RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Report on the content and technical structure of the

ETER

Infrastructure

USI: Università della Svizzera italiana

RISIS "Research infrastructure for research and innovation policy studies"

FP7, Grant agreement no: 313082

Task 6, Workpackage 6, coordinated by AIT Austrian Institute of Technology Gmbh

SEVENTH FRAMEWORK PROGRAMME

Report on the content and technical structure of the European Tertiary Education Register
ETER-RISIS

Benedetto Lepori, Università della Svizzera italiana
Michael Ploder and Daniel Wagner-Schuster, JOANNEUM RESEARCH
Hebe Gunnes, NIFU

Version 16.06.2016

Contents

List of Tables

List of figures

TABLE 1. LIST OF ABBREVIATIONS

Abbreviation	Full Name
AQUAMETH	Advanced Quantitative Methods for the Evaluation of the performance of public sector research
DG EAC	Directorate General Education and Culture
EC	European Commission
EEA	European Economic Area
EFTA	European Free Trade Agreement
ERA	European Research Area
ETER	European Tertiary Education Register
EUMIDA	European Microdata Project
EUROSTAT	European Statistical Office
FTE	Full Time Equivalents
HC	Head Counts
HEI	Higher Education Institutions
ICT	Information and Communications Technology
IMDAS	Integrated Manufacturing Data Administration System
ISCED	International Standard Classification of Educational Degrees
NE	National Experts
NIFU	Nordic Institute for Studies in Innovation, Research and Education
NSA	National Statistical Authority
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Co-operation and Development
RISIS	Research Infrastructure for the Assessment of Science, Technology and Innovation Policy
UAS	Universities of Applied Sciences
UOE	Unesco OECD Eurostat handbook on educational statistics
USI	Università della Svizzera italiana

1 Basic characteristics

1.1 Overview of the facility

The ETER-EUMIDA facility is a set of databases providing a register of European Higher Education Institutions and containing basic statistical information on them, including descriptors, geographical information, students and graduates, personnel, finances, and research activities. These databases, created by merging data from national statistical authorities, are the only available comprehensive information on European higher education, and thus are of fundamental value for analytical purposes.

The ETER-EUMIDA facility is composed by three components:

- The ETER main database including data on the years 2011, 2012 and 2013 (2014 to be collected). It is available on-line at www.eter-project.com. Data can be downloaded in different formats.
- The EUMIDA database including data on the year 2008. It is available as off-line file (SPSS format).
- A file including additional data for the years 2008, 2011 and 2012, including the numbers of scientific publications, impact, participation to European programs. It is also available off-line as SPSS data file.

The three files share the same identifier system (see below 2.2) and therefore can be easily matched and combined for purposes of analysis.

Coverage is very extensive as both databases include not only doctorate-awarding HEIs, but also second-tier HEIs in binary systems, as well as a large number of specialized schools at the tertiary level. The estimated coverage of tertiary education at the bachelor, master and PhD is around 100%. Coverage currently includes all EU-28 countries, Iceland, Norway, Liechtenstein, Serbia, Switzerland, and FYORM (some countries are missing for some years).

The two datasets are built on the same basic approach, but also display some relevant differences in coverage, data sources, methodology and especially, engineering.

This report presents the characteristics of the ETER dataset and of its infrastructure. We will also highlight differences with EUMIDA, since a major issue for RISIS is to provide joint access to both datasets and present the additional data provided together with ETER and EUMIDA for purposes of scholarly research.

This report refers to the status of ETER in July 2015, after completing the first two waves of data collection have been completed. It draws heavily on the ETER handbook (Lepori, Bonaccorsi, Daraio, et al 2015a), as well as on the accompanying report on data collection (Lepori, Bonaccorsi, Daraio, et al 2015b).

This report reflects the status of the facility at the end of June 2016. It will be subsequently updated as soon as additional releases will be available.

1.2 Context and aims

EUMIDA and ETER have to be seen in the context of the movement towards the provision of micro-data at the level of individual Higher Education Institutions, which started with the PRIME-AQUAMETH project (Bonaccorsi and Daraio 2007). PRIME-AQUAMETH was an experimental project, which proved the feasibility of collecting HEI-level data from different countries and publicly available sources. Further, the project proved that, despite existing comparability problems (Bonaccorsi, Daraio, Lepori and Slipersaeter 2007), these data could be used to provide interesting analytical insights on European higher education (Lepori, Probst and Baschung 2010, Daraio, Bonaccorsi, Geuna, Lepori and et. al. 2011).

Following these results, the European Commission (DG EAC and EUROSTAT) funded in 2010-2011 a large-scale feasibility study for a European Register of Tertiary Education. The EUropean MIcro DAta (EUMIDA) project developed a consistent methodology for the delineation of a census of HEIs in Europe and for the data collection from official sources. Further, the project managed to provide the first register of European Higher Education and to collect a large number of data (Lepori and Bonaccorsi 2013, Bonaccorsi 2014).

The European Tertiary Education Register (ETER) was launched in 2013 in order to consolidate and establish the EUMIDA methodology and to prepare for a regular data collection on European Higher Education. It is a service contract from the Directorate General of Education, Audiovisual and Culture (DG-EAC) and realized by a

consortium of four partners (Università della Svizzera italiana, JOANNEUM RESEARCH, NIFU and Università Roma La Sapienza) in cooperation with a number of national experts and national statistical authorities.

ETER builds on the results and experience of the EUMIDA (EUropean MIcroDAta collection) study, and has the following goals:

- To further develop the indicators tested in EUMIDA towards a set of more complete indicators while characterizing HEIs in the ERA by following their main activity dimensions.
- To extend the coverage of the EUMIDA dataset and consolidate the European Higher Education perimeter, i.e. the list of HEIs officially included in the dataset.
- To collect, validate, and publish the data for the defined perimeter for the years 2011 and 2012.
- To document the methodology developed for the project in a methodological handbook, while providing suggestions for the consolidation of the European Tertiary Education Register and for a regular data collection on European HEIs.

As of July 2015, the contract has completed the first two waves of data collection, referring to the year 2011 (respectively the academic year 2011/2012) and 2012 (academic year 2012/2013), finalized the data cleaning, and the dataset was made available on-line.

In July 2015, the European Commission awarded a follow-up contract for two additional years to the same consortium (plus the University of Pisa). The new contract foresees two additional waves of data collection for the year 2013 (to be published in June 2016) and 2014 (to be published in June 2017); some additional variables are foreseen, as well as a stronger focus on policy analysis and dissemination.

1.3 Legal name of the operating organization

The current ETER database is operated by JOANNEUM RESEARCH, Institute of Information and Communication Technology, Graz on behalf of the ETER consortium composed by Università della Svizzera italiana, JOANNEUM RESEARCH, NIFU and Università Roma la Sapienza. While the dataset has been collected through a service contract of the European Commission, and therefore is owned by the EU, the database system itself is proprietary to JOANNEUM RESEARCH.

1.4 Database location and type of access

The ETER database is located at JOANNEUM RESEARCH and is accessible on-line through a web application: www.eter-project.com.

Users can perform searches on the datasets and then download either the whole dataset or parts of it in an Excel format. Additionally, it is possible to download metadata and demographic information on the included HEIs, as well as methodological information on the project (methodological handbook). Public access is available for most of the data. Some data is available for research purposes only. For this data, restricted access under a non-disclosure agreement is possible.

The EUMIDA dataset and the ETER-EUMIDA additional data are available as stand-alone files, which can be provided to interested users under the signature of a non-disclosure agreement. They are available to registered users from the RISIS website.

2 Data content of ETER

2.1 Definition and description of observations

A key task of EUMIDA and ETER was to provide a clear definition of the *observation units* and of the *perimeter* to be considered, i.e. which units should be included in the register and in the data collection.

The unit of observation in ETER are Higher Education Institutions (HEIs). These are defined as entities

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

In practice, the definition of HEIs included universities (doctorate-awarding), non-university higher education (like Fachhochschulen) and specialized schools delivering education at the tertiary level. We provide below more detailed information on coverage in respect to the whole of national tertiary education.

The focus on HEIs implies that all data in ETER are aggregated at the level of the whole HEI and no subunit data are provided (even in case of multi-site organizations). The only relevant exceptions are foreign campuses, which follow standard practices in educational statistics and are treated as stand-alone organizations. ETER includes only descriptive information on the existence and localization of other education sites, but no disaggregated data at this level.

The definition of the organization is taken as adopted by national statistical authorities. In most cases, it corresponds to the legal entity, but some departures are possible for associated units and research centres. Different national practices in handling resources, staff, and students, meaning the actual definition may slightly differ by country.

A threshold has been implemented in the perimeter. This implies that institutions must have at least 200 students or 30 employees to be included in the ETER dataset. EUMIDA did not have such a threshold, and all research active HEIs, regardless of size, were included in the EUMIDA perimeter; as a matter of fact, practices in this respect differed between countries, with a few also including very small HEIs, while others were more restrictive.

2.2 Number of units and demography

ETER includes 2675 unique units referring to the year 2011, while EUMIDA includes 2471 units referring to the year 2009. However, the EUMIDA-ETER additional file includes only the 1378 HEIs in the so-called restricted set, for which enough data are available.

ETER introduced stable IDs across time for HEIs, as well as demographic notations to handle processes like births, deaths and mergers. In particular, HEIs common to EUMIDA and ETER carry the same ID and thus can be easily matched: more precisely, 1974 of 2471 HEIs in EUMIDA have been carried over in ETER. Most new cases concern countries not covered by EUMIDA. About 350 HEIs included in EUMIDA have been excluded in ETER because of not reaching the size threshold.

Importantly, no other variable is associated uniquely to IDs, but all HEI characteristics are collected and recorded by individual years, including variables such as HEI name, country, and location.

Technically, *unique records* in the ETER database are characterized by the pair HEI-ID and YEAR. Additionally, consistency across time is maintained as IDs are carried over from year to year when HEIs do not witness demographic changes.

Rules for handling demographic events correspond to the approach of the Business Units Register. For instance, in the case of mergers, the parent organizations ID's are reserved, but become non active, whereas the new organization receives a new ID. A separate record of demographic events is maintained in order to be able to track dependencies (see the ETER handbook for full reference).

The table below provides an overview of demographic events between EUMIDA and ETER. Note that changes in the name of an organization do not necessarily constitute a demographic event and imply a change of ID, which is also the case when an institution changes its legal status.

TABLE 2. OVERVIEW OF DEMOGRAPHIC EVENTS BETWEEN 2009 AND 2011

Event code	Description	N EUMIDA	N ETER	Remarks
0	no demographic event	1974	1974	
1	entry	0	557	Related to the integration of new countries in the dataset
1	entry	0	80	Entries in countries already covered by EUMIDA
2	exit	382	0	Most cases because HEIs are below the size threshold of ETER
3	birth	0	24	
4	death	36	0	
5	merger	32	13	
6	split	1	2	
7	take-over	42	21	
8	spin-out (spin-off)	0	2	Two cases of foreign colleges created by existing HEIs in the perimeter

Source: Lepori et al. (2014b).

2.3 Data acquisition and processing (e.g. data cleaning)

As a general rule, ETER data are provided from official sources, like the National Statistical Authorities and/or Research and Higher Education Ministries, meaning data should be considered as officially validated. Primary sources are in most cases administrative data collected at the national level for management and statistical purposes (like the preparation of the EUROSTAT educational statistics).

There are a few exceptions to this rule: in many cases, organizational descriptors and geographical information were provided by the consortium itself based on Internet sources. In a few countries, statistical data were also elaborated by the consortium based on data published online by the NSA. A full documentation of sources is provided in the metadata.

The process was organized in the following main steps.

a) *Definition of the perimeter*. For the countries participating in EUMIDA, the National Statistical Authorities received a pre-filled sheet including all EUMIDA HEIs in their respective countries and were requested to update the sheet by indicating changes which occurred in the period 2009-2011. A systematic notation of demographic events was provided. The perimeter also included the official name of the HEI (in the national language and English) in 2011 as reference.

b) *Data collection*. Based on the perimeter, the consortium provided master MS-Excel files for the data collection. These sheets include all variables, as well as remark fields. Moreover, metadata were collected in separate sheets for each individual variable. NSAs are also requested to provide detailed methodological explanations on some variables for which comparability is expected, specifically financial and staff data.

c) *Checking of the data*. Data sheets received were subject to a set of extensive checks for inconsistencies and problematic values by the consortium (see table 2 below). This was important in order to address issues related to underlying problems in the data, misinterpretation of guidelines, as well as simple mistakes in the handling of the data. A list of checks which have been performed can be found in the ETER handbook.

TABLE 3. DESCRIPTION OF DATA VALIDATION PROCEDURES IN ETER

Type of checks	Description	Procedure
Accuracy checks	Accuracy checks verify that data entered have the right format foreseen by the handbook (for example years as a 4-digit numeric code) and that no logically impossible values are found (foundation year>2011).	Accuracy checks are performed in the data collection sheet and on delivered data. Simple mistakes are corrected directly, whereas unclear cases are reported back to NSAs/NEs for clarification and correction.
Completeness checks	Particular attention in ETER has been devoted to the correct coding of blanks and missing values. As a general rule, ETER does not include any blank cells and the null value ("0") is reserved for the case where it is known that the corresponding value is null. All other cases are coded with special codes like missing ("m"), not applicable ("a"), included in totals ("x") or in another column ("xc") or row ("xr").	Blank cells are highlighted automatically. Clear cases are recoded directly and ambiguous cases (for example between missing and not applicable) are reported back to national experts and NSAs.
Consistency checks	a) These checks control for logical consistency between different variables, for example when the highest degree delivered is at ISCED 7 level, all values for students and graduates at ISCED8 level should be not applicable. Rules in this respect are stipulated in the handbook. b) Further, these checks also control whether the sums of breakdowns by subcategories equals the total and numerical relationships between values (example R&D expenditures lower than total expenditures).	Clear cases are recoded directly and ambiguous cases (for example between missing and not applicable) are reported back to national experts and NSAs. Remaining inconsistencies are flagged.
Deviant cases	Standard ratios are calculated (for example students to graduates) and compared to the national averages. See the handbook for the list of ratios.	Deviant values are identified and checked. In case there are specific reasons, an explanation is added to the metadata for that specific HEI.
Check of missing data	An analysis of missing data is performed (also including issues of breakdowns by subcategories).	When it is expected that data should be available, possibly with some limitations, this is requested to NE/NSAs.
Control of metadata completeness	Metadata are systematically controlled for completeness, taking into account also issues emerging from the checks on the data.	When metadata are missing or incomplete, further information is requested. Quality of metadata is critical for the exploitation of the database.
Expert checks	Expert checks based on knowledge of national systems, as well on information available on the Web and EUMIDA data, are performed in order to ensure that provided data are realistic.	Potentially problematic cases are notified back to national experts and NSAs. When these are related to methodological issues, the corresponding remarks are integrated into the metadata.

2.4 Cleaning of the EUMIDA datasets

The original dataset for the year 2008 produced by the EUMIDA project displayed a number of differences in respect to ETER. The most important ones are:

- The list of variables is partially different, not only concerning their names, but also less variables were included especially for what concerns descriptors and geographical information.
- Some classifications differed, as EUMIDA used the old ISCED-1997 classification, as well as FOE-2007 for the fields of education.
- No systematic coding of missing data was available, they were included as blank cells in the database.
- No flags and less remarks were included in the dataset.

- The documentation of the dataset and metadata is less systematic and less usable, as it was provided in the form of a large word file.

The cleaning of the dataset had the goal of making EUMIDA as similar as possible to ETER in order to allow users to combine in a smooth way the two datasets for the purposes of longitudinal analysis. Following main changes were made:

- Descriptors and geographical information was systematically compared between ETER and EUMIDA and inconsistent cases were cleaned. To the extent of possible, the additional variables introduced in ETER have been integrated in EUMIDA as well (for example postcode and geographical coordinates).
- All EUMIDA variables have been renamed identically to ETER. Variable codes and labels have also been made uniform.
- Blank cells have been coded to ETER standard codes.
- Classifications have been updated. For example ISCED-F-2011 is adopted (the two missing fields not included in FOE-1997 consistently recoded as "xc").
- Flags and remarks columns have been introduced in order to be able to track any specific issues.
- Accuracy and consistency checks have been run.

The revised EUMIDA file will be available in SPSS format, accompanied by the original metadata file and by a file specifying changes which have been made.

2.5 Information all variables/indicators

Both ETER and EUMIDA include the following groups of variables:

- Institutional descriptors, e.g. the name of the institution and the foundation year.
- Geographical descriptors like the NUTS region, the city of the main seat and its postcode.
- Data on numbers of students and graduates divided by ISCED-2011 level, by gender, fields of education, citizenship and mobility.
- Data on HEI expenditures and revenues.
- Data on the number of staff, divided between academic and non-academic, available both as headcount (HC) and full-time-equivalents (FTE), as well as on the number of professors (HC).
- Data on research activities (PhD students, R&D expenditures).

Differences between EUMIDA and ETER are as follows. ETER includes a few additional variables, especially concerning geographical information and staff. Moreover, in EUMIDA, many variables and breakdowns are only available for the so-called research-active HEIs. These institutions comprise about half of the EUMIDA sample, but most of the students and research activities. Finally, ETER adopts the more recent release of classifications of educational levels (ISCED-2011) and educational fields (ISCED-F).

To the extent of possible, the current version of EUMIDA has been updated and extended to match ETER definitions (see details below). Variables in ETER and EUMIDA also have identical labels in order to make a merge of both datasets easier.

For full reference on the revision and cleaning of the EUMIDA dataset the reader is requested to read the corresponding document.

2.5.1 List of variables

The table below provides a full list of ETER variables, with additional information on their availability in EUMIDA. For complete information and definitions the reader should refer to the ETER handbook. In general, most definitions concerning students, graduates, staff and finances are derived from the UOE data collection handbook (UOE 2006), respectively with the OECD 2002 Frascati manual for R&D expenditures and thus comply with official statistics at EUROSTAT and OECD.

TABLE 4. FULL LIST OF ETER VARIABLES

Dimension	Variables	Differences in EUMIDA
Identifiers	ETER ID Country code National identifier (optional) Institution name (in own language) English institution name (if available) Year	Same variables in EUMIDA.
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	The same variables are also included in EUMIDA.
Geographic information	Region of establishment, NUTS2 code Region of establishment, NUTS3 code Name of the city Postcode Multi-site institution	The same variables are also included in EUMIDA.
Educational activities	Highest degree delivered Number of enrolled students at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility Number of graduates at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility Distance education institution	Number of students: only aggregated values for ISCED5-7 are available (no breakdown between levels 5,6 and 7). The 1997 FOE classification is used, hence ISCED-F04 included in ISCEDF03 and ISCED-F06 included in ISCED-F05.
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 expenditure	The 1997 FOE classification is used, hence ISCED-F04 included in ISCEDF03 and ISCED-F06 included in ISCED-F05.
Expenditures	Personnel expenditure Non-personnel expenditure Capital expenditure Accounting of capital expenditure	
Revenues	Core budget Third party funding Private funding Tuition fees	

	Student fees funding	
Staff	Number of academic staff in FTE Number of academic staff in headcounts Number of academic staff, by fields of education gender and citizenship in headcounts Number of administrative staff in FTE Number of administrative staff in headcounts Number of full professors (HC) Inclusion of PhD students Number of total staff in FTE Number of total staff in HC	Staff data only in FTEs. No data by fields and on numbers of professors. No data on gender.
Indicators	Share of women among students ISCED6 and ISCED7. Share of women among graduates ISCED6 and ISCED7. Share of women among PhD students and graduates (ISCED8). Share of women among academic staff and full professors. Share of foreigners among students ISCED6 and ISCED7. Share of foreigners among graduates ISCED6 and ISCED7. Share of foreigners among PhD students and graduates (ISCED8). Share of mobile among students ISCED6 and ISCED7. Share of mobile among graduates ISCED6 and ISCED7. Share of mobile among PhD students and graduates (ISCED8). Subject concentration of undergraduate students (ISCED5-7). Subject concentration of PhD graduates (ISCED8). Subject concentration of academic staff. PhD intensity Full professors as share of academic staff (headcounts). Academic staff as share of total staff (headcounts). Core budget as a share of total budget. Third-party funds as a share of total budget. Students' revenues as a share of total budget.	Not currently available in EUMIDA
Erasmus mobility data	Number of incoming Erasmus students Number of outgoing Erasmus students	Not currently available in EUMIDA

2.5.2 Additional variables

The additional dataset for ETER-EUMIDA includes variables not covered by the core dataset, but relevant for purposes of analysis of higher education. The current set of variables includes:

- The number of participations of EU Framework Programmes derived from the EUPRO database.
- The number of publications of HEIs derived from the Leiden rankings.
- A correspondence table with the OECD/EUROSTAT functional urban areas, which allows matching ETER with data on regional development (http://www.oecd.org/regional/redefiningurbananewwaytomeasuremetropolitanareas.htm).

This set of variables might be expanded in the future. For full reference on the additional dataset the reader should consult the corresponding document with detailed description of variables and sources.

2.6 Temporal, geographical and sectoral coverage

2.6.1 Temporal coverage

EUMIDA-ETER data are in principle collected for every year. EUMIDA provides data for the year 2008, whereas ETER for the year 2011 and 2012 and 2013. Data for the year 2014 will be available in summer 2017, respectively 2017.

Depending on the nature of the data and the practices of data collection, individual data refer to slightly different periods as detailed in Table 5. Departures from these reference periods for individual countries are recorded in the metadata.

TABLE 5. REFERENCE PERIODS

Variable	Reference period/date
Descriptors and geographical information	Last day of calendar year (31st of December).
Expenditures	Calendar year (1st January – 31st of December).
Revenues	Calendar year (1st January – 31st of December).
Personnel FTE	Calendar year based on person-years.
Personnel headcount	End of first month of beginning of academic year.
Students	End of first month of beginning of academic year.
Degrees (including PhD degrees)	Academic year or calendar year (to be specified).

2.6.2 Geographical coverage

ETER covers all 28 European Union member states, EEA-EFTA countries (Iceland, Liechtenstein, Norway and Switzerland), as well as candidate countries (the Former Yugoslav Republic of Macedonia, Montenegro, Serbia and Turkey), for a total of 36 countries.

Data for the following countries are currently missing:

EUMIDA 2008: France, Iceland, Liechtenstein, the Former Yugoslav Republic of Macedonia, Montenegro, Serbia and Turkey.

ETER 2011: Belgium (French part), Romania, Slovenia, Montenegro and Turkey.

ETER 2012: Belgium (French part), Romania, Slovenia, the Former Yugoslav Republic of Macedonia, Liechtenstein, Montenegro, Turkey, Iceland and Hungary.

ETER 2012: Belgium (French part), Romania, Slovenia, the Former Yugoslav Republic of Macedonia, Liechtenstein, Montenegro, Slovakia, Turkey, Iceland and Hungary.

2.6.3 Sectorial coverage

ETER aims to provide a fairly complete coverage of the European Higher Education sector (tertiary education), as defined by graduation of at least the ISCED-2011 level 5 (diploma at the tertiary level). The criterion that tertiary education should be a major activity excludes professional organizations (delivering only some curricula), as well as public research organizations (even if they employ PhD students as researchers).

Almost all ETER organizations belong to the higher education sector. There are a few cases of research organizations included in ETER either because they award a large number of PhD degrees or because they are associated with a university.

In general, the ETER coverage matches fairly well the official list of HEIs considered as part of the higher education system at the national level, with the exclusion of very small HEIs. As previously mentioned, a threshold has been implemented; meaning institutions must have at least 200 students and 30 staff (FTE) to be included in the ETER dataset. Coverage at the bachelor, master and PhD level in ETER should be considered as quite complete. The same applies for research activities in the higher education sector. In respect to all tertiary

education (graduating from ISCED level 5 onwards), ETER covers educational providers delivering short diplomas of less than 3 years only to a limited extent (ISCED level 5), like professional schools in the vocational sector or preparatory classes before university studies. In terms of the number of students, ETER includes 85% of the tertiary education students in the countries that delivered data, respectively 65% in the countries that provided information about the perimeter, i.e. the number of higher education institutions.

2.7 Classifications used in the database

2.7.1 Type of HEI

This variable specifies a European-level standardized classification of Higher Education Institutions. It is relevant in order to provide comparative analysis of higher education systems and analyze subgroups. It is available in ETER only.

The following categories are used:

- UNI (university). These HEIs display a largely academic orientation (without excluding some focus on applied research), they have the right to award doctorate degrees, and can bear the full name of "University" (including variants like technological university, etc.). In general, awarding doctorates should be the main criterion to classify HEIs in this category, even if a few doctoral-awarding HEIs might be included in the two following categories.
- UAS (university of applied sciences). These institutions are officially recognized as a part of higher education, though not as universities (see definition above). Commonly these institutions have a focus on professional education. In most cases they do not have the right to award a doctorate (exceptions are possible). Examples are Fachhochschulen (Austria, Germany), Hogescholen (Netherlands), colleges (Norway), and Polytechnics (Finland). This institutional category applies only to countries that have a binary HE system, where these institutions are given a specific legal status. Examples include Norway, Switzerland and the Netherlands.
- Other. All institutions that do not fit the description of university/university of applied science will be categorized as "other." This may apply to institutions such as art academies and military schools. Also technological and professional schools in countries without a binary system (like the UK or France) should be classified in this way.

2.7.2 Legal status

Consistent with the UOE manual, a classification of HEIs by legal status is provided both in ETER and EUMIDA.

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

Private institutions are further divided between *government dependent* – which either receives more that 50% of their core funding from government agencies or whose teaching staff is paid by a government agency – and *independent private*.

Thus, this classification includes three categories: public, private, private government-dependent.

2.7.3 Geographical classification

ETER provides extensive geographical information on HEIs. For each HEI, the following geographical data is included:

- Country code.
- NUTS 2 and NUTS 3 code of the main seat (only NUTS included in EUMIDA).
- The name of the city and the postcode of the legal seat (not included in EUMIDA).
- Geographical coordinates (derived from the postcode) are foreseen in one of the next releases (not included in EUMIDA).

Additional information is provided for multi-site institutions, i.e. HEIs having establishments in different cities. This includes additional NUTS codes and descriptive information on the locations.

2.7.4 Levels of education

The International Standard Classification of Educational Degrees (ISCED)[1] provides information on the level of curricula and is therefore used to break down numbers of students and graduates. ETER adopts the more detailed ISCED-2011 classification, which includes separate classifications for bachelor (ISCED6), master (ISCED7) and PhD (ISCED8). The classification scheme used in ETER singles out the so-called long degrees (ISCED7 long), i.e. 4-5 year curricula without an intermediate classification. EUMIDA used a simpler distinction between diploma, bachelor, master and PhD, which can, in principle, be matched to ISCED-2011.

The table below provides detailed information on the classification used in ETER.

TABLE 6. ISCED-2011 CLASSIFICATION

ISCED-2011 level	Definition	Criteria
ISCED 5 short-cycle tertiary education	Programs at ISCED level 5, or short-cycle tertiary education, are often designed to provide participants with professional knowledge, skills and competencies. Typically, they are practically based, occupationally specific and prepare students to enter the labor market. However, these programs may also provide a pathway to other tertiary education programs. Academic tertiary education programs below the level of a Bachelor's program or equivalent are also classified as ISCED level 5.	Duration: 2-3 years Entry requirements: ISCED 3 or 4
ISCED 6 Bachelor's or equivalent levels	Programs at ISCED level 6, bachelor's or equivalent level, are often designed to provide participants with intermediate academic and/or professional knowledge, skills and competencies, leading to a first degree or equivalent qualification. Programs at this level are typically theoretically-based but may include practical components and are informed by state of the art research and/or best professional practice. They are traditionally offered by universities and equivalent tertiary educational institutions.	Duration: 2-3 years Entry requirements: ISCED 3 or 4 Usually: first degree at tertiary level
ISCED 7 Master of equivalent level	Programs at ISCED level 7, master's or equivalent level, are often designed to provide participants with advanced academic and/or professional knowledge, skills and competencies, leading to a second degree or equivalent qualification. Programs at this level may have a substantial research component but do not yet lead to the award of a doctoral qualification. Typically, programs at this level are theoretically-based but may include practical components and are informed by state of the art research and/or best professional practice. They are traditionally offered by universities and other tertiary educational institutions.	Duration: 2-3 years Entry requirements: ISCED 6 Usually: second degree at tertiary level Direct access to ISCED 8 level
ISCED 7long Master or equivalent level long degrees	Long first degree program at a master's or equivalent level with a cumulative theoretical duration (at the tertiary level) of at least five years (that does not require prior tertiary education). These programs will be singled out in ETER when possible, given their different characteristics and their impact on the number of diplomas.	Duration: at least 5 years Entry requirements: ISCED 3 or ISCED 4 Usually: first degree at tertiary level Direct access to ISCED 8 level
ISCED 8 Doctoral or Equivalent level	Programs at ISCED level 8, or doctoral or equivalent level, are designed primarily to lead to an advanced research qualification. Programs at this ISCED level are devoted to advanced study and original research and are typically offered only by research-oriented tertiary educational institutions such as universities. Doctoral	Duration: at least 3 years Entry requirements: ISCED 7 Research-based

[1] ISCED: http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

	programs exist in both academic and professional fields.	programs (not only courses).

2.7.5 Fields of education

The Fields of Education and Training classification allows breaking down numbers of students and graduates by field of study. It is envisaged to introduce this classification also for academic staff.

ETER adopts the most recent version of the classification, i.e. the Fields of Education and Training 2013 classification[2]. EUMIDA used the previous Fields of Education classification 1997: the main difference between the two classification schemes is that in ISCED-F 2013 two new fields have been distinguished, namely Business, Administration and Law (included in social sciences in the previous schemes) and ICT (included in natural sciences in FOE-1997).

For students and graduates, the breakdown by field of education is provided separately by level of education.

TABLE 7. FIELDS OF EDUCATION AND TRAINING CLASSIFICATION

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

[2] ISCED-F; http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

			84 Transport services

2.7.6 Citizenship and mobility

ETER includes classification of students, graduates and staff by:

- Citizenship, distinguishing between nationals and foreigners.
- Mobility (students and graduates only), distinguishing between residents (nationals and foreigners who earned their qualifying education in the country) and mobile students.

Both classifications are compliant with standard EUROSTAT definitions from the UOE manual (UOE 2006).

These breakdowns are provided separately by ISCED level (but not by field of education).

3 Quality and accuracy of data

ETER data have been subject to an extensive data validation and quality control procedure, which has been coordinated by the University of Rome La Sapienza together with the other consortium partners. Since ETER has no control over the primary sources, much of the data quality process is concerned with documenting methodological departures and comparability problems in order to make the users aware of potential problems.

EUMIDA data have been subject to some data quality controls, but in a somewhat less systematic way. Documentation of methodological and comparability issues is provided in an accompanying document, but currently not systematically linked to the data themselves.

3.1 Overall 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 (see also chapter 2), and

2. a 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 and improve the quality of the collection in the second wave. In this phase the quality control is performed on data at both a "global" and "local" level.

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* by estimation of the distribution of observations according to a model distribution (mostly a lognormal distribution) and by identification of extreme values which do not fit into this distribution (see the ETER handbook and Ruocco and Daraio 2013, for full details).
- Checks of external validity through comparisons with other data sources or with information available from the Internet; this was particularly useful in order to explain observed inconsistencies and deviant cases, which were due to specific characteristics of the considered HEI.

3.2 Accuracy and consistency

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, and student and graduate data being integer variables, among others. This also includes the correct coding of missing and null values.

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 expenditure 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 expenditure should be lower than total expenditures. The ancestor year should precede the foundation year of the actual HEI (which, according to the definition in ETER, should precede or be the same as the legal status year).

These dimensions have been systematically checked during the data collection process. A final check on the complete dataset has been performed before publication.

Overall, the current version of the ETER dataset reaches a very good level of accuracy and consistency. Very few remaining cases are due to national specificities or simply to rounding errors. These have been flagged in the dataset.

3.3 Completeness

Completeness of data strongly varies by variable and by country. Overall, we can summarize the situation as follows:

- 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 only about one-third of the countries in 2011.
- Staff data are generally available in most countries. The main exceptions are countries which provided for the time being only the descriptors and no statistical information. However, the breakdown of academic staff between national and foreign citizenship is available for a much smaller number of countries.
- Students and graduates data are available for most countries, including breakdowns by gender, nationality, and fields of education. The breakdown by mobile students is less widely available. Since this is a breakdown requested by EUROSTAT, but not yet implemented in all countries, the number of countries providing this breakdown is believed to increase.
- The situation is similar for PhD students. Some major countries, such as Spain and the UK, did not deliver data on PhD-students at all.

The level of completeness varies largely by country. In the analyzed dataset there is a group of countries with a very high level of completeness (over 85%) including CH, CY, DE, DK, IE, IT, LI, LU, MT, NO, SE, a second group with an acceptable level (50%-85%) including AT, BG, CZ, EE, ES, FI, GR, IS, LT, LV, MK, NL, PL, PT, and a third group with minor data availability (below 50%) including BE (for French part only descriptors are available), FR, HR, UK.

3.4 Ratios and outlier identification

A central component of the data quality process in ETER was a careful analysis of ratios between variables. This involved two main types of checks:

a) whether ratios are in a reasonable range which could be expected from the process considered (for example whether ratio between students and graduates is near the usual length of the curriculum), and

b) whether ratios for individual HEIs and for whole countries are within a reasonable range in the overall distribution. The latter analysis was performed by comparing the distribution of ratios with hypothetic distributions (notably, lognormal) in order to ascertain whether observations are “unlikely” to be generated from the empirically (robustly) estimated distribution.

This analysis identified a number of deviant cases, which were then checked manually. Most cases could be explained by specificities of the observed HEI and have been documented within the dataset. Other cases revealed mistakes which were corrected in agreement with NSAs.

3.5 Metadata and comparability problems

In order to highlight problems of comparability between 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, metadata are an essential resource in order to understand problems highlighted by the quality control of the data, as in many cases revealed data problems are already explained by the providing NSA. In this respect, data quality and metadata analysis are complementary.

Some emerging issues revealed by the metadata are the following:

- Total expenditure is not perfectly comparable for countries which do not include capital expenditures or with a definition of capital expenditures different from the others.
- The breakdown of revenues by categories, although not always recalled in metadata, may hide different classification choices which can have a minor impact on the comparability of figures.
- Minor specificities about inclusion and classification of staff across countries and within countries among HEI categories (typically universities vs. colleges) may impact full comparability.
- Staff data are available in some countries only in headcounts, in others in Full-Time-Equivalents (FTE).
- Classifications of students by new ISCED levels of education are not straightforward in every country. Nevertheless the problem was solved for the majority of cases, with a few exceptions where no disaggregation was possible.
- Similar consideration applies for classification by field of education: in several countries the ISCED-97 classification was used.
- A breakdown of students and graduates by mobility status is not fully comparable among countries, since different criteria to identify mobile students are adopted (residence, place of prior education).
- Information on R&D expenditure per HEI is available only in a subset of countries.

Metadata can be accessed directly on the ETER website and are sorted by variable and country. Concerning EUMIDA, metadata have been summarized in a written document that can be downloaded together with the data.

3.6 Flags

In order to alert users concerning data and comparability issues, data flags are introduced directly in the dataset (in separate columns alongside the data columns). The dataset includes also short remarks explaining the flag and referring to metadata for full explanations. For example a flag like "d" (definition differs) might be accompanied by the remark that mobile students are counted based on residence (rather than on place of prior education).

The table below provides a full list of flags for the ETER dataset.

TABLE 8. LIST OF FLAGS

Code	Description	Definition
b	Break in time series	When changes in definitions or data collection procedures imply that the data are not comparable across years. This flag will be relevant in the framework of multi-annual data collection.
d	Definition differs	Differences in definitions adopted for data collection imply that the value of the marked cells differs significantly from those complying with the ETER methodology.
i	See metadata	There are specific conditions which imply that the value of a cell should be interpreted in a different way or not directly compared with others.
ic	Inconsistent	Either when the sum of break down differs from the total or another semantic rule is violated.
rd	Rounding differences	When a sum of data does not fully correspond to the total because of rounding differences.
c	Confidential	When data are available, but restricted to public access (this flag is relevant only for user with unrestricted access).
ms	Missing subcategory	This flag is applied to totals in order to warn users that the total does not include one relevant subcategory (for example total expenditures not including capital expenditures).

3.7 Special codes

Special codes replace blank cells, which are not allowed in the dataset, providing more information on why a numerical value cannot be provided. These special codes largely follow standard conventions by Eurostat.

TABLE 9. SPECIAL CODES

Code	Description	Definition
m	Missing	The data is not available.
a	Not applicable	This variable is not applicable for the specific case. For example, number of PhD students in non-doctorate awarding HEIs is coded in this way.
x	Included in totals	The data are not available, but it is included in the total. This applies for example when data on revenues from student's fees are not available, but nevertheless these amounts are included in the total revenues of the HEI.
xc	Included in another column	The data are not available, but are included in the value for another column. This applies for examples when educational fields cannot be split.
xr	Included in another row	The data are not available but are included in another row. This happens in the rare case that two HEIs are part of a holding and have a common budget.
c	Confidential	When data are available, but restricted to public access, this code is displayed in place of the data.
s	Below threshold	This code is displayed in the public dataset when the count of a cell is so low that data protection issues might arise as individuals can be identified (for example number of students below 3).

4 Legal issues encountered and access conditions

4.1 Owner of the raw data

National Statistical Agencies and/or Ministries of Research and Higher Education are the owners of most of the raw data in ETER, and in particular, almost all statistical data.

The consortium itself has collected descriptors from public sources.

Since ETER is under a service contract of the European Commission, the owner of the dataset is the European Commission.

The situation concerning EUMIDA is similar, as the dataset is legally owned by the European Commission. There is however no systematic documentation of data sources and thus it is impossible to track ownership of the raw data (even if most probably originates from the NSAs).

4.2 Access to the data and conditions of use

In many countries, raw data reported in ETER are publicly available at the national level. The NSAs in these countries have confirmed this to the ETER consortium. For other countries with restricted access, the NSAs have signed an agreement for data disclosure, and delivered this to the ETER consortium. These agreements allow public access for the largest part of the data, while a few data (mostly financial) are under restricted access for research purposes (under the condition of non-disclosure of individual data points).

Access to restricted data is possible under signature of a non-disclosure agreement.

Additional datasets and the EUMIDA dataset (as soon as it is available) will be available through the RISIS projects. To this aim, potential users need to register to the RISIS system, accept the RISIS code of conduct and to sign a non-disclosure agreement (which covers both the ETER confidential data and additional data in RISIS).

5 Technical structure of the ETER database

5.1 Information on the database system

The ETER project provides an infrastructure for data collection, which allows for standardization and systematization of the process. This infrastructure 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.
- A master database including an upload interface and documentation of database activities (with time and active person).
- A web application, which allows for individual data exports (downloads of all variables for all countries is also possible, such as downloads of staff data for a specific country).

The specific requirements of the ETER project suggest a centralized web server based collection tool, which will be specified for the needs of the two data collection rounds in ETER. The advantage of such a solution is a closer linkage between data collection, feedback and revision (identification of problems and coordinated support) and finally, integration to a raw data set ready for advanced quality, consistency checks and analysis.

The data management tool has been derived from existing tools (imdas/archivis) engaged by JOANNEUM RESEARCH. The existing model of a data management system and centralized data collection was adapted according to the specific needs of the ETER project. Imdas was programmed in Gupta and C# and accesses a relational database, namely an MS SQL Server database. The database is managed via direct and central access by JOANNEUM RESEARCH, which guarantees data security, consistency and quality.

5.2 Technical variable definition

The following table summarizes the set of variables used in ETER and the format the delivered data should have. As the ETER data collection provides special codes, for example missing data, in reality all data are types of "text." Detailed definitions are provided in the ETER handbook.

TABLE 10: LIST OF VARIABLES IN DATASET

Dimension	Variables	Format
Identifiers	ETER ID	Text
	National identifier (optional)	Text
	Institution name (in own language)	Text
	English institution name (if available)	Text
	Year	Integer
Basic institutional descriptors	Country Code	ISO code
	Legal status	Nominal
	Institution category, national definition (in own language)	Text
	Institution category, national definition (in English, if available)	Text
	Institution category standardized	Nominal
	Foreign campus	Binary
	Foundation year	Integer
	Legal status year	Integer
	Ancestor year	Integer
	University hospital	Nominal
	Institutional website	Website

Geographic information	Region of establishment, NUTS2 code	Text
	Region of establishment, NUTS3 code	Text
	Name of the city	Text
	Postcode	Text
	Multi-site institution	Binary
Educational activities	Highest degree delivered	Nominal
	Number of enrolled students at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility	Integer
	Number of graduates at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility	Integer
	Distance education institution	Binary
Research	Research active institution	Binary
	Number of enrolled students at ISCED levels 8, by fields of education, gender, citizenship and mobility	Integer
	Number of graduates at ISCED levels 8 (doctorates), by fields of education, gender, citizenship and mobility	Integer
	R&D expenditures	Numeric
Expenditures	Personnel expenditure	Numeric
	Non-personnel expenditure	Numeric
	Capital expenditure	Numeric
	Accounting of capital expenditures	Nominal
Revenues	Core budget	Numeric
	Third party funding	Numeric
	Private funding	Numeric
	Tuition fees	Nominal
	Student fees funding	Numeric
Staff	Number of academic staff in FTEs and headcounts	Numeric/Integer
	Number of academic staff, by fields of education gender and citizenship in headcounts	Numeric/Integer
	Number of administrative staff in FTEs and headcounts	Numeric/Integer
	Number of professors	Integer
	Inclusion of PhD students	Binary
	Number of total staff in FTE and HC	Numeric/Integer
Erasmus students	Number of incoming Erasmus students	Integer
	Number of outgoing Erasmus students	Integer

The ETER IDs in combination with the respective reference year is used as a unique identifier, since each ETER ID can only be found once a year.

Besides data on the institutional level, the database includes metadata information in order to consider institutional and country specific characteristics in the data. Metadata include:

- metadata at the institutional level,
- metadata 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,

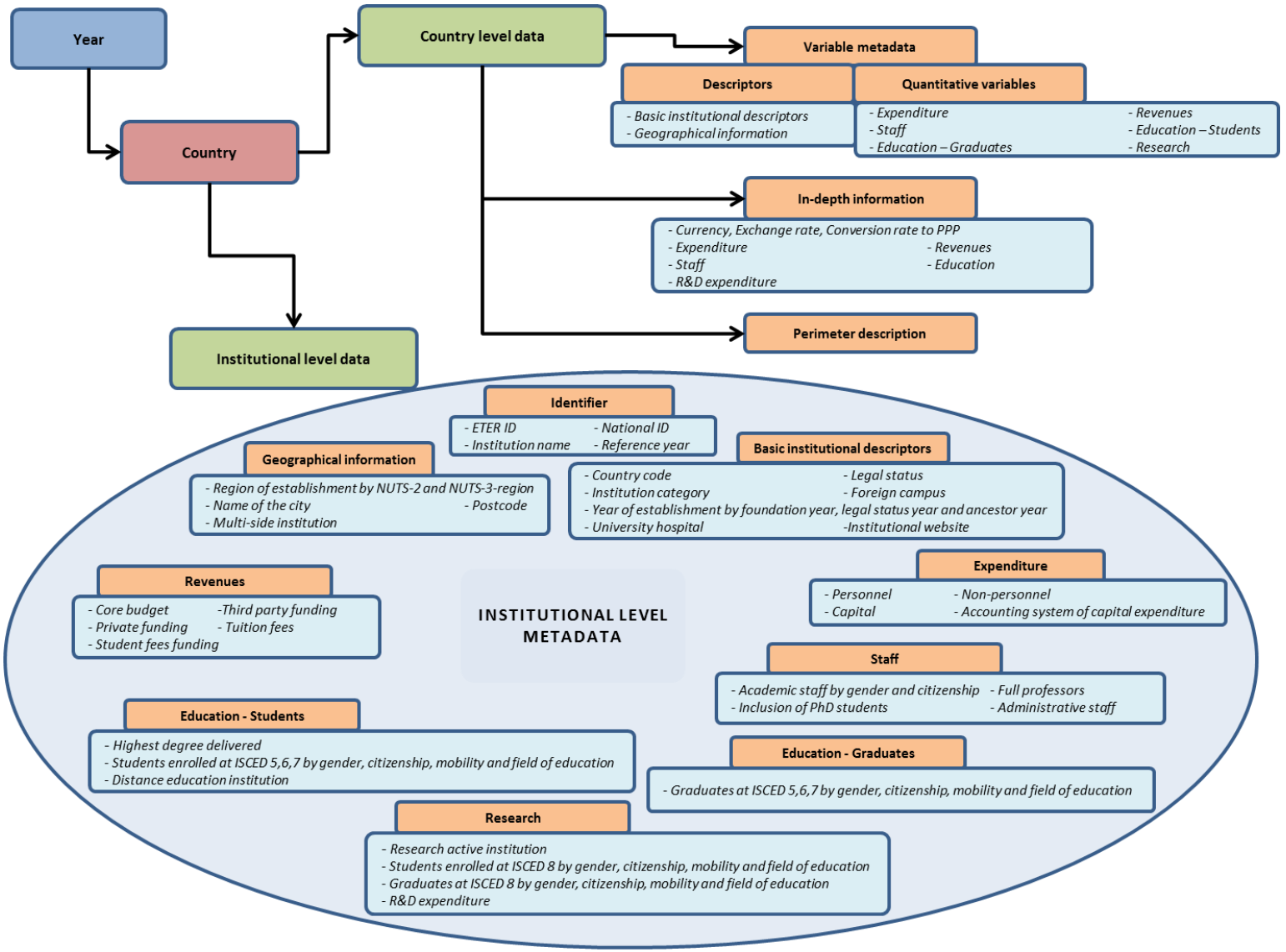
- the reference period, and
- data sources.

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

5.3 Description of the Entity Relationship Model

The following figure shows the structure of the database and their 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 the possibility of detailed data descriptions for institutions. Additionally, the database provides the opportunity to flag data, e.g. in the case of incomparable data, after quality control, and where flags are available for all quantitative variables. 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.

FIGURE 1: STRUCTURE OF THE ETER DATABASE

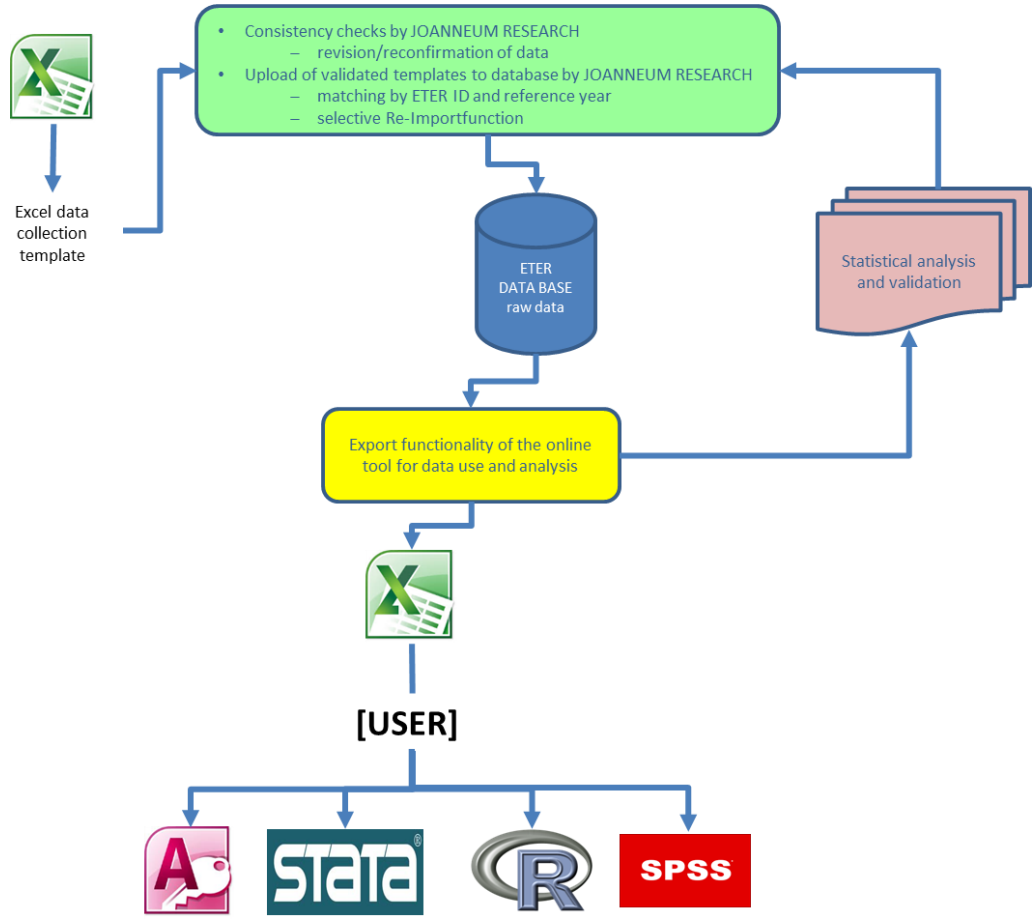


5.4 Interfaces for access and to other infrastructures

Completed and validated data collection sheets are imported into the database *imdas pro*, and all collected data are hosted in a secured server environment by JOANNEUM RESEARCH, which is also responsible for data management (cleaning, preparing for upload, uploading, updating) and backup tasks in the process. The existing model of a data management system and decentralized data collection has been adapted according to the specific needs of the ETER project. This will be done by JOANNEUM RESEARCH coordinated with the core partners.

The database also provides the foundation for ETER web application developed from JOANNEUM RESEARCH. The web application enables the user to retrieve data from the whole ETER data set in order to conduct research on micro data of the European higher education sector. The following figure gives an overview about the process starting from collected data until their exploitation in other facilities.

FIGURE 2: FROM COLLECTED DATA TO STATISTICAL ANALYSIS AND VALIDATION



The ETER web application 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.

5.4.1 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", and "Country Code".

5.4.2 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 one to switch back directly from the metadata area to the last search by using "Last Search Result".

5.4.3 Export data from the ETER database

In order to follow path3 the user needs to access "Download ETER Data", where two different types of access will be provided:

- An open public access, where small numbers and all data, for which public access was restricted by a national statistical authority, are coded.
- A restricted access, where accredited users receive access to the entire dataset for research purposes under the condition that individual data are not disclosed. To this aim, it is envisaged that ETER makes use of the accreditation system which will be established by the EU-FP7 Infrastructure Initiative on Research Infrastructure for the Assessment of Science, Technology and Innovation Policy (RISIS).

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/or country. While choosing a year is mandatory to get a search result, the user will get information for all countries at once if she/he does not select a country.

The user can choose to view:

- the entire data set, or
- data sets for one or several specific countries 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.

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 or related variables (e.g. all revenues and expenditures).

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.

While the export function "Export Full Spreadsheet" enables the export of all variables included in ETER at once, all data for a specific group of variables (see groups and included variables in chapter **Fehler! Verweisquelle konnte nicht gefunden werden.**) can be exported by selecting their name in the export function. Additionally,

four more export possibilities are provided, which cover related variables and present a useful extension for research purposes. The following table will give an overview of these export possibilities and the included variables.

TABLE 11: ADDITIONAL EXPORT POSSIBILITIES

Export possibility	Variables
Export All Expenditures and Revenues	Personnel expenditure Non-personnel expenditure Capital expenditure Expenditure unclassified Total expenditure Accounting of capital expenditures Core budget Third party funding Private funding Tuition fees Student fees funding Revenue unclassified Total revenues Research active institution R&D Expenditure
Export Student Graduates and Research	Highest degree delivered Number of enrolled students at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility Distance education institution Number of graduates at ISCED levels 5, 6, 7, by fields of education, gender, citizenship and mobility Research active institution Number of enrolled students at ISCED level 8 by fields of education, gender, citizenship and mobility Number of graduates at ISCED level 8 by fields of education, gender, citizenship and mobility R&D Expenditure
Export All Students	Highest degree delivered Number of enrolled students at ISCED levels 5, 6, 7, 8 by fields of education, gender, citizenship and mobility
Export All Graduates	Highest degree delivered Number of graduates at ISCED levels 5, 6, 7, 8 by fields of education, gender, citizenship and mobility

The user can access directly from the search mask to the country level metadata.

5.4.4 Additional functions

The selection field “Last search result” enables the user to call their last query and presents a quick link to the results if necessary. Subscribed users can change their assigned password by using selection field “Settings” and the function “Change Password.” There are no restrictions for the chosen password.

6 References

Bonaccorsi, A. (2014). *Knoweldge, 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.

Bonaccorsi, A., Daraio, C., Lepori, B. & Slipersaeter, S. (2007). Indicators on individual higher education institutions: addressing data problems and comparability issues. *Research Evaluation, 16(2)*, 66-78.

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.

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

Lepori, B., Bonaccorsi, A., Daraio, A., Daraio, C., Gunnes, H., Hovdhaugen, E., Ploder, M., Scannapieco, M. & Wagner-Schuster, D. (2014a). *ETER project. Handbook for data collection* Brussels: .

Lepori, B., Bonaccorsi, A., Daraio, A., Daraio, C., Gunnes, H., Hovdhaugen, E., Ploder, M., Scannapieco, M. & Wagner-Schuster, D. (2014b). *ETER project. Report on the first wave of data collection, June 2014.* Brussels: .

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.

OECD (2002). *Frascati Manual. Proposed Standard Practice for Surveys on Research and Experimental Development* Paris: OECD.

Ruocco, G. & Daraio, C. (2013). An empirical approach to compare the performance of heterogeneous academic fields. *Scientometrics, 97(3)*, 601-625.

UOE (2006). *UOE data collection on education systems. Manual. Concepts, definitions, classifications* Montreal, Paris, Luxembourg: UNESCO, OECD, Eurostat.

Bonaccorsi, A. (2014). *Knoweldge, 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.

Bonaccorsi, A., Daraio, C., Lepori, B. & Slipersaeter, S. (2007). Indicators on individual higher education institutions: addressing data problems and comparability issues. *Research Evaluation, 16(2)*, 66-78.

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.

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

Lepori, B., Bonaccorsi, A., Daraio, A., Daraio, C., Gunnes, H., Hovdhaugen, E., Ploder, M., Scannapieco, M. & Wagner-Schuster, D. (2015a). *ETER project. Handbook for data collection* Brussels: .

Lepori, B., Bonaccorsi, A., Daraio, A., Daraio, C., Gunnes, H., Hovdhaugen, E., Ploder, M., Scannapieco, M. & Wagner-Schuster, D. (2015b). *ETER project. Technical report on the data collection* Brussels: .

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.

OECD (2002). *Frascati Manual. Proposed Standard Practice for Surveys on Research and Experimental Development* Paris: OECD.

Ruocco, G. & Daraio, C. (2013). An empirical approach to compare the performance of heterogeneous academic fields. *Scientometrics, 97(3)*, 601-625.

UOE (2006). *UOE data collection on education systems. Manual. Concepts, definitions, classifications* Montreal, Paris, Luxembourg: UNESCO, OECD, Eurostat.