortium or the original data providers assume no liability for any interpretation of the data provided, nor for uses which do not take into account any or all of the underlying methodological issues.

- *Should I inform the ETER consortium of the use of the data?*
  Yes. Users should provide to the ETER consortium a copy of reports and scientific publications issues from these data. Additionally users may provide feedback on the value of ETER and advice on how the value of the database could be enhanced. Comments and suggestions could be sent to eter@joanneum.at.

- *How can I get access to restricted data?*
  Due to national confidentiality requirements, some ETER data are available only for research purposes on the condition that individual data points are not disclosed publicly. This concerns mostly financial data for a small number of countries, as well as data for some private HEIs. These data are coded with "c" on the database available on this platform.
  Access to the restricted data will be provided upon signature of a non-disclosure agreement. To get access to restricted data, please contact the ETER project team.

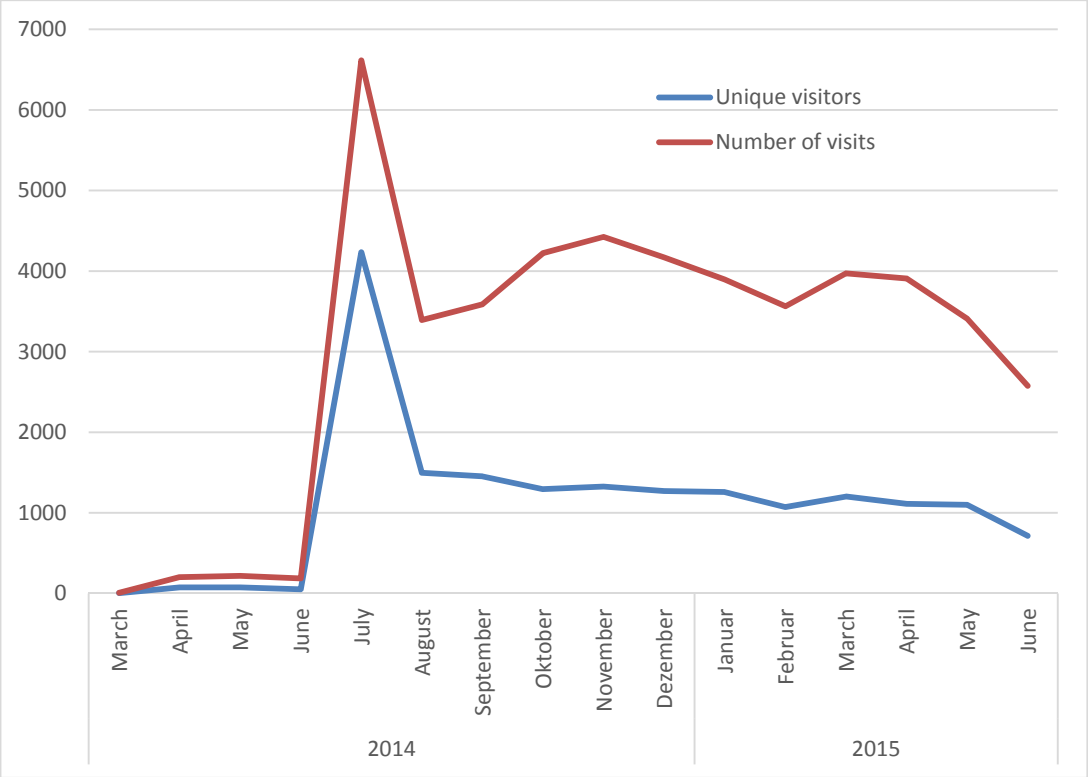
Access to restricted data is possible after users sign a data disclosure agreement, by which users agree to only make use of the data for analytical purposes and to publish results in a way that individual data points cannot be identified. As of March 2015, two requests have been received, the disclosure agreements have been signed and the data has been delivered to the users.

We conclude that the approach followed by the ETER consortium allowed for the smooth publication of data, in agreement with the NSAs, by allowing them to restrict data that are not public at the national level and where the national legal basis is too limited. The extent of restriction is fairly limited and it is likely to be further reduced in the future.

2.5 Website and data usage

The first version of ETER was made public in June 2015 and raised a high level of interest, as witnessed by media reports. As shown by Figure 3. Usage statistics of the ETER website, the ETER website had about 4,000 visitors in July 2014, immediately after the publication of the data, but interest has remained high since then with about 1,200 unique visitors per month, which is quite a sizeable number for a statistical database like ETER. This supports the insight that ETER is achieving a number of regular users. Since the publication of the data, the full ETER dataset has been downloaded more than 250 times.

Figure 3. Usage statistics of the ETER website



A further sign of interest and use of ETER is the fact that the consortium already received inquiries for delivery of restricted data, one from a university institute and one from an international organization.

To foster use of ETER, consortium partners have undertaken a number of dissemination activities. Since these activities were not foreseen in the ETER contract, they have been funded through other sources and particularly, from the EU-FP7 infrastructure project on Research Infrastructure for Research and Innovation Policy Studies (RISIS). These activities included:

- Two panel sessions presenting the ETER and MULTIRANK project at the two international conferences STI2014 (Leiden, September 2014) and CHER2014 (Rome, September 2014).
- The preparation of a poster presenting the ETER project, which has been displayed at STI2014, CHER2014 and at the RISIS week in Rome (January 2015).
- The organization of a specific information and discussion session on the usage of ETER at the RISIS week in Rome.

A major dissemination activity in spring 2015 will be the organization of a 1-day training course on ETER supported by the RISIS project; the course will present the ETER database, provide examples of usage and organize a laboratory session where users will be guided on how to properly exploit the RISIS dataset (http://risis.eu/training/). The course will be offered in five editions in Rome (16.04.2015), Oslo (16.06.2015), Leiden (18.06.2015), Wien (30.06.2015) and Lugano (01.09.2015). A similar course has also taken place at the European Commission on 25.06.2015.

Exploitation of ETER for specific analysis in higher education is also emerging. Particularly, a number of papers using ETER data will be presented to the Science and Technology Indicators Conference in 2015 (http://www.sti2015.usi.ch/). This type of ETER usage is expected to grow rapidly in the next few years.

3 Coverage and HEI demography

The analytical unit of ETER are individual Higher Education Institutions (HEIs), broadly defined as organizations for which delivering degrees at the tertiary level (at least ISCED-2011 level 5) is a major activity. However, no commonly accepted definition of HEIs exists within educational statistics, whose primary units for classification purposes are educational programs (and relatedly, enrolments of students and degrees awarded).

Therefore, a major issue for ETER was to provide such a definition and criteria for including and excluding organizations within the perimeter for data collection, based on conceptual, but also on practical reasons.

This chapter reviews the ETER definition of perimeter and its coverage, comparing it with the whole of tertiary education as defined in the EUROSTAT data collection. We also provide information on national perimeters and evidence of changes across time, so-called organizational demography.

3.1 ETER definition of the perimeter

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

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

A HEI is nationally recognised if it is officially accredited as such by a legitimate organization in a country. Being recognisable as a distinct organisation means that the perimeter of these institutions can be identified rather unambiguously, they have an internal organisational structure and, at least in principle, their own budget. The major activity criterion is excluded from perimeter organizations, which deliver training at the tertiary level as a side-activity, such as professional associations who provide education in their specific economic sector, or secondary education institutions who offer preparatory classes before entering university, which are classified at the tertiary level.

In order to ensure both a manageable data collection process and uniformity across countries, ETER adopted a formal size criterion. More precisely, institutions with less than 30 FTEs of academic staff and less than 200 students should only be included in exceptional cases, specifically for institutions graduating students mostly at the ISCED 8 level.

For those countries who participated in EUMIDA, the basis for the construction of the perimeter was the list included in EUMIDA. This list was further checked with the use of available information on websites; particularly those of the educational ministries, since these in most cases publish a list of nationally accredited HEIs. This list was then sent to the NSA in the concerned country for validation and to decide on unclear cases; at this stage, exclusion based on size was also decided for a number of HEIs.

3.2 Perimeter and exclusion cases

Table 7 provides a short description of the national perimeters for ETER. These descriptions can also be downloaded from the ETER database, since they represent central information for ETER users. Overall, they show that in almost all countries coverage is quite extensive, and includes not only universities, but also second-tier HEIs (Universities of Applied Sciences), as well as a number of specialized institutions. In many countries, the perimeter exactly matches the national-level definition of Higher Education Institutions, with the exception of HEIs, which do not reach the ETER size threshold. On the contrary, professional schools graduating at ISCED 5, as well as

secondary education institutions providing some preparatory course at the tertiary level, are routinely excluded.

Table 7. National perimeter descriptions

Country	Perimeter description
Austria	Higher education is provided by 22 Public Universities, 21 Universities of Applied Sciences (Fachhochschulen, FH, introduced in 1994), 13 Private Universities (introduced in 2000), and 17 University Colleges of Teacher Education (Pädagogische Hochschulen, PH, introduced in 2007). All these categories are included, there are a few exclusions because of size.
Belgium (French)	Belgium higher education is composed by three types of institutions for a total of 42 HEIs in the academic year 2011/2012. The six universities, 20 higher education institutions (from the academic year 2011/2011, after a few mergers), as well as 16 art schools. From the academic year 2010/2011, architectural schools (which are included in EUMIDA as self-standing HEIs) have been merged as faculties inside the universities, thus their data are included with those of the universities. Additionally, the Belgian system includes two federal HEIs, namely the military academy and the Evangelical Theological School (included in the perimeter for the Flemish part of the country). Academies have been excluded since these are a consortia of universities coordinating their doctoral training and do not directly enrol students. This is a difference with EUMIDA.
Belgium (Flanders)	The perimeter includes all institutions officially recognized as part of the Higher Education System by law in Flanders, including universities and HogeScholen. A number of institutes are also officially registered and recognized HEIs, which are allowed to grant degrees. They however do not receive any funding from the Flemish Government and do not submit to the same regulations. They are therefore not part of the DHO (Department of Higher Education from the Flemish Government). No data are available at the DHO for these institutes. For more info on the perimeter, see http://www.hogeronderwijsregister.be/het-hoger-onderwijs/de-instellingen.
Bulgaria	Currently the tertiary education system includes 51 higher schools, which under the Higher Education Act are state owned and private, including universities, specialized higher schools, and self-contained colleges. There are 37 public (25 universities, 11 specialized higher schools, and one self-contained college) and 14 private higher schools (five universities, two specialized higher schools, and seven self-contained colleges). A few institutions have been excluded because of their small size.
Croatia	The list of HEIs includes all officially recognized institutions, as part of the Higher Education System by law, including public and private universities, public and private polytechnics, and public and private schools of higher education.
Cyprus	Operating today in the Republic of Cyprus, there are three public and five private universities, five public institutes of tertiary education (only Cyprus Police Academy meets ETER criteria in academic year 2011-2012), and 30 private (13 of them do not meet ETER criteria in academic year 2011-2012).

Country	Perimeter description
Czech Republic	The perimeter includes all officially recognized institutions, as part of the Higher Education system, i.e. those that have state approval based on the accreditation of at least one study program of the official Accreditation Commission of the Czech Republic. Not having this approval means that an educational institution cannot be called a "higher education institution" ("vysoká škola"). It comprises all public as well as private institutions of both the university and non-university type. We exclude "tertiary professional schools" because they are excluded from Czech law on higher education institutions. They are mostly rather small institutions, and they fall under the law of secondary education. They offer study programmes on level 5B (ISCED 97) and according to the new classification ISCED 2011 they will be categorized on the 6th level, which is the level of bachelor programmes (with professional orientation). There are 179 of them with a 7% share on tertiary students in total.
Denmark	The perimeter includes all institutions, as part of the Higher Education System by law, including universities, universities of applied sciences, and teacher-training institutions. Some small music and art academies are not included, as they have less than 200 students and less than 30 employees, such as Aarhus School of Marine and Technical Engineering, Copenhagen School of Marine Engineering, and Technology Management and Maritime Training and Education Centre (MARTEC). The perimeter includes the University of Greenland and the University of Faroe Islands, but these two institutions are not part of the Danish higher education sector. There are no data for these two institutions in the 2011 dataset.
Estonia	All operative higher education institutions are included, even though they do not pass the threshold. As Estonia is a small country, excluding these small institutions would lead to excluding the field of theology in this country.
Finland	All official institutions are included.
Former Yugoslav Republic of Macedonia	The list of HEIs includes five public universities, 15 private universities and faculties. Of the latter, only four are included, the others are excluded because of size reasons. In addition to the state universities, a significant number of private higher education institutions were opened in the country and are private higher education institutions with representative numbers of students: the FON University (2002), the European University – Republic of Macedonia, the American College University. A new type of public-private non-profitable higher education institution has been introduced. Currently there is only one university of this type, the South-East European University established in Tetovo in 2001 with financial contributions from numerous international donors and foreign higher education institutions.

Country	Perimeter description
France	The perimeter includes all of the HEIs officially recognized as delivering Higher Education by the French governmental authorities, whether private or public, in 2011. Geographically it covers the entire French territory including the "départements et territoires ou collectivités d'outre-mer (Polynésie française, Nouvelle-Calédonie)". The perimeter excludes institutions with less than 200 students and/or less than 30 permanent staff. It also excludes all institutions or establishments delivering training in only very specific fields mainly disconnected from any research and sometimes integrated into the secondary school system (like 'Classe preparatoire aux grandes écoles', Instituts de soins infirmiers'…). It excludes all private business schools. However some institutions have been included due to their specific and significant role in the French Higher Education System: Collège de France, Ecole Pratique des Hautes Etudes, Ecole Nationale des Chartes, and the Regional Institut of Administration. The perimeter represents 71% of the student population in 2011/12; but with 100% of the student population at universities and National Polytechnic Institutes, and 97% of Engineering students.
Germany	All HEIs included in the perimeter, regardless of their size. In Germany, higher education is offered at three types of Higher Education Institutions. Universities (Universitäten), Universities of Applied Sciences (Fachhochschulen), and Universities of Art/Music (Kunst- und Musikhochschulen). Besides these three main types, which may be either state institutions or private institutions recognised by the state, some special categories exist, like church-maintained colleges, universities of cooperative education (Berufsakademien), colleges of education, or colleges of public administration. In their operations, including the organisation of studies and the designation and awarding of degrees, all types are subject to higher education legislation.
Greece	In Greece, after the last reforms of the national system of higher education, 22 public universities operate today. After this reform, 14 Technological Educational Institutes are also operating, as TEI of Lamia merged with TEI of Chalkida and TEI of Messolonghi with TEI of Patras in the academic year 2013-14. As far as other institutes of tertiary education are concerned, 10 Merchant Marine Academies, two Advanced Schools of Tourism, two Drama Schools, one School of Theatre, and four Higher Ecclesiastic Academies also operate. Of them, only six Merchant Marine Academies and two Higher Ecclesiastic Academies are eligible, according to quantitative ETER criteria.
Hungary	The Hungarian Higher Education system is composed by universities and colleges (non-university higher education institutions). Two smaller categories of HEIs exist as well: foreign institutions of tertiary education operating in Hungary and international higher education institutions. Private higher education institutions can also operate as international higher education institutions, if the majority of degree courses it offers are held in a foreign language and, on the basis of an agreement concluded with a foreign higher education institution, it provides several degree courses that are certified or recognised in another EEA or OECD country. There are currently two such institutions: Andrássy University and Central European University. Foreign higher education institutions can only operate in Hungary if they are recognised as a higher education institutions in their home country and the degree it awards is recognised as a degree appropriate to the type of higher education institution it is. At present there are 25 such institutions, e.g. Champlain College, McDaniel College, FernUniversität in Hagen and École d'Art Maryse Eloy. About 10 HEIs are excluded because of their size.

Country	Perimeter description
Iceland	The perimeter includes the seven Icelandic universities officially recognized as part of the Higher Education System.
Ireland	Ireland has 7 universities legislated under the Universities Act 1997, 13 institutes of technology legislated under the Institute of Technology Act 2006, and the Dublin Institute of Technology which is legislated under the Dublin Institute of Technology Act 2006. There are also a significant number of private and not for profit independent colleges offering higher education programmes with awarding power delegated by the Higher Education and Training Awards Council (HETAC). The proposed perimeter corresponds to the list of HEIs included in the annual statistical returns of the HEA's Student Records System. The perimeter excludes private independent colleges (no data are collected by NSA) and several institutes with delegated authority from HETAC to make awards (no data are collected by NSA).
Italy	Higher education in Italy is based on two systems, both of which lead to first, second, and third-cycle qualifications (ISCED levels 5-8): University education: it is provided by 96 university institutes, sub-divided as follows: 61 state universities, 18 non-state legally recognised universities, six higher schools or institutes called 'special system higher schools, and 11 on-line universities. The following two types of institutions release qualifications equal to those issued by universities, but are not included in the perimeter: Higher schools for language mediators offering three-year courses and issuing qualifications equivalent to the laurea (first-cycle degree) on language mediation sciences released by universities; Specialisation institutes for psychotherapists, issuing qualifications equal to the Specialisation diploma (third-cycle qualification) released by Italian universities in the broader field of psychology. High Level Arts and Music Education (Afam): it is provided by 20 states whose main purpose is preparing students for the practice of the arts (painting, sculpture, decorating and scenery); four Istituti superiori per le industrie artistiche, ISIA (Higher institutes for Artistic Industries) offering specific training in the design field (graphic design); one Accademia nazionale di arte drammatica (National Academy of Drama) in Rome; 54 Conservatori di musica (Conservatoires) aimed at the teaching of music, one Accademia nazionale di danza (National Dance Academy), in Rome; 23 legally recognised Accademie di belle arti (Academies of fine arts) together with 19 officially recognised music institutes (the last two categories are not included in the ETER perimeter). There are several other institutions offering higher education that are not included in the present perimeter, both because they all have small dimensions, and because they currently do not issue first, second, and third cycle certifications, and they fall under the competences of Ministries other than the Ministry of education.
Latvia	The perimeter includes all officially recognized institutions as part of the Higher Education System by law, including universities, universities of applied sciences, and teacher-training institutions. Nine small institutions are excluded, as they have less than 200 students and less than 30 employees and no students on the ISCED 8 level (and thus are considered research intensive): Latvijas Nacionālā Aizsardzības akadēmija, Starptautiskā Praktiskās Psiholoģijas Augstskola, Laterāna Pontifikālās Universitātes filiāle Rīgas Augstākais reliģijas zinātņu institūts, Laterāna Pontifikālās Universitātes filiāle Rīgas Teoloģijas institūts, Rīgas Uzņēmējdarbības koledža, Olaines Mehānikas un tehnoloģijas koledža, Ugunsdrošības un civilās aizsardzības koledža, Valsts Robežsardzes koledža, Kristīgās vadības koledža, Lutera akadēmija.

Country	Perimeter description
Liechtenstein	The perimeter includes the only university in the country. Excluded institutions: Private University in the Principality of Liechtenstein has less than 100 students, Internationale Akademie für Philosophie im Fürstentum Liechtenstein is very small and has only offered a diploma since 2013, Liechtenstein Institute only performs research and is extremely small.
Lithuania	All HEIs that are part of the official statistics are included, with the exception of two institutions that do not reach the size threshold.
Luxembourg	Higher Education in Luxembourg includes the University of Luxembourg, ISCED 5 technical diplomas offered as a side activity of secondary schools and a number of campuses of foreign universities, mostly offering postgraduate education. Only the University of Luxembourg is included in the perimeter.
Malta	Only the University of Malta is included. Malta college of arts, science and technology, and the Institute of Tourism Studies are excluded because of size; non-state sector institutions are also excluded.
Montenegro	The list of HEIs includes one public university, two private universities, and seven independent faculties. A few HEIs have been excluded because of size.
Netherlands	Higher education in the Netherlands is provided by three types of institutions. Government-funded institutions receive funding from the Ministry of Education, Culture and Science or the Ministry of Economic Affairs, Agriculture and Innovation, and charge their students government-approved fees. There are 36 higher professional education institutions and 14 universities, including the Open University. 'Legal entities providing higher education' are covered by the terms of the Higher Education and Research Act but are not funded by the Ministry of Education, Culture and Science. They are free to decide on their own fees and admissions policy, although students must be in possession of a HAVO, MBO or VWO certificate. There are more than 70 private providers of higher education in the NL. The ETER perimeter instructions require higher education to be a major activity of the institution. It is very difficult to assess whether the activity is really "major," because some privates will also offer other types of (post-secondary) education. In terms of the two additional criteria (size and continuity): most are very small (much less than the ETER thresholds of 30 FTEs in terms of staff and less than 200 students). Most don't have English names. Others are offshoots of public institutions. Most are not very visible, aside from two or three.
Norway	The perimeter includes most institutions, as part of the Higher Education System by law, including universities, universities of applied sciences, and teacher-training institutions. A total of nine private higher education institutions that are part of the official statistics are excluded due to a limited number of students or staff. These institutions enrolled approximately 1400 students in 2011. The Norwegian Defence University College would not provide data due to security reasons, but according to Statistics Norway, they had a total of 791 students in 2011. The institution is included in the perimeter, but not in the dataset. The Norwegian Police University College had a total of 2223 students, but no breakdowns on ISCED levels are available for this school. The institution is thus only included in the perimeter, and not in the dataset.

Country	Perimeter description
Poland	The perimeter includes all institutions in Poland's HE system, both public and non-public (private). Among them are: universities, technical universities, agricultural academies (including agricultural universities), academies of economics (including universities of economics), higher teacher education schools (including pedagogical universities together with Jesuit University "Ignatianum" in Cracow), medical universities, maritime universities, physical universities, theological academies, fine art academies together with Fryderyk Chopin University of Music in Warsaw and University of Arts in PoznańHEIs, academies of the Ministry of National Defence, and other HEIs (including public vocational schools). About 150 HEIs have been excluded for size reasons.
Portugal	The perimeter includes all official institutions recognized by law as part of the Portuguese Higher Education System. This includes Universities and Polytechnics, irrespective of their ownership (private vs. public). Portugal has a majority of private institutions, which differ from the public sector because these have, on average, a lower dimension and are more specialized in a few fields. As a result, the large majority of students are enrolled in the public sector. There are currently 133 recognized institutions in Portugal, but 43 of these have been removed from the perimeter due to their small size, measured in terms of the number of enrolled students and/or their number of full-time equivalent staff. Most of these institutions are smaller private institutions, notably several specialized in a few fields such as Teacher Training or Nursing Education. The recent trend in the system is a decrease in the number of enrolled students, both due to demographic and financial reasons. This has changed the system in the sense that those smaller private institutions have been struggling to achieve the number of students required for their academic and financial survival.
Romania	Higher education in Romania is offered in both public and private higher education institutions. These include universities, academies, and colleges organized in specialized departments. In accordance with its objectives, university education comprises: short university education offered by university colleges (3 years), long university education (4 to 6 years), and postgraduate university education (1 to 2 years). Public higher education institutions are coordinated by the Ministry of Education and Research. University autonomy is fully guaranteed. Private higher education is an alternative to public education. It is subject to an accreditation process. Accredited private institutions may obtain state support. A number of small HEIs have been excluded because of size.

Country	Perimeter description
Serbia	Serbia has 8 public and 10 private universities. In general there are three types of higher education institutions in Serbia – universities (univerzitet), colleges of applied sciences (visoka škola strukovnih studija), and colleges of academic studies (visoka škola). Faculties usually have the status of a legal body– but, they cannot exist independently, i.e. they need to be a constituent part of a university. The law also allows colleges of applied sciences to come together and form universities of applied sciences, but until now such institutions basically do not exist. Universities that serve as the only teaching and research higher education institutions in the country, also provide all three cycles of higher education as well as some forms of lifelong learning. Colleges of applied sciences, of which 69 have been accredited since October 2009, provide a first and some sort of second cycle of higher education. This second cycle is limited to 60 credits of applied specialisation that is not the same as a Master of applied sciences. Private universities are located in Belgrade and Novi Sad. Serbia accepted European Credit Transfer and Accumulation System (ECTS) which gave its universities an opportunity to make partnerships with other European universities and to make better study abroad programs and exchanges. The Slovakian Higher Education law knows three types of HEIs: universities, higher education institutions, and professional higher education institutions. In 2014, 23 public and 13 private HEIs could be identified. The ETER perimeter includes 32 HEIs, four are excluded because of their small size.
Slovakia	The Slovakian Higher Education law includes three types of HEIs: universities, higher education institutions, and professional higher education institutions. In 2014, 23 public and 13 private HEIs could be identified. The ETER perimeter includes 32 HEIs, four are excluded because of their small size.
Slovenia	Tertiary education in Slovenia consists of short-cycle higher vocational education and higher education. Higher vocational education is provided by law, namely the Higher Vocational Education Act. It is provided by both public and private higher vocational colleges, where students develop vocational competencies according to vocational standards. It gives an education at ISCED 2011 5 level. Higher education (supported by Higher Education Act) on the other hand is provided by universities and other higher education institutions, namely by faculties, higher professional colleges, and art academies. Universities, faculties, and art academies may provide study programmes of all cycles. Higher professional colleges provide, as a rule, the first cycle programmes (undergraduate); if they meet special stipulations, they may provide the second cycle programmes (graduate). Faculties and art academies provide first and second, as well as third cycle study programmes (ISCED 2011 6, 7, 8 level).
Spain	The Spanish university system consists mainly of universities recognized by the Ministry of Education (public, catholic and private). There are additional private institutions that are not formally approved by the Ministry but that have association agreements with the universities and are not included in the perimeter.
Sweden	All official public institutions that are part of the official statistics are included, with one exception: Operahögskolan i Stockholm. Some additional private institutions are included.

Country	Perimeter description
Switzerland	The perimeter includes all institutions which are officially recognized as part of the Higher Education system by law, including universities, universities of applied sciences, and teacher-training institutions; a few very small-scale HEIs are excluded (Institut de hautes études en administration publique, Lausanne, IDHEAP, Facoltà di Teologia di Lugano, Frankin College Switzerland, Theologische Hoschule Chur, Graduate Institute of International and development Studies, Geneva). The perimeter does not include professional schools outside the higher education sector, which enrolled in 2011 about 20% of total students at the tertiary level (however with a strongly decreasing share of tertiary education in the last ten years). About half of these students are enrolled in public higher professional schools (largely full time), the rest in part-time professional courses. There are about 150 higher professional schools with an average number of students below 200; no disaggregated data are available at the national level.
Turkey	Types of higher education institutions in Turkey are universities (üniversite), technical universities (teknik üniversite), vocational schools (meslek yüksekokulu), institutes (enstitü), and higher institutions of technology (yüksek teknoloji enstitüsü. Higher education institutions are established in Turkey under the law of The Turkish Grand National Assembly. Within the Turkish Higher Education System, higher education institutions (both public and non-profit foundations) are tied to the Council of Higher Education. Therefore, higher education institutions are recognized by the Council of Higher Education of Turkey since its foundation. Since study programs of higher education institutions are approved by the Council of Higher Education, those study programs are also recognized.
United Kingdom	Since the Further and Higher Education Act of 1992 ended the 'binary divide' between universities and polytechnics, all publicly funded higher education institutions have been part of a single higher education sector, subject to the same funding and quality assurance arrangements. There are 160 institutions in the UK that are permitted to award degrees by the UK authorities (UK and Scottish Parliament, Welsh and Northern Ireland Assemblies). In addition to institutions awarding degrees, there are also over 700 colleges and other institutions that do not have degree-awarding powers but provide complete courses leading to recognised UK degrees. Institutions that have degree awarding powers validate courses at these institutions. There are also a small number of independent private institutions that receive no direct government funding and are not included in the perimeter. University of London is a confederal organisation. The colleges of the university are shown separately in the HESA reference volumes and in the ETER perimeter.

3.3 Analysis of the coverage

By its construction, ETER includes a subset of the programs delivering degrees at the tertiary level, covered by the EUROSTAT data collection. Since HEIs that enrol large numbers of students are included in ETER, we expect that, in terms of numbers of students and of graduates, the coverage when compared to the nationally aggregated data published by EUROSTAT, is quite large. Table 8 provides an overview by country of the coverage of ETER as compared with EUMIDA on the one hand, and the total number of students at the tertiary level from EUROSTAT on the other hand.

In terms of the number of HEIs, ETER includes slightly more HEIs than EUMIDA – the total number of units increased from 2,471 to 2,678. The increase is due to the fact that ETER covers more countries. Additionally, France did not deliver any data to EUMIDA, as well as European Economic Area countries (Iceland, Liechtenstein), and candidate countries (Croatia, EU member state since 2014; FYRM, Montenegro, Serbia and Turkey). Especially Turkey, with 182 HEIs, counts among the largest higher education systems in the ERA.

When considering the EUMIDA countries, the total number of HEIs decreased from 2,471 to 2,202. This decrease is almost exclusively due to the exclusion of small HEIs in a number of countries. The largest drops are in Poland (from 457 to 273), Italy (from 243 to 176), and Norway (from 68 to 50). A substantial increase has been witnessed in Spain (from 47 to 80), where the EUMIDA perimeter included only the public universities and thus was clearly too small.

It is important to underscore that this number of HEIs is contingent on the specific definition of an HEI according to ETER, and on the introduction of a size threshold, and thus cannot be directly compared with other estimates of the number of Higher Education Institutions in Europe. As a matter of fact, the total number of tertiary education providers in Europe is likely to be in the tens of thousands, but most of them are small-scale educational providers at the ISCED 5 level, with no research mission.

In terms of the number of students, ETER currently includes 16.5 million tertiary education students, compared to 15.5 million in the EUMIDA dataset. The current coverage of the entire field of tertiary education in the perimeter in both databases is at 68%. With 88% of total tertiary education students, ETER coverage for the countries who have delivered data until now are at nearly the same level as EUMIDA, despite the fact that data are incomplete for two countries: Spain (no data on ISCED 8 students) and the UK (only ISCED 6 students).

When looking at countries individually, coverage of tertiary education slightly increased in many cases. The coverage is substantially better in Spain (from 65% to 79%) and Switzerland (from 74% to 82%), two of the countries showing low coverage in EUMIDA. Coverage dropped slightly for some countries, the most notable case being Italy (from 100% to 94%), where the EUMIDA perimeter was extremely extensive. Some large variances are explained by differences in definitions, like in the case of Greece, where the number of students in ETER is much lower than those reported to EUROSTAT and in EUMIDA; this is explained by the application of the new higher education act, which only allows students within a maximum duration of study to be counted. Most differences for Germany, Spain and the UK are explained by the non-or partial inclusion of PhD student data.

Overall, these data show that the strategy adopted in ETER for the definition of the perimeter was a sensible one: the reduction of roughly 10% of the HEIs included in the perimeter did not entail a significant drop in coverage by number of students, with the possible exception of Italy. Known issues in a few countries (especially Spain) could be addressed in order to make coverage uniform across all countries, both in terms of the number of observations and the number of countries.

Table 8. Overview of the ETER coverage and comparison with EUMIDA

Country	Number of HEIs		Coverage of tertiary education		Total students at tertiary level				Remarks
	EUMIDA	ETER	EUMIDA	ETER	EUMIDA	EUROSTAT 2009	ETER	EUROSTAT 2012	
Austria	68	68	0.88	0.93	270'131	308'150	349'545	376'499.00	
Belgium (French)	87	41		-				271'193.52	
Belgium (Flemish)		28	1.20	1.14	260'400	216'830	236'077	206'518.48	
Bulgaria	59	52	0.95	1.00	259'945	274'247	283'655	284'995.00	
Croatia	-	33	-	1.03		139'069	162'459	157'289.00	
Cyprus	37	25	0.75	0.95	23'086	30'986	30'227	31'772.00	
Czech Republic	73	71	0.91	0.89	379'220	416'847	392'760	440'230.00	
Denmark	14	33	-	1.09	-	234'574	282'668	258'932.00	2011
Estonia	34	29	1.00	0.96	68'168	68'399	64'806	67'607.00	
Finland	49	44	1.00	1.00	296'569	296'691	308'917	308'924.00	
Former Yugoslav Republic of Macedonia	-	10	-	0.87	-	65'200	58'520	67'490.00	2011
France	-	286	-	0.74	-	2'172'855	1'694'841	2'296'306.00	
Germany	410	387	0.83	0.85	2'025'307	2'438'600	2'494'876	2'939'463.00	Data on PhD students are only partial
Greece	60	50	0.92	0.54	584'274	637'623	360'597	663'698.00	Data only include students within the maximum duration of studies by law
Hungary	72	53	0.96	0.93	381'033	397'679	355'555	380'757.00	
Iceland	-	7	-	0.99	-	16'919	18'711	18'845.00	2011
Ireland	21	27	0.85	0.99	155'036	182'609	189'801	192'647.00	
Italy	243	176	1.00	0.94	2'005'240	2'011'713	1'816'448	1'925'930.00	
Latvia	61	49	0.90	0.95	112'567	125'360	92'007	97'041.00	
Liechtenstein	-	1	-	0.81	-	754	782	960.00	
Lithuania	46	43	0.96	0.90	202'509	210'744	156'899	175'066.00	
Luxembourg	1	1	0.66	0.88	3'526	5'376	4'710	5'376.00	2011
Malta	4	1	1.02	1.09	10'515	10'352	13'326	12'203.00	
Montenegro	-	7	-	-	-		-	-	
Netherlands	59	55	1.02	0.87	629'551	618'502	689'646	793'678.00	
Norway	68	50	0.98	1.05	215'851	219'282	249'342	238'224.00	
Poland	457	273	0.91	0.82	1'954'469	2'149'998	1'644'927	2'007'212.00	
Portugal	138	106	1.00	0.94	373'002	373'002	368'576	390'273.00	
Romania	82	82	0.94	-	1'034'214	1'098'188	-	705'333.00	
Serbia	-	16	-	-			-	-	
Slovakia	33	32	0.99	0.90	232'936	234'997	204'478	226'305.00	2011
Slovenia	13	47	0.84	-	96'323	114'391	-	104'003.00	
Spain	47	80	0.65	0.79	1'167'901	1'800'834	1'548'569	1'965'829.00	No data on ISCED8 students
Sweden	49	39	0.86	1.06	365'162	422'580	482'595	453'328.00	2012 data very low, to be checked
Switzerland	36	35	0.74	0.82	172'595	233'488	222'168	269'573.00	
Turkey	-	182	-	-	-	2'924'281	-	4'353'542.00	
United Kingdom	150	152	0.93	0.93	2'248'624	2'415'223	2'324'000	2'495'780.00	No data on ISCED8 students
Total	2471	2671	0.68	0.68	15'528'154	22'866'343	17'102'488	25'182'822	
Countries included in data collection			0.90	0.88	15'528'154	17'312'691	17'102'488	19'367'993	
Greece: data for 2008 instead of 2009; Luxembourg data for 2011									

A more precise analysis can be done by using data on students and graduates disaggregated by the ISCED-2011 level; this is relevant as there are good reasons to expect that the coverage of ETER is much better at levels 6 (bachelor), 7 (master) and 8 (PhD), than at level 5 (professional diplomas of less than three years).

In terms of the highest degree delivered, the current ETER database includes only 23 HEIs whose highest degree is at ISCED 5 level, 19 of them being in a single country (Latvia). Additionally, 413 HEIs are included in the database that also award ISCED 5 level degrees, as well as higher level degrees. These figures imply that, at the organizational level, ISCED 5 degrees are largely offered by distinct providers, which are in most cases not included in ETER.

This observation is confirmed by a comparison of the number of students with EUROSTAT aggregates (Table 9), since only few countries reported ISCED 5 students in ETER and the total corresponds to only 11% of ISCED 5B students in the EUROSTAT data. We notice that some deviant figures are a remnant of the fact that ISCED-1997

5B does not correspond exactly with ISCED-2011 5: for example, in Norway, most short diplomas were probably classified under ISCED-1997 5A. Some issues might be generated by problems in the mapping of the programs to ISCED-2011.
On the contrary, the coverage of ETER is nearly complete for students at levels 6 (bachelor) and 7 (master) of the new ISCED classification, as the aggregate figure reaches 96% of the EUROSTAT totals and, for most countries, the figure is quite close to 1. Most figures exceeding 1 are likely to be due to classification differences between ISCED-1997 and ISCED-2011. The lower figures for Greece are explained by a modified perimeter of students (only students within the maximum duration of studies are included in ETER), whereas data for Sweden will be rechecked.
Coverage for PhD students reaches 84% for the countries delivering data and most figures are very near to EUROSTAT totals. The large difference for Germany is due to the lack of complete enrolment statistics for PhD students, where national totals include survey-based estimates. Some differences are also due to different counting methods, since PhD numbers in some countries in ETER are based on employment, thus resulting in substantially lower figures.
Comparisons for graduates are slightly more precise since EUROSTAT data already distinguishes between first-level and second-level qualification, but there is still a distinction between professional and general degrees (5A vs. 5B). This is likely to generate some of the differences noted in countries like Norway. Data on PhD graduates are very near to the Eurostat aggregate, demonstrating that ETER coverage at this level is almost complete.

Table 9 Comparison between ETER and EUROSTAT for students (above) and graduates (below)

	ETER						EUROSTAT			Coverage				Remarks
	ISCED5	ISCED6	ISCED7	ISCED7long	ISCED5-7	ISCED 8	ISCED5B	ISCED5A	ISCED6	ISCED 5	ISCED 6-7	ISCED5-7	ISCED 8	
Austria	-	188'113	46'750	88'288	323'151	26'394	40'436	310'011	26'052	-	1.04	0.92	1.01	
Belgium (Flemish)		177'957	46'976		224'933	11'144	134'194	128'969	8'031	-	1.74	0.85	1.39	
Bulgaria		195'535	51'621	31'315	278'471	5'184	16'210	264'082	4'703	-	1.05	0.99	1.10	
Croatia					158'827	3'632	49'398	104'656	3'235			-	1.12	
Cyprus	2'393	19'218	7'260	527	29'398	829	8'458	22'604	710	0.28	1.19	0.95	1.17	
Czech Republic		240'513	92'223	34'740	367'476	25'284	32'870	381'255	26'105	-	0.96	0.89	0.97	
Denmark	30551.0	175886.0	66435.0	279.0	273151.0	9517	33'422	216'653	8'857	0.91	1.12	1.09	1.07	
Estonia		44'758	13'055	3'949	61'762	3'044	20'791	43'765	3'051	-	1.41	0.96	1.00	
Finland		228'273	60'051		288'324	20'593	84	288'645	20'195	-	1.00	1.00	1.02	
Former Yugoslav Republic of Mace		54'750	3'314			456	2'203	65'287	:	-	0.89	0.86		2011
France	142'402	727'605	394'014	323'894	1'622'748	72'093	576'668	1'649'057	70'581	0.25	0.88	0.71	1.02	
Germany		1455178.0	319318.0	609769.0	2384265.0	110611	490'360	2'240'603	208'500	-	1.06	0.87	0.53	
Greece		299'640	37'070		336'710	23'887	224'478	415'773	23'447	-	0.81	0.53	1.02	
Hungary	21079.0	247035.0	80241.0		348355.0	7200	42'048	331'455	7'254	0.50	0.99	0.93	0.99	
Iceland	303	13'396	4'197	363	18'259	452	359	18'008	478	0.84	1.00	0.99	0.95	2011
Ireland	39'568	120'163	21'907		181'638	8'163	39'780	143'937	8'930	0.99	0.99	0.99	0.91	
Italy		1'100'293	343'887	337'215	1'781'395	35'053	4'263	1'887'038	34'629	-	0.94	0.94	1.01	
Latvia	16'465	61'770	11'253		89'488	2'519	16'821	77'697	2'523	0.98	0.94	0.95	1.00	
Liechtenstein	-	528	234	-	762	20		854	106		0.89	0.89	0.19	
Lithuania		123'880	21'919	8'644	154'443	2'456	49'777	122'414	2'875	-	1.26	0.90	0.85	
Luxembourg		3'222	1'098		4'320	390	984	4'034	358	-	1.07	0.86	1.09	2011
Malta	2'466	6'914	3'116	752	13'248	78	1'628	10'498	77	1.51	1.03	1.09	1.01	
Netherlands	2'469	577'187	101'280		680'936	8'710	15'884	765'252	12'542	0.16	0.89	0.87	0.69	
Norway	43'438	135'380	34'122	26'870	239'810	9'532	871	229'135	8'218	49.87	0.86	1.04	1.16	
Poland		1'035'404	427'569	142'602	1'605'575	39'352	13'253	1'953'696	40'263	-	0.82	0.82	0.98	
Portugal		229'411	56'645	63'050	349'106	19'470	74	370'972	19'227	-	0.94	0.94	1.01	
Slovakia		120633.0	62482.0	10410.0	193525.0	10953	2'851	206'231	12'145	-	0.94	0.93	0.90	
Spain		1'068'205	113'805	366'559	1'548'569		321'392	1'621'895	22'542		0.95	0.80	-	
Sweden	6030.0	309888.0	58369.0	86730.0	461017.0	21578	27'494	404'482	21'352		1.12	1.07	1.01	
Switzerland		142'986	53'365	3'101	199'452	22'716	55'717	191'844	22'012	-	1.04	0.81	1.03	
United Kingdom		1'789'475	534'525		2'324'000		390'792	2'010'039	94'949	-	1.16	0.97	-	
All countries	307'164	10'893'196	3'068'101	2'139'057		501'310	2'613'560	16'480'841	596'456	0.12	0.98	0.86	0.84	

	ETER						EUROSTAT				Coverage					Remarks
	ISCED5	ISCED6	ISCED7	ISCED7long	ISCED5-7	ISCED 8	ISCED5	ISCED6	ISCED7	ISCED8	ISCED 5	ISCED 6	ISCED 7	ISCED5-7	ISCED 8	
Austria	-	28244	10871	13182	77963	2246	15440	38'735	12'798	2'412	-	0.73	1.88	0.78	0.93	
Belgium (Flemish)		37637	20364		62152	1678	25666	14'076	22'021	1'424	-	2.67	0.92	0.94	1.18	
Bulgaria		38285	22467	4451	73019	1153	4151	33'804	25'157	979	-	1.13	1.07	1.03	1.18	
Croatia					0	1338	11504	25'460	1'518	1'338	-	-	-	-	1.00	
Cyprus	771	3283	1728	86	5868	52		2'841	1'086	48		1.16	1.67	1.49	1.08	
Czech Republic		53162	32312	7178	92652	2712	7816	60'195	37'079	2'683	-	0.88	1.07	0.88	1.01	
Denmark	8493.0	37089.0	17232.0	7.0	62821.0	1865.0	8335	32'643	16'137	1'552	1.02	1.14	1.07	1.10	1.20	
Estonia		7522	2627	485	10634	233	3886	4'747	2'674	190	-	1.58	1.16	0.94	1.23	
Finland		35345	16092		51437	1834	25	35'345	16'092	1'834	-	1.00	1.00	1.00	1.00	
Former Yugoslav Republic of Mace		8944	1085		10029	197	215	10'177	1'567	146	-	0.88	0.69	0.84	1.35	2011
France	56560	192700	166328	54982	470570		224364	311'026	148'615	13'188	0.25	0.62	1.49	0.69	-	
Germany		208492	58292	118795	385579	26563	149160	309'808	68'440	26'807	-	0.67	2.59	0.73	0.99	
Greece		51940	9361		61301	1522	20184	32'809	11'606	1'734	-	1.58	0.81	0.95	0.88	
Hungary	4276.0	36430.0	19440.0		59839.0	1232.0	11'163	38'456	19'056	1'242	0.38	0.95	1.02	0.88	0.99	2011
Iceland							91	2'828	1'137	40	-	-	-	-	-	2011
Ireland	13707	30520	12991		57218	1532	15007	27'368	16'200	1'447	0.91	1.12	0.80	0.98	1.06	
Italy		196096	113813	39983	349892	11616	2611	213013	156250	11458	-	0.92	0.98	0.94	1.01	
Latvia	3274	13647	3694		20615	267	3460	12'886	4'859	267	0.95	1.06	0.76	0.97	1.00	
Liechtenstein		213	86		299	2		60	145	6		3.55	0.59	1.46	0.33	
Lithuania		32475	7228	1225	40928	382	12698	21'510	7'773	398	-	1.51	1.09	0.97	0.96	
Luxembourg		530	273		803	57	435	635	162	57	-	0.83	1.69	0.65	1.00	2011
Malta	1209	2153	987	0	4349	24	537	1'872	1'041	13	2.25	1.15	0.95	1.26	1.85	
Netherlands	1791	92600	40416		134807	4136	1156	102'042	44'811	4'040	1.55	0.91	0.90	0.91	1.02	
Norway	20016	29210	10015	3768	63009	1461	239	30'251	8'448	1408	83.75	0.97	1.63	1.62	1.04	
Poland		249963	161017	44716	455696		3980	318'912	312'475	3'590	-	0.78	0.66	0.72	-	
Portugal		50982	31356	7700	90038	4155	55	51'010	40'290	2'909	-	1.00	0.97	0.99	1.43	
Slovakia		36751.0	31080.0	1524.0	69355.0	2178.0	823	36'263	33'107	2'181	-	1.01	0.98	0.99	1.00	
Spain		115301	60420	98609	274330	10504	107330	208'961	66'182	9'483	-	0.55	2.40	0.72	1.11	
Sweden	1921	34270	15017	11719	62927	3347	8731	43'113	13'953	3'343	0.22	0.79	1.92	0.96	1.00	
Switzerland		29675	18105	989	48769	3639	26680	28'733	20'097	3'638	-	1.03	0.95	0.65	1.00	
United Kingdom	120945	400480	239145		760570	22115	128432	389'296	242'440	20'438	0.94	1.03	0.99	1.00	1.08	
All countries	232'963	2'053'939	1'123'842	409'399		108'040	794'174	2'438'875	1'353'216	120'293	0.29	0.84	1.13	0.83	0.90	

This analysis can be summarized as follows:

- In general, ETER provides good coverage of tertiary education from the bachelor onwards, while coverage of short professional diplomas (ISCED-2011 level 5) is very limited. This is mostly due to the institutional fragmentation of the ISCED 5 level, characterized by a very large number of small institutions, and by reduced data availability for this subsector at the national level.
- In most cases, there is good consistency between ETER and EUROSTAT aggregated figures for students, even if a more precise analysis can be done only when EUROSTAT statistics are published using ISCED-2011. Large departures are explained by the different definitions adopted (Greece) and issues with counting PhD students.

3.4 Demography

An important issue in the analysis of higher education systems concerns organizational demography, i.e. the fact that new organizations are continuously created over time, some may close, while events such as mergers may also take place. In the case of higher education, there is clear evidence of a long-term process of creation of new HEIs under the pressure of the expansion of higher education; data derived from EUMIDA foundation years display successive waves of foundations of different types of HEIs, with the emergence since the 60s of non-university institutions, as well as of private institutions. Further, policy studies of higher education document processes of restructuring within the national higher education system, and especially, consolidation of the non-university sector through mergers. This highlights the fact that, unlike in the private sector, public regulation is a major driving factor of institutional demography in higher education.

Tracking demography is relevant both in terms of data management and in order to fully exploit the availability of multi-annual data for analytical purposes. This obviously requires the ability to trace HEI data back to their ancestors in the case of mergers, dealing with implications of entries and exits for other HEIs etc.

Accordingly in ETER, demographic changes have been tracked between 2008 (EUMIDA) and 2011, respectively between the two waves of ETER data collection. While of course changes in this period are relatively limited, this information will become more relevant over a longer period of time.

Table 10 provides an overview of the changes in the HEI population between EUMIDA and ETER. Accordingly, in ETER 2011, 1993 out of the 2467 HEIs included in EUMIDA are still present, while 474 were excluded (mostly due to the size threshold) and 682 were added (mostly due to the inclusion of new countries). When looking to the 2012 data collection, 1754 HEIs are the same as in EUMIDA, while 446 are new. For about 500 HEIs there is no information, as the 2012 perimeter was not updated for the countries not collecting data.

Table 10. Changes in the HEI population between EUMIDA and ETER

		ETER 2011		ETER 2012		
		0	1	0	1	No information
EUMIDA 2009	0	45	682	23	446	258
	1	474	1993	459	1754	254

Table 11 provides a summary of the observed demographic event, while Table 12 includes a more detailed breakdown by country. 561 of the 639 entry events from EUMIDA to ETER concern countries which were not included in EUMIDA and thus only reflect the extension of the country's coverage; among EUMIDA countries, we found only two cases where the revision of the perimeter led to the inclusion of a substantial number of already existing HEIs: Spain (27 entries) due to the inclusion of private universities, and Slovenia (34 entries), due to the inclusion of a large number of HEIs

graduating at level ISCED 5. Entries between 2011 and 2012 concerns HEIs (mostly recently founded) that in the meantime reached the ETER size threshold.
The exits are in almost all cases due to the fact that the concerned HEIs did not reach the ETER size threshold: the largest number concerns Poland (162) and Italy (63). A few other cases were motivated by more careful analysis. For example, the Belgian Academies have been excluded from the perimeter, since these are a consortia of universities offering graduate education, but not employing staff and enrolling students. Exits from 2011 to 2012 are related to the size threshold as well and are concentrated in a few countries.

Table 11. Overview of demographic events between 2009 and 2011

Event code	Description	2009-2011	2011-2012	Remarks
1	entry	561	0	Related to the integration of new countries in the dataset
1	entry	78	23	Entries in countries already covered by EUMIDA
2	exit	382	28	Most cases because HEIs are below the size threshold of ETER
3	birth	24	7	
4	death	36	2	
5	merger	45	6	
6	split	3	4	
7	take-over	43	11	
8	spin-out (spin-off)	2	0	Two cases of foreign colleges created by existing HEIs in the perimeter
	no information	0	512	Countries which did not update the perimeter in 2012

This cursory analysis displays that in the considered period, no country underwent an organizational restructuring of their national higher education systems; demographic changes have rather been individual events concerning specific cases of HEIs.

Birth and death. ETER data displays a non-negligible number of new HEI foundations (24 from 2009 to 2011 and 7 from 2011 to 2012), as well as closures of HEIs existing in 2009 or 2011 (36 and 2). In Germany, these cases almost exclusively concern private universities of applied sciences, many of which are specialized in fields such as arts or management. Similar remarks apply to Poland. While a more detailed analysis of the characteristics of these HEIs will be provided at a later stage, there is sufficient evidence that the foundation of new higher education institutions, or their closure, is a phenomenon concerning mostly private and specialized HEIs, which don't have the right to award a doctorate.

Mergers and take-overs. Between 2009 and 2012, we observe 15 mergers and 23 take-overs. The difference between the two events is that, in the case of merger, a new HEI is founded, in most cases bearing a new name and organizational structure, whereas in a take-over an HEI is integrated as a department within an existing HEI. A remarkable feature of these events is that they are distributed across countries, showing that they are a rather isolated and sectorial event rather than the outcome of large-scale restructuring of HEI systems.

An analysis of the list of mergers (see Table 12) displays that most cases are located in Nordic or Central eastern countries and they respond to a logic of consolidation, either between schools in the same scientific domain (like the Stockholm University of Arts) or geographical region (Aalto University in Helsinki, the University colleges in Oslo). Some of the merged institutions are quite large and internationally visible, like in the case of Aalto.

Table 12. Mergers in higher education 2009-2011

Institution name	Country	Description
Haute Ecole de Namur-Liège-Luxembourg (HENALLUX)	Belgium	Merger of two professional higher education institutions in the Namur region of Belgium
Ecole supérieure des Arts de la Fédération Wallonie-Bruxelles Arts2	Belgium	Merger of two art schools in the French community of Belgium
The Royal Danish Academy of Fine Arts, Schools of Architecture, Design and Conservation	Denmark	Merger of two Danish art schools
Aalto university	Finland	Merger between the University of Art and Design, Helsinki School of Economics, and Helsinki University of Technology
University of Eastern Finland	Finland	Merger of the Universities of Joensuu and Kuopio
National University of Public Service	Hungary	Merger of Zrínyi Miklós National Defence University, the Police College, and the Faculty of Public Administration of Corvinus University of Budapest
Edutus College	Hungary	Merger of the Harsányi János College, Budapest and of the College of Modern Business Studies, Tatabánya
Lithuanian University of Health Sciences	Lithuania	Merger of the Kaunas University of Medicine and the Lithuanian Veterinary Academy
Oslo and Akershus university college of applied sciences	Norway	Merger of Oslo and Akershus university colleges
West Pomeranian University of Technology, Szczecin	Poland	Merger of the university of Agriculture in Szczecin and Szczecin University of Technology
Linnaeus University	Sweden	Merger between the University of Kalmar and Växjö University
Stockholm University of Arts	Sweden	Merger of the University of Dance and Circus, the University College of Opera, and Stockholm Academy of Dramatic Arts
University of Wales Trinity Saint David	UK	Merger of the University of Wales, Lampeter, and Trinity College, Carmarthen
Vilentum University of Applied Sciences	NL	Merger of CAH Dronten Professional Agricultural University and Stoas Professional University for Teacher Training
Institute of Art, Design and Enterprise – University	PT	Merge of a design and marketing school.

The 21 take-overs are cases where small specialized schools have been absorbed by a large university or college in the same region and thus respond to an overall process of consolidation of the higher education systems in a smaller number of generalist higher education units. In the Belgian case, they are the outcome of a political decision to integrate all formerly independent faculties of architecture in the universities, whereas in other countries they are isolated events.

Table 13. Demographic events by country

Events are reported by affected HEIs, thus the number of effective events is smaller since some event involve multiple HEIs, like in the case of mergers

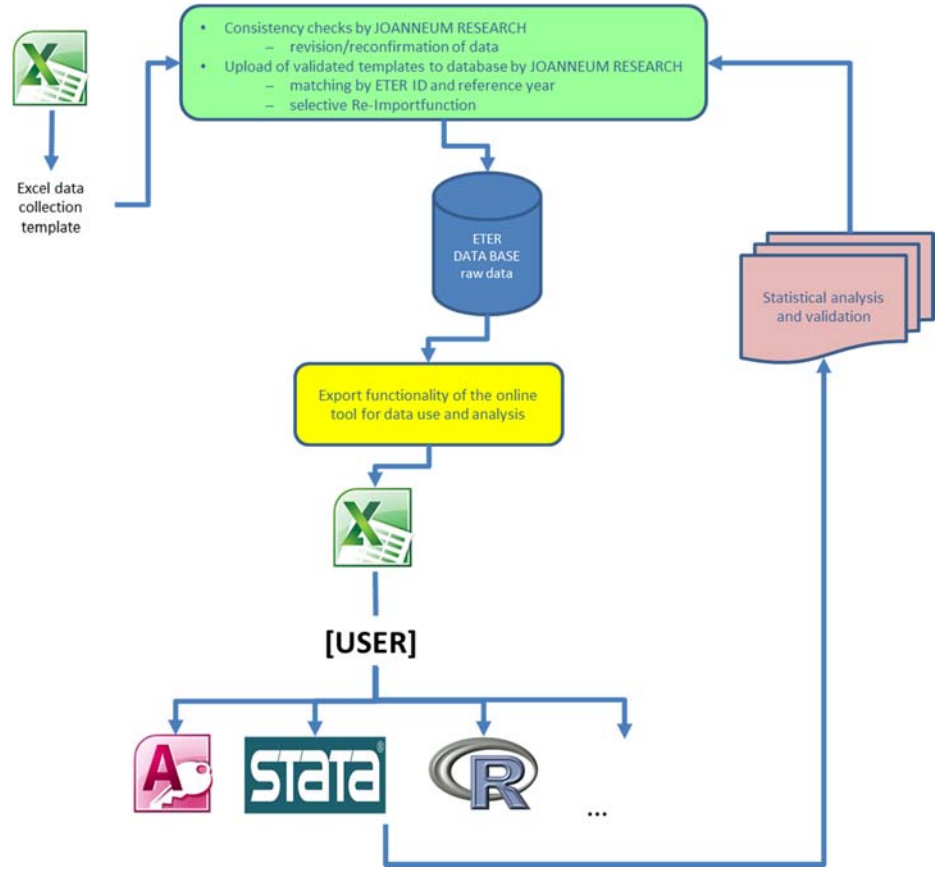
	No change	Entry	Exit	Birth	Death	Merger	Split	Take-over	Spin-out
AT	66	1	1	1	1				
BE	62		9			6		12	
BG	51		7	1	1				
CH	35		2					2	
CY	21	2	13	2				4	
CZ	72			1	2				
DE	390		37	12	11		3	2	
DK	11	21				4			
EE	29				3			2	
ES	49	27		4					
FI	40					7		4	
FR	1	285							
GR	48	2	12						
HR	3	30						1	
HU	47		15	1		6		6	
IE	21	6							
IS		7							
IT	175		63		1	4			1
LI		1							
LT	41		3			3			1
LU	1								
LV	47	1	10		2			2	
ME		7							
MK		10							
MT	1								
NL	57		1					2	
NO	47		15			3		4	
PL	284		162	2	10	3			
PT	113		19		5			2	
RO	78	4	4						
RS		16							
SE	37		8			6			
SI	13	34							
SK	32		1						
TR		182							
UK	148	3				3			
Grand Total	**2020**	**639**	**382**	**24**	**36**	**45**	**3**	**43**	**2**

4 Data collection and data management

The ETER project has developed an infrastructure that allows for standardization and systematization of the process of data collection, preliminary validation and data management. This is highly important in order to master the complexity of a data collection process that requires cooperation with the National Statistical Authorities and to ensure a sufficient level of uniformity in the data.

Figure 4 presents the overview of the whole process, starting from data collection and control to data export for data quality analysis (see chapter 5) and data analysis.

Figure 4: Data collection interfaces



The infrastructure used in ETER includes:

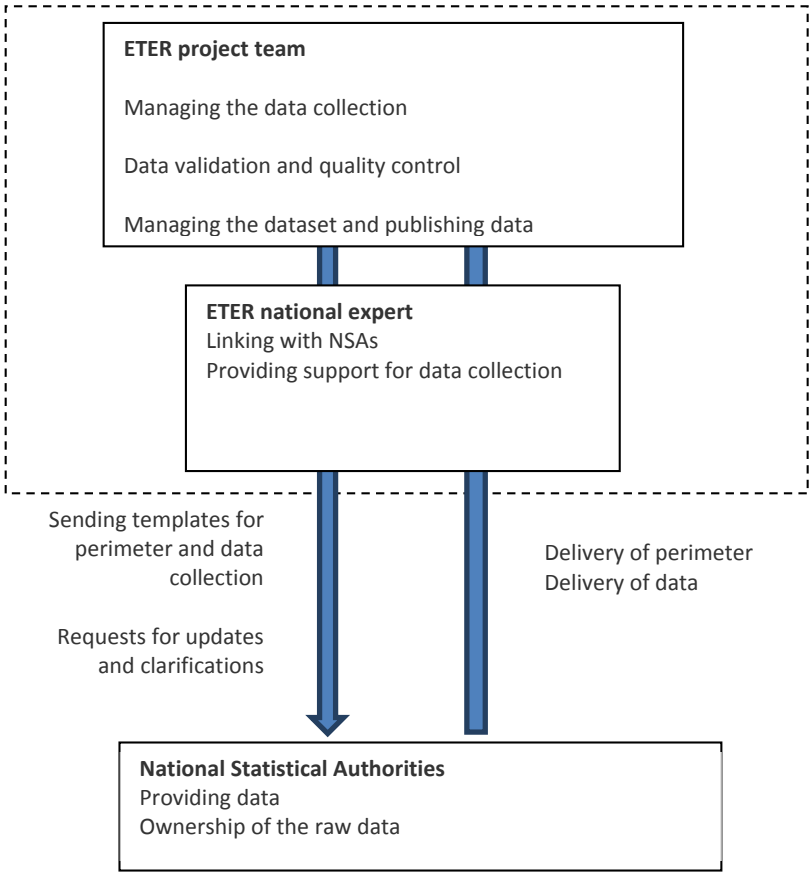
- Templates for data collection, including documentation (e.g. flags and special values as commonly used for EUROSTAT-statistics), which guide national data sources (statistical offices, national authorities, other sources) and country experts addressing and supporting national data sources.
- Procedures for the preliminary data validation in order to detect mistakes and inconsistencies.
- A master database including upload interface and documentation of the database, which also constitutes the basis of the online tool for access to data by final users.

Following this chapter, we will shortly present the different components of the data collection, validation and integration process. For a more technical description, the reader should refer to the ETER handbook.

4.1 Data collection organization

The data collection process is managed in close cooperation between three levels, namely the ETER project team, national experts and the National Statistical Authorities (Figure 5).

Figure 5. Organization of the data collection process



The *project core team* manages the entire ETER infrastructure, including methodology, the data collection template and the data management system. The project core team is also in charge of preliminary validation of the collected data and for quality control.
National experts (NE) are dedicated persons who manage the contacts to each individual country, finding the right contact points and solving practical problems with NSAs. National experts might also help for some parts of data collection, particularly for descriptors and geographic information. Most national experts are actually part of the core team's organization (see the full list in the annex), however in a few countries where particular issues were expected, ad hoc national experts were nominated (for example because of their knowledge of the national language and higher education system).
National Statistical Authorities (NSAs) are the main source of data for ETER, with the partial exception of descriptors and geographic information. NSAs fill in the data in the ETER data collection sheets (in Excel), provide most of the metadata, as well as consent on the use of data, since they are the primary owners of most data.
Within the ETER project, cooperation with the NSAs was excellent: NSAs not only provided data, but were also readily available to advise and respond concerning unclear cases and potential problems with the data detected by the data validation system and quality control. Besides correcting some figures, this process involved a large deal of flagging and commenting of special cases, which need to be taken into

account when using the data (see further in chapter 5 concerning issues of data quality, flagging and comparability problems).
In order to achieve smooth cooperation with NSAs, a suitable timeline taking into account other reporting duties to EUROSTAT (UOE data collection) was important. Therefore, in the second wave of data collection, the process began in the summer of 2014 to avoid overlap with UOE. This allowed most NSAs to deliver data by Christmas. Also, taking into account data releases, the optimal timeline for future data collection would then be as follows:

- Start of data collection: July of year X+2 (for example July 2015 for data collection 2013).
- Deadline for delivering data: November of year X+2 (for example November 2015 for data collection 2013).
- Publication of data: spring of year X+3 (for example April 2016 for data collection 2013).

Despite generally good cooperation, there was some difficulty when collecting data from a few countries, and in some cases, data could not be delivered. The main reasons for problems were:

- Availability issues of data at the national level, respectively the fact that micro-data are confidential within the country, like in the case of Slovenia.
- Lack of NSA resources for participating in the project.
- Communication problems when the NSA does not own the micro-data, which are managed by other entities like the higher education ministry.

Therefore, most issues are not due to a lack of commitment by the NSAs, but rather to internal organizational questions within the concerned countries, particularly limitations on the public availability of data on HEIs. Conversely, the success of ETER reflects the fact that in the vast majority of European countries statistical data on HEIs (particularly on those in the public sector) are made systematically available to the public, thus implementing the goals of the transparency policy.
To address problematic situations, the ETER consortium has deployed a set of measures in close coordination with the European Commission: these included repeated contacts with the involved countries, the use of national contacts within the university and policy layer, help by the consortium to prepare the data sheets, and finally, contacts from the Commission and particularly EUROSTAT. This close cooperation between the consortium and the Commission (respectively EUROSTAT) in addressing NSAs proved to be effective in order to address problematic cases, and thus should be continued in the future.
Table 14 summarizes the situation of the problematic countries, as well as the measures adopted by the consortium in order to address problems and their outcome. In most cases, solutions could be worked out and data collection should be completed or planned in the next months. There are however a few countries where the situation is uncertain and may go beyond the possibilities of the consortium to collect data in the current phase. The consortium remains optimistic that, given the wide visibility of the project, the remaining countries will join the data collection in the future.

Table 14. Situation of specific countries

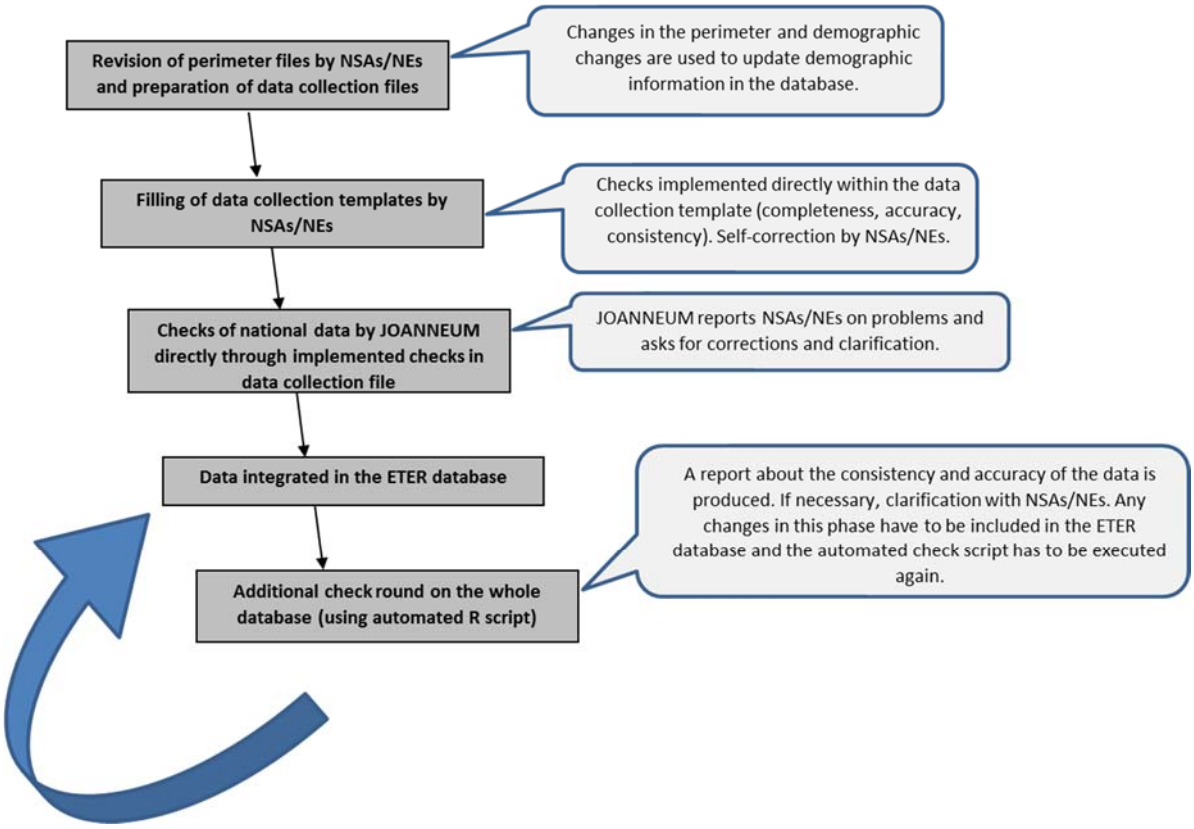
Country	Issue	Actions	Outcome and current status
Belgium (French)	A reorganization of the Belgian government led to the closure of the higher education observatory and, therefore, there was no authority in charge of collecting and managing data both for EUROSTAT and ETER.	Contacts have been made with the new Academy of Research and Higher Education (ARES), in close cooperation with EUROSTAT.	ARES decided in principle to participate in ETER and began the procedure to receive authorization from the Belgian government. Once this has been done, data collection can begin and data can be integrated into ETER.
Iceland	Data for 2011 have been delivered, but partially were compiled directly from the consortium. The NSA did not participate in the 2012 data collection.	No further action taken.	Only data for 2011 available.
Former Yugoslav Republic of Macedonia	NSA has no resources to participate in data collection.	Data for 2011 were prepared by consortium from data available on the NSA website.	Decision not to further collect data by the consortium, data for 2012 not available.
Montenegro	Very difficult case, very hard to find a responsive contact at the NSA. They are not part of the EUROSTAT data collection.	No specific action taken beyond direct contacts from the consortium.	No data have been delivered.
Romania	Repeated contacts with the NSA, which declared that they cannot participate and suggested to outsource data collection.	Contacts by the consortium through the Higher Education Evaluation Agency, as well as by EUROSTAT.	No response.
Serbia	The statistical office informed the consortium that they do not hold data on individual HEIs and suggested we address the ministry.	The Ministry of Education, Science and Technological Development was contacted by the European Commission in December 2013, but no answer has been received.	No response.

Slovenia	Slovenia informed the consortium in December 2013 that the NSA is not allowed to deliver individual-level data.	Repeated contacts by the consortium, by the European commission and through a contact at the University of Ljubljana.	Data has been promised repeatedly, only the perimeter is currently available. The main problem is that there is a general decision within the country that HEI-level data are confidential and should not be published. At least a core set of data should be available by late March 2015.
Turkey	The first contact with NSA was positive, but data collection was delayed by the Higher Education Council. We received a letter on 27/02/2014 from the NSA that they cannot deliver any data because they have not received a reply from the Higher Education Council.	No further action.	No data delivered.
UK	Despite frequent contacts, the NSA announced that they couldn't provide data because of a lack of resources, although about 60% of the data would be available. Publicly available data has been collected by the consortium, but these are extremely limited. Some additional data have been provided by the NSA.	Repeated attempts by the consortium. Unfortunately, most UK data are available only under payment of a fee and, therefore, cannot be made available for ETER.	Limited data available in ETER. This is particularly problematic because of the importance of the country. Contacts are on-going with the Higher Education Statistical Authority in order to improve the situation with UK.

4.2 Data collection and preliminary validation

Figure 6 presents the main steps of the data collection process.

Figure 6. Data collection, validation and integration



1) Data collection starts with a revision of each country's perimeter. The process is as follows:

- The filled perimeter files for the last reference year are adapted to serve as templates. Any changes in the perimeter and demographic changes are recorded by NSAs/NEs in these files.
- Returned files are used to update the information about demographic events and the database. Also, the updated perimeters are transformed into the data collection files in order to start data collection immediately after validation of the perimeter files.

2) The data collection is performed in Excel and can be started after revising the perimeter, since Excel files are a common and practical way to share data for the ETER partners. The bases of the data collection templates for the reference year are the files from the preceding year.

The excel files already include a number of checks in order to call attention to NSAs/NEs and thus enable them to correct the data immediately. The implemented checks alert the deliverer of data about:

- incomplete data and blank cells,
- accuracy problems, i.e. wrong data formats or irregularities, and
- mistakes in sums and inconsistencies between variables (see the list of checks in the annex of the ETER handbook).

The data collection files include prefilled cells with the previous year's values in order to support NSAs/NEs and reduce the burden for data collection. These include information that is not expected to change systematically from year to year, like descriptors, nominal variables, which are not likely to change, and the resulting not

applicable values (e.g. no ISCED 5 level in a country, then "a" is prefilled in the respective categories), flags (except inconsistence "ic") and metadata.
Prefilled nominal variables, flags and metadata are labelled in order to draw the attention of NSAs/NEs. Data are seen as confirmed if the responsible persons remove the labelling. This simultaneously safeguards that the people completing the data receive some help during the process, while also ensuring that users are alerted to the data types.
3) The incoming data collection files are screened for any mistakes detected by the implemented automatic checks in the Excel files in order to provide reasonably clean data for the first upload into the database. Problematic cases are analysed by the core team of experts who are responsible for the supervised countries under their responsibilities and decide whether to correct the data (when the source of the problem is clear), to flag the data, when for example the reason of the problem is already known or explained in the metadata or to report back to NEs/NSAs when the explanation of problems are not straightforward.
4) The resulting data sets are to be uploaded into the database in order to track any changes. In a further step, the imported data are separately checked, country-by-country, using an R script, which produces a report about the consistency and accuracy of the data. The R script detects the following potential problems:

- missing values,
- inconsistencies regarding accuracy (data formats, irregularities),
- mistakes in sums, and
- inconsistencies between variables and noticeable ratios, which will be calculated from the data (see the list of checks and noticeable ratios in the annex of the ETER handbook).

Using the data- consistency and accuracy report, obvious mistakes are corrected by the ETER project team, while all other issues are clarified in cooperation with the NSAs/NEs. Additionally, the data- consistency and accuracy report forms the basis of the initial flagging of data. After corrections to the data are made, our new data is delivered and the data are again uploaded to the database.
This approach ensures that checks are easily reproducible running the R script, reasons for changes can be traced by the quality report, and changes in the data can be tracked directly in the database.
The final version of the data is uploaded and published when the pre-validation phase by the R script shows no further un-flagged inconsistencies. The pre-validated data can then be exported in order to perform quality checks and add additional flags according to the outcome of the quality analysis.
Data collection in the first round started in November 2013 and ended in March 2014 whereas in the second round data collection was launched in August 2014 and concluded for most countries in December 2015. This timeline proved to be optimal to allow NSAs to work in parallel in the provision of EUROSTAT/UOE data and ETER data. Therefore, it is suggested for the future to adopt the following calendar for data collection:

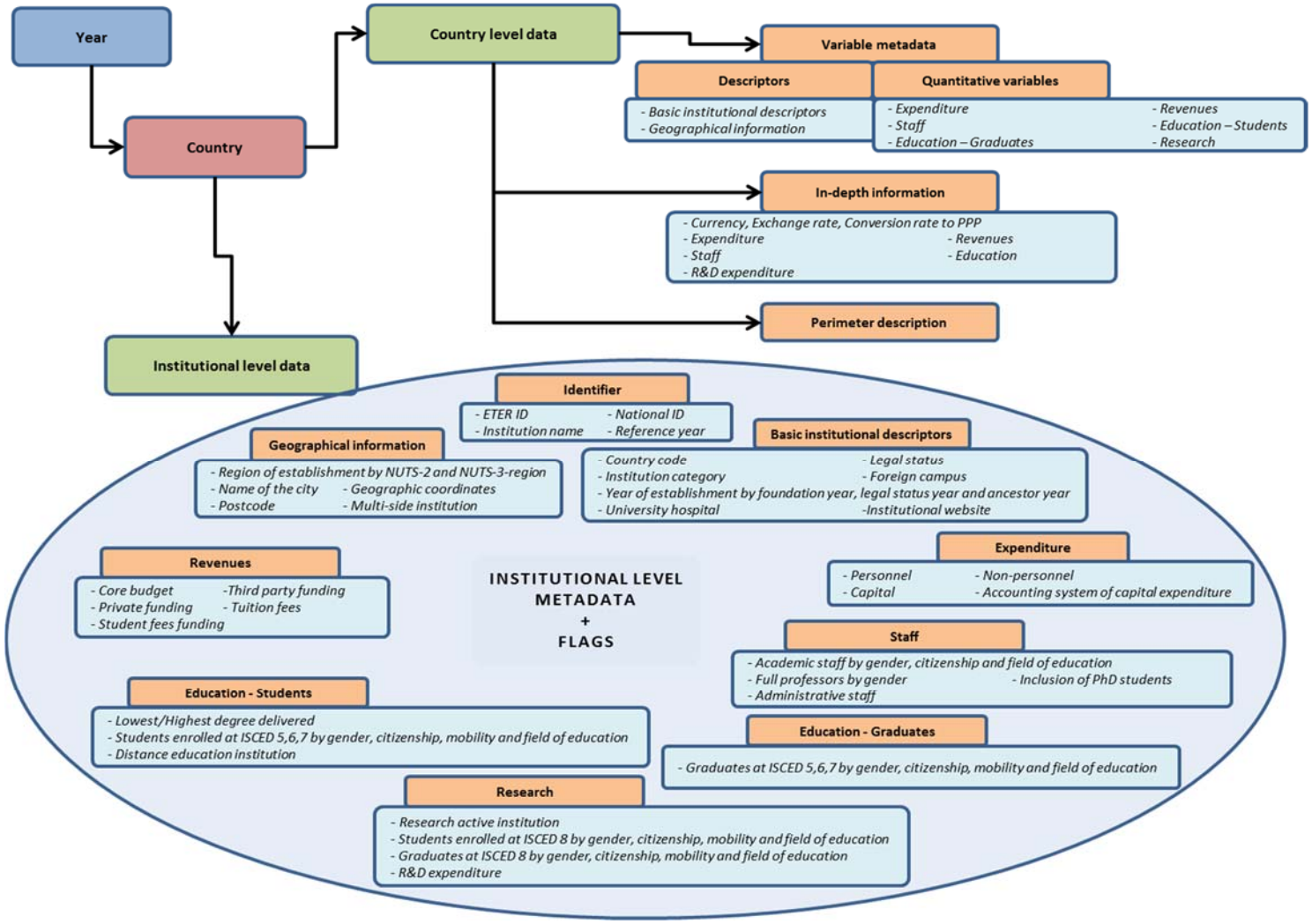
- launch in June/July of each year, in parallel with EUROSTAT/UOE.
- Delivery by end of November.
- Data publication in late spring.

4.3 Data management tool

The data management tool is derived from already existing tools (imdas/archivis) engaged by JOANNEUM RESEARCH. The existing model of a data management system and centralized data collection has been adapted according to the specific needs of the ETER project. This has been done by JOANNEUM RESEARCH and coordinated with the core-partners, selected country experts, and the client.
The following figure shows the structure of the database and the first data level, the reference year. After the reference year, data are structured by country. For each country, variables are collected at the institutional level, which also includes a set of institutional metadata in order to provide a detailed data description for institutions.

Additionally, the data set provides flags for all quantitative variables, e.g. in the case of incomparable data after quality control. In the case of corresponding sub-categories of variables (e.g. male and female students), the flag marking a statistical footnote will apply for all corresponding sub-categories. Flags are provided as separate columns in the data set.

Figure 7: Structure of the database



Besides data on the institutional level, the database includes metadata information in order to consider institutional and country specific characteristics in the data (see chapter7 of the handbook for a description of the metadata system). Metadata are both at the HEI level and at the country level. Country level metadata are structured into metadata for descriptor and quantitative variables, which contain information about:

- the content and deviation from ETER definitions,
- the reference period, and
- data sources.

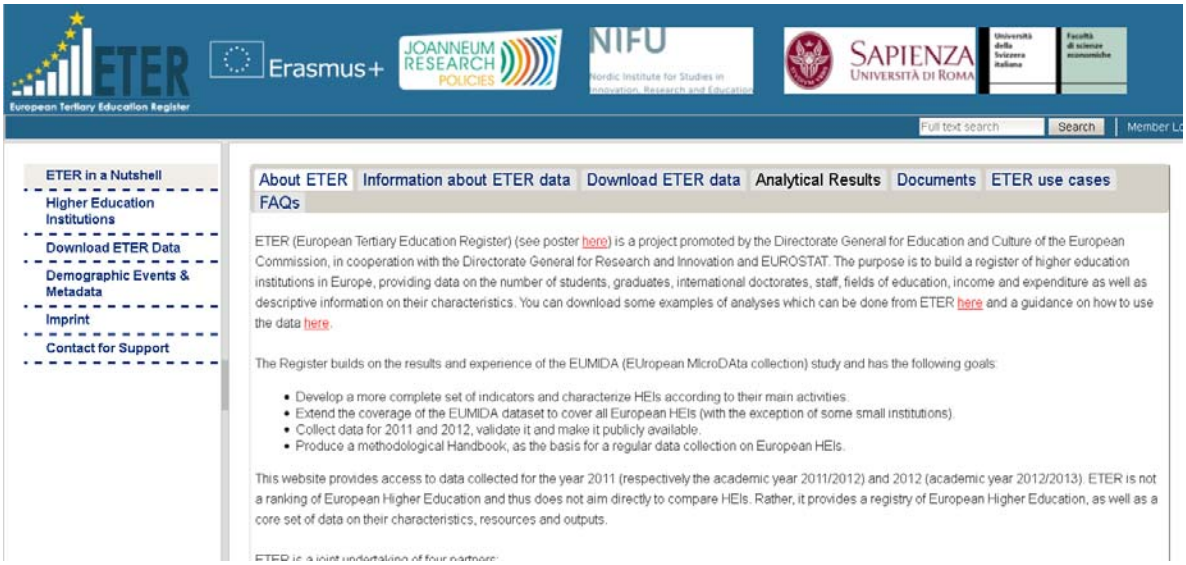
"In-depth information" contains more detailed additional information about the quantitative variables covered in ETER. Finally, country level data also the perimeter descriptions and thus provides information about the higher education landscape in the observed countries.

The database is managed via direct and central access by JOANNEUM RESEARCH, which guarantees data security, consistency and quality.

4.4 The on-line web access

The database also provides the foundation for the ETER web application developed by JOANNEUM RESEARCH, which can be considered as the output of the ETER project. The web application enables the user to retrieve data from the entire ETER data set in order to conduct research on micro data of the European higher education sector. This chapter will therefore focus on the structure and functionalities of the web application, whose start page is shown in the following figure.

Figure 8: Start page of the ETER web application



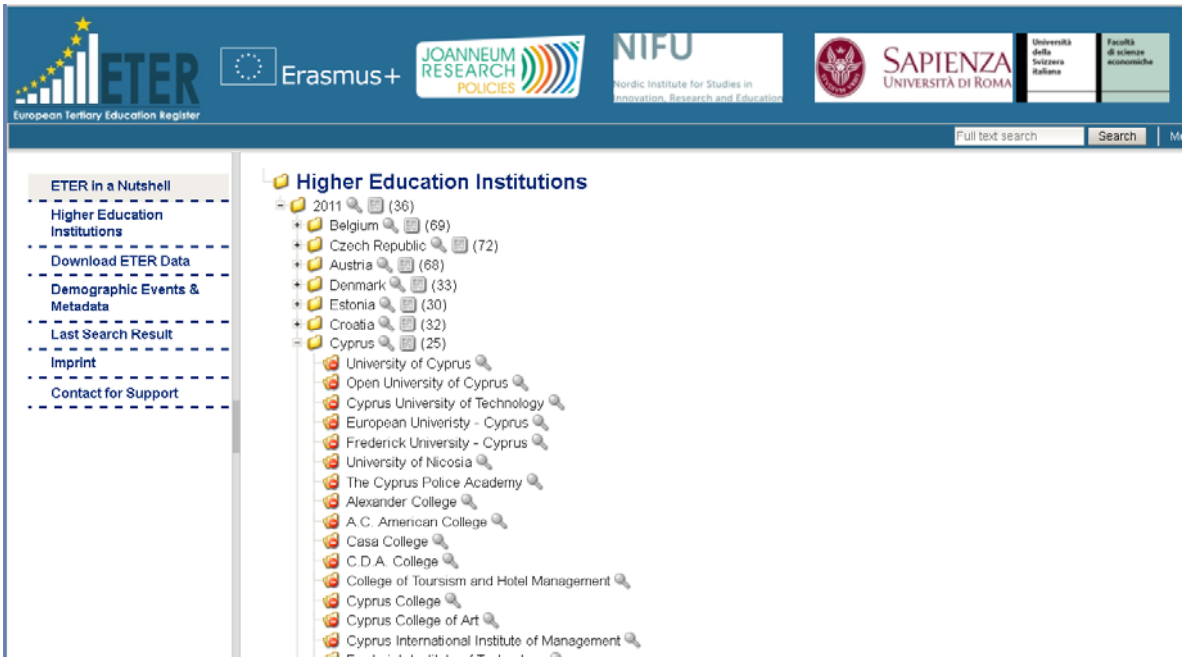
The homepage includes a short description of the ETER project and the performing consortium members. Starting from the homepage of the ETER web application, three paths are prepared for the user to define an individual query (depending on the information required):

- path1: the user wants to get an overview of the included higher education institutions,
- path2: the user wants to export demographic information or metadata on the country level, or
- path3: the user wants to export data from the ETER micro data set.

4.4.1 Getting an overview about higher education institutions in ETER

The selection field "Higher Education Institutions" offers the possibility to have a closer look at the included higher education institutions in the ETER project. The data are prepared by year and country, and by using the plus sign, a list of all included HEIs in a specific country's perimeter will appear. Using the magnifier symbol leads the user to a detailed view of the variable (year, country or institution), which includes "ETER ID," "English Institution Name," "Year," "Country Code" and an additional export possibility of the data for the selected institution.

Figure 9: Overview of institutions included in ETER



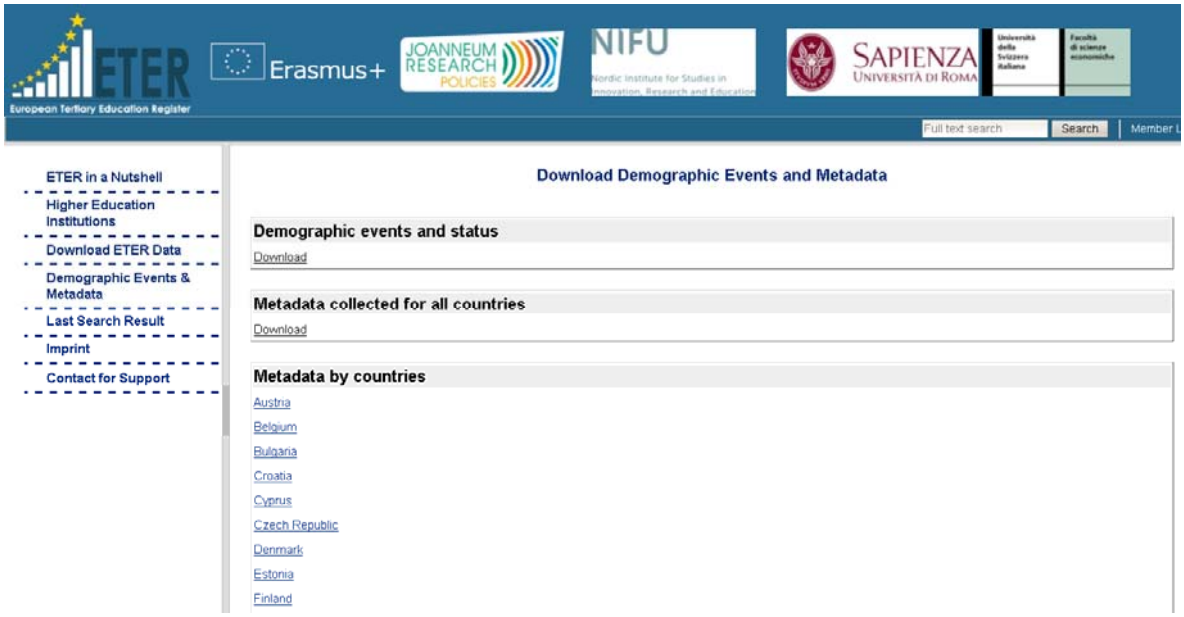
4.4.2 Export country level metadata and information about demographic events

By choosing the selection field "Demographic Events & Metadata", users can gather information about demographic events or country level metadata. The user is forwarded to a platform where the following information can be downloaded.

- an Excel-file where all demographic events for all countries and all institutions since 2009 were collected,
- one Excel-file for every country, where country level metadata for all variables and perimeter descriptions are collected in four unified sheets, and
- an Excel file with four sheets (descriptors, quantitative metadata, in-depth information, perimeter description), where all corresponding metadata are collected and organized by year, country and variable. Using a filter enables the user to choose the desired metadata by year, country and/or metadata variable.

To simplify the handling of institutional data and metadata, the application enables switching back directly from the metadata area to the last search by using "Last Search Result."

Figure 10: Download demographic events and metadata



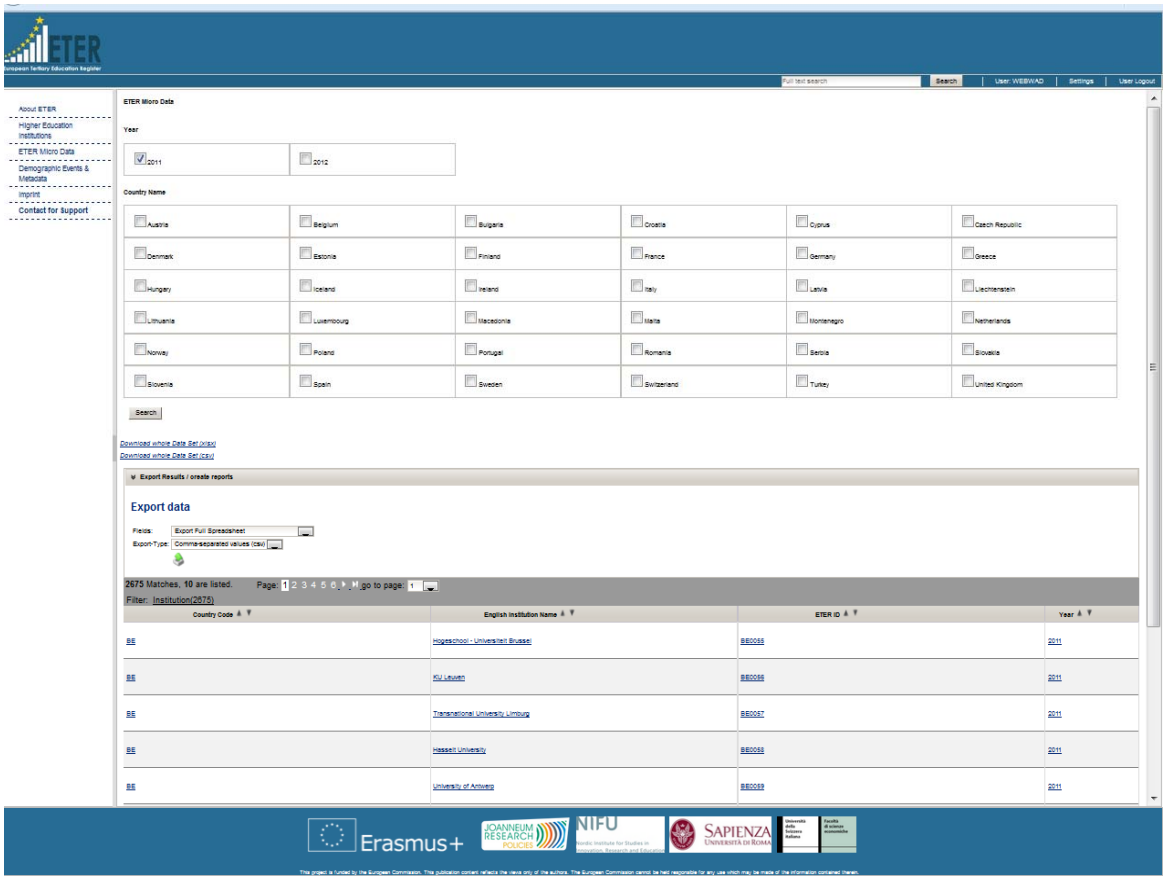
4.4.3 Export data from the ETER database

In order to follow path3, the user needs to access "ETER Micro Data," which leads to a mask where the data can be selected by year and country.
This means that the user can choose to view:

- the entire data set, or
- data sets for a specific country in one or more years.

After choosing the required information, the ETER web application displays the search result, which can be exported by using the selection field "Export Results / create reports." The following figure shows the result of a search for the year 2011, which leads to all data in the data set for this year.

Figure 11: Research results for the year 2011



The result shows all institutions in the data set, which can now be exported to Excel. The export function enables an export of:

- all variables at once, using "Export Full Spreadsheet,"
- a specific group of variables (e.g. staff), or
- a predetermined group of related variables (e.g. all revenues and expenditures). Four sets of variables are currently provided: "all expenditures and revenues data"; "students, graduates and research", "all student's data", "all graduates data".

An export of a specific group includes the basic variables:

- "ETER ID",
- "National Identifier",
- "Institution Name",
- "English Institution Name", and
- "Year" plus all variables assigned to a group.

4.5 Data sources

Information on data sources have been collected within ETER metadata to ensure traceability of data. Even if the situation slightly improved compared to the 2011 collection, indications provided in metadata have a different level of detail among countries. Generally speaking, data come from official sources recorded by statistical agencies or relevant Ministries. On the contrary, descriptors in the large majority of cases come from institutions' websites and are not recorded in official national data collections. In most cases they have been collected directly by the ETER consortium or by national experts.

The table below summarizes the sources of quantitative variables for the 27 countries which delivered data for 2012.

In most countries, even when data are provided by the same NSA they come from different collections, possibly performed in different ways. As shown in the table, in the majority of countries, quantitative information come from administrative sources. Bulgaria, and the United Kingdom are exceptions, with all data coming from surveys, while a few other countries (CY, DK, HR, IT, LT, NL, SE) have a mixed situation. Information on R&D expenditures are collected either by surveys or through administrative data.

Table 15. Data sources by typology

ETER variable	AT	BE*	BG	CH	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HR	IE	IS	IT	LI	LT	LV	MT	NL	NO	PL	PT	SE	SK	UK
Total expenditure	m	1	m	1	0,1	1	1	2	m	m	1	1		1	1	m	1	m	0	m	1	1	1	-	1	0	1	0
Personnel expenditure	m	1	m	1	0,1	1	1	2	m	m	1	1		1	1	m	1	m	0	m	1	1	1	-	1	0	1	m
Non-personnel expenditure	m	1	m	1	0,1	1	1	2	m	m	1	1		1	1	m	1	m	0	m	1	1	1	-	1	0	1	m
Capital expenditure	m	1	m	m	0,1	1	1	2	m	m	m	1	m	1	1	m	1	m	0	m	1	2	1	-	1	0	1	m
Total income	m	1	m	1	0,1	1	1	2	m	m	1	1	m	1	1	m	1	m	0	m	1	1	1	-	1	0	1	0
Core budget	m	1	m	1	0,1	1	1	2	m	m	m	1	m	1	1	m	1	m	0	m	1	1	1	-	1	0	1	m
Third party funding	m	1	m	1	0,1	m	1	2	m	m	m	1	m	1	1	m	1	m	0	m	1	1	1	-	1	0	1	m
Private funding	m	1	m	1	0,1	m	1	2	m	m	m	1	m	1	1	m	1	m	0	m	1	1	1	-	1	0	1	m
Student fees funding	m	1	m	1	0,1	m	1	2	m	m	m	1	m	1	1	m	1	m	0	m	1	1	1	-	1	0	1	m
Number of academic staff (FTE)	m	2	0	1	0,1	1	1	0	m	1	1	m	m	0	1	m	0,1	1	1	m	1	0	1	-	1	1	1	m
Number of academic staff (HC)	m	2	0	1	0,1	1	1	m	m	1	1	m	1	0	1	m	0,1	1	1	1	1	0	1	-	1	1	1	0
Number of full professors	m	2	0	1	0,1	1	1	0	m	1	1	m	1	0	1	m	1	1	1	1	1	0	1	-	1	1	1	m
Number of administrative staff (FTE)	m	2	0	1	0,1	1	1	0	m	1	1	m	m	m	1	m	0	1	1	m	1	0	1	-	1	1	1	m
Number of administrative staff (HC)	m	2	0	1	0,1	1	1	m	m	1	1	m	1	m	1	m	0	1	1	1	1	0	1	-	1	1	1	0
Total staff (FTE)	m	2	0	1	0,1	1	1	0	m	1	1	m	m	m	1	m	0	1	1	m	1	0	1	-	1	1	1	m
Total staff (HC)	m	2	0	1	0,1	1	1	m	m	1	1	m	1	m	1	m	0,1	1	1	1	1	0	1	-	1	1	1	0
Students and graduates ISCED 5-7	1	2	0	1	0,1	1	1	1	1	1	1	1	1	0	1	m	0	1	1	1	1	1	1	-	1	1	1	0
Students and graduates ISCED 8	1	2	0	1	0,1	1	1	1	1	1	1	1	1	0	1	m	0	1	1	1	1	0	1	-	1	1	1	0
R&D Expenditures	m	m	m	2	0,1	m	m	0	m	m	1	m	-	0	m	m	m	m	0	1	1	m	0	-	m	0	m	m

0=survery, 1=administrative, 2=other

* Flanders part

For further reference, the following table presents an overview of the sources of data by country and by variable domain. It can be observed that in most cases data come from National Statistical Offices and/or from Ministry of Education and Research, according to national governance settlements. However, for very small countries with one or very few HEIs, the data come directly from their institutional database.

Table 16. Data sources by country

Country	Expenditure	Income	Academic and non academic staff	Students and graduates ISCED 5-7	Students and graduates ISCED 8	R&D Expenditures
AT	m	m	m	Statistics Austria	Statistics Austria	m
BE*	yearly financial reports		VLIR (2013). Statistische gegevens betreffende het personeel van de Vlaamse universiteiten. + administrative data for university colleges	-	-	m
BG	m		Republic of Bulgaria National Statistical Institute			m
CH	Federal Statistical Office					
CY	CYSTA: Services Survey for Private		CYSTAT: Annual Survey on Education			CYSTAT: R&D survey
CZ	Ministry of Education, Youth and Sports					m
DE	statistics of finances of universities		statistics of personnel of universities	statistics of students of universities	statistics of examinations of universities	m
DK	Public finances		Danske Universiteter + Uddannelses og Forskningsministeriet (UFM)	Student register		Annual questionnaire
EE	m	m	m	-	-	m
ES	m	m	-	-	-	m
FI	Ministry of Education and Culture/Finnish National Board of		Ministry of Education and Culture + Statistics Finland	Statistics Finland	Statistics Finland	Statistics Finland
FR	Ministère de l'Éducation nationale, de		m	Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la		m
GR	m	m	Hellenic Statistical Autoritiy			National Documentation Center
HR	-	-	-	-	-	-
IE	Funding Statements of higher education		Quarterly staffing returns to HEA	HEA database (Student Record System)		m
IT	MIUR Ufficio Statistico		MIUR database + Survey of Contract academic staff and Technical and Administrative staff + Survey on AFAM	MIUR Ufficio Statistico		m
LI	m	m	-	-	-	m
LT	School report		Centre of Information Technologies of Education			School report
LV	m	m	Central Statistical Bureau data collection on Higher Educatiob Institutions			Central Statistical Bureau R&D data collection
MT	University of Malta					
NL	DUO (Dienst Uitvoering Onderwijs)		Vereniging Hogescholen + VSNU (Association of Universities in the Netherlands)	DUO (Dienst Uitvoering Onderwijs)	VSNU (Association of Universities in the Netherlands)	m
NO	DBH - Database on Higher education		NIFU/Register og research personnel + DBH	DBH - Database on Higher education	DBH + NIFU/Doctoral degree register	NIFU/R&D statistics
PL	Ministry of Science and HE		-	Ministry of Science and HE	Ministry of Science and HE	Ministry of Science and HE
PT	DGEEC- Direção-Geral de Estatísticas da Educação e Ciência					m
SE	Swedish Higher Education Authority		Swedish Higher Education Authority + Statistics Sweden	data not published	data not published	Statistics Sweden
SK	Ministry of Education, Science, Research and Sport of the Slovak Republic					m
UK	Higher Education Statistics Agency (HESA)					m

* Flanders part

5 Validation and Data quality

Validation and data quality control are central tasks in ETER. They respond to the requirements of ensuring the best possible quality of the data collected and to limit inconsistencies as much as possible.
Institution-level databases, like ETER, raise particularly difficult issues concerning data quality given the lower level of aggregation, which makes them particularly sensible to quality issues. Further, given the fact that ETER is not only intended for research purposes, but also for providing transparency and visibility of the activities of individual HEIs, it is not sufficient that the dataset display a sufficient level aggregated data quality. Instead, care must also be taken to guarantee the correctness of individual data points and the comparability between individual observations.
A central issue for data quality is represented by the high level of heterogeneity in the HEI sample considered: not only are there systematic differences in the organization of national systems of higher education, which also translates to differences in indicators. Additionally, HEIs themselves are highly heterogeneous organizations, ranging from large and internationally reputed research universities to small-scale, mostly teaching oriented, specialized schools in arts, theology or teacher training. This significant level of heterogeneity makes the identification and interpretation of deviant observations highly problematic.
This chapter is structured in two sections. First, we present the main components of the ETER data quality approach. Second, we summarize the results of the analyses performed on the current 2012 dataset and their implications. The data quality analysis refers to the data as they were available at the mid of May 2015. Moreover, many quality problems detected have been already been corrected.

5.1 Data quality approach

The ETER approach to data quality is based on the combination of two integrated processes:

1. preliminary level quality and validation checks performed within the data collection phase on a country basis in order to allow for an easy return on the respondents and the correction of data before online integration;
2. overall final quality and validation phase which has the role of performing more complex controls that can provide hints to use data in the appropriate way or possibly to improve the quality of future collections. In this phase, quality control is performed on data at both the "global" and "local" levels.

Different methods are applied:

- A systematic analysis of *internal data quality*, more specifically referring to four dimensions: format accuracy, completeness, consistency, and timeliness (see the ETER handbook for details).
- Advanced statistical methods for *outlier detection*.
- Checks of external validity by comparing with other data sources or with information available elsewhere.

We shortly summarize the methods in the subsections below; the reader should refer to the data quality chapter of the ETER handbook for a full description.

5.1.1 Internal data quality

In order to characterize the internal quality of data collected in ETER, we evaluated a set of dimensions belonging to the ISO 25012 view related to Internal Data Quality, which are a subset of those proposed by Eurostat:
a) 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 financial values, student and graduate data being integer variables, and so on.

For a dataset like ETER, a central dimension of accuracy consists of monitoring the adherence to the rules concerning missing values, their coding, and the correct distinction between "0" values and missing values (respectively not applicable values).
b) For each variable, dimension, and dataset, *completeness* evaluates the share of missing values that are present in the considered dataset. In ETER, completeness is analysed by variable, by country, and by individual HEI.
c) Consistency verifies possible violations of semantic rules defined over the involved data, and specifically between different variables.
Given the nature of the ETER dataset, there is a high number of mutual dependencies between variables, which can be exploited for purposes of data quality analysis. In broad terms, they can be regrouped in the following categories (see a complete list in the ETER handbook):

- Logical dependencies between categorical variables and values. For example, when the highest degree delivered is ISCED 5, all numbers of students and graduates at levels 6-8 have to be coded as "not applicable." Similarly, if an HEI is non-research active, R&D expenditures should be "not applicable." Most of these rules are already stipulated in the definition of these variables.
- Sums of breakdowns of variables equal to the total, for example the sum of male, female, and gender unclassified students should be equal to the total.
- Relationships between valued variables. For example, R&D expenditures should be lower than total expenditures; the ancestor year should precede the foundation year of the actual HEI (which should, in turn, precede the legal status year).

d) For each variable, *timeliness* evaluates the time lag between the ETER collection date and the Source Release date. Ideally, it should be envisaged to reduce as much as possible this lag, but in the case of ETER, this might come at the price of lowering completeness, since the source may not have been released for some countries.

5.1.2 Outliers detection

An outlier can indicate an observation or processing error, or a special element of the observed population that needs to be treated differently from the bulk in the subsequent processes.
In ETER, the outlier detection phase is part of the quality validation, and has been performed with two different objectives: (i) to identify possible errors in data collected; (ii) to better understand the collected data in order to perform a proper analysis.
This phase has been performed in two different steps. In the first step, a set of quantitative indicators (variables or ratios) are identified as candidate inputs in the outlier detection process. The second step involves the usage of an outlier detection method in which the distribution of the bulk of observations is robustly estimated according to a suitable model distribution.
The method we used performs parameter estimation by regression on QQ plot positions[2].
For the purpose of outlier detection, the method assumes that the observations y_i are generated by a model probability density, with cumulative density function (cdf) $F(Y|\theta)$ where θ is a vector of parameters specifying F. The value of θ can be estimated robustly from the bulk of the observations by minimizing the sum of squares:

$$\theta = \arg\min \sum_{i\epsilon\Lambda}\left[g(y_i) - g(F^{-1}(\hat{F}|\theta))\right]^2 \quad \text{[EQ 1]}$$

Where:

- Λ indexes a subset of the observations y_i , i.e. all the observations that will be used by the fitting phase
- g is a monotonic function, differentiable on the range of Y .

[2] A Q–Q (Quantile-Quantile) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.

- $\widehat{F_i}$ are plot positions as used in QQ plots.

The method can be applied with different distributions. We used the implementation of the method proposed by the R package "extremevalues."[3] In particular we compare the empirical data with the lognormal distribution, a widely empirically validated distribution suitable for non-normal or non-symmetric variables as the micro university data are (Ruocco and Daraio, 2013). In addition, for the lognormal distribution solving [EQ1] with a suitable transformation g yields linear regression equations of the form:
$b = (A'A)^{-1}A'x$ [EQ2]
Where:

- b is a 2-dimensional vector containing functions of the distribution parameters
- A is a $|\Lambda| \times 2$ matrix containing functions of $\widehat{F_i}$ and
- x is a $|\Lambda|$dimensional vector containing functions of yi.

By solving EQ2, it is possible to estimate if the observed values are likely to be generated by the lognormal distribution or not.
Hence the method allows us to compare real observations with the estimated ones, and thus obtain a robust estimate for the bulk of the distribution.

5.1.3 External validity checks

External validity checks refer to information derived from comparing the ETER data with other available information in order to assess whether the provided data are correct. It also has the essential advantage of being able to detect issues that do not influence internal data quality.
External validity checks have been used to a limited extent until now in ETER, but this is a domain where further development could be envisaged. The external checks performed include:

- Comparison of country-level totals of students and graduates with the corresponding UOE data provided by EUROSTAT.
- Direct check of descriptors and categorical variables, for example research active or distance education HEI, through consultation of the institutional website or of sources like Wikipedia (for example concerning foundation year).
- Feedback from users of the ETER database based on their specific knowledge.

External validity checks are particularly important when other methods detect potential problems for checking whether they are justified by the specificities of the considered HEI (and thus simply reflect sample heterogeneity rather than quality problems).

5.2 Results of data quality analysis

In the following sections, we present the outcome of the data quality checks for the ETER second wave of data collection referring to the year 2012. The analysis is based on data collected and pre-validated on May 8,' 2015, and includes 27 countries: Austria, Belgium (Flemish part), Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, Spain, Slovakia, Sweden, Switzerland, and the United Kingdom.
Problems detected by the data quality analysis have been analysed by the consortium in order to ascertain the source of the problems; corrections are made in the case of mistakes, however when problems are due to issues which cannot be resolved (for example underlying differences in national data sources), they are flagged in order to alert users.

5.2.1 Format accuracy

[3]See URL: http://cran.r-project.org/web/packages/extremevalues/extremevalues.pdf

This control has been performed at the national level, on the data provided by each NSA during the preliminary validation phase. After that, it has been performed again on the overall dataset. The identified deviant cases have been either directly corrected or reported back to the national experts for checking and correction.
A few cases of format inaccuracy are still present in the dataset, namely:

- in the metadata, the date of source release and reference period/date are reported in a variety of formats, even because often they are only a proxy of the exact date;
- in quantitative variables, few financial values are negative. In a few cases (Belgium) they are connected with the classification decisions (reduction of special reserves reported as negative expenditure). In Germany several core-funding values are negative and the explanation provided in the metadata seems unsatisfactory.
  It is important to underline that in the latter case, problems found are present in official national data, and are not derived from the ETER collection process itself.

5.2.2 Completeness

The analysed dataset is fairly complete, with an overall completeness index of 78.45%, meaning that the number of missing and confidential data is just above the threshold of 20% (it includes also data reported in other rows/columns, so that the real missing data are less than 15%). The level is comparable with the previous data collection, taking into account that new variables have been introduced. There are 142,193 total missing values, articulated as follows:

- In 38.1% of the cases, both total value and breakdowns of the variable are missing (i.e. the level of total expenditures and the amount of all expenditure breakdowns);
- In 52.9% of the cases, the total value is provided but values for sub-categories are missing (i.e. the data on total income is available but the distinction between core budget, third party funding and student fees funding is not);
- In 8.1% of the cases, the unavailable subcategory data is included in another sub-category (i.e. when breakdown by field is made according to the old ISCED classification, the field of ICT is included in Natural sciences, mathematics and statistics);
- In less than 1% of the cases, missing data are actually included in another row (i.e. when there is an on-going merger process).

The level of completeness largely varies by domain and variable, as shown in the table reported below. On average it is higher for data on students and graduates, although the breakdown by mobility status is more problematic. The introduction of new standard information requested by Eurostat on this topic will improve the availability and comparability of mobility information in the coming years. Completeness is lower for financial data on income and expenditures (58% for totals and below 50% for most of categories breakdowns). The lack of availability of this information is due to the absence of standardized collection procedures at the national level in some countries, while it is expected that the administrative data on institutions' balance sheets are available. An effort in this direction could significantly improve the situation. Data on R&D expenditures is also quite poor, and is only available for 40% of research active institutions. Data on staff are in an intermediate position above the 60% threshold. Data on FTEs are slightly more complete than data in HC (even if there is variability on the estimation procedure), and there is more data on academic staff than technical and administrative staff. The information on the total number of full professors is the most complete, reaching 71%. On the contrary, the availability of academic staff breakdowns by citizenship (35%), and by field (26%, the lowest level in the dataset), are quite poor. About the latter case, in some countries, a breakdown by field is available, but it does not match with the ISCED-F classification (which indeed refers to education programmes).

The availability of descriptors is fairly complete (they have been collected directly by the consortium in several cases), with the noticeable exception of information on complementary data about the foundation (legal status year is available only or 53% of HEIs and ancestor year for 36%). These descriptors were not mandatory in the data collection.

Table 17. Completeness by variable

Variable	Completeness indicator	Variable	Completeness indicator
Legal.status	1,00	Students.enrolled.at.ISCED.6 by sex	0,89
Institution.Category.	1,00	Students.at.ISCED.6 by citizenship	0,82
Institution.Category (English)	1,00	Students.at.ISCED.6 by mobility	0,59
Institution.Category.standardized	1,00	Students.at.ISCED.6 by ISCED-F	0,83
Foreign.Campus	0,87	Total.students.enrolled.at.ISCED.6	0,94
Name.of.foreign.institution	0,87	Students.at.ISCED.7 by sex	0,88
Country.of.foreign.institution	0,87	Students.at.ISCED.7 by citizenship	0,81
Foundation.year	0,99	Students.at.ISCED.7 by mobility	0,62
Legal.status.year	0,53	Students.at.ISCED.7 by ISCED-F	0,85
Ancestor.year	*0,36*	Total.students.enrolled.at.ISCED.7	0,95
University.hospital	1,00	Students.ISCED.7.l.d. by sex	0,84
Institutional.website	1,00	Students.ISCED.7.l.d. by citizenship	0,83
Region.of.establishment..NUTS.2.	1,00	Students.ISCED.7.l.d. by mobility	0,61
Region.of.establishment..NUTS.3.	1,00	Students.ISCED.7.l.d. by ISCED-F	0,82
Name.of.the.city	1,00	Total.students ISCED.7 long.degree	0,84
Postcode	1,00	Total.students.enrolled.ISCED.5.7	0,99
Multi.site.institution	0,98	Distance.education.institution	1,00
NUTS.3.codes.of.other.campuses	0,98	Graduates.at.ISCED.5 by sex	0,90
Personnel.expenditure..NC.	0,51	Graduates.at.ISCED.5 by citizenship	0,85
Non.personnel.expenditure..NC.	*0,47*	Graduates.at.ISCED.5 by mobility	0,86
Capital.expenditure..NC.	*0,41*	Graduates.at.ISCED.5 by ISCED-F	0,96
Expenditure.unclassified..NC.	0,52	Total.graduates.at.ISCED.5	0,96
Total.expenditure..NC.	0,58	Graduates.at.ISCED.6 by sex	0,88
Accounting.system.of.capital.exp.	*0,39*	Graduates.at.ISCED.6 by citizenship	0,79
Core.budget..NC.	0,50	*Graduates.at.ISCED.6 by mobility*	*0,50*
Third.party.funding..NC.	*0,40*	Graduates.at.ISCED.6 by ISCED-F	0,89
Private.funding..NC.	*0,38*	Total.graduates.at.ISCED.6	0,93
Tuition.fees	*0,44*	Graduates.at.ISCED.7 by sex	0,87
Student.fees.funding..NC.	*0,46*	Graduates.at.ISCED.7 by citizenship	0,80
Revenue.unclassified..NC.	*0,49*	Graduates.at.ISCED.7 by mobility	0,54
Total.revenues..NC.	0,58	Graduates.at.ISCED.7 by ISCED-F	0,90
Total.academic.staff..FTE.	0,67	Total.graduates.at.ISCED.7	0,94
Academic.staff by sex.HC.	0,59	Graduates.at.ISCED.7.l.d. by sex	0,83
Academic.staff by citizenship.HC.	*0,35*	Graduates.at.ISCED.7.l.d. by citiz.	0,82
Academic.staff by ISCED-F	*0,26*	Graduates ISCED.7.l.d. by mobility	0,59
Total.academic.staff..HC.	0,63	Graduates.at.ISCED.7.l.d. by ISCED-F	0,80
Number.of.full.professors by sex	0,66	Total.graduates.at.ISCED.7.l.d.	0,83
Total.number.of.full.professors	0,71	Total.graduates.ISCED.5.7	0,97
Inclusion.of.PhD.students	0,74	Research.active.institution	0,97
Administrative.staff.FTE.	0,61	Students.at.ISCED.8 by sex	0,86
Administrative.staff.HC.	0,57	Students.at.ISCED.8 by citizenship	0,82
Total.staff..FTE.	0,66	Students.at.ISCED.8 by mobility	0,70
Total.staff..HC.	0,62	Students.at.ISCED.8 by ISCED-F	0,79
Highest.degree.delivered	0,97	Total.students.enrolled.at.ISCED.8	0,86
Lowest.degree.delivered	0,97	Graduates.at.ISCED.8 by sex	0,83
Students.at.ISCED.5 by sex	0,90	Graduates.at.ISCED.8 by citizenship	0,81
Students.at.ISCED.5 by citizenship	0,84	Graduates.at.ISCED.8 by mobility	0,68
Students.at.ISCED.5 by mobility	0,86	Graduates.at.ISCED.8 by ISCED-F	0,88
Students.at.ISCED.5 by ISCED-F	0,90	Total.graduates.at.ISCED.8	0,90
Total.students.enrolled.at.ISCED.5	0,90	*R.D.Expenditure..NC.*	*0,40*

The level of completeness also varies largely by country. In the analysed dataset there is a group of countries with a high level of completeness (over 85%) including BE, BG, CH, CY, DE, DK, ES, IE, IT, LI, MT, PT, SE, a second group with an good level (70%-

84%) including CZ, EE, FI, GR, LT, LV, NL, NO, PL, SK and a third group with poorer data availability (50%-66%) including AT, FR, HR. UK is the only country with average completeness below 50%.
Also the degree of homogeneity of completeness within the country is quite diversified in the dataset, ranging from countries providing more or less the same information for all HEIs in the perimeter (BE, BG, CH, EE, ES, FI, IE, PT, SE, SK) to countries with high internal variation (AT, DE, DK, FR, HR, GR, IT, LV, NO).

Table 18. Completeness by indicator and by country

Country	Number of HEIs included	Average completeness	Min	Max	Range
AT	68	**0,58**	0,45	0,85	0,4054
BE	28	**0,92**	0,86	0,94	0,0743
BG	52	**0,86**	0,84	0,89	0,0473
CH	35	**0,99**	0,95	1,00	0,0439
CY	25	**0,92**	0,70	1,00	0,2973
CZ	71	**0,80**	0,68	0,89	0,2094
DE	386	**0,97**	0,38	0,99	0,6081
DK	33	**0,86**	0,10	0,93	0,8311
EE	29	**0,83**	0,80	0,84	0,0439
ES	80	**0,86**	0,82	0,95	0,1283
FI	44	**0,72**	0,66	0,74	0,0777
FR	286	**0,66**	0,07	0,90	0,8210
GR	50	**0,76**	0,41	0,87	0,4628
HR	33	**0,57**	0,27	0,82	0,5473
IE	27	**0,87**	0,81	0,89	0,0777
IT	176	**0,86**	0,47	0,94	0,4696
LI	1	**0,95**	0,95	0,95	0,0000
LT	43	**0,70**	0,53	0,84	0,3040
LV	49	**0,76**	0,09	0,88	0,7804
MT	1	**0,90**	0,90	0,90	0,0000
NL	55	**0,74**	0,48	0,82	0,3412
NO	50	**0,82**	0,23	0,89	0,6520
PL	272	**0,79**	0,70	0,89	0,1892
PT	106	**0,89**	0,85	0,93	0,0743
SE	39	**0,97**	0,97	0,99	0,0169
SK	32	**0,84**	0,77	0,88	0,1014
UK	151	**0,34**	0,10	0,52	0,4222

5.2.3 Consistency

A first round of checks was performed during the data collection phase in order to check with national respondents for possible mistakes.
It has been performed again during the validation phase, through the calculation of a set of consistency indicators (see ETER handbook for further reference).
On average, data are very consistent for all countries, accordingly the data validation process has been very effective in detecting problems.

Table 19. Main data consistency issues

Country / HEI	Inconsistency	Explanation	Action
several HEIs in CH, FR	minimal inconsistencies in expenditure breakdown	rounding	none, the difference is below € 0.10
IE	inconsistency in expenditure breakdown	figures reported under heading "capital expenditures" refers to capital grants and are not included in the total	none, explained in metadata
several HEIs in BE, CH, FR, NO, PT	minimal inconsistencies in income breakdown	rounding	none, the difference is below € 0.10
BE0072	large inconsistency in income breakdown	negative unclassified has been replaced by 'x' thus sum inconsistent	none, explained in metadata
NO0030	large inconsistency in income breakdown		feedback to NSA/NE
FI, NL	inconsistency in academic staff (HC) breakdowns	breakdowns only available in FTE, not comparable with aggregate academic staff in HC	none, explained in metadata
DE	minimal inconsistencies in staff FTE breakdown	rounding	none, the difference is +/-1
UK	small inconsistencies in students and graduates breakdown	rounding	none, explained in metadata
GR	students by ISCED-F not consistent	data on ISCED 6 - F are available only for the end of academic year 2012/13	none, explained in metadata
several HEIs in DE	highest degree inconsistency	if highest degree= 1 then students/graduates ISCED7 should be 'a'	none, explained in metadata
FR0074, FR0075, FR0076	highest degree inconsistency	if highest degree= 2 then students/graduates ISCED8 should be 'a'	ISCED8 enrolled in umbrella HEI
NO0013	lowest degree inconsistency	lowest degree = 1 but 29 graduates ISCED5	none, explained in metadata
PT0048	lowest degree inconsistency	lowest degree = 2 but 24 graduates ISCED6	none, explained in metadata
several HEIs in NO	lowest degree inconsistency	lowest degree = 1 but students ISCED5 reported	none, explained in metadata

There are however a few remaining inconsistencies. Some cases are just small-scale differences due to rounding errors and are thus not relevant. There are also a number of cases with inconsistent coding of variables according to the rules defined in the ETER handbook, because of different rules and decisions taken at national level. Specifically, it happens with HEIs classified as non research active which report research expenditure (small amount), or when the lowest/highest degree level is not consistent with the presence of students/graduates. Indeed, in principle HEIs are institutions offering educational programs and awarding correspondent degrees, but in some cases HEIs enrol students while the correspondent degree is awarded by another institution with the relative power (i.e. Norwich University of the Arts in UK enrols

ISCED 8 students but degrees are awarded by the University of the Arts, London). In addition, the institutional perimeter is changing and growing in complexity in some countries in accordance with national reforms of the higher education systems (i.e. in France where community institutions have been created in recent years) making the application of ETER rules less straightforward (ie. University of Grenoble in FR is a community of three universities -FR0074, FR0075, FR0076- which offers common phd degree programs). At present these issues involve only a very small number of cases, which can be flagged in the dataset and a rethinking of the coding rules is not necessary.
Once again, problems summarised in table 19 in most cases are not due to mistakes in data collection, but point to more fundamental problems with the available data. In any case they have a low or minimal impact on data quality, referring to rounding or missing data coding rules. These cases are all flagged and explained in the notes and metadata.

5.2.4 Timeliness

Information on the source release date at the national level has been provided by most of the countries. As it could be expected, the situation is diversified among countries and within them among different domains. Only AT, BE, BG, FR, LV, PT, and SE released all of their quantitative data on the same date; but it should to be noted that these countries are still missing some variables.
Differences in release dates among countries are wide, so a trade-off should be defined when operating a European collection.
If the ETER data collection is published by mid-2015, it would have an average lag of 12-18 months with respect to individual national releases. Only a small share of data became available at the national level when the ETER collection had already begun (September-December 2014), suggesting that there was only a small window to anticipate the collection. In any case, the analysis related to timeliness is an important indication for the usage of ETER collected data. Indeed, analysis across countries must be time-aware in order to produce consistent results.

Table 20. Dates of source delivery

ETER variable	AT	BE	BG	CH	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HR	IE	IT	LI	LT	LV	MT	NL	NO	PL	PT	SE	SK	UK
Total expenditure	-	m	-	30/11/13	30/11/14	31/1/13	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	6/3/14
Personnel expenditure	-	m	-	30/11/13	30/11/14	31/1/13	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Non-personnel expenditure	-	m	-	30/11/13	30/11/14	31/1/13	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Capital expenditure	-	m	-	-	30/11/14	31/1/13	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Total income	-	m	-	30/11/13	30/11/14	31/1/13	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	6/3/14
Core budget	-	m	-	30/11/13	30/11/14	31/1/13	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Third party funding	-	m	-	30/11/13	30/11/14	-	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Private funding	-	m	-	30/11/13	30/11/14	-	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Student fees funding	-	m	-	30/11/13	30/11/14	-	1/6/14	2014	m	m	m	1/7/13	-	m	m	-	m	29/6/13	-	30/6/14	1/10/13	m	-	12/3/14	1/4/13	-	-
Number of academic staff (FTE)	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	2014	m	31/12/12	m	-	-	m	m	31/8/13	m	26/6/14	-	31/5/14	20/12/13	m	m	12/3/14	15/4/13	-	-
Number of academic staff (HC)	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	-	m	31/12/12	m	-	m	m	m	31/8/13	m	26/6/14	1/10/12	31/5/14	20/12/13	m	m	12/3/14	15/4/13	-	23/1/14
Number of full professors	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	2014	m	31/12/12	m	-	m	m	m	31/8/13	m	26/6/14	1/10/12	31/5/14	20/12/13	m	m	1/2/14	15/4/13	-	-
Number of administrative staff (FTE)	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	2014	m	31/12/12	m	-	-	m	m	31/8/13	m	26/6/14	-	31/5/14	20/12/13	m	m	12/3/14	15/4/13	-	-
Number of administrative staff (HC)	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	-	m	31/12/12	m	-	m	m	m	31/8/13	m	26/6/14	1/10/12	31/5/14	20/12/13	m	m	12/3/14	15/4/13	-	23/1/14
Total staff (FTE)	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	2014	m	31/12/12	m	-	-	m	m	31/8/13	m	26/6/14	-	31/5/14	20/12/13	m	m	12/3/14	15/4/13	-	-
Total staff (HC)	-	31/12/13	30/4/13	31/12/13	30/9/14	31/1/13	1/6/13	-	m	31/12/12	m	-	m	m	m	31/8/13	m	26/6/14	1/10/12	31/5/14	20/12/13	m	m	12/3/14	15/4/13	-	23/1/14
Number of students at ISCED levels 5, 6, 7	9/9/14		30/4/13	31/3/13	30/9/14	31/12/14	1/6/13	30/3/14	m	31/10/13	m	1/7/13	m	m	m	11/4/14	m	17/5/13	1/10/12	30/6/14	1/10/13	m	m	1/2/14	m	-	16/1/14
Number of graduates at ISCED levels 5, 6, 7	9/9/14		30/4/14	30/6/13	30/9/14	31/12/14	1/6/13	30/3/14	m	31/10/13	m	-	m	m	m	17/10/13	m	17/5/13	1/10/12	31/1/14	1/10/13	m	m	1/2/14	m	-	16/1/14
Number of students at ISCED level 8	9/9/14		30/4/13	31/3/13	30/9/14	31/12/14	1/6/13	30/3/14	m		m	1/7/13	m	m	m	28/7/14	m	17/5/13	1/10/12	31/1/14	1/7/13	m	m	1/2/14	m	-	16/1/14
Number of graduates at ISCED level 8	9/9/14		30/4/14	30/6/13	30/9/14	31/12/14	1/6/13	30/3/14	m	31/12/12	m	-	m	m	m	28/7/14	m	17/5/13	1/10/12	31/1/14	m	m	m	1/2/14	m	-	16/1/14
R&D Expenditures	-		-	30/11/13	30/9/14	-	-	26/3/14	m	-	m	-	5/11/14	m	m	-	m	20/6/13	m	30/6/14	-	m	m	-	m	-	-

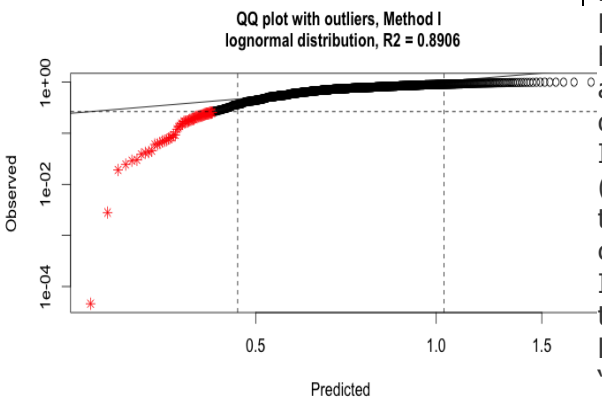
5.2.5 Outlier detection

For a selected subset of variables, a statistical analysis to highlight the presence of outliers has been applied. The distribution that generates our observations has been robustly estimated (combined in the form of ratios listed in the handbook: see the relative chapter on Data Validation and Data Quality), and "outliers" have been identified as the observations that are "unlikely" to be generated from the empirically (robustly) estimated distribution.

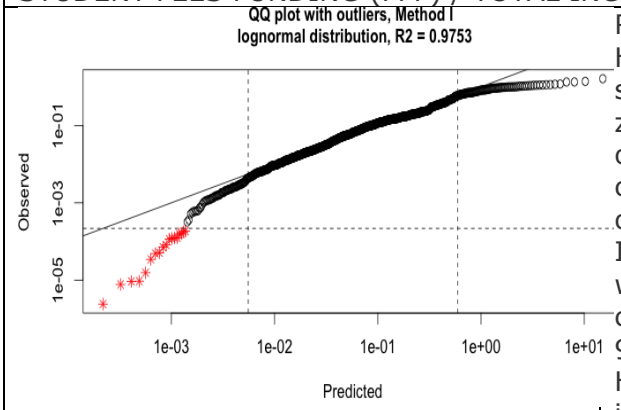
In the following table (for the most representative ratios), we report a chart showing the observed values on the y-axis and the predicted values (by a lognormal distribution) on the x-axis; moreover, the r-square of fit is reported for each indicator analysed. Overall, 35 have been controlled, involving all relevant quantitative variables contained in the ETER dataset.

It is worth it to underline that the aim of the outlier detection for data quality is the identification of "abnormal" values that may point to errors in the data collection process (i.e. figures inputted in a different unit of measure for a whole country), particularly important in ETER, given the decentralised data collection process.

Therefore, the consortium carefully examines the list of HEIs detected as outliers by country and by ratio, but only cases suspected to derive from mistakes or poor data collection quality are highlighted here. Most of the outliers detected indeed refer to HEIs with very special features (i.e. postgraduate institutions have a ratio between undergraduate and postgraduate students which will be "abnormal" but it is related to their special nature and not to problems in the data). It is not an easy process and it required the combination of statistical techniques with expert knowledge and contextualised expertise.

CORE BUDGET (PPP) / TOTAL INCOME (PPP)	
	R^2=0.8906 HEIs with an amount of core budget below 25.5% of total income are detected as outliers. 58 outliers have been detected in 8 countries (BE, DE, HR, IE, IT, LT, NO, SE) but concentrated in DE (34). In 21 cases the core budget is less than 10% of total income, and in 10 cases is even below 5%. In most cases there are issues related to the classification of income, which for a large majority, is reported under "unclassified." In DE, income category classifications probably do not match the ETER criteria (negative core budget) and the explanation in the metadata is not clear. Other cases are HEIs (mainly private) heavily relying on student fees funding or contract service activities. In IE, the share of fees is relevant for all HEIs.

STUDENT FEES FUNDING (PPP) / TOTAL INCOME (PPP)

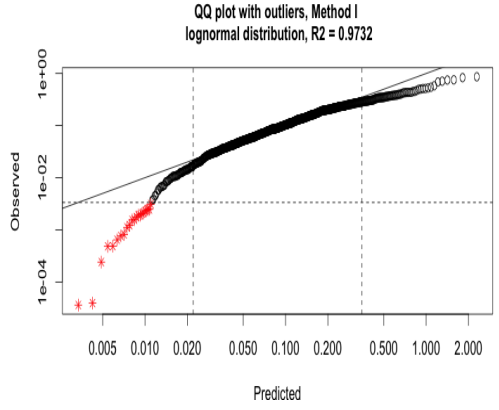


R^2=0.9753

HEIs with a marginal contribution from students (close to 0, but not equal to zero) are highlighted as outliers. Only 18 outliers have been detected in five countries (DE, FR, HR, NO, SE) but most of them are located in Germany.

In most cases, outliers are public HEIs with a very high share of core budget close to 100% (this share is often over 90% in DE). Nevertheless there are also HEIs with a large share of unclassified income in DE, pointing to an imperfect match with ETER categories. In FR, the case is an HEI that declares they have no tuition fees, while NO fees are only applied to postgraduate courses and in SE the amount of fees is always very low.

THIRD PARTY FUNDING (PPP) / TOTAL REVENUES (PPP)

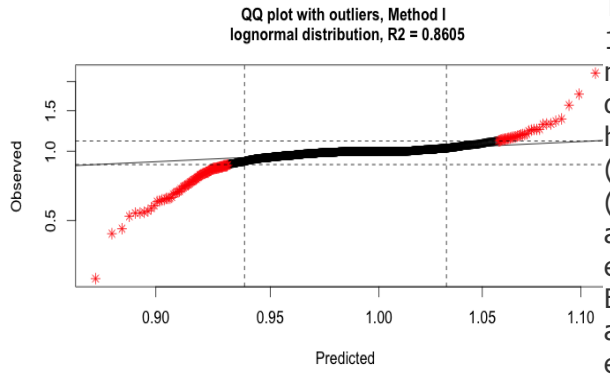


R^2=0.9732

22 HEIs with a marginal amount of funding coming from third parties (close to zero) are highlighted as outliers. Cases are distributed in 6 (CH, DE, HR, IT, LT, NO) but concentrated in DE (10).

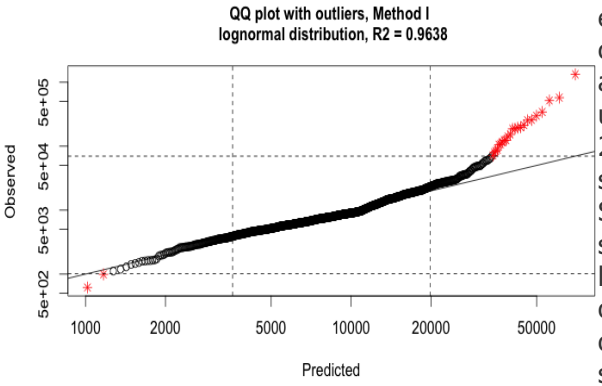
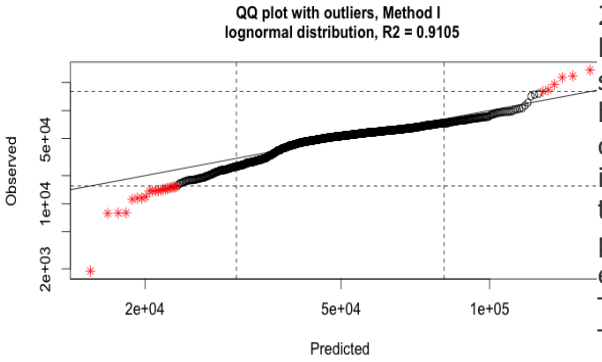
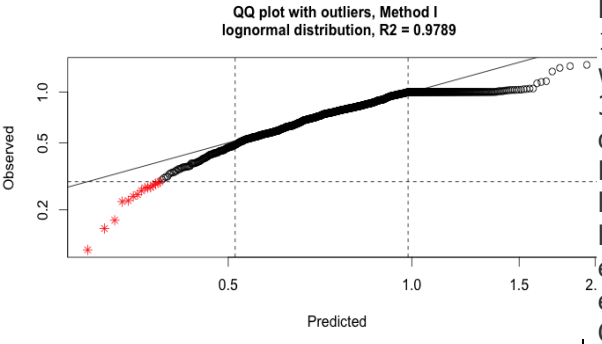
The low share of third party funding is not suspect by itself, especially when a detailed classification of income categories is not known. There are also several cases in DE but also in IT with no third party funding at all. In other countries outliers are HEIs for which there is no expectation of a large share of third-party funds (like public teaching-oriented HEIs).

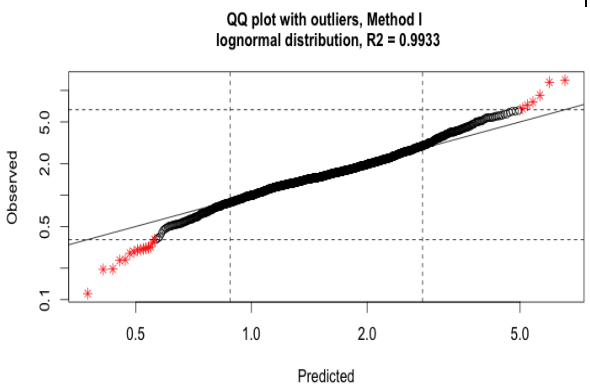
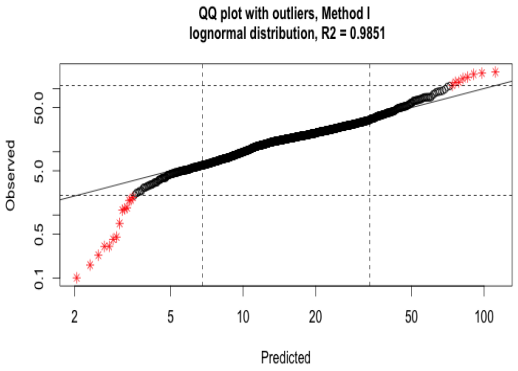
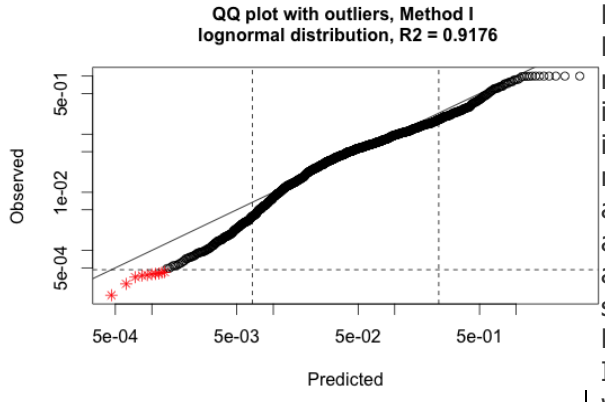
TOTAL EXPENDITURE (PPP) / TOTAL INCOME (PPP)

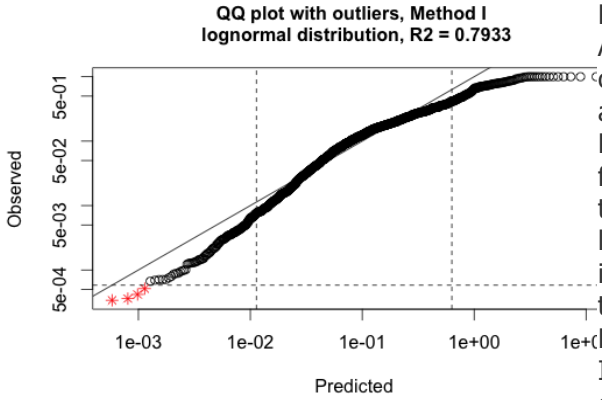


R^2=0.8605

This ratio seems to be more problematic. 122 outliers are detected, the highest number in the whole analysis. Cases are distributed in several countries with higher incidence in CH (7), CY (10), DE (24), FR (7), HR (9), IE (7), IT (18), LT (7), PT (18), UK (8). With financial accounting systems, the classification of expenditures and income required by ETER, divergence between the two aggregates are allowed and possible. The equivalence between income and expenditure does not necessarily happen in a single year, since it is possible to register income in year t and spend it in year t+x. A more accurate control would be possible with adequate time series data.

TOTAL EXPENDITURES (PPP) / TOTAL ENROLLED STUDENTS ISCED LEVELS 5-8	
	R^2=0.9638 25 cases are detected as outliers. Two are lower bound outliers with unitary expenditure below 1,000 PPP. They are closer to average and are non-PhD awarding (the Italian case is an online university) 23 upper bound outliers are distributed in seven countries (CY, DE, FR, HR, IT, NL, SE, UK) excluding country level systematic problems in the data. The level of expenditures goes from 70,000 to over 1 million PPP units per student. In one case, data on students ISCED 5-7 are simply missing, while the majority of cases refer to graduate schools or other institutions with a partial range of educational offer. There are also special HEIs, like defence colleges, theological schools, and medical campuses. Also the University of Cambridge is detected as an outlier, as it has a very high (but not abnormal) level of per capita expenditures.
PERSONNEL EXPENDITURE (PPP) / TOTAL STAFF	
	R^2=0.9105 27 outliers are detected for this ratio. Most of them are lower bound and are sufficiently clear by the metadata (CY) or by the national average in the corresponding category of HEIs (colleges in LT). The lowest case is in DE and refers to an HEI with a very low a share of personnel expenditure on total expenditure (flagged). There are also six upper bound cases. The highest level is registered for The Cyprus Institute a graduate school where the education department is only a branch.
TOTAL STAFF FTE / TOTAL STAFF HEAD COUNT	
	R^2=0.9789 15 lower bound outliers are detected, with a ratio between FTE and HC below 30%. This could be the consequence of a diffuse practice of part time jobs in some HEIs. This is the case in CH where the highest number of outliers (8) are located: UAS have a large number of external teachers with low percentage employment ratios. Other cases in other countries (DE and 1 in CZ) referring to universities of applied sciences or colleges, may have a similar explanation.

TOTAL ACADEMIC STAFF (FTE) / NUMBER OF ADMINISTRATIVE STAFF (FTE)	
	R^2=0.9933 21 outliers have been detected: 15 because of a number of academic staff below the threshold of 38% of administrative staff, and six with a high proportion of academic staff, from six to 12 times higher than technical staff. Cases are spread across six countries (BE, BG, CY, DE, IT, LT, SK). The classification of staff categories can be problematic, but no systematic large distortions emerge for whole countries. There are four cases of medium and large HEIs that are specialised in health and medical studies, while the remaining outliers are small HEIs, sometimes very small where the proportion is less meaningful. In addition, none are a traditional university model.
ENROLLED STUDENTS ISCED 5-8 / TOTAL ACADEMIC STAFF (FTE)	
	R^2=0.9851 Half of them show a very low ratio with less than 2 students per academic staff; on the contrary, in the other half of cases there are more than 100 students per staff. In most cases the explanation is directly connected with the nature of the HEI (graduate school, medical university, distance university, dance academy, police of military academy). In others there are problems with missing or estimated data.
FOREIGN STUDENTS / TOTAL ENROLLED STUDENTS – ISCED LEVEL 6	
	R^2=0.9176 Nine lower bound outliers are detected, mostly located in Poland (plus one each in BG, DE, IE). They are private institutions or typologies of schools with a national scope (i.e. the school for public administration in DE). Also in Poland, the average share of foreign students is low and there are cases with only national students. Few outlier cases are detected also for ISCED 5 level (not shown here), which were not considered suspicious.
FOREIGN STUDENTS / TOTAL ENROLLED STUDENTS – ISCED LEVEL 7	

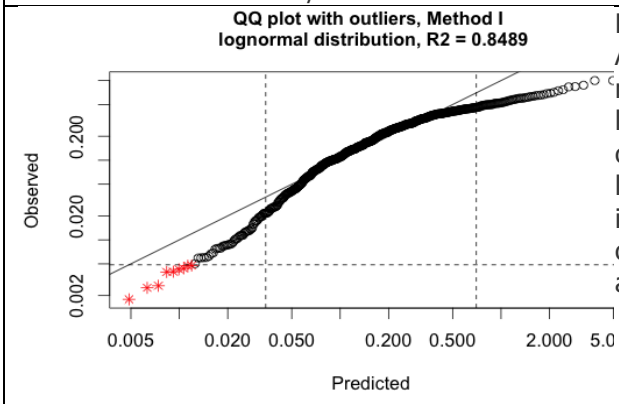


R^2=0.7933

Although the distribution only fits the observation fairly well, just nine outliers are detected, all except one located in Poland (institutions with less than five foreign students each). As said before, the share of foreign students in Poland is low at all levels and in the majority of institutions, especially non-university types.

No outliers are detected at the level ISCED 7 long degree, even if the number of observations here is lower.

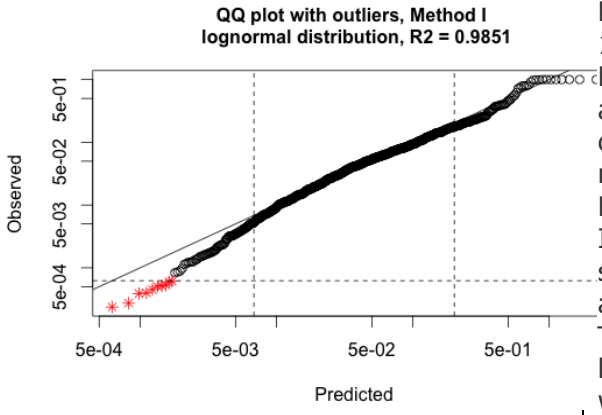
FOREIGN STUDENTS / TOTAL ENROLLED STUDENTS – ISCED LEVEL 8



R^2=0.8489

Also at the ISCED 8 level, the large majority of outliers (10 out of 11) are located in Poland, which has a low level of international attractiveness at all levels. The other case is Vilnius University in Lithuania, which has the lowest share of foreign students (but also the highest absolute number equal to five).

MOBILE STUDENTS / TOTAL ENROLLED STUDENTS – ISCED LEVEL 6

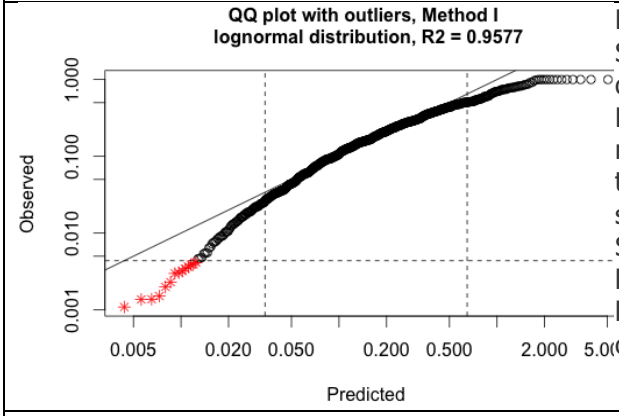


R^2=0.9851

15 lower bound outliers were detected. Half of them are located in ES where the average incidence is low. Also cases in other countries (BG, DE, EE, and IE) are not suspect, since they are in line with HEIs in the same category and country. In addition, the definition of mobile students is not the same as ETER for EE and IE.

There are also five outliers at ISCED 5 level and refer to colleges in CYPRUS where students at level 5 are mostly or totally from abroad.

MOBILE STUDENTS / TOTAL ENROLLED STUDENTS – ISCED LEVEL 7

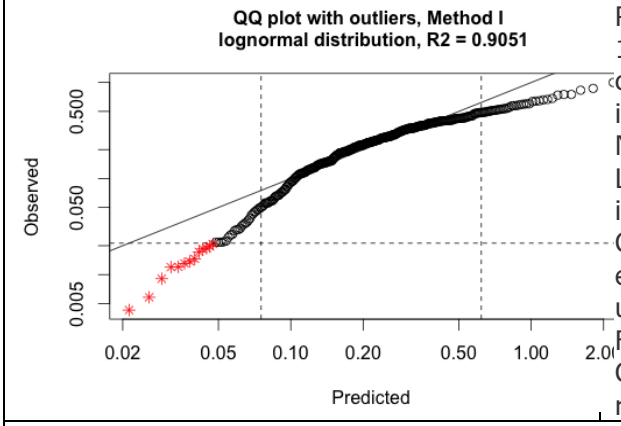
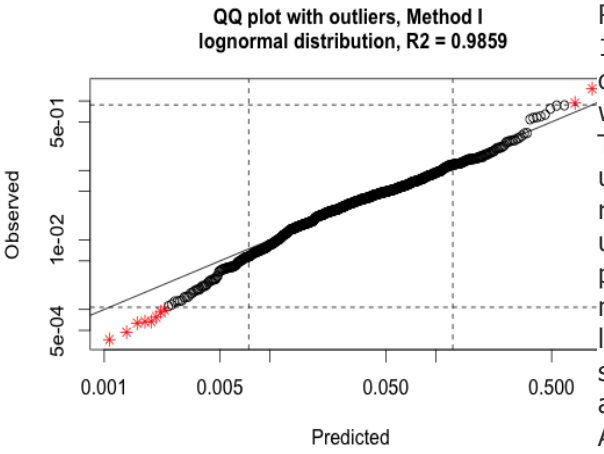
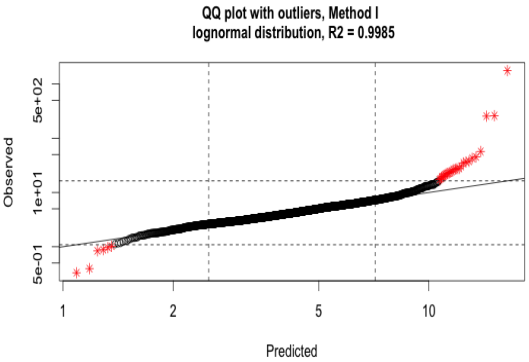


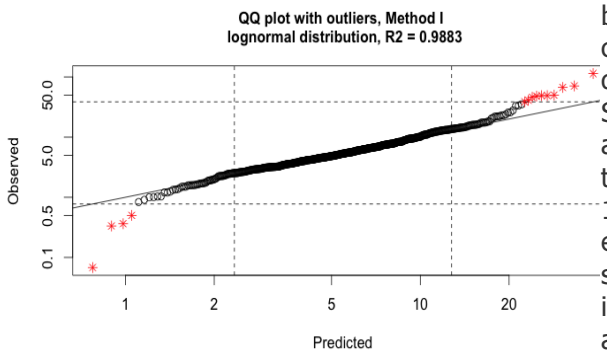
R^2=0.9577

Seven lower bound outliers are detected, distributed in five countries (BE, CY, DE, LV, and PT). They are institutions with no more than one mobile student out of the total enrolled, but the figures do not look suspect for the reasons explained above.

Similar is the situation for the ISCED 7 long degree course where, due to the lower number of observations, only one outlier is detected (ES).

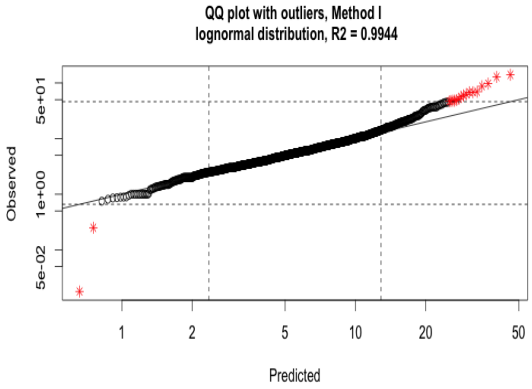
MOBILE STUDENTS / TOTAL ENROLLED STUDENTS – ISCED LEVEL 8

	R^2=0.9051 16 lower bound outliers are detected, concentrated in BG (11) and also involving FR, HR, LV, and PT. None of them looks suspect: in BG and LV the share of mobile doctoral students is always low or zero; it is also low in Croatia, with a couple of partial exceptions. The PT's case is a private university with low numbers, and the French case is the University of New Caledonia in one of the outermost regions.
GRADUATES ICED LEVEL 8 / GRADUATES ISCED LEVELS 5-7	
	R^2=0.9859 10 outliers are detected in the distribution of this ratio, which fits very well into the lognormal model. Two upper bound outliers: a graduate university in Italy with a mistake in reported figures and a very small university of theology in DE (below ETER perimeter threshold). This HEI has a number of ISCED 5-7 graduates much lower than the number of enrolled students (part of a diffused problem already highlighted). Among lower bound outliers, we find cases of HEIs with a short PhD awarding history (number of ISCED 8 students larger than graduates) and without a focus on the third cycle of higher education in CZ, in IT (online without focus on research), in NO, in SE, and the UK.
TOTAL ENROLLED STUDENTS ISCED LEVELS 5-7 / GRADUATES ISCED LEVELS 5-7	
	R^2=0.9985 Lognormal distribution fits the observations almost perfectly. Nevertheless 31 outliers are detected: six lower bound with a number of students equal or lower than the number of graduates, and 25 upper bound with the number of students 17 times higher than the graduates and more. On the contrary, upper bound outliers could be very recent HEIs where degree programmes are still not completed (AT, DE). Several cases refer to very small-specialized HEIs (theological school, arts academy, police college, young distance education universities).
TOTAL ENROLLED STUDENTS ISCED LEVELS 8 / GRADUATES ISCED LEVELS 8	
	R^2=0.9883 14 outliers are detected. Four are lower bound with fewer students than gradates,



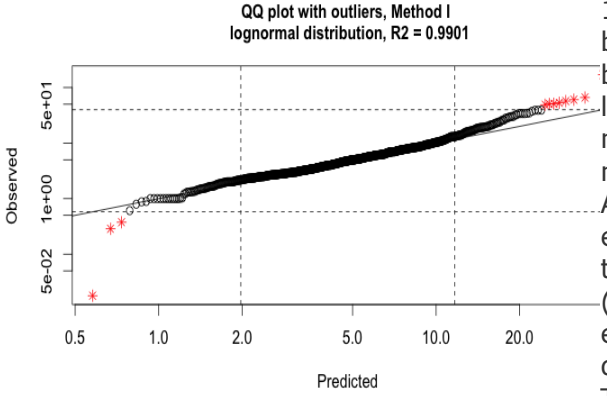
but the numbers are very small and it is clear that there is no focus on the doctoral cycle. The most extreme case is Selye János University in SK with an abnormal number of graduates compared to students and to previous year.
10 upper bound cases point to recent engagements in doctoral education by small and special categories of institutions (art academy, police academy, distance university, colleges). The number of graduates is therefore very small (one in most cases). The only exception is University of Ioannina in GR, which a low graduation rate also in 2011.

TOTAL FOREING ENROLLED STUDENTS ISCED LEVELS 5-7 / FOREIGN GRADUATES ISCED LEVELS 5-7



R^2=0.9944
Two lower bound outliers are detected, one is an HEI with a very low presence of foreign students and the ratio is quite meaningless. The second one (in PT) was checked and confirmed as correct.
16 upper bound outliers are HEIs where the number of foreign students is significantly high compared to foreign graduates. They seem to be institutions improving their international attractiveness starting from very low levels. (The number of foreign graduates ranged from 1 to 4) so even a small change in absolute numbers has a high impact on the ratio.
Similar results are obtained with the analysis at the ISCED 8 level.

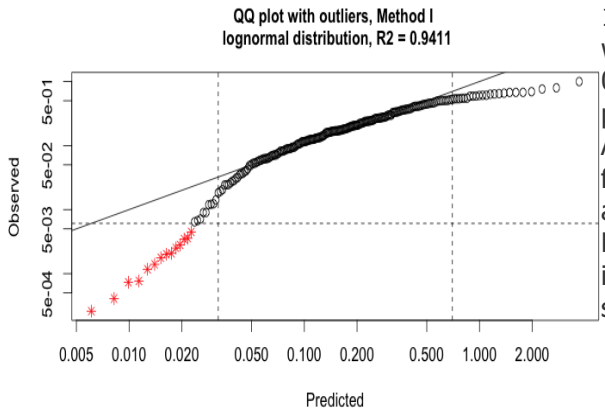
TOTAL MOBILE ENROLLED STUDENTS ISCED LEVELS 5-7 / MOBILE GRADUATES ISCED LEVELS 5-7



R^2=0.9901
11 outliers are detected: Three lower bound cases are all in PT and eight upper bound outliers in various countries. The latter are HEIs that show only one to two mobile graduates, but have increased the number of students coming from abroad. A very high ratio may suggest that exchange students have been included in the figures, but the low number of cases (spread throughout four countries) excludes systematic problems with the definition and data.
The disaggregated analysis for each ISCED level points to similar conclusions, although most outliers are always located in the same countries (DE, PT)

R&D EXPENDITURE (PPP) / TOTAL EXPENDITURE (PPP)

R^2=0.9411



14 lower bound outliers are detected, with a share of R&D expenditure below 0.5%. Cases are located in LT and PL, plus one in HR.
All LT cases refer to college model HEIs focused on undergraduate education, and are not suspect.
Polish cases refer to non-research active institutions that nevertheless report a small amount of R&D expenditure.

The results of the outlier detection process are in line with the ones obtained during the first wave of data collection, even if the number of cases has decreased. Indeed some problems highlighted in the first wave of data collection have been solved with a positive impact, also on the second phase. In addition, most problems detected in the first round of data quality in February 2015 were corrected so that the quality of data has further improved.
Most of the cases detected refer to individual institutions that are outstanding in the sample because of their specific features (category, legal status, history, presence of hospital, distance education, highest/lowest degree awarded, subject specialization, etc.). In other cases, the explanation referred to peculiar data and other caveats already flagged and explained in the notes or metadata.
Overall, the main scope of the outlier analysis is the identification of systematic problems in the data at the national level, which are not anticipated in the metadata. It was the case of mistakes in data reporting for CZ and FI, which were easily corrected during the first round. The analysis confirms that the overall quality of the dataset is good and the collaboration with national providers has been proficient.

5.2.6 Cross year comparison

The availability of the results of two data collections and, in perspective, of a time series dataset requires a further quality control of data across years. When new waves of data are to be integrated in the dataset, a quality control check will determine which of the new observations are likely to be erroneous.
Several methods and automatic software and algorithms have been introduced by specialised literature, although they usually work with aggregate statistics more than with micro-data.
At the individual institution level, the volatility of data increases exponentially, either because of reform process concerning the higher education sector in one or more countries or because of demographic events and growth processes, and evolution concerning single HEIs.
The availability of just two waves of data collection (namely 2011 and 2012) is not sufficient to test advanced statistical methods and to automatically detect possible errors on a large scale. Even when an observation seems to be an "outlier" (i.e. total enrolling students reducing by 90% from a year to the next) it is not possible to recognize which of the two observations is more likely to be incorrect.
In the next phases of ETER implementation, it will be important to develop ad-hoc methodologies to control for cross year changes by comparing the evolution over time of related variables (i.e. if HEIs staff increase sharply from a year to another also personnel expenditure is expected to increase proportionally). At the present stage, we performed a simplified check with the aim of detecting systematic errors or material mistakes (i.e. swap of columns/rows during the manual input of data in the excel data collection sheet; change of unit of measure, etc.).
We followed two different approaches:

- A systematic and complete control of descriptors that, even if collected yearly, is likely to be stable over time (i.e. the institutional category or legal status of a HEI will not change unless of a reform process which should be registered also by the 'legal status year' update). No problems emerged in this respect (the second wave data collection procedure helped to make the occurrence of these problems unlikely) with the exception of geographical information and especially postcode changes for less than 30 cases in the overall dataset. These cases were fixed.
- An overview control of quantitative variables, with specific focus on aggregate variables. Even when aggregate figures at the European or national level are stable, there is high volatility at the individual level. A change in total income/expenditure, academic staff or enrolled students by +/- 50% is more diffused than expected. This happens more frequently with small institutions and a low number of variables (i.e. if enrolled doctoral students go from 1 to 2 there is a +100% increase). Therefore sharp variations from 2011 and 2012 are not enough to consider these cases suspicious. In most cases there is a common trend at the country level (i.e. the level of income/expenditure decreased for all Italian institutions: -18% on average and it is not suspicious if some HEIs experienced a reduction by -40%). In other cases there is no clear pattern at the national level, but micro level variability remains large (i.e. enrolled students in most of observed countries). In any case no systematic problems pointing to input errors were detected.

Within the cross year comparison, we also introduced a new set of flags highlighting a break in the time series for HEIs interested by demographic events impacting their perimeter. Indeed, according to ETER perimeter guidelines in the case of a take-over (ie. incorporation of a HEI in another pre-existing one) the dominant HEI maintains its ID, but the institutional perimeter changes. Similarly in the case of a spin-off, the original HEI maintains its ID but the perimeter shrinks. These cases are already recorded in the demographic event register available on the ETER website. The flag in the dataset reminds the user to compare figures with caution and to refer to demographic events for explanations.

The following cases are flagged in the dataset:

- CZ0047 College of Information Management and Business Administration from 2012 includes also data for Central Bohemian Institute of Higher Education (take-over in 2012);
- DE0094 Technische Universität Dresden from 2012 includes also data for International Graduate School (IHI) Zittau (take-over in 2012);
- EE0003 Tallinn University from 2012 includes also data for Tallinn Pedagogical College (take-over in 26-08-2012);
- NL0044 University of Applied Sciences Leiden from 2012 includes also data for Hogeschool Helicon (take-over in January 2013);
- PL0276 Vistula University from 2012 includes also data for Academy of Finance in Warsaw (merger in 01-11-2012);
- UK0123 University College London: starting from 2012/2013 includes also data for The School of Pharmacy which became part of UK0123 in January 2012;
- UK0170 University of Wales Trinity Saint David: starting from 2012/2013 includes also data for Swansea Metropolitan University (which became part of UK0170 only in August 2013).

5.2.7 Cross country comparability

In order to highlight problems of comparability of ETER figures across countries, specific metadata have been collected together with quantitative variables. Although the degree of completeness of metadata is lower than the average level of the dataset and information is sometimes incomplete, this has proved very useful to better understand the concrete situation. Indeed metadata are an essential resource in order to understand problems highlighted by the quality control of the data, as in many

cases data problems are already explained by the providing source. In this respect, data quality and metadata analysis are complementary.
The main emerging issues are the following:

- Total expenditure is not perfectly comparable for countries that do not include capital expenditures (CH, FI, IE) or with a different definition of capital expenditures that may impact the level (NL, NO, SE). In addition, different perimeters of HEIs expenditure (i.e. inclusion/exclusion of ancillary services to students) hamper full comparability;
- The breakdown of income by categories, although not always recalled in metadata, may hide different classification choices, which can have a minor impact on the comparability of figures. The presence of large unclassified income categories hamper the comparability of other breakdowns;
- Minor specificities about the inclusion and classification of staff across countries and within countries among HEI categories (typically university vs. colleges) may impact full comparability;
- Availability of FTE and HC measures of staff are jeopardised;
- Classification of students and graduates according to the new ISCED levels has not been enforced in every country. The problem has been solved with an ad hoc concordance table for ETER with two exceptions (HR, UK student data);
- Similar consideration applies for classifications by field of education: in several countries the ISCED-97 classification was used;
- Breakdowns of students and graduates by mobility status is not fully comparable among countries;
- Information on R&D expenditures are available only in a sub set of countries.

In addition, there are a number of departures from the reference period/date required or suggested by ETER, according to the UOE manual rules. Sometime this exception also applies to aggregate reporting to UOE, while in other cases there is a discrepancy due to the different availability of data when going down to the individual micro data.
The table below summarizes departures from ETER definitions or national specifications for each variable. Cases have been flagged individually in the dataset.

Table 21. Comparability issues and departures from ETER definition

ETER variable	Main departures from ETER definition by country
Total expenditure	CH: Capital expenditures not included
	CZ: only data available at institutional level
	FI: Capital expenditures not included
	IE: Capital expenditures not included
	LT: HEIs have large discrepancies between total revenues and total expenditures - this is because those HEIs spent only a part of their revenues in financial year 2012
Personnel expenditure	CY: Professors or other personnel may be paid for their services but not included as personnel expenditure but as “other expenditure"
	IE: Total pay expenditure from funding statements 2011-12 (calendar year accounts not available)
Non-personnel expenditure	IE: Total non-pay expenditure from funding statements 2011-12 (calendar year accounts not available)
	LT: Current expenditures (as defined in UOE manual) plus financial aid to students
	NL: Non-personnel expenditure is the sum of Housing Costs
Capital expenditure	CZ: only data available at institutional level
	IE: Not expenditures but total capital grants received in 2011-12. (not included in total)
	NL: Depreciation. Investments are not included
	NO: depreciation
	SE: Depreciation
Total income	CZ: only data available at institutional level
	IE: Total Core budget + third party funding + student fees funding
Core budget	CZ: only data available at institutional level
	IE: Total state grants from funding statements 2011-12 (calendar year accounts not available)
	LT: Sum of current and investment transfers, contributes from the State and other public bodies, contributes from the private sector, revenues from commercial activities, income from assets rents and sales
Third party funding	BE: not comparable with data collection 2011 because the data which was delivered there were estimates
	IE: Total research grants and contracts, and other income from funding statements 2011-12 (calendar year accounts not available)
Private funding	IT: includes revenues from contracts with other subjects (part of third party funding) + commercial activities (core budget) + tranfers from households, public and private firms, no-profit organisations. Distinction not possible for non state HEIs
	NL: includes revenues from business, non-profit organisations, donations and sponsoring (but excludes any third party funding for which it is not immediately whether the source is private or public)
Student fees funding	IE: Total exchequer fees from funding statements 2011-12 (calendar year accounts not available)
Number of academic staff (FTE)	BE: for universities: professors, assistants, researchers, and scholarships; for university colleges: teaching and policy staff
	ES: Academic staff exclused personnel whose primary assignment is only research and graduate students employed for teaching assistance or research
	IE: Includes Exchequer and Non-Exchequer funded staff. Includes core-funded and research staff.
	IT: estimation. FT is the standard for academic staff
	NL: all staff of the universities' medical faculties has been formally transferred to the academic hospitals and therefore it is not anymore included in the staff (academic and non-academic) of the universities. This is the case for the eight universities that have academic hospitals affiliated to them Health Field is not anymore covered
	PT: data on researchers included only for 30 public institutions

Number of academic staff (HC)	BE: for universities: professors, assistants, researchers, and scholarships; for university colleges: teaching and policy staff
	ES: Academic staff exclused personnel whose primary assignment is only research and graduate students employed for teaching assistance or research
	FI: academic staff breakdowm based on person-workyears (FTE)
	IE: only Core-funded staff (perimeter narrower than FTE)
	LV: Personnel elected to academic position and visiting scholars are included
	NO: A person can only be registered once in the registry. II-positions are excluded, as only persons with an employment of minimum 40 per cent are included.
	PT: data on researchers included only for 30 public institutions
Number of full professors	IE: FTEs available only
Number of administrative staff (FTE)	IE: Includes Exchequer and Non-Exchequer funded staff. Includes core-funded and research staff
	IT: estimation (assumed equal to HC)
Number of administrative staff (HC)	GR: relevant data available only for the end of the academic year 2012-2013
	IE: Available for Exchequer funded staff
	NO: Only persons with an employment of minimum 40% are included
	UK: does not include non academic atypical staff
Total staff (FTE)	IT:estimation (assumed equal to HC)
Total staff (HC)	NO: Only persons with an employment of minimum 40 per cent are included.
	UK: does not include non academic atypical staff
Number of enrolled students at ISCED levels 5, 6, 7	AT: Not classified and therefore missing: All "Universitätslehrgänge" of public universities (including "Master-Lehrgänge") as well as all "Lehrgänge" of private universities, universities of applied sciences and universities of teacher education that are not "Master-Lehrgänge"
	BE: students are actively enrolled with a diploma contract on October 31th, 2012. Students with a credit contract are not included in these numbers. For ISCED level 6, we only included Professional Bachelors, Academich Bachelors and Bachelor after Bachelor courses. Preparatory programs (schakelprogramma's en voorbereidingsprogramma's) or teacher training programs are not included. **This is different from the UOE data collection where teacher training programs are included in the ISCED 6**.
	CZ: Number of students in headcounts, double-counting from different institutions is not avoided
	FR: Level 5 include Capacité en droit and Préparation aux concours administratifs niveau bac
	GR: relevant data available only for the end of the academic year 2012-2013 (also graduates and ISCED8)
	HR: All levels 5-7 summarized under ISCED 5
	LV: ISCED 7 long degrees included in ISCED 7 (also graduates)
	SE: individuals are counted once per higher education institution (In UOE indivduals are counted only once)
	UK: All undergraduated students listed under ISCED6; all postgraduate students (including doctorate ones) listed under ISCED7
breakdown by internationalisation/ mobility	EE: definition based on country of permanent or usual residence (also graduates)
	IE: based on domiciilary data (also graduates and ISCED8)
	LI: refers to residence at the time of immatriculation (also graduates and ISCED8)
breakdown by fields of education	CZ: according to ISCED97 (also graduates)
	EE: based on FoE 1997, field 04 included in field 03, and field 06 included in field 05 (also graduates)
	FI: according to ISCED-97 (also graduates)
	IE: based on FoE 1997. F 04 is included in F 03 and F 06 is included in F 05
	SE: based on FoE 1997. F 04 is included in F 03 and F 06 is included in F 05 (also graduates and ISCED8)

Number of graduates at ISCED levels 5, 6, 7	AT: see enrolled students
	BE: The registration of graduates for academic year 2012-2013
	EE: graduates during the academic year (1.10.2012-30.09.2013) (also ISCED8)
	PT: students enrolled in an ISCED 7 Long Degree ("Mestrado Integrado") are accounted in ISCED 7 long degree student in terms of enrollments. They are included in graduates ISCED 6 whenever they complete 180 ECTS, then they are accounted as an ISCED 7 Long Degree graduate when they finish
	SE: The FOET-columns show degrees and not graduates
breakdown by internationalisation/ mobility	BE: mobility: was not enrolled in a Flemish school in secondary education and does not have the nationality of the reporting country
Students at ISCED level 8	BE: students are actively enrolled with a diploma contract on October 31th, 2012
	DE: students cover 54% of UOE doctoral students since ETER includes matriculated students and UOE data base on sample survey
	FR: include HDR (Habilitation à Diriger des Recherches) (also graduates)
	NL: data in FTE
	NO: PhD students are counted by the number of PhD-agreements, with activity in the fall semester. The number includes research fellows (employed at the institutions), physicians that do their PhD as part of their R&D time etc.
Graduates at ISCED level 8	BE: registration of graduates for academic year 2011-2012, due on December 15th, 2012
R&D Expenditures	LT: R&D expenditure are all expenditures for R&D performed within a statistical unit during a specific period, whatever the source of funds

5.3 Implications and conclusions

Overall, the approach to data quality adopted by ETER has proved to be quite effective in managing a complex process of data collection from different countries and sources. The current version of the database reaches very good levels of internal accuracy and consistency and most of the problems encountered found an explanation and have been systematically documented. The following practices have been particularly helpful in this respect:

- First, the two-stage approach to data validation and quality, with a first set of checks being performed when data is delivered and a second set (including also statistical analyses) once data have been integrated. This two-stage approach avoided major revisions of the datasets.
- Second, the multi-method approach combining simpler internal validity checks with quality indicators, statistical testing and external checks. This makes it unlikely that problems are undetected.
- Third, the use of data flags to specifically annotate data problems and deviations from definitions is a suitable way to make users aware of the problems detected, but which cannot be corrected in the database (for example, because they are already present in the source data).
- Fourth, the systematic use of metadata to explain observed problems, which are due to the underlying characteristics of data sources. This requires in some cases an update to the collected metadata, since their quality and level of detail varies between countries.

Lessons learned and recommendations about data quality will be delivered within the final quality report accompanying the two waves of data collection.

6 Indicators in ETER

This chapter will be dealing with:

- The rationale for indicators on HEIs.
- The logic for their selection in ETER.
- How the indicators have been calculated (with reference to the ETER handbook for technical details).
- Distributional analysis of indicators in the database and their interpretation.

The indicators calculated in ETER can be seen as an added value in the project. The indicators highlight traits and characteristics in HEIs that are in many cases not available through other sources. Hence, it is likely that indicators will draw a lot of attention from final users, and will possibly inspire research and further use of the ETER database.

The variables and indicators available in ETER cover 31 countries and 2293 HEIs. The description in this chapter is based on 2012-data, but for a few countries and indicators 2011 data are included in order to maximize coverage.

6.1 General approach

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

All indicators are available on the institutional level, and it is possible to compare different types of institutions within a country or between countries. It is further possible to select a certain type of institution, either by size or by type (for example private/public or university/university of applied science) and compare this subgroup. This is what makes the ETER indicators unique, compared to official statistics from Eurostat, which are conducted on the national level.

The indicators discussed in this document are all derived from the variables present in the ETER database or the variables that will be included in the second wave of data collection in ETER. Restricted data will not be published in any form, and this implies that indicators using restricted data will be not be presented (as they are restricted).

Some of the variables stand alone as indicators, without having to be calculated. For instance the total number of students, total number of graduates etc. at an institution give an indication of the size of the HEI compared to others in the country. It also makes it possible to compare institutions of approximately the same size throughout Europe.

Most indicators used are calculated using variables available in the data set. These are indictors such as gender balance (share of women), share of foreigners, share of internationally mobile individuals and subject mix at institution/among students/graduates/staff and degree orientation of the institution. In addition, indicators such as full professor's share of total academic staff and share of researchers among staff, as well as the core budget's share of the total budget will be included among indicators.

The first wave of data collection in ETER has managed to collect data from 31 countries, but not all countries delivered all variables and breakdowns. When a variable or breakdown is missing for a country, this generally applies to all HEIs in this country. It is important that the indicator is available for a meaningful number of HEIs and countries in order to provide meaningful results.

Figure 12. Availability of varaibles in ETER by number of countres

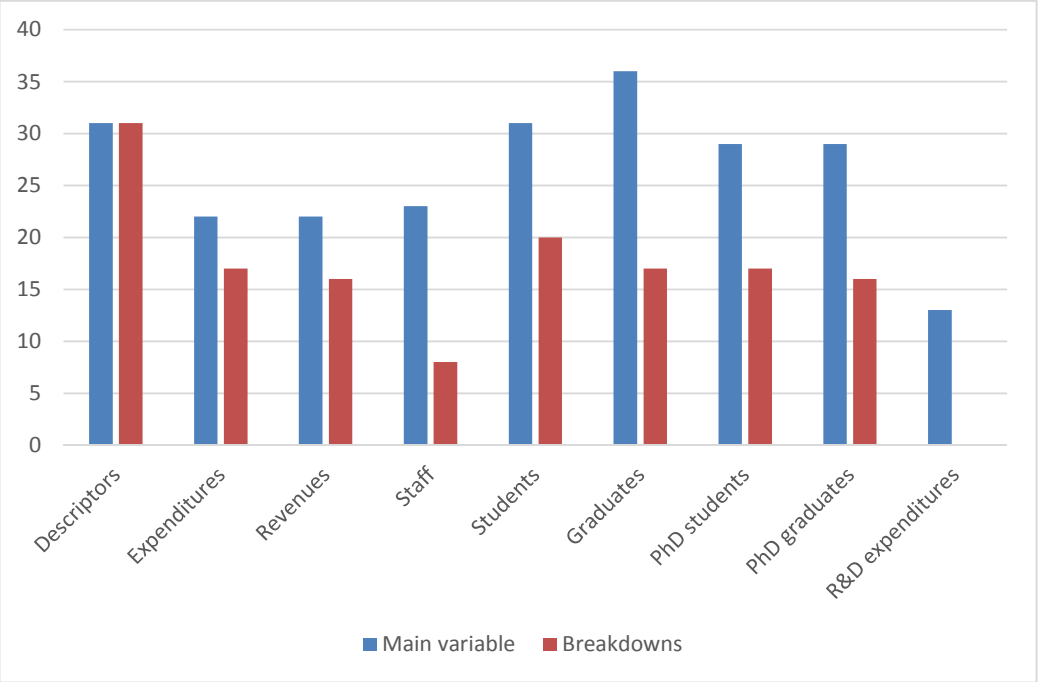
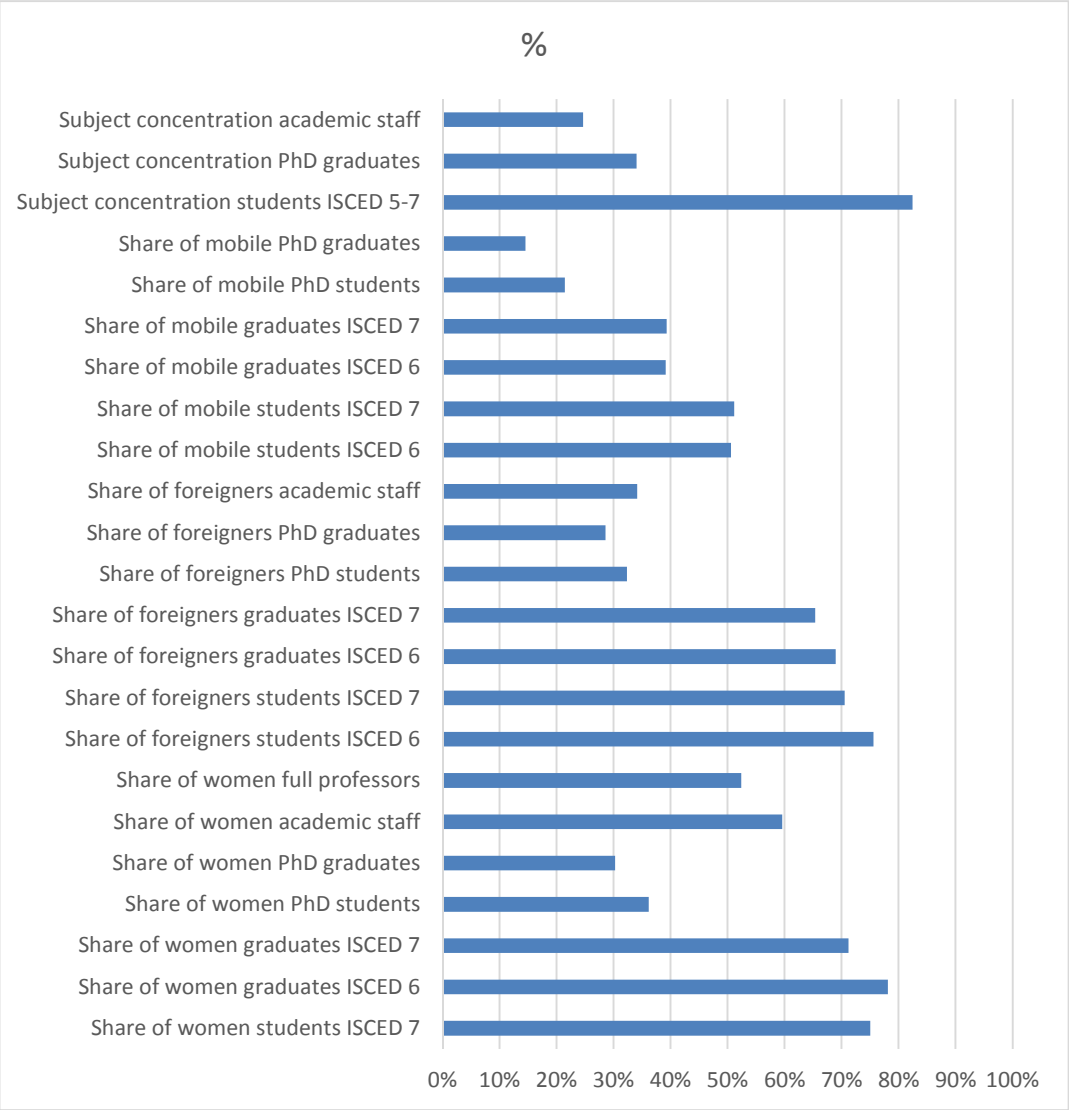


Figure 1 indicates that ETER has a good coverage of variables on students (ISCED level 5-8), graduates (ISCED level 5-8) and staff (with the exception of the breakdown of staff by field). Approximately two-thirds of the countries have delivered data on revenues and expenditures and half of them for R&D expenditure. In the initial phase of ETER we suggest only using data from the areas where the coverage is good for the calculation of indicators. This implies that we currently have few indicators related to research, and more indicators related to gender balance and teaching.

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

Figure 13. Availability of breakdowns as a percentage of the HEIs



As shown by Figure 13, when counting individual HEIs rather than countries, the situation does not change substantially: breakdowns of students by sex, citizenships and field of study are available in most cases, slightly less so for mobility. The situation is more problematic concerning breakdowns of staff data (even when aggregates are available in most cases) and for revenues and R&D expenditures.
The direct integration of indicators in the ETER database presents advantages in terms of usability, but also of robustness, since the ETER consortium holds the needed competence to deal with problematic cases. Drawing on this competence, the consortium has and will continue to discuss which indicators should and should not be included in the ETER database.

6.2 List of indicators

There is a range of indicators that are included in ETER. Most of these indicators are related to the students, graduates, PhD-students, PhD-graduates and staff in HEIs, since these are the types of variables where most countries have delivered data. The majority of these indicators are also non-controversial, as they only provide further descriptions of the institutions/countries, to facilitate comparisons.

Table 22. List of indicators in the ETER database by dimension

Dimension	Variable	Note	N.
Gender in HEIs	Students (ISCED 6 and 7)		2
(share of women)	Graduates (ISCED 6 and 7)		2
	PhD-students		1
	PhD-graduates		1
	Staff		1
	Full professors	Included in 2012	1
Nationality in HEIs	Students (ISCED 6 and 7)		2
(share of	Graduates (ISCED 6 and 7)		2
foreigners)	PhD-students		1
	PhD-graduates		1
Mobility in HEIs	Students (ISCED 6 and 7)		2
Share of	Graduates (ISCED 6 and 7)		2
mobile individuals)	PhD-students		1
	PhD-graduates		1
Subject mix	Students (ISCED 5-7)		1
(Herfindahl index)	PhD-graduates		1
	Staff	Included in 2012	1
Degree orientation	PhD-intensity		1
Staff	Full professors, as share of academic staff (HC)		1
	Academic staff (FTE), as share of total staff		1
Revenues	Core budget as share of total budget		1
	Third party funding as share of total budget		1
	Tuition fees as share of total budget		1

The numbers for students and graduates are somewhat incomplete for ISCED 5, as only 436 out of 2249 HEIs in the current ETER database graduate at level ISCED 5. Thus, we suggest creating only the indicators for students and graduates on ISCED levels 6 to 8. All indicators for the ISCED level 7 include both regular master degrees and integrated (or long) master degrees.

6.3 Calculation of indicators

6.3.1 Overall rules and special cases

For countries that have reported restricted data, the indicator using that kind of data will be set to restricted as well (special code "c"). Hence, this implies that indicators calculated using restricted data will be restricted as well and will be coded as "c" in the public databases.

In a few cases, the sum of breakdowns does not correspond to the total. These cases are flagged and explained in the database. To avoid inconsistencies, such as shares greater than 1, the number of cases at the denominator has to be derived from the

sum of breakdown values, and not from the total. Cases where the indicator is clearly unreliable because of problems in the underlying data, such as cases where HEI's core budget is negative, will also labelled as "missing" (a specific explanation will be added).
In general, values that are unclassified will not be included in the calculation of indicators. This applies for example if the gender is unknown, the share of women will be calculated as the share of women among men and women, where any unclassified person (missing) is excluded. A similar case would be where nationality is unknown, in these cases the share of foreigners will only be calculated as the share of foreigners among foreigners and nationals. However, indicators computed from breakdowns where the share of unclassified cases exceeds 25% will be flagged with "ic".
Flags related to data will also be applied to the indicators using that data.
When data used for calculating an indicator are missing (codes "m", "x" or "xc"), the indicator will be set to missing.
When data used for calculating indicators are not applicable ("a"), the indicator will be set to not applicable as well ("a).
When a total is "0", the value for the indicator referring to a breakdown will be set to "a" (not applicable). For example, this applies when an HEI has "0" students at a specific ISCED level.

6.3.2 Gender

Calculation of share of women for students, graduates, staff and full professors:

$$\text{Share of women in HEI} = \frac{\text{number of women}}{\text{number of women + men}}$$

"Unclassified" is not included in the denominator.
Share of women among students at ISCED-level 6, 7. This indicator is a calculation of share of female students in relationship to all students at that specific level for a HEI. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both the number of female students and total number of students at that level is present in the data set per HEI. Of the 29 countries that delivered data on students, 27 also had the gender breakdown for all HEIs – missing only the UK and Luxembourg. The student data with a breakdown for gender accounted for 88 per cent of all students in the ETER database. We compute the indicator separately for ISCED 6 and ISCED 7 (including both normal and log ISCED 7 degrees).
Share of women among students at ISCED-level 8 (PhD). This indicator is a calculation of the share of female PhD-students in relationship to all PhD-students at the institution. The only requirement of this indicator is that the data has been delivered for PhD-students, and that both the number of female PhD-students and total number of PhD-students is in the data set. Units with no PhD-students will display the value a. (not available). Of 27 countries that delivered data on ISCED 8 level students, all had the breakdown on gender, so the gender variable is accounted for in all countries with data on ISCED 8 level. The UK and Spain did not provide data on PhD-students.
Share of women among graduates at ISCED-level 6, 7. This indicator is a calculation of the share of female graduates in relationship to all students at that specific level for a HEI. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both the number of female students and total number of students at that level is present in the data set per HEI. Of the 29 countries that delivered data on students, 27 also had the gender breakdown for all HEIs – missing only the UK and Luxembourg. The student data with a breakdown for gender accounted for 88 per cent of all students in the ETER database. We compute the indicator separately for ISCED 6 and ISCED 7 (including both normal and log ISCED 7 degrees).

Share of women among graduates at ISCED-level 8 (PhD). This indicator is a calculation of the share of female PhD-graduates in relationship to all PhD-graduates. The only requirement of this indicator is that the data has been delivered for PhD-graduates, and that both the number of female PhD-students and total number of PhD-students is in the data set. Units with no PhD graduates will display the value "a" (not applicable). 26 countries delivered data on PhD graduates, and of these, 24 had the breakdown on gender. UK and the Netherlands only reported the total number of PhD-graduates, but did not provide the breakdown, while France, Spain and Iceland did not provide data on PhD graduates at all.
Share of female academic staff. This indicator is a calculation of the share of female academic staff (HC) in relationship to total academic staff (HC). The only requirement of this indicator is that the data has been delivered for academic staff, and that both the number of female academic staff and total academic staff is in the data set. 23 countries provided data on head count for academic staff. Of these, only the UK did not have the gender breakdown. In addition to this, Finland provided a gender breakdown based on FTE. Countries that did not report a head count for academic staff are Austria, Denmark, Estonia, France and Poland.
Share of female full professors. This indicator is a calculation of the share of female full professors in relationship to total full professors. The only requirement of this indicator is that the data has been delivered for full professors, and that both the number of female full professors and total full professors is in the data set. For 2011, 22 countries delivered data for full professors, and we assume that most of them will be able to provide the gender breakdown.

6.3.3 Share of foreigners

Calculation of the indicator:

$$\text{Share of foreigners} = \frac{\text{number of foreigners}}{\text{number of foreigners + nationals}}$$

"Unclassified" is not included in the total.
Share of foreign students at ISCED-level 5, 6, 7. This indicator is a calculation of share of foreign students in relationship to all students at that specific level. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both number of foreign students and total number of students at that level is present in the data set. Of the 29 countries that delivered data on students, 26 also had the nationality breakdown – missing Croatia, Iceland and the Netherlands. The student data with a breakdown for nationality accounted for 95 per cent of all students in the ETER database. We compute the indicator separately for ISCED 6 and ISCE7, while the indicator is not computed for ISCED 5 given the low coverage of ETER at that level.
Share of foreign students at ISCED-level 8 (PhD). This indicator is a calculation of the share of foreign students in relationship to all students at that specific level. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both the number of foreign students and total number of students at that level is in the data set. Of 27 countries that delivered data on ISCED 8 level students, 20 had the breakdown on nationality. The countries missing were Greece, Croatia, Iceland, Latvia, the Netherlands and Norway. The UK and Spain did not provide data on PhD-students.
Share of foreign graduates at ISCED-level 8 (PhD). This indicator is a calculation of share of foreign graduates in relationship to all graduates at that specific level. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both the number of foreign graduates and the total number of graduates at that level is in the data set. 26 countries delivered data on PhD graduates, and of these, 19 had the breakdown on nationality. The UK and the Netherlands reported the total number of PhD-graduates, but did not provide the

breakdown. Greece, Croatia, Latvia and Macedonia did not provide this variable. France, Spain and Iceland did not provide data on PhD graduates at all.

6.3.4 Share of mobile students and graduates

Calculation of the indicator:

$$\text{Share of mobile students} = \frac{\text{number of mobile students}}{\text{number of mobile + not mobile students}}$$

"Unclassified" is not included in the total

Share of mobile students at ISCED-level 5, 6, 7. This indicator is a calculation of share of mobile students in relationship to all students at that specific level. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both number of mobile students and total number of students at that level is in the data set. Of the 29 countries that delivered data on students, 16 also had the mobility breakdown – missing Austria, Czech Republic, Finland, Greece, Croatia, Iceland, Italy, Latvia, the Netherlands, Norway, Poland and the UK. The student data with a breakdown for mobility accounted for 48 per cent of all students in the ETER database, and is thus below the suggested threshold of 50 per cent. But as stated above, the consortium assumes that more countries will implement this variable in their reports, so that this share will probably already increase in the 2012 data set. We compute the indicator separately for ISCED 6 and ISCE7, while the indicator is not computed for ISCED 5 given the low coverage of ETER at that level.

Share of mobile students at ISCED-level 8 (PhD). This indicator is a calculation of the share of mobile students in relationship to all students at that specific level. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both number of mobile students and total number of students at that level is in the data set. Of 27 countries that delivered data on ISCED 8 level students, 15 had the breakdown on mobility. This accounted for 60 per cent of the PhD students reported in the dataset – please note that UK and Spain did not provide data on PhD-students. The countries missing were Austria, Czech Republic, Finland, Greece, Iceland, Italy, Lithuania, Macedonia, Malta, the Netherlands, Norway and Poland.

Share of mobile graduates at ISCED level 8 (PhD). This indicator is a calculation of the share of mobile graduates in relationship to all graduates at the ISCED 8 level. The only requirement of this indicator is that the data has been delivered following the ISCED-levels, and that both the number of mobile graduates and total number of graduates is in the data set. Of the 30 countries that delivered data on ISCED 8 level graduates in 2011, 13 had the breakdown on mobility (this goes for 11 of 23 countries in 2012). This accounted for 57 per cent of the PhD graduates reported in the dataset in 2011 (and 44 per cent in 2012). The countries missing the breakdown in 2011 were Austria, Bulgaria, Czech Republic, Spain, Finland, Greece, Croatia, Iceland, Italy, Lithuania, Macedonia, Malta, the Netherlands, Norway, Poland, Slovakia and the UK.

6.3.5 Subject concentration in education and graduate education

Calculation of the index:

$$Herfindal = \frac{1}{n^2} * \sum_{1}^{10} n_j^2$$

Where n_j^2 is the number of students or graduates in field j and n is the total number of students or graduates for that level within the HEI. The sum runs over the 10 fields of education. The index runs from 1, when all students are in the same field, to 0.1 when the students are equally distributed across fields in an HEI.

Subject concentration in education is computed using the distribution of the total number of students at levels ISCED 5-7 by field of education (ISCED-F).

Subject concentration for graduate education is computed using the distribution of PhD graduates (ISCED 8) by field of education (ISCED-F). Graduates are preferred to students, as data are considered more reliable. Since PhD education is closely

associated with research, this can also be considered as an indicator of subject concentration in research.
When data are computed using ISCED-1997 FOET (8 fields), a flag will be added.

6.3.6 PhD intensity

Calculation of the indicator:

$$\text{PhD Intensity} = \frac{\text{number of graduates on ISCED 8 level}}{\text{number of graduates on ISCED-levels 5, 6 and 7}}$$

PhD intensity. This indicator expresses the orientation of an HEI towards graduate education and, indirectly, also to research, as normalized to the size of its educational activities. It is a commonly used indicator of the orientation of an HEI towards research.

6.3.7 Indicators related to staff

Calculation of the indicator:

$$\text{Share of academic staff per HEI} = \frac{\text{number of academic staff (FTE)}}{\text{Total number of staff (FTE)}}$$

Share of academic staff of total staff (FTE). This indicator is calculated as the number of FTE performed by academic staff in relationship to the total number of FTEs at the institution. Hence, the indicator requires that there is data on FTEs for academic staff, administrative staff and total staff. 21 countries reported FTEs for total staff, and all of them reported FTEs for academic staff. In addition, Croatia delivered data on FTEs for academic staff. Austria, Estonia, France, Greece, Latvia, Macedonia and the UK did not provide data on FTEs for either type of staff.
Calculation of the indicator:

$$\text{Share of full professors} = \frac{\text{number of full professors (HC)}}{\text{Total number of academic staff (HC)}}$$

Share of professors among academic staff. This indicator is calculated as the number of professors in relationship to the total number of academic staff at the institution. Hence, the indicator requires that there is data both on the number of full professors, in addition to the total number of academic staff, and in both cases there should be formal definitions for who is counted in both variables. For 2011, 22 countries delivered data for full professors. Austria, Estonia, France, Ireland, Iceland, Macedonia, and the UK did not provide this variable.

6.3.8 Indicators related to revenues

Calculation of the indicator:

$$\text{Share of core budget} = \frac{\text{core budget}}{\text{total revenues}}$$

Core budget as share of total revenues. This indicator is calculated as the share of core budget funding in relationship to total revenues at the institution. Hence, the indicator requires that there is data both on core funding and total revenues. 18 of 30 countries have provided data on revenues, and of these, 15 have reported data by the breakdown of the source of funds. This means that the number of countries present is right on the threshold of half the countries. These 15 countries account for two thirds of the reported revenues.
Calculation of the indicator:

$$\text{Share of third party funding} = \frac{\text{third party funding}}{\text{total revenues}}$$

Third party funding as a share of the total budget. This indicator is calculated as the share of third party funding in relationship to total revenues at the institution. Hence, the indicator requires that there is data both on third party funding and on total revenues that are available (not restricted or missing).
Calculation of the indicator:

$$\text{Share of tuition fees} = \frac{\text{tuition fees}}{\text{total revenues}}$$

Tuition fees as share of total budget. This indicator is calculated as the share of funding from tuition fees, in relationship to total revenues at the institution. Hence, the indicator requires that there is data both on tuition fees and on total revenue available in the data set.

6.4 Descriptive statistics and discussion

This section displays the descriptive statistics for all the indicators available in ETER. All indicators are computed for individual HEIs. It is possible then to analyse the distribution of indicators' values across the whole sample or restricted to individual countries or groups of HEIs.
In this section, we provide general descriptive statistics on the whole sample of HEIs. The descriptive statistics used are mean, median, standard deviation and the range of values an indicator can take. Since all the indicators described are shares the minimum values should be 0 and 1. Therefore the calculation and display of descriptive statistics for the indicators also indirectly function as a quality check of the data, if there are inconsistencies that create values beyond the accepted range 0-1.

As displayed by

Figure 14, there are variations in the number of HEIs for which an indicator can be computed in the ETER database. Ranging from more than 1’700 HEIs (71% of the cases) for the gender breakdown at level ISCED 6 to 469 HEIs for the mobility breakdown at level 8 (20% of the cases). There are two reasons for these differences: Firstly, only half of the HEIs in ETER are entitled to award degrees at ISCED level 8. These institutions will not have indicators related to PhD students or PhD graduates. Secondly, for some HEIs the indicator cannot be computed since the corresponding breakdown is missing or restricted in the dataset. This applies particularly to financial data and to data on revenues.

Figure 14. Availability of indicators (N=2293)

6.4.1 Share of women (gender)

The dimension of gender is related to students, graduates and staff, where the categories are men, women and unclassified. Most of the countries that delivered data on students, staff and graduates provided a breakdown on gender.

The rationale for including the share of women among students, graduates and staff is that it enables us to make comparisons between similar institutions or countries that have similar HEI-systems. Historically, men, especially professors, have dominated higher education. However, during the last 20 to 30 years women are now the majority of students in most countries, so in time this pattern will probably change. This indicator can be used to compare institutions or countries, to examine how gender-equal the institutions or the higher education systems are.

Table 23. Indicators on gender. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Share of women students ISCED 6	1856	.54	.17	.00	.44	.55	.64	1.00
Share of women students ISCED 7	1719	.54	.18	.00	.43	.57	.67	1.00
Share of women graduates ISCED 6	1797	.58	.19	.00	.48	.60	.71	1.00
Share of women graduates ISCED 7	1638	.55	.20	.00	.44	.58	.68	1.00
Share of women PhD students	829	.48	.14	.00	.39	.49	.57	1.00
Share of women PhD graduates	697	.47	.18	.00	.38	.48	.56	1.00
Share of women academic staff	1364	.41	.14	.00	.31	.40	.48	1.00
Share of women full professors	1206	.22	.15	.00	.12	.20	.29	1.00

Table 1 shows that the indicators on gender are available for a large number of HEIs – the lower numbers for ISCED 8 students and graduates derive from the fact that many HEIs do not award degrees at this level. In addition, availability is lower for the gender breakdown of staff data.

The distribution of these indicators is rather smooth, with averages near 50 per cent and a moderate standard deviation for all indicators related to students and graduates. For academic staff the mean and median is around 40 per cent women, but for professors the average is only 24 per cent women. This illustrates a well-known and documented phenomenon in studies of gender composition of academic staff, that there are fewer women in top positions in academia (Bain and Cummings 2000, Vabø et. al. 2012).

6.4.2 Share of foreigners (Nationality)

ETER will currently have two indicators using information on nationality, for students and graduates. This indicator could also have been applied to staff, but for the time being only a few countries have delivered data on nationality for staff, and creating the indicator for staff is therefore postponed.

The rationale for including shares of foreigners as an indicator is that this signals if the country recruits many students from abroad or not, and if this varies with level of education. Commonly there are more foreign applicants among PhD-students compared to students at lower levels of education. This indicator can be used to compare institutions or countries, and to examine how internationalized the institutions or the higher education systems are.

Table 24. Indicators on nationality. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Share of foreigners students ISCED 6	1740	.08	.13	.00	.01	.04	.09	1.00
Share of foreigners students ISCED 7	1623	.13	.15	.00	.02	.08	.18	1.00
Share of foreigners graduates ISCED 6	1587	.07	.13	.00	.00	.03	.07	1.00
Share of foreigners graduates ISCED 7	1504	.11	.15	.00	.01	.06	.15	1.00
Share of foreigners PhD students	744	.20	.18	.00	.04	.15	.31	1.00
Share of foreigners PhD graduates	658	.21	.20	.00	.05	.16	.32	1.00
Share of foreigners academic staff	786	.07	.10	.00	.02	.04	.10	.90

Table 2 show that information on nationality is available for many HEIs. The mean and median shares of students from one country studying in other countries are low: 7-8 per cent among students and graduates at ISECD level 6. However, there are differences between levels, the higher the level of studies (ISCED-level) the higher the share of students with other nationalities. This indicates that mobility is more common at higher levels of study, and especially at the PhD-level.

6.4.3 Mobility

Mobility among students and graduates is currently reported in Eurostat. But it is a new variable, and not all countries have implemented a good system for the collection of this type of data. Few countries delivered data on mobility for 2011, but the consortium assumes that the proportion of countries able to provide this information will rise within the next few years. Even though the number of countries and students/graduates covered are below the suggested threshold, we still suggest that this indicator should be included in the ETER database.

Table 25. Indicators on mobility. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Share of mobile students ISCED 6	1164	.08	.13	.00	.01	.04	.08	1.00
Share of mobile students ISCED 7	1177	.15	.17	.00	.03	.10	.21	1.00
Share of mobile graduates ISCED 6	900	.05	.12	.00	.00	.02	.06	1.00
Share of mobile graduates ISCED 7	905	.12	.15	.00	.02	.07	.16	1.00
Share of mobile PhD students	493	.22	.18	.00	.07	.20	.34	.87
Share of mobile PhD graduates	335	.24	.20	.00	.08	.23	.34	1.00

Table 3 shows that even though not all countries delivered data that can be used to calculate mobility indicators, we still have information for about half of the HEIs present in ETER. The statistics show the same pattern as for nationality, that the share of mobile students and graduates is low for ISCED-level 6, but rises with level. Hence, also here can we observe that mobility is more common the higher the level of studies are.

6.4.4 Subject concentration in education and graduate education

Using student numbers, ETER can construct an indicator on subject composition for education, while the subject distribution of PhD graduates is used as the bases for an indicator on subject composition for research. The index is constructed as a Herfindahl index of the distribution of students by the ten fields of education and runs between 0.1 (students distributed equally by field) and 1 (all students in a single field).

Table 26. Indicators on subject concentration. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Herfindahl index students ISCED 5-7	1897	.54	.32	.13	.24	.46	.92	1.00
Herfindahl index PhD graduates	783	.46	.30	.14	.23	.33	.63	1.00
Herfindahl index academic staff	567	.48	.31	.13	.22	.35	.78	1.00

This indicator is available for about three-quarters of HEIs for education and 60% of (doctorate-awarding) HEIs. On the contrary, subject concentration of academic staff is available only for a minority of the HEIs in the sample. This indicator shows a wide dispersion, displaying that the European HE system is composed both by specialists and generalists HEIs. The lower scores for graduate education is due to the fact that many specialized HEIs do not award the doctorate.

6.4.5 PhD intensity

This indicator, inspired by the U-MAP project, characterizes HEIs in terms of their (relative) orientation to PhD degrees. Institutions that do not award degrees at ISCED level 8 will therefore not have this indicator. For the institutions that award PhD degrees, the aim is to investigate the number of graduates at ISCED level 8 compared to the total number of graduates at ISCED level 6-7. Therefore, this indicator runs from 0 to a score which can be greater than 1.

Table 27. Indicators on degree orientation. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
PhD intensity	841	.06	.10	.00	.02	.03	.06	1.50

This indicator has been calculated for a third of the institutions in the dataset, those that award PhD degrees. The mean and median values are quite low, but the standard deviation indicates that there is great variation among institutions and the database includes a few HEIs with more graduates at the PhD level than at the graduate level.

6.4.6 Indicators related to staff

In addition to gender and nationality of the staff, ETER has also collected information on other variables that can be used for calculating indicators related to staff. We suggest two indicators related to head count: full professors as a share of academic staff, and academic staff as share of total staff.

Table 28. Indicators on staff. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Academic staff / total staff (FTE)	1402	.60	.12	.08	.54	.61	.68	.93
Full professors / academic staff (HC)	977	.18	.14	.00	.08	.14	.24	1.00

Table 28 shows that the share of academic staff as a share of all staff is available for quite a few institutions, more than half of the ETER dataset. On average, academic staff constitutes just over 60 per cent of all staff at an institution, and the standard deviation is fairly low.

The share of full professors of academic staff is available for less than half of all institutions present in ETER. The mean indicates that full professors usually constitute a minor share of all academic staff, as the mean is 20 and the median is 15. Hence, the distribution is also skewed towards the left. However, the standard deviation indicates that there is great variation between institutions.

6.4.7 Indicators related to revenues

The definitions of revenues and the categories of sources of funds seem to be clearer than those related to expenditures, as we have detected fewer "special cases" in the dataset. Nevertheless, calculations of indicators related to revenues undergo the same concerns as indicators related to expenditure. There are, however, some indicators based on revenues that the consortium, in agreement with the NSA's, propose to include in the ETER database, and these are core budget as a share of total revenues, third party funding as a share of total revenues and tuition fees as a share of total revenues.

Table 29. Indicators on revenues. Descriptive statistics

	Valid N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Core budget / total budget	1107	.66	.26	-.64	.55	.73	.83	1.00
Third party budget / total budget	830	.13	.12	.00	.04	.09	.19	.86
Tuition fees / total budget	891	.15	.22	.00	.02	.06	.17	1.00

Table 7 show that indicators on revenues are available for a very limited number of institutions. Core budget as a share of total budget is only available for less than 40 per cent of institutions, while third party funds and tuition fees as part of the total budget are available for a third of institutions in the database.

Core budget as a share of total budget is on average 78 per cent, with a median of 85, indicating a slightly right skewed distribution. The standard deviation is moderately large, but the range displayed by minimum and maximum indicate great variations at the level of the institution and special cases in the data set. ETER suggests that all special cases should be removed from the display of indicators, so that the minimum and maximum values are within the range of 0-1.

Third party funding as a share of total revenues is on average 13 per cent, with a median of 9. The standard deviation is comparatively large, indicating great variation among institutions. In a few individual cases, most revenues are composed by third-party funds.

ETER suggests that all special cases, especially those generating negative values should be removed from the display of indicators, so that the minimum and maximum values are within the range of 0-1.

Tuition fees as the share of total revenues is on average 18 per cent, with a median of 6, indicating a highly left-skewed distribution. The standard deviation is large, indicating great variation among institutions. ETER suggests that all special cases, especially those generating negative values should be removed from the display of indicators, so that the minimum and maximum values are within the range of 0-1.

A few cases of German HEIs where the core budget is negative provide deviant scores of this indicator (below 0 or above 1). These cases are flagged in the database and the corresponding financial indicator has been removed from the database.

7 References

Arellano-Gault, D., Demortain, D., Rouillard, C. & Thoenig, J. (2013). Bringing Public Organization and Organizing Back In. *Organization Studies, 34(2)*, 145-167.

Bain, O. & W. Cummings. (2000). Academe´s glass ceiling: Societal, professional-organizational and institutional barriers to the career advancement of academic women. *Comparative Education Review*, Vol. 44, No. 4: 493-514.

Blickenstaff, J.C. (2005). Women and science careers: leaky pipeline or gender filter? Gender and Education, 17(4): 369-386

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

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

Bozeman, B. & Bretschneider, S. (1994). The "publicness puzzle" in organization theory: A test of alternative explanations of differences between public and private organizations. *Journal of Public Administration Research and Theory, 4(2)*, 197-224.

Daraio, C., Bonaccorsi, A., Geuna, A., Lepori, B. & et. al. (2011). The European university landscape: a micro characterization based on evidence from the Aquameth project. *Research Policy, 40(1)*, 148-164.

Deiaco, E., Holmén, M. & McKelvey, M. (2010). What does it mean conceptually that universities compete? In M. McKelvey & M. Holmén(Eds.) *Learning to compete in European Universities* (pp. 300-328). Cheltenam: Edward Elgar.

European Commission. (2012). She Figures 2012. Gender in Research and Innovation. Statistics and indicators. Available online: http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf

Fumasoli, T. & Lepori, B. (2011). Patterns of strategies in Swiss Higher Education Institutions. *Higher Education, 61(2)*, 157-178.

Jongbloed, B. (2003). Marketization in Higher Education: Clark's Triangle and the Essential Ingredients of Markets. *Higher Education Quarterly, 57(2)*, 110-135.

Jongbloed, B. (2008). Creating public-private dynamics in higher education funding. A discussion of three options. In J. Enders & B. Jongbloed(Eds.) *Public-Private Dynamics in Higher Education Funding* (pp. 113-138). Bielefeld: Transcript.

Kyvik, S. (2009). The Dynamics of Change in Higher Education Springer.

Kyvik, S. (2006). Change processes in non-university higher education in Western Europe. *Paper for the CHER 19th Annual Conference, University of Kassel, 2006*.

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

Kyvik, S. & Lepori, B. (2010). Research in the non-university higher education sector in Europe Dordrecht: Springer.

Lepori, B. & Kyvik, S. (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.

Laufer, J. (2002). "L'approche différenciée selon les sexes: Comparaison internationale", Management International, vol. 7, n° 1.

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

Lepori, B., Benninghoff, M., Jongbloed, B., Salerno, C. & Slipersaeter, S. (2007). Changing models and patterns of higher education funding: Some empirical evidence. In A. Bonaccorsi & C. Daraio(Eds.) *Universities and Strategic Knowledge Creation. Specialization and Performance in Europe* (pp. 85-111). Bodmin, Cornwall: MPG Books Limited.

Lepori, B., Dinges, M., Reale, E., Slipersaeter, S., Theves, J. & Van den Besselaar, P. (2007). Comparing the evolution of national research policies: what patterns of change? *Science and Public Policy, 34(6)*, 372-388.

Meek, V. L., Goedegebuure, Kivinen & Rinne (1996). The Mockers and Mocked: Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education.

Rossi, F. (2009). Increased Competition and Diversity in Higher Education: An Empirical Analysis of the Italian University System. *Higher Education Policy, 22(4)*, 389-413.

Theves, J. & Esterle (2005). National reports on data/indicators production. Country report France.

Vabø, A, Gunnes, H. & Tømte C. (2012). Women and men in top academic positions and management. State of the art. Paper presented at the *7th Gender Equality in Higher Education conference* in Bergen.

Wolfinger, N.H, M. A. Mason & M. Goulden. (2008). Problems in the Pipeline: Gender, Marriage and Fertility in the Ivory Tower. Journal of Higher Education, 79(4): 388-405.